## www.mymatnscioud.com

## © MyMathsCloud

	Sampling Methods		Linear Bagrossion		Distributions
Simple	Every sample of certain size has the same chance of being selected (as every	Variables	r _evplanatory/independent	Binomial Distribution	$x \sim B(n, p)$
Random	other sample of the same size)	vanabies	v = response/dependent	Binompdf (=), Binomcdf (≤)	$E(X)=Mean=np, Var(X)=np(1-p), P(X = x)=\binom{n}{x}p^{x}(1-p)^{n-x}$
Sample (SRS)	Label cars 001-150 since want blocks of 3		y is the variable that you're testing to see whether it	Normal Distribution	$x \sim N(\mu, \sigma^2)$
	<ul> <li>If using a number line table: 456124321150240984312102364219814</li> </ul>		gets affected by x ( y on x or y versus x)	Normcdf (given x, want prob) Invnorm (given prob, want x, $\mu$ , $\sigma$ )	Standardised variable $z = \frac{x - \mu}{\sigma}$
	Split the number line up into blocks of 3 and select 3 cars 456 124 321 150 240 984 312 102 364 219 814	Line of best fit aka least squares	Way 1: calculator (raw data)	Geometric Distribution	X~Geo(p)
	<ul> <li>If using calculator: Randint (1, 150, 3) will select 3 cars</li> </ul>	regression line (LSRL)	Enter the data	(how long until 1st success)	Mean = $\frac{1}{p}$ , variance = $\frac{1-p}{p^2}$
Stratified	Split population into smaller groups called strata (subgroups) based on common	Slope interpretation: for every 1	Stat,Calc, LineReg	Geopdf (=)	$P(X = x) = p(1 - p)^{x-1}$ $P(X > r) = (1 - p)^{x}$
222 222	characteristics/shared attributes. e.g. A company has 320 drivers. 80 office workers and 40 mechanics. We want	increase in unit of x, y	rieq ist. leave blank	Geocdf (≤)	$F(aX + b) = aF(X) + b VAR(aX + b) = a^{2}Var(X)$
888 888	to select a committee of 11 to represent all the employees.	increases/decreases by b (increases	Way 2: easier formula not given raw data but given	Expectation Algebra	If X and Y independent:
	number for each category = $\frac{category \text{ total}}{complete \text{ total}} \times \text{ sample size required}$	is + slope and decreases if - slope)	$y = a + bx$ , where $a = \overline{y} - b\overline{x}$ and $b = r \frac{sy}{2}$		$E(XY) = E(X)E(Y), Var(aX \pm bY) = a^{2}Var(X) + b^{2}Var(Y)$
	$\text{Drivers} = \frac{320}{440} \times 11 = 8$ , Office $= \frac{80}{440} \times 11 = 2$ , $\text{Mech} = \frac{40}{440} \times 11 = 1$		Note:	Expected Value Discrete	$E(X) = \sum xP(X = x)$ . Multiply x's by their prob and add
	Label: Drivers 001-320, Office 321-400, Mechanics 401-440 Drivers: Randint (1 320.8) will select 8 drivers	y intercept interpretation:	$\bar{x}$ and $\bar{y}$ = means	Variance Discrete	$Var(X) = \sum x^{-}P(X = X) - E(X)^{-}$ Multiply $x^{2'}s$ by their prob and add and then take away $E(X)^{2}$
	Office workers: Randint (321,400,2) will select 2 office workers	when x=0 then y=a i.e. the starting	$s_x$ and $s_y$ = standard deviations	Normal Approximation To	n large, p close to $\frac{1}{2}$ OR $n\hat{p} \ge 10$ , $n(1 - \hat{p}) \ge 10$
	Mechanics: Randint (1,320,1) will select 1 mechanic Choose subjects in a systematic orderly/logical way by sampling every kth	value		Binomial (don't forget continuity	$X \sim B(n, p) \xrightarrow{2} X \sim N(np, np(1-p))$
Systematic	element/value	To plot line of best fit on TI 84:	Way 3: harder formula	correction before finding the prob)	Mean is $np$ and s.d. is $\sqrt{np(1-p)}$
0000000000	e.g. Select a sample of 15 people from a business survey of 1200 people	Enter data and go Linreg	$y - \overline{y} = \frac{s_y}{s_{xx}}(x - \overline{x})$ , where,		Sampling Distributions
	1200 = 80 i.e. we'll take every 80 <sup>th</sup> value	Store RegEq : y vars, function	$s_{m} = \sum r^2 - \frac{(\sum x)^2}{2} = \sum r^2 - n\bar{r}^2$	SAMPLE Mean Distribution	$\overline{x} \sim N(\mu, \frac{\sigma^2}{\sigma})$ . Mean is $\mu$ and s.d. is $\sqrt{\sigma}$
	<ul> <li>If using number line: 12436289461234560124</li> </ul>	Y <sub>1</sub> ,calculate	$S_{XX} = \sum_{n} n \sum_{(X,Y) \in Y} nX$	(Central Limit Theorem)	$\sqrt{n}$
	Split this up into blocks of 4: 1243 6289 4612 3456 0124	stat plot, select plot 1, x list. L <sub>1</sub> ,	$s_{xy} = \sum xy - \frac{(2xy)(2yy)}{n} = \sum y^2 - nxy$	SAMPLE Proportion	$\binom{n(1-n)}{n}$
	124 is our starting point 124+80=204, 204+80=284, 284+80=364 etc	Press zoom. (9:zoomstat if can't see		Distribution	$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$ . Mean is $p$ and s.d. is $\sqrt{\frac{p(1-p)}{n}}$
	124,204,284,364,444, 524, 604,684, 764,844,924,1004,1084, 1164, 044	graph		Condition: $n\hat{p} \ge 10$ , $n(1 - \hat{p}) \ge 10$ , $\hat{p} = n$	use POPULATION value $p$ , not sample value $\hat{p}$ which is $\frac{x}{n}$
	If using a calculator: Randint (1,1200,1) gives 24 so start at 24 24+80=104, 104+ 80 = 184, 184+80=264 etc	Making Predictions	we can plug x values into the line of best fit and solve	p = p	(we only use $\hat{p}$ for the conditions, intervals and test statistic)
	24,104,184,264,344,424,504,584,664,744,824,904,984,1064,1144		for y and vice versa	Hypot	hesis Testing – Test Statistics
Cluster	Split population into smaller groups called clusters and sample <u>EVERYONE</u> from randomly chosen subgroups. Normally we split up based on area/geographical	Interpolation	plugging values into line of best fit that ARE within	Assumptions	See confidence interval section
	location. This is similar to stratified sampling, but with stratified sampling the	The second states	the data range (sate/accurate)	Hypothesis Templates	1 sample:
888	randomness is applied when selecting within each subgroup whereas with cluster sampling the randomness is applied when selecting between each	Extrapolation	plugging values into line of best fit that aren't within the data range (pot cafe/pot accurate)	Lise population values and	$H_1: p > \cdots$ or $\mu > \cdots$
	subgroup. When we select the subgroup, we choose <u>EVERY</u> number inside.	Correlation coefficient $(r)$	Way 1: calculator (if have raw data)	ask yourself	$p < \cdots$ or $\mu < \cdots$
	Therefore not all subgroups will be chosen and therefore won't be represented in the cample unlike with stratified	contraction coefficient (7)	Stat.Calc LineReg(ax+b)	<ul> <li>whether 1 or 2 sample</li> </ul>	$p \neq \cdots$ or $\mu \neq \cdots$
0058 8888 0058 8888	in the sample unine with stratmed.	Note:	if r is not appearing go to:Catalog, Diagnostic on	whether Z or T	2 sample: $H_{a}: n_{a} - n_{a} = 0$ i.e. $n_{a} = n_{a}$ or $\mu_{a} - \mu_{a} = 0$ i.e. $\mu_{a} = \mu_{a}$
	Confidence Intervals (CI)	$s_x$ and $s_y$ = standard deviations			$H_0(p_1^{-1}, p_2^{-1}) = p_1(p_2^{-1}) = p_2(p_2^{-1}) = p_$
Terminelem			Way 2: easier formula		$p_1 - p_2$ > 0 i.e. $p_1$ > $p_2$ or $\mu_1 - \mu_2$ > 0 i.e $\mu_1$ > $\mu_2$
and Form	$\widehat{p}, \mu$ =population proportion/mean (parameters) $\widehat{p}, \overline{x}$ =sample proportion/mean (statistics/point estimates)	U=no correlation	$r = \frac{-s_y}{\sqrt{s_{xx}s_{yy}}}$ or $r = \frac{s_x}{s_y}b$ where $y = a + bx$		$p_1 - p_2 \neq 0$ i.e. $p_1 \neq p_2$ or $\mu_1 - \mu_2 = 0$ i.e
	lower limit + upper limit	-1=perfect pegative correlation			2 sample paired: $\mu_1 \neq \mu_2$
$\hat{\mathbf{p}} = \frac{x}{2} = \frac{subset}{2}$	2	2 percet legative correlation	Way 3: harder formula		$H_0: \mu_d = 0$
• n sample		0.00-0.19 very weak	$r = \frac{1}{n-1} \sum \left( \frac{x_i - \hat{x}}{s_i} \right) \left( \frac{y_i - \hat{y}}{s_i} \right)$		$H_1/H_a: \mu_d > 0$ $\mu_d < 0$
$\overline{x}$ =sample		0.20-0.39 weak	$n-1 = (s_x) (s_y)$ Where		$\mu_d \leq 0$ $\mu_d \neq 0$
mean	MOE MOE	0.40-0.59 moderate	$S_{xx} = \sum r^2 - \frac{(\sum x)^2}{2} - \sum r^2 - m\bar{z}^2$		Slope of regression line:
		0.60-0.79 strong	$S_{XX} = \Delta X = \frac{1}{n} = \Delta X = nX$		$H_0: \beta = 0$ $H: \beta \neq 0$
	lower statistic upper	0.8-1 very strong	$s_{yy} = \sum y^2 - \frac{(2y)}{n} = \sum y^2 - n\overline{y}^2$		$\mu_1: p \neq 0$ $\beta < 0$
	limit <b>p</b> or x limit		$s_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{x} = \sum xy - n\bar{x}\bar{y}$		$\beta > 0$
	ci = { statistic ± margin of error	Determination of	Way 1:	Test Statistic (TS)	1 prop TS: $Z = \frac{\hat{p} - p}{p}$
	statistic ± (critical value)(standard error)	variation/proportion of	Square r value found above	AP Stats: $\sigma$ unknown	$\frac{p(1-p)}{2}$
	Note: The standard error is just the S.D. of the sampling distribution	variation $(r^2)$		always, so can always use	2 prop TS: $7 - (\widehat{p_1} - \widehat{p_2}) - (p_1 - p_2)$
	= (statistic - margin of error, statistic + margin of error)		Way 2: use a formula (never use this)	I for mean tests.	$2 \mu 0 \mu 13. 2 - \frac{1}{\hat{p}(1-\hat{p})(\frac{1}{2}+\frac{1}{2})}$ where $p_1 - p_2 = 0$
	= (lower limit, upper limit) = (a, b)	This tells us the percent of	$r^2 = 1 - \frac{m}{SST}$	$\sigma_{\rm e} \approx \sigma_{\rm e}$ but you will also	$\sqrt{(n_1 n_2)}$ $\approx -x_1 \approx -x_2 \approx -x_1+x_2$
Chabiatia /	(Y	variation in y that is due to x	$SSE = \sum (y_i - \hat{y})^2 = \sum (residuals)^2$	$\sigma_1 \sim \sigma_2$ , but you will also notice that the mark scheme	$p_1 = \frac{1}{n_1}, p_2 = \frac{1}{n_2}, p = \frac{1}{n_1 + n_2}$
Point Estimate	$\left(\frac{x}{n}\right)$ if given subset x of the sample size n		SST=sum of squares for total (sum of squared	normally doesn't pool	1 mean TS: Z = $\frac{x-\mu}{S}$ , T= $\frac{x-\mu}{S}$ df =n-1
i onit Estimate	$=\begin{cases} a + b \\ a + b \end{cases}$ if given confidence interval (a, b)	Posidual	actual data value – predicted value from line	regardless of size of $\sigma$	<b>2</b> mean <b>TC</b> : <b>7</b> = $\frac{\sqrt{n}}{X_1 - X_2 - (\mu_1 - \mu_2)} \frac{\sqrt{n}}{T_1 - X_2 - (\mu_1 - \mu_2)}$
-		Residual	actual data value predicted value nom me		Z mean 13. Z = $\frac{s_1^2 + s_2^2}{s_1^2 + s_2^2}$ , 1= $\frac{s_p}{s_1 + \frac{1}{n_2} + \frac{1}{n_2}}$
Margin Of	$\left(\frac{b-a}{2}\right)$ if given interval (a, b)	To draw residual plot on TI 84	This is the vertical distance from the points to the	Either use TS or P value to	$(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2$
Error (MOE)	$\hat{\mathbf{p}}(1-\hat{\mathbf{p}})$	Stat plot, select plot 1, $x$ list; $L_1$ .	line. The less the gap between points and line, the	decide whether to reject:	where $\mu_1 - \mu_2 = 0$ and $s_p = \sqrt{\frac{n_1 + n_2 - 2}{n_1 + n_2 - 2}}$
This is the part	$z_c = \frac{p(x - p)}{n}$ (if proportion)	Y List: Resid (2 <sup>nd</sup> , stat, Resid)	lower the residual value	<: TS< CV reject	If don't pool: T= $\frac{\overline{X_1} - \overline{X_2} - (\mu_1 - \mu_2)}{(\mu_1 - \mu_2)}$
of the formula	$=\begin{cases} v \\ r \\ s \\ r \\ r$	Make sure you have already	Deinte lie about line means positive socidual	>: TS> CV reject	$\sqrt{\frac{s_1^*}{n_1} + \frac{s_2^*}{n_2}}$
after the $\pm$ in	$Z_c \sqrt{n}$ (if mean and using Z)	stored the regression equation	Points lie below line means negative residual	≠: TS< CV or TS>CV reject	$df=n_1 + n_2 - 2$ (if pool)/min $(n_1 - 1, n_2 - 1)$ (if no pool)
the confidence	$T_c \frac{S}{T_c}$ (if mean and using T)	in line of best fit section	Formes the below time means negative residual		2 means paired TS: Z/T = $\frac{3}{2}$ df =n-1
interval	( Vn		The mean of the residuals is always zero	P value Method:	Slope of regression line TS: $T = \frac{slope}{\beta} = \frac{b-\beta}{\beta} = \frac{b}{\beta}$
	βq ms in s	Computer Printout Analysis	Perposte ustisble it	If P value $< \alpha$ reject $H_0$	
	To solve for n: Use either $z_c \sqrt{\frac{n}{n}}$ or $T_c \sqrt{n} / Z_c \sqrt{n}$	(you are asked to locate and	Variable Coefficient Std Dev T P		where $\beta = 0$ , df = n-2
	If don't know $\hat{p}$ then use $\hat{p}(1-\hat{p})=0.25$ unless told otherwise	interpret the values in red on	Constant INTERCEPT Amount. SLOPE SE Slope	Conclusion	There is sufficient/insufficient evidence at the% level to reject $H_{\theta}$ & we can conclude
	Interpretation: If we repeated this procedure multiple times	the table to the right)	S = SD residuals R-Sq = $r^2$ value	P value	If p values is than $\alpha$ reject $H_0$ i.e $P < \alpha$
	then statistic would be within(WOE) of the true value (%) of the time			To calculate: Find TS First	
Standard Error		Standard Error of Slope	Standard deviation of the estimated slope for	<: P(Z <ts)< th=""><th>Interpretation:</th></ts)<>	Interpretation:
Standard Error	Proportion: $\sqrt{\frac{p \cdot q}{p}}$ , Mean: $\frac{s}{\sqrt{p}}$	Parameter (SE)	predicting y by using an amount in x	>: P(Z>TS)	Assume H <sub>0</sub> is true, then prob of being in the
Formulae	<u>(6)</u>	Standard Deviation of Residuals	vertical from the actual data value (observed v	≠: 2P(Z>TS) or 2P(Z <ts)< th=""><th>correct tail is the p value (in lower tail if &lt; test, in</th></ts)<>	correct tail is the p value (in lower tail if < test, in
All these	• 1 prop: $\hat{\mathbf{p}} \pm \mathbf{z}_{c_1} / \frac{\mathbf{p} \cdot \mathbf{q} \cdot \mathbf{p}}{\mathbf{p}}$ where $\hat{p} = \hat{\mathbf{p}}, \hat{\mathbf{q}} = 1 - \hat{\mathbf{p}}$		vertical nom the actual data value (observed)	Find prop 5 using normcur	1000000000000000000000000000000000000
formulae have			variable) to the regression line i.e measure of	with $\mu$ and $\frac{\sigma}{\sigma}$ or Todf with df	upper tail if > test and in double either tail if $\neq$ test) the p value
	Tests, 1 Prop Z interval		variable) to the regression line i.e measure of variation in y variable for a given amount of x	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcdf with df	test) the p value test and in double either tail if $\neq$
STATISTIC/	Tests, 1 Prop Z interval • 2 prop: $\widehat{p_1} - \widehat{p_2} + z_{-} \sqrt{p_1 \widehat{q_1} + p_2 \widehat{q_2}}$ where $\widehat{p_2} = \frac{x_1}{2}, \widehat{p_2} = \frac{x_2}{2}$		variable) to the regression line i.e measure of variation in <i>y</i> variable for a given amount of <i>x</i> variable	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcdf with df (see test statistics section above for df for Tcdf)	upper tail if > test and in double either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value
STATISTIC/ SAMPLE	• 2 prop: $\widehat{p}_1 - \widehat{p}_2 \pm z_{c_2} \sqrt{\frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2}}$ where $\widehat{p}_1 = \frac{x_1}{n_1}, \widehat{p}_2 = \frac{x_2}{n_2}$	To Check Whether A Good Model	variable) to the regression line i.e measure of variation in y variable for a given amount of x variable Scatter plot looks like a straight line	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcdf with df (see test statistics section above for df for Tcdf) Chi-Souared	upper tail if > test and in double either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value
STATISTIC/ SAMPLE values as their values not	$ \begin{array}{c} \mathbf{v}_{1} \\ \text{Tests, 1 Prop Z Interval} \\ \bullet & 2 \operatorname{prop}: \widehat{p_{1}} - \widehat{p_{2}} \pm z_{c_{1}} \left( \frac{ \widehat{p_{1}} \widehat{p_{1}} + p_{2} }{n_{1}} + \frac{ \widehat{p_{2}} }{n_{2}} \right) \text{ where } \widehat{p_{1}} = \frac{x_{1}}{n_{1}}, \widehat{p_{2}} = \frac{x_{2}}{n_{2}} \\ \text{Tests, 2 Prop Z Interval} \\ \bullet & \text{Tests, 2 Prop Z Interval} \end{array} $	To Check Whether A Good Model	variable) to the regression line i.e measure of variable to the regression line i.e measure of x variable Scatter plot looks like a straight line High r value	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcdf with df (see test statistics section above for df for Tcdf) Chi- Squared	upper tail in 2 test and in bouble either tail in $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2_{calc} = \sum_{E} \frac{(0-E)^2}{E}$ Reject: $x^2_{alc} > x^2_{critical}$
STATISTIC/ SAMPLE values as their values, not population	$ \begin{array}{c} \mathbf{V}_{n} \\ \text{Tests, 1 Prop Z interval} \\ \bullet  \mathbf{2 prop;} \ \widehat{p_{1}} - \widehat{p_{2}} \pm z_{e_{v}} \sqrt{\frac{p_{1}e_{1}}{n_{1}} + \frac{p_{2}e_{2}}{n_{2}}} \text{ where } \ \widehat{p_{1}} = \frac{x_{1}}{n_{1}}, \ \widehat{p_{2}} = \frac{x_{2}}{n_{2}} \\ \text{Tests, 2 Prop Z interval} \\ \bullet  \mathbf{1 mean:} \ \overline{X} \pm Z_{e_{v}} \frac{d_{v}}{d_{v}} \text{ OR } \overline{X} \pm T_{e_{v}} \frac{d_{v}}{d_{v}} (\sigma \text{ unknown}/ \text{ small sample}) \end{array} $	To Check Whether A Good Model	variable) to the regression line i.e measure of variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r <sup>2</sup> value Becidual let has an orther with wildow variation	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcdf with df (see test statistics section above for df for Tcdf) Chi- Squared	upper tail in 2 test and in bouble either tail in $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2_{calc} = \sum \frac{(0-E)^2}{E}$ . Reject: $x^2_{calc} > x^2_{critical}$ Hypothesis $H_a$ : are independent/in the ratio/distributed $H_a$ : are not independent/in the ratio/distributed
STATISTIC/ SAMPLE values as their values, not population values	Tests, 1 Prop Z interval • 2 prop: $\widehat{p_1} - \widehat{p_2} \pm z_c \sqrt{\frac{\beta_1 \hat{s}_1}{n_1} + \frac{\beta_2 \hat{s}_2}{n_2}}$ where $\widehat{p_1} = \frac{x_1}{n_1}, \widehat{p_2} = \frac{x_2}{n_2}$ Tests, 2 Prop Z interval • 1 mean: $\overline{X} \pm Z_c \frac{\sigma}{m} \text{OR } \overline{X} \pm T_c \frac{\sigma}{\sqrt{m}} (\sigma \text{ unknown/ small sample})$ Tests, 2 OR T Interval	To Check Whether A Good Model	variable) to the regression line i.e measure of variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value High r <sup>2</sup> value Residual plot has no pattern with uniform variation agross r (tarty night)	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcdf with df (see test statistics section above for df for Tcdf) Chi- Squared	upper tail if > test and in bouble either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2_{calc} = \sum_{i=1}^{(G-E)^2}$ . Reject: $x^2_{calc} > x^2_{critical}$ Hypothesis.H <sub>i</sub> : are independent/in the ratiodistributed H <sub>i</sub> : are not independent/or the ratiodistributed H <sub>i</sub> : are not independent to others
STATISTIC/ SAMPLE values as their values, not population values Ask yourself	Tests, J Prop Z Interval • 2 prop: $\widehat{p_1} - \widehat{p_2} \pm z_c \sqrt{\frac{p_1 \widehat{v_1}}{n_1} + \frac{p_2 \widehat{v_2}}{n_2}}$ where $\widehat{p_1} = \frac{x_1}{n_1}, \widehat{p_2} = \frac{x_2}{n_2}$ Tests, 2 Prop Z Interval • 1 mean: $\overline{X} \pm Z_c \frac{\sigma_1}{\sigma_1} OR \ \overline{X} \pm T_c \frac{\pi}{\sigma_1}$ ( $\sigma$ unknown/ small sample) Tests, Z OR T Interval $z_c$ : Invnorm with area $\frac{1}{2}, \mu = 0, \sigma = 1$	To Check Whether A Good Model	variable) to the regression line i.e measure of variable) to the regression line i.e measure of x variable Scatter plot looks like a straight line High r <sup>2</sup> value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations	with µ and رَبَّتُ or Tcdf with df (see test statistics section above for df for Tcdf) Chi- Squared	upper tail in 2 test and in bouble either tail in $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2_{calc} = \sum_{i=1}^{(Q-E)^2}$ . Reject: $x^2_{calc} > x^2_{critical}$ HypothesisH <sub>i</sub> : are independent/in the ratio/ distributed H <sub>i</sub> : are not independent/in the ratio/ distributed dif(row:)1(columns:1) for independence and n-1 for others df=(no-2)1(columns:1) for independence and n-1 for others df=(no-2)1(columns) for negative for others and the negative for the nend negative for the nend negative for the n
STATISTIC/ SAMPLE values as their values, not population values Ask yourself first: Is proportion or	$\begin{split} \mathbf{v}_{1}^{\mathbf{v}_{1}} & \mathbf{v}_{1}^{\mathbf{n}_{1}} \\ \text{Tests, 1 Prop Z Interval} \\ \bullet & 2 \operatorname{prop}: \widehat{p_{1}} - \widehat{p_{2}} \pm x_{c_{1}} \left( \frac{\beta_{1} q_{1}}{m_{1}} + \frac{\beta_{2} q_{2}}{m_{2}} \right) \\ \text{Tests, 2 Prop Z Interval} \\ \bullet & 1 \operatorname{mean}: \widetilde{X} \pm Z_{c} \frac{\sigma_{1}}{m_{c}} \otimes \widetilde{X} \pm T_{c} \frac{\sigma_{1}}{m_{c}} (\sigma \operatorname{unknown}) \\ \text{Tests, 2 OR T Interval} \\ & \operatorname{Tests, 2 OR T Interval} \\ & Z_{c}: \operatorname{Invnorm with area} \frac{1-q_{1}}{m_{c}}, \mu = 0, \sigma = 1 \\ & T_{c}: \operatorname{InvT with area} \frac{1-q_{1}}{m_{c}} \frac{1}{m_{c}} dc = n-1 \end{split}$	To Check Whether A Good Model Linear Re	variable) to the regression line i.e measure of variation in y variable for a given amount of x variable Scatter plot looks like a straight line High $\tau^2$ value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations $y = ax^6$ where $x \to logx, y \to logy$	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcdf with df (see test statistics section above for df for Tcdf) Chi- Squared	upper tail in 2 test and in bouble either tail in $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2_{calc} = \sum \frac{(\partial_z - E)^2}{E}$ . Reject: $x^2_{calc} > x^2_{critical}$ HypothesisH_z are independent/in the ratio/_distributed Hi; are not independent in the ratio/_distributed Hi; are not independent in the ratio/_distributed Hi; are not independent in the ratio/_distributed H; are not independent in the ratio/_distributed H; are not independent in the ratio/_distributed H;
STATISTIC/ SAMPLE values as their values, not population values Ask yourself first: Is proportion or mean and then	$\begin{split} \mathbf{v}_{1}^{\mathbf{v}} & \mathbf{v}_{1}^{\mathbf{v}} \\ \text{Tests, 1 Prop Z interval} \\ \bullet & 2 \text{ prop: } \widehat{p_{1}} - \widehat{p_{2}} \pm z_{c_{1}} \sqrt{\frac{p_{1} \cdot q_{1}}{n_{1}} + \frac{p_{2} \cdot q_{2}}{n_{2}}} \text{ where } \widehat{p_{1}} = \frac{x_{1}}{n_{1}}, \widehat{p_{2}} = \frac{x_{2}}{n_{2}} \\ \text{Tests, 2 Prop Z interval} \\ \bullet & 1 \text{ mean: } \overline{X} \pm Z_{c_{1}} - \overline{q_{1}} \text{ (} \sigma \text{ unknown}/ \text{ small sample} ) \\ \text{Tests, 2 OR T interval} \\ z_{c}: \text{ invnorm with area} \frac{1-q_{0}}{p_{2}}, \mu = 0, \sigma = 1 \\ T_{c_{1}}: \text{ inv T with area} \frac{1-q_{0}}{q_{2}}, \frac{d(r - n - 1)}{Z} \\ \bullet & 2 \text{ means: } \overline{X}_{1} - \overline{X}_{0} \pm Z_{c_{1}} + \frac{q_{2}^{2}}{q_{2}} + \frac{q_{2}^{2}}{Z} / \overline{X}_{c_{1}} - \overline{X}_{c_{1}} + \frac{1}{q_{2}} \end{split}$	To Check Whether A Good Model	variable) to the regression line i.e measure of variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations $y = ax^{b}$ where $x \rightarrow logx, y \rightarrow log y$ On calculator. "writing((log L, log L, )	with µ and <sup>2</sup> / <sub>vin</sub> or Tcdf with df       (see test statistics section above for of for Tcdf)       Chi- Squared       Errors       Type 2 Error Steps:	upper tail if > test and in bouble either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2_{calc} = \sum_{acc} \frac{(o-x)^2}{2}$ . Reject: $x^2_{calc} > x^2_{critical}$ Hypothesis. $u_a$ re is dependent/in the ratio/_ distributed df=(row-1)(columns-1) for independence and n=1 for others df=n = 2.1 approximating p or $\mu/a$ but this never comes up <b>Definitions:</b> <b>Type</b> 1: $H_a$ true, but we say it is false i.e. reject it <b>Type</b> 2: $H_a$ false, but we say true i.e. accept it
STATISTIC/ SAMPLE values as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2	$\frac{\nabla q^{n}}{\operatorname{Tests}, 1 \operatorname{Prop Z} \operatorname{interval}}$ • 2 prop: $\widehat{p_{1}} - \widehat{p_{2}} \pm z_{e_{0}} \frac{\widehat{p_{1}e_{1}}}{n_{1}} + \frac{\widehat{p_{2}e_{2}}}{n_{2}}$ where $\widehat{p_{1}} = \frac{x_{1}}{n_{1}}, \widehat{p_{2}} = \frac{x_{2}}{n_{2}}$ Tests, 2 Prop Z interval • 1 mean: $\overline{X} \pm Z_{e} \frac{d}{q_{1}} \operatorname{OR} \overline{X} \pm T_{e} \frac{d}{q_{1}}$ ( $\sigma$ unknown/ small sample) Tests, 2 OR T Interval $z_{e}$ : Invorm with area $\frac{1-w_{e}}{2}, \mu = 0, \sigma = 1$ $T_{e^{2}}$ : InvT with area $\frac{1-w_{e}}{2}, \frac{de_{1}}{d1} = n-1$ • 2 means: $\overline{X_{1}} - \overline{X_{2} \pm Z_{e}} \frac{d^{2}}{q_{1}} + \frac{w_{2}}{q_{2}} / \overline{X_{1}} - \overline{X_{2} \pm T_{e}} s_{p} \sqrt{\frac{1}{n_{1}} + \frac{1}{n_{2}}}$	To Check Whether A Good Model Linear Re Power Exponential	variable) to the regression line i.e measure of variable) to the regression line i.e measure of x variable Scatter plot looks like a straight line High $r^2$ value Residual plot has no pattern with uniform variation across x (starry night) egression - Transformations $y = ax^5$ where $x \rightarrow log_{X_1} y \rightarrow log_Y$ On calculator. PumBeg(L, jo timBeg (log L, log L_2) $y = ab^5$ where $x \rightarrow x_1 y \rightarrow log_Y$ On calculator. PumBeg(L, i $y \rightarrow top_X$ )	with µ and <sup>4</sup> / <sub>vin</sub> or Tcdf with df       (see test statistics section above for df for Tcdf)       Chi- Squared       Errors       Type 2 Error Steps: Step 1: Find CV (using invnorm)	upper tail if > test and in double either tail if $\neq$ test) the p value Larger sample size $\Longrightarrow$ smaller p value If $x^2_{callc} = \sum_{i=1}^{(G-E)^2}$ . Reject: $x^2_{callc} > x^2_{critical}$ HypothesisH <sub>i</sub> : are independent/in the ratio/ distributed H <sub>i</sub> : are not independent/in the ratio/ distributed H <sub>i</sub> : are not independent/in the ratio/ distributed H <sub>i</sub> : are not independent/in the ratio/ distributed <b>Definitions</b> . Type 1: H <sub>0</sub> true, but we say it is false i.e. reject it <b>Type 2:</b> H <sub>0</sub> false, but we say it is false i.e. reject it <b>Calculations</b> :
STATISTIC/ SAMPLE values as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use 2 if or	$\begin{split} \mathbf{v}_{1}^{\mathbf{v}_{1}} & \mathbf{v}_{1}^{\mathbf{v}_{1}} \\ \text{Tests, 1 Prop Z interval} \\ \bullet & 2 \operatorname{prop}: \widehat{p_{1}} - \widehat{p_{2}} \pm z_{c}, \left( \frac{\beta_{1} q_{1}}{n_{1}} + \frac{\beta_{2} q_{2}}{n_{1}} \right) \\ \text{main } \widehat{p_{1}} = \frac{x_{1}}{n_{1}}, \widehat{p_{2}} = \frac{x_{2}}{n_{2}} \\ \text{Tests, 2 Prop Z interval} \\ \bullet & 1 \operatorname{mean}: \overline{X} \pm Z_{c} \frac{\sigma_{1}}{\sigma_{1}} \\ \operatorname{ord}: \overline{X} \pm Z_{c} \frac{\sigma_{1}}{\sigma_{1}}, \left( \overline{\sigma} \operatorname{urknown} \right) \\ \text{small sample} \\ \text{Tests, 2 OR T interval} \\ z_{c}: \operatorname{invnorm with area} \frac{1-q_{0}}{2}, \mu = 0, \sigma = 1 \\ T_{c}: \operatorname{invT with area} \frac{1-q_{0}}{2}, \frac{\sigma_{1}}{n_{1}} + \frac{\sigma_{2}}{n_{2}}, \left( \overline{x_{1}} - \overline{x_{2}} \pm T_{c} S_{p} \sqrt{\frac{1}{n_{1}} + \frac{1}{n_{2}}} \right) \\ \bullet & 2 \operatorname{means}: \overline{X_{1}} - \overline{X_{2}} \pm T_{c} \sqrt{\frac{\sigma_{1}^{2}}{n_{1}} + \frac{\sigma_{2}}{n_{2}}} (\overline{x_{1}} - \overline{x_{2}} \pm T_{c} S_{p} \sqrt{\frac{1}{n_{1}} + \frac{1}{n_{2}}} \\ \text{where } S_{p} = \sqrt{\frac{(\alpha_{1} + \beta_{2} \pm (\alpha_{2} - 1)s_{2}^{2}}{n_{1} + (\alpha_{2} - 1)s_{2}^{2}}} (\text{pooled formula}) \end{split}$	To Check Whether A Good Model Linear Re Power Exponential Interpretations	variable) to the regression line i.e measure of variation in y variable for a given amount of x variable Scatter plot looks like a straight line High $r^2$ value Residual plot has no pattern with uniform variation across x (starry night) seression - Transformations $y = \alpha x^0$ where $x \to logx, y \to log y$ On calculator: Pwrkeg( $L_i, L_j$ ) or Linkeg (log $L_i \log L_j$ ) $y = ab^x$ where $x \to x, y \to \log y$ On calculator: Explegit, $L_i$ or Schege ( $L_i \log L_j$ ) Transformed scatter must look linear	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcdf with df (see test statistics section above for df for Tcdf) Chi- Squared Errors Type 2 Error Steps: Step 1: Find CV (using invnorm) $<: area = \alpha$ (left), $\mu = 0$ , $\sigma = 1$	upper tail if > test and in double either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2_{callc} = \sum_{i=1}^{(Q-E)^2}$ . Reject: $x^2_{callc} > x^2_{critical}$ HyothesisH, are independent/in the ratio/ distributed H <sub>i</sub> : are not independent/in the ratio/ distributed dif(row:).(Columns:)1 for independence and n-1 for others df = n - 2 if approximating p or µ/a but this never comes up <b>Definitions</b> : <b>Type 1:</b> H <sub>0</sub> true, but we say it is false i.e. reject it <b>Type 2:</b> H <sub>0</sub> false, but we say true i.e accept it <u>Calculations</u> : <b>Type 1:</b> This is a (prob of rejecting the test i.e. being in the critical region)
STATISTIC/ SAMPLE values as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use 2 if or known OR $n \ge$	Tests, 1 Prop Z interval • 2 prop: $\widehat{p_1} - \widehat{p_2} \pm z_{<2} \left[ \frac{\widehat{p_1} \widehat{q_1}}{n_1} + \frac{\widehat{p_2} \widehat{q_2}}{n_2} \text{ where } \widehat{p_1} = \frac{x_1}{n_1} \cdot \widehat{p_2} = \frac{x_2}{n_2} \right]$ Tests, 2 Prop Z interval • 1 mean: $\overline{X} \pm Z_{<2} \subset \overline{q_2} \text{ OR } \overline{X} \pm T_{<2} \subset \overline{q_1} \text{ for unknown}/ small sample)$ Tests, 2 OR T Interval $z_{<1}$ : Invorm with area $\frac{1-y_0}{n_1}, \mu = 0, \sigma = 1$ $T_{<1}$ : InvT with area $\frac{1-y_0}{n_1}, \frac{d = n - 1}{2}$ • 2 means: $\overline{X}_1 - \overline{X}_2 \pm Z_{<2} \left( \frac{q_2^2 + q_2^2}{n_1 + q_2^2} / \overline{X}_1 - \overline{X}_2 \pm T_{<2} S_{p} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$ where $S_{p} = \sqrt{\frac{(n_1-1)q^2 + (n_2-1)s^2}{n_1 + n_2-2}}$ (pooled formula) We pool where $\sigma_{<2} \sigma_{<2}$ $S_{<2} < 2 < s_{<2} < S_{<2}$	To Check Whether A Good Model Linear Ro Power Exponential Interpretations	variable) to the regression line i.e measure of variation in y variable for a given amount of x variable. Scatter plot looks like a straight line High r value High r value High r value Across x (starry night) gression Transformation across x (starry night) gression Transformation $x_1(x_1) \propto 1.0g_{X_1} \rightarrow 1.0g_{X_2} \rightarrow 1.0g_{X_2}$	$\label{eq:constraints} \begin{split} & \mbox{with } \mu \mbox{ and } \mbox{$_{\sqrt{n}}$} \mbox{or Tcdf with df} \\ & (se test statistics section \\ & \mbox{ above for fcf} \mbox{$_{\sqrt{n}}$} \mbo$	upper tail if > test and in double either tail if ≠ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2_{calc} = \sum_{ac} \frac{(o-z)^2}{2}$ . Reject: $x^2_{calc} > x^2_{critical}$ Hyothesis $H_{ac}$ are independent/in the rais/_ distributed df=(row 1)(columns-1) for independence and $n-1$ for others df= $n - 21$ agrowinating p or $\mu/a$ but this newer comes up Definitions: Type 1: $H_{a}$ true, but we say it is false i.e. reject it Type 2: $H_{a}$ false, but we say true i.e accept it Calculations: Type 2: $H_{a}$ (prob of NOT being in the critical region-see
STATISTIC/ SAMPLE values as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use 2 if $\sigma$ known OR $n \ge 30$	$\sum_{k=1}^{n} \sum_{j=1}^{n} \sum_{j$	To Check Whether A Good Model Linear Re Power Exponential Interpretations	variable) to the regression line i.e measure of variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r <sup>2</sup> value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations $y = ax^b$ where $x \to logx, y \to log y$ On calculator: PwrReg(L,L_b) or LinReg (log L,log L_2) $y = ab^c$ where $x \to x, y \to log y$ On calculator: ExpReg(L,L_b) or LinReg (log L,log L_2) $y = ab^c$ where $x \to x, y \to log y$ On calculator: ExpReg(L,L_b) or LinReg (log L,log L_2) Transformed scatter must look linear Residual plot of TRANSFORMED no pattern starry night <b>Probability</b>		upper tail if > test and in double either tail if $\neq$ test) the p value Larger sample size $\Longrightarrow$ smaller p value If $x^2 calc = \sum \frac{(D-E)^2}{E}$ . Reject: $x^2 calc > x^2 critical$ HypothesisH, are independent/in the ratio/_ distributed $H_1$ ; are not independent/in the ratio/_ distributed $H_2$ ; are not independent/in the ratio/_ distributed $H_1$ ; are not independent/in the ratio/_ distributed $H_2$ ; are not independent/in the ratio/_ distributed $H_2$ . The transformed is the ratio of the <b>Definitions</b> . Type 1: $H_2$ true, but we say it is false i.e. reject it <u>Calculations</u> : Type 2: $H_3$ false, but we say rue is a cacept it <u>Calculations</u> : Type 2: $H_2$ is for bot of rejecting the test i.e. being in the critical region) Type 2: $H_2$ is a platerative method which is harder.
STATISTIC/ SAMPLE values, not population values, not values, not values Ask yourself first Is proportion or mean and then is it 1 or 2 samples Use 2 if $\sigma$ known OR $n \ge$ 00 Use 2 if $\sigma$ upgenue 09	$\begin{split} \mathbf{y}_{1}^{\mathbf{y}_{1}} & \mathbf{x}_{1}^{\mathbf{y}_{1}} & \mathbf{x}_{1}^{\mathbf{y}_{1}} \\ \mathbf{z}_{1}^{\mathbf{x}_{1}} = \mathbf{x}_{2} + \mathbf{z}_{c} \left( \frac{\beta_{1} q_{1}}{n_{1}} + \frac{\beta_{1} q_{2}}{n_{1}} + \frac{\beta_{1} q_{2}}{n_{2}} \right) \\ \mathbf{x}_{1}^{\mathbf{x}_{1}} = \mathbf{z}_{2} + \mathbf{z}_{c} \left( \frac{\beta_{1} q_{1}}{n_{1}} + \frac{\beta_{1} q_{2}}{n_{2}} \right) \\ \mathbf{x}_{1}^{\mathbf{x}_{2}} = \mathbf{z}_{1}^{\mathbf{x}_{2}} \\ \mathbf{x}_{2}^{\mathbf{x}_{1}} = \mathbf{z}_{1}^{\mathbf{x}_{1}} \mathbf{Q} \mathbf{x}_{1}^{\mathbf{x}_{1}} + \frac{\sigma_{1}}{q_{1}} \left( \mathbf{\sigma} \text{ unknown} / \text{ small sample} \right) \\ \mathbf{x}_{2}^{\mathbf{x}_{1}} = \mathbf{i} \text{ norm with area} \frac{1-q_{0}}{n_{2}}, \mu = 0, \sigma = 1 \\ \mathbf{x}_{c}^{\mathbf{x}_{1}} + \mathbf{n} \mathbf{x}_{2}^{\mathbf{x}_{1}} - \mathbf{x}_{2}^{\mathbf{x}_{1}} \mathbf{x}_{1} \mathbf{x}_{2}^{\mathbf{x}_{1}} \mathbf{x}_{1} \\ \mathbf{z}_{2}^{\mathbf{x}_{1}} \text{ invit with area} \frac{1-q_{0}}{n_{2}}, \mathbf{d} = \mathbf{n} - 1 \\ \mathbf{z}_{1}^{\mathbf{x}_{1}} = \mathbf{x}_{2}^{\mathbf{x}_{2}} \mathbf{z}_{2} \frac{q_{1}^{\mathbf{x}_{1}}}{n_{1} + n_{2}^{\mathbf{x}_{2}}} \mathbf{d} \mathbf{x}_{1} \mathbf{x}_{2} \mathbf{x}_{1}^{\mathbf{x}_{2}} \mathbf{x}_{1} \mathbf{x}_{2} \mathbf{x}_{1}^{\mathbf{x}_{2}} \mathbf{x}_{1} \mathbf{x}_{2} \mathbf{x}_{2} \mathbf{x}_{1} \mathbf{x}_{2} \mathbf{x}_{1} \mathbf{x}_{2} \mathbf{x}_{2} \mathbf{x}_{1} \mathbf{x}_{2} \mathbf{x}_{1} \mathbf{x}_{2} \mathbf{x}_{1} \mathbf{x}_{2} \mathbf{x}_{1} \mathbf{x}_{2} \mathbf{x}_{1} \mathbf{x}_{2} \mathbf{x}_{2} \mathbf{x}_{2} \mathbf{x}_{1} \mathbf{x}_{2} \mathbf{x}_{1} \mathbf{x}_{2} \mathbf{x}_{2} \mathbf{x}_{2} \mathbf{x}_{1} \mathbf{x}_{2} \mathbf{x}_{$	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A	variable) to the regression line i.e measure of variable) to the regression line i.e measure of x variable Scatter plot looks like a straight line High $r^2$ value Residual plot has no pattern with uniform variation across x (stary night) egression - Transformations $y = ax^5$ where $x \rightarrow logx$ , $y \rightarrow logy$ On calculator: Pwreg(L, log t, log (lg, log lg, l) $y = ab^5$ where $x \rightarrow x, y \rightarrow \log y$ On calculator: Pwreg(L, l, l) or timeg( log l, log lg, l) Transformed scatter must look linear Residual plot of TRANSFORMED no pattern stary night <b>Probability</b> $P(A) = \frac{n(A)}{n(A)} = \frac{number of favourable outcomes}{n(A)}$		upper tail if > test and in double either tail if $\neq$ test) the p value Larger sample size $\Longrightarrow$ smaller p value If $x^2_{callc} = \sum_{i=1}^{(G-E)^2}$ . Reject: $x^2_{callc} > x^2_{critical}$ HypothesixH <sub>i</sub> : are independent/in the ratio/ distributed $H_i$ : are not independent/in the ratio/ distributed $H_i$ : Type 1: $H_i$ true, but we say it is false i.e. reject it <b>Calculations:</b> Type 1: This is a (prob of rejecting the test i.e. being in the critical region). Type 2 = $\beta$ : Prob of NOI being in the critical region-see blue on left or the method of ow to find this Note: There is an alternative method which is harder Only use this method if you're forced to
STATISTIC/ SAMPLE values, as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use 2 if $\sigma$ known OR $n \ge$ 30 Use 1 if $\sigma$ unknown OR n < 30	Tests, 1 Prop Z interval 2 prop: $\widehat{p_1} - \widehat{p_2} \pm x_c \int_{\overline{p_1}}^{\overline{p_1}} \frac{\overline{p_1} - \overline{p_2}}{\overline{p_1}} \frac{\overline{y_1}}{\overline{p_1}} + \frac{\overline{p_2}}{\overline{p_2}}$ where $\widehat{p_1} = \frac{\overline{x_1}}{\overline{n_1}}, \widehat{p_2} = \frac{\overline{x_2}}{n_2}$ Tests, 2 Prop Z interval 1 mean: $\overline{X} \pm Z_c \frac{\sigma_1}{\sigma_1} \otimes \overline{X} \pm \overline{C_c} \frac{\sigma_1}{\sigma_1}$ ( $\sigma$ unknown/ small sample) Tests, 2 OR T Interval $Z_c$ : invorm with area $\frac{1-y_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-y_0}{2}$ df = $n - 1$ 2 means: $\overline{X}_1 - \overline{X}_2 \pm Z_c \int_{\overline{\sigma_1}}^{\overline{\sigma_2}} \frac{\sigma_2}{\overline{\sigma_1}} + \frac{\sigma_2}{\sigma_2}$ ( $\overline{X}_1 - \overline{X}_2 \pm T_c S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ where $S_p = \sqrt{\frac{(n_1-y_2)^2 + (n_2-1)s_2^2}{2}}$ (pooled formula) We pool when $n_1 \approx \sigma_2$ : $S_1 < 2S_2$ or $S_2 < 2S_1$ If don't pool use formula: $\overline{X}_1 - \overline{X}_2 \pm T_c \frac{s_1^2}{n_1^2} + \frac{s_2^2}{n_2}$ $z_c$ : invnorm with area $\frac{1-y_0}{2}, \mu = 0, \sigma = 1$	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events	variation in y variable for a given amount of x variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations $y = x^{0}$ where $x \rightarrow logx, y \rightarrow logy$ On calculator: Pwrkeg( $L_{1,L}$ ) or Linkeg (log $L_{1,log} L_{2}$ ) $y = ab^{2}$ where $x \rightarrow x, y \rightarrow log y$ On calculator: Pwrkeg( $L_{1,L}$ ) or Enkeg (log $L_{2}$ ) $y = ab^{2}$ where $x \rightarrow x, y \rightarrow log y$ On calculator: Pwrkeg( $L_{1,L}$ ) or Enkeg (log $L_{2}$ ) Transformed scatter must look linear Residual plot of TRANSFORMED no pattern starry night <b>Probability</b> $P(A) = \frac{n(A)}{n(D)} = \frac{number of favourable outcomes}{number of passible outcomes}$	$\label{eq:constraints} \begin{split} & \mbox{with } \mu \mbox{ and } \pi_0^{-1}\mbox{Cdf with df} \\ & \mbox{(see test statistics section} \\ & \mbox{above for of for Tcdf)} \\ \hline & \mbox{Chi- Squared} \\ \hline & \mbox{Errors} \\ \hline & \mbox{Errors} \\ & \mbox{Step 1: Find CV} (using invorm) \\ & \mbox{carea} = \alpha \ (left), u=0, 0, \sigma=1 \\ & \mbox{carea} = \alpha \ (left), u=0, 0, \sigma=1 \\ & \mbox{carea} = \alpha \ (left), u=0, \sigma=1 \\ & care$	upper tail if > test and in double either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2_{calc} = \sum \frac{(d-E)^2}{E}$ . Reject: $x^2_{calc} > x^2_{critical}$ HypothesisH <sub>a</sub> : are independent/in the ratio/ distributed H <sub>1</sub> : are not independent/in the ratio/ distributed H <sub>1</sub> : are not independent/in the ratio/ distributed H <sub>2</sub> : are not independent/in the ratio distributed H <sub>2</sub> : are not independent in the ratio distributed H <sub>2</sub> : Type 2: H <sub>0</sub> folds, but we say it is false i.e. reject it Type 2: H <sub>0</sub> false, but we say it is false i.e. reject it H <sub>2</sub> : Type 2: Firob of NOI being in the critical region-see blue on left for the method of wo find this Note: There is an alternative method which is harder Only use this method if you're forced to. Re-arrange $z = \frac{x + \mu}{2}$ tog et $x = \mu + z_{c} \frac{\pi}{2}$
STATISTIC/ SAMPLE values as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use 2 if or uknown OR $n \ge$ 30 Use 7 if or uknown OR n < 30 AP Stats:	$\begin{split} \mathbf{v}_{n} & \mathbf{v}_{n} \\ \text{Tests, 1 Prop Z interval} \\ \bullet & 2 \text{ prop: } \widehat{p_{1}} - \widehat{p_{2}} \pm z_{<} \left( \frac{\widehat{p_{1}} \widehat{q_{1}}}{n_{1}} + \frac{\widehat{p_{2}} \widehat{q_{2}}}{n_{2}} \text{ where } \widehat{p_{1}} = \frac{x_{1}}{n_{1}} , \widehat{p_{2}} = \frac{x_{2}}{n_{2}} \\ \text{Tests, 2 Prop Z interval} \\ \bullet & 1 \text{ mean: } \overline{X} \pm Z_{<} \frac{\sigma_{2}}{\sigma_{2}} \text{ OR } \overline{X} \pm T_{<} \frac{\sigma_{1}}{\sigma_{1}} (\sigma \text{ unknown} \text{ small sample}) \\ \text{Tests, 2 OR T interval} \\ z_{<} : \text{invorm with area} \frac{1-q_{0}}{n_{2}},  d = n - 1 \\ \hline T_{<} : \text{ inv T with area} \frac{1-q_{0}}{n_{1}},  d = n - 1 \\ \bullet & 2 \text{ means: } \overline{X}_{1} - \overline{X}_{2} \pm T_{<} \sqrt{\frac{q_{2}^{2}}{n_{1}} + \frac{q_{2}^{2}}{n_{2}^{2}}} / \overline{X}_{1} - \overline{X}_{2} \pm T_{<} S_{p} \sqrt{\frac{1}{n_{1}} + \frac{1}{n_{2}}} \\ \text{ where } S_{p} = \sqrt{\frac{(n_{1}-1)n^{2}+(n_{2}-1)n^{2}}{(n_{1}+n_{2}-2)}} \text{ (pooled formula)} \\ \text{ We pool when } \sigma_{1} = \sigma_{2} : S_{1} < 2S_{2} \text{ or } S_{2} < 2S_{1} \\ \text{ if don't pool use formula: } \overline{X}_{1} - \overline{X}_{2} \pm T_{<} \frac{s_{1}^{2}}{n_{1}^{2}} + \frac{q_{2}^{2}}{n_{1}^{2}} \\ z_{<} : \text{(invnorm with area} \frac{1-q_{0}}{2},  d = n - 1 \\ \overline{T}_{<} : : \text{ inv T with area} \frac{1-q_{0}}{2},  d = n - 1 \\ \end{array}$	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule)	variable) to the regression line i.e measure of variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations $y = ab^2$ where $x \to log x, y \to log y$ On calculator: PumReg(L, L, ) or LinReg (log L, log L,) $y = ab^2$ where $x \to x, y \to log y$ On calculator: PumReg(L, L, ) or LinReg (log L, log L,) $y = ab^2$ where $x \to x, y \to log y$ On calculator: Exped(L, L, ) or transformed scatter must look linear Residual plot of TRANSFORMED no pattern starry night <b>Probability</b> $P(A) = \frac{n(L)}{n(U)} = \frac{muther of favourable outcomes}{number of possible outcomes}$ P(A')=1-P(A) i.e. probabilities add to 1 $P(AUB)=P(A)+P(B)-P(A \cap B)$	$\label{eq:constraints} \begin{split} & \mbox{with } \mu \mbox{ and } \pi_{\sqrt{n}}^{0} \sigma^{-} \mbox{Tcdf with } df \\ & \mbox{ (see test statistics section } \\ & \mbox{ above for } for \mbox{Tcdf}) \\ \hline & \mbox{ Chi- Squared } \\ \hline & \mbox{ Errors } \\ \hline & \mbox{ Frors } \\ \hline & \mbox{ Type 2 Error Steps: } \\ & \mbox{ Step 1: Find CV (usy in (snorm)) } \\ & \mbox{ : area } = \alpha \ (left), u=0, \ \sigma=1 \\ & \mbox{ : area } = \alpha \ (left), u=0, \ \sigma=1 \\ & \mbox{ : area } = \alpha \ (left), u=0, \ \sigma=1 \\ & \mbox{ : area } = \alpha \ (left), u=0, \ \sigma=1 \\ & \mbox{ : area } = \alpha \ (left), u=0, \ \sigma=1 \\ & \mbox{ : area } = \alpha \ (left), u=0, \ \sigma=1 \\ & \mbox{ : area } = \alpha \ (left), u=0, \ \sigma=1 \\ & \mbox{ : area } = \alpha \ (left), u=0, \ \sigma=1 \\ & \mbox{ : area } = \alpha \ (left), u=0, \ \sigma=1 \\ & \mbox{ : step 2: Find error (norm cdf) } \\ & \mbox{ : covers - 100, upper-CV } \\ & \mbox{ : unper } u \ \sigma=2 \\ & \mbox{ : area } u $	upper tail if > test and in double either tail if $\neq$ test) the p value Larger sample size $\Longrightarrow$ smaller p value If $x^2 calc = \sum_{i=1}^{i} \frac{(g-E)^2}{n}$ . Reject: $x^2 calc > x^2 criticalHypothesis, t_{i=1}^{i=1} in dependent/in the ratio/$
STATISTIC/ SAMPLE values as their values, not population values Ask yourself first is proportion or mean and then is it 1 or 2 samples Use 2 if $\sigma$ known OR $n \ge$ 00 Use T if $\sigma$ unknown OR n < 30 AP Stats: $\sigma$ unknown always, so can	Tests, 1 Prop Z interval Tests, 1 Prop Z interval 2 <b>2 prop</b> : $\widehat{p_1} - \widehat{p_2} \pm x_c \int_{\overline{p_1}} \frac{ \widehat{p_1} + \widehat{p_2} + \widehat{p_2} + \widehat{p_1} + \widehat{p_2} - \widehat{p_2} + \widehat{p_2} + \widehat{p_1} + \widehat{p_2} - \widehat{p_2} + $	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events	variable) to the regression line i.e measure of variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r <sup>2</sup> value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations $y = ax^b$ where $x \rightarrow logx, y \rightarrow log y$ On calculator: PwrReg(L,L, ) or timReg (log L, log L_2) $y = ab^c$ where $x \rightarrow x, y \rightarrow log y$ On calculator: PwrReg(L,L_2) or timReg (log L, log L_2) $y = ab^c$ where $x \rightarrow x, y \rightarrow log y$ On calculator: PwrReg(L,L_2) or timReg (log L, log L_2) Transformed scatter must look linear Residual plot of TRANSFORMED no pattern starry night <b>Probabily</b> $P(A) = \frac{n(A)}{n(D)} = \frac{number of favourable outcomes}{number of possible outcomes}$ P(A') = 1 - P(A) i.e. probabilities add to 1 $P(A \cup B) = 0$	with $\mu$ and $\frac{\sigma}{\sqrt{n}} \sigma^{-1}$ Cdf with df (see test statistics section above for df for Tcdf) Chi-Squared Errors Type 2 Error Steps: Step 1: Find CV (using invnorm) $c: area = \alpha$ (left), $u=0, \sigma=1$ $\Rightarrow: area = \alpha$ (left), $u=0, \sigma=1$ $\Rightarrow: area = \frac{\sigma}{2}$ (left), $u=0, \sigma=1$ $\Rightarrow: \frac{\sigma}{2}$ (left), $u=0, \sigma=1$ c: Lower-100, upper-CV $u=new \mu, \sigma=\frac{\sigma}{2}$	upper tail if > test and in double either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2_{callc} = \sum_{i=1}^{(D-E)^2}$ . Reject: $x^2_{callc} > x^2_{critical}$ HypothesisH <sub>i</sub> are independent/in the ratio/ distributed $H_{i}$ : are not independent in the ratio independent in the ratio only use this method of out to find this Note: There is an alternative method which is harder $H_{i}$ : are not independent in the ratio if $y$ are ratio to. $H_{i}$ : are not independent in the ratio independent in the rat
STATISTIC/ SAMPLE values, as their values, not population values Ask yourself first Is proportion or mean and then is it 1 or 2 samples Use 2 if $\sigma$ known OR $n \ge$ 30 Use 7 if $\sigma$ unknown OR n < 30 AP Stats: $\sigma$ unknown always, so can always, so c	Tests, 1 Prop Z interval Tests, 1 Prop Z interval 2 <b>prop</b> : $\overline{p_1} - \overline{p_2} \pm x_c \int_{\overline{p_1}}^{\overline{p_1}} \frac{\overline{p_1} - \overline{p_2}}{\overline{p_1}} \frac{\overline{w_1}}{\overline{m_1}} + \frac{\overline{p_2}}{\overline{p_2}}$ where $\overline{p_1} = \frac{\overline{w_1}}{\overline{m_1}}, \overline{p_2} = \frac{\overline{w_2}}{\overline{n_2}}$ Tests, 2 Prop Z interval 1 <b>mean</b> : $\overline{X} \pm Z_c \frac{\sigma_1}{\sigma_1} \otimes \overline{X} \pm \overline{C_c} \frac{\sigma_1}{\overline{m_1}} (\sigma \text{ unknown}/ \text{ small sample})$ Tests, 2 OR T Interval $z_c : \text{Invnorm with area} \frac{1-w_0}{2}, \mu = 0, \sigma = 1$ $T_c : \text{ invT with area} \frac{1-w_0}{2}, \overline{x_1} - \overline{x_2} \pm T_c S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ where $S_p = \sqrt{\frac{(n_1 + 1)^2 + (n_2 - 1)^2 + 2}{n_1 + n_2 - 2}}$ (pooled formula) We pool when $\sigma_1 \approx \sigma_2 : S_1 < 2S_2$ or $S_2 < 2S_1$ If don't pool use formula: $\overline{X_1} - \overline{X_2} \pm T_c \left(\frac{\overline{s_1}^2}{\overline{s_1}^2} + \frac{w_2^2}{\overline{s_2}}\right)$ $z_c : \text{Invnorm with area} \frac{1-w_0}{2}, \mu = 0, \sigma = 1$ $T_c: \text{ invT with area} \frac{1-w_0}{2}, d(\overline{m_1} + n_2 - 2) \text{ (no pool)}$ $\overline{dfmin(n_1-1, n_2-1)} \text{ (no pool)}$	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events	variation in y variable for a given amount of x variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value High r <sup>2</sup> value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations $y = x^{5}$ where $x \rightarrow logx, y \rightarrow logyOn calculator: Pwrleg(L, L) or Lindeg (log L, log L_)y = ab^{2}$ where $x \rightarrow x, y \rightarrow log yOn calculator: Expleg(L, L) or Expleg(L, log L_)Transformed scatter must look linearResidual plot of TRANSFORMED no pattern starry nightP(A) = \frac{n(A)}{n(D)} = \frac{number of favourable outcomes}{number of possible outcomes}P(A') = -P(A) L = P(A) L = P(A) B = P(A) + P(B) = P(A \cap B)P(A \cap B) = 0addition rule becomes: P(A(B) = P(A) + P(B))$	$\label{eq:constraints} \begin{split} & \mbox{with } \mu \mbox{ and } \pi_{0,\overline{m}} \sigma^{-} Tcdf \mbox{ with } df \\ & (se test statistics section \\ & \mbox{ above for } for Tcdf) \\ \hline & \mbox{ chi} \mbox{ Squared} \\ \hline \\ \hline & \mbox{ Frors } \\ \hline & \mbox{ Step 1: Find CV}(using invorm) \\ & \mbox{ carea} = \alpha \ (left), u=0, 0, \sigma=1 \\ & \mbox{ carea} = \alpha \ (left), u=0, \sigma=1 \\ & \mbox{ carea} = \alpha \ (left), u=0, \sigma=1 \\ & \mbox{ carea} = \alpha \ (left), u=0, \sigma=1 \\ & \mbox{ carea} = \alpha \ (left), u=0, \sigma=1 \\ & \mbox{ carea} = \alpha \ (left), u=0, \sigma=1 \\ & \mbox{ carea} = \alpha \ (left), u=0, \sigma=1 \\ & \mbox{ carea} = \alpha \ (left), u=0, \sigma=1 \\ & \mbox{ carea} = \alpha \ (left), u=0, \sigma=1 \\ & \mbox{ carea} = \alpha \ (left), u=0, \sigma=1 \\ & \mbox{ carea} = \alpha \ (left), u=0, \sigma=1 \\ & \mbox{ carea} = \alpha \ (left), u=0, \sigma=1 \\ & \mbox{ carea} = \alpha \ (left), u=0, \sigma=1 \\ & \mbox{ carea} = \alpha \ (left), u=0, \sigma=1 \\ & \mbox{ carea} = \alpha \ (left), u=0, \sigma=1 \\ & \mbox{ carea} = \alpha \ (left), u=0, \sigma=1 \\ & \mbox{ carea} = \alpha \ (left), u=0, \sigma=1 \\ & \mbox{ carea} = \alpha \ (left), u=0, \sigma=1 \\ & \mbox{ carea} = \alpha \ (left), u=0, \sigma=1 \\ & \mbox{ carea} = \alpha \ (left), $	Upper tail if > test and in double either tail if $\neq$ test) the p value Larger sample size $\Longrightarrow$ smaller p value If $x^2_{calc} = \sum_{i=1}^{(O-E)^2}$ . Reject: $x^2_{calc} > x^2_{critical}$ HypothesisH <sub>i</sub> : are independent/in the ratio/distributed H <sub>i</sub> : are not independent/in the independent in the critical region. See J: Prob of NOT being in the critical region-see blue on left for the method of wo to find this Note: There is an alternative method which is harder. Only use this method if you're forced to. Re-arrange $z = \frac{T_{H}}{2\pi}$ to get $\overline{x} = \mu + z_c \frac{\sigma_H}{\sigma_H}$ $< P(\overline{x} < \mu + z_c \frac{\sigma_H}{\sigma_H})$ where $\mu$ -original mean, $z_c = CV$ Lower = -100, upper: $\mu + z_c \frac{\sigma_H}{\sigma_H}$ unew $\mu, \sigma \frac{\sigma_H}{\sigma_H}$
STATISTIC/ SAMPLE values as their values, not population values Ask yourself first: Is proportion or mean and then is it or 2 samples Use 2 if or unknown OR $n \ge$ 30 Use 7 if or unknown OR AP Stats: or unknown always, so can always, ure T for mean tests.	Tests, 1 Prop Z interval Tests, 1 Prop Z interval 2 <b>2 prop</b> : $\widehat{p_1} - \widehat{p_2} \pm z_c \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$ where $\widehat{p_1} = \frac{x_1}{n_1}, \widehat{p_2} = \frac{x_2}{n_2}$ Tests, 2 Prop Z interval 1 <b>1 mean</b> : $\overline{X} \pm \overline{z}, \frac{d}{q_1} \otimes \overline{X} \pm \overline{z}, \frac{d}{q_1} \otimes \overline{w}$ for unknown/small sample) Tests, Z OR T Interval 2 <b>c</b> : Invrtom with area $\frac{1-q_2}{n_2}, de = n - 1$ 7 <b>c</b> : InvT with area $\frac{1-q_2}{n_1}, \frac{de}{q_2}, \overline{X}_1 - \overline{X}_2 \pm \overline{z}_1, \overline{sp}, \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ where $\overline{s_p} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2^2}}$ (pooled formula) We pool when $\sigma_1 \approx \sigma_2: S_1 < 2S_2$ or $S_2 < 2S_1$ If don't pool use formula: $\overline{X}_1 - \overline{X}_2 \pm \overline{z}_1, \frac{s_1^2 + s_2^2}{n_1}, \frac{s_1^2 + s_2^2}{n_2}$ $z_c: Invnorm with area \frac{1-q_0}{2}, \mu = 0, \sigma = 1T_c: InvT with area \frac{1-q_0}{2}, \mu = 0, \sigma = 1T_c: InvT with area \frac{1-q_0}{2}, den_1 + n_2 - 2 (no pool)Tests, 2 Samp Z or T Interval2 2 means paired: \overline{d} \pm T_c; \frac{d}{q_1}$	To Check Whether A Good Model Linear Ro Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events	variation in y variable in wraitable in y variable in y variable in y variable for a given amount of x variable Scatter plot looks like a straight line High r value High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations y a ax <sup>2</sup> where x - logx, y - log y On calculator: Purkleg(L, L) or Linkeg (log, L, log L, log L) (L, log L, log L) - Calculator: Expleg(L, L, l) or Linkeg (log, L, log L, log L) - Transformed scatter must look linear Residual plot of TRANSFORMED to pattern starry night Probability P(A) = $\frac{n(A)}{n(B)} = \frac{number of favourable outcomes}{number of passible outcomes}$ P(A) = 0 A (A) + P(B) - P(A ∩ B) - P(A ∩ B) - 0 addition rule becomes: P(AU)B=P(A)+P(B) - P(A ∩ B) - 0 (A ∩ B) - P(A ∩ B) -	$\label{eq:constraints} \begin{split} & \mbox{with } \mu \mbox{ and } \pi_0^{-1} \mbox{ for Tcdf with df} \\ & \mbox{ (see test statistics section \\ \mbox{ above for of for Tcdf)} \\ \hline & \mbox{ Chi- Squared} \\ \hline & \mbox{ Frors } \\ \hline & \mbox{ Frors } \\ \hline & \mbox{ Type 2 Error Steps: } \\ & \mbox{ Step 1: Find CV (upt 0, 0, \sigma = 1) \\ & \mbox{ are a } \alpha \ (left), u=0, 0, \sigma = 1 \\ & \mbox{ are a } \alpha \ (left), u=0, 0, \sigma = 1 \\ & \mbox{ are a } \alpha \ (left), u=0, 0, \sigma = 1 \\ & \mbox{ are a } \alpha \ (left), u=0, 0, \sigma = 1 \\ & \mbox{ are a } \alpha \ (left), u=0, 0, \sigma = 1 \\ & \mbox{ are a } \alpha \ (left), u=0, 0, \sigma = 1 \\ & \mbox{ are a } \alpha \ (left), u=0, 0, \sigma = 1 \\ & \mbox{ step 2: Find error (normcdf) } \\ & \mbox{ clowers - 100, upper=CV } \\ & \mbox{ arewe } \mu, \sigma = \pi^{-1} \\ & \mbox{ step CV, upper=100 } \\ & \mbox{ arewe } \mu, \sigma = \pi^{-1} \\ & \mbox{ step CV, upper=CV } \\ & \mbox{ are erc V, upper=CV } \\ & \mbox{ are erc erc V, upper=CV } \\ &  are erc erc erc erc erc erc erc erc erc e$	Upper tail if > test and in double either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2_{calc} = \sum_{i=1}^{(\underline{0}-\underline{E})^2}$ . Reject: $x^2_{calc} > x^2_{critical}$ Hypothesis: $u_{i}$ are independent/in the ratio/_distributed dif( $x_{i} = 1$ ) if $x_{i}$ is independent of $x_{i}$ is a standard of the observation of the ob
STATISTIC/ SAMPLE values as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use Z if $\sigma$ known OR $n \ge 300$ Use Z if $\sigma$ unknown OR $n \ge 300$ Use T if $\sigma$ unknown OR n < 30 AP Stats: $\sigma$ unknown always, so can always, so can	Tests, 1 Prop Z interval Tests, 1 Prop Z interval 2 <b>prop</b> : $\widehat{p_1} - \widehat{p_2} \pm x_c \int_{\overline{p_1}}^{\overline{p_1} + \overline{p_2}} \sum_{\overline{p_2}}^{\overline{p_2}}$ where $\widehat{p_1} = \frac{x_1}{n_1}, \widehat{p_2} = \frac{x_2}{n_2}$ Tests, 2 Prop Z interval 1 <b>mean</b> : $\overline{X} \pm Z_c \frac{\sigma_1}{\sigma_1} \cos \overline{X} \pm T_c \frac{s_1}{\sigma_1}$ (or unnown/ small sample) Tests, 2 OR T Interval $z_c$ : Invnorm with area $\frac{1-\alpha_0}{p_2}, \mu = 0, \sigma = 1$ $T_{c^2}$ InvT with area $\frac{1-\alpha_0}{p_2}, \frac{d(n-n-1)}{d(n-n-1)}$ 2 <b>means</b> : $\overline{X_1} - \overline{X_2} \pm Z_c, \frac{\sigma_{n^2}^2 + \sigma_{n^2}^2}{n_1 + \frac{m_2}{p_2}}$ (pooled formula) We pool when $\sigma_1 \approx \sigma_2$ : $S_1 < 2S_2$ or $S_2 < 2S_1$ If don't pool use formula: $\overline{X_1} - \overline{X_2} \pm T_c \frac{S_1^2 + \frac{m_2}{p_2}}{n_2}$ $z_c$ : Invnorm with area $\frac{1-\alpha_0}{p_2}, d=n = 1$ $T_{c^2}$ : Inv1 with area $\frac{1-\alpha_0}{p_2}, d=n = 1$ $T_c$ : S 2 Samp Z or T interval 2 <b>means</b> paired: $d \pm T_c; \frac{z_1}{2}$ Tests, 2 ORT Interval. We usually use T	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Constituent	variable) to the regression line i.e measure of variable) to the regression line i.e measure of x variable Scatter plot looks like a straight line High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations $y = ab^x$ where $x \rightarrow logx, y \rightarrow log y$ On calculator: Fupfleg(L,L,L) or timfleg (log L,log L,) $y = ab^x$ where $x \rightarrow x, y \rightarrow log y$ On calculator: Fupfleg(L,L,L) or timfleg (log L,log L,) Transformed scatter must look linear Residual plot of TRANSFORMED no pattern starry night Probability $P(A) = \frac{\pi(A)}{\pi(D)} = \frac{number of favourable automes}{number of possible automes}$ P(A')=1-P(A) i.e. probabilities add to 1 $P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) P(B)$ Addition rule becomes: $P(A \cup B) = P(A) + P(B)$ addition rule becomes: $P(A \cup B) = P(A) + P(B)$	$\label{eq:constraints} \begin{split} & \mbox{with } \mu \mbox{ and } \pi_{\sqrt{n}}^{0} \mbox{or Cfd with } df \\ & \mbox{(see test statistics section } \\ & \mbox{ above for d ffor Tcdf)} \\ \hline & \mbox{Ch- Squared} \\ \hline & \mbox{Errors} \\ \hline & \mbox{Errors} \\ \hline & \mbox{Type 2 Error Steps:} \\ & \mbox{Step 1: Find CV (using invnorm)} \\ & \mbox{crass} = \alpha \ (left), u=0, \alpha=1 \\ & $x$: ares = \alpha$ (left), u=0, \alpha=1$ \\ & \mbox{$x$: ares$	upper tail if > test and in double either tail if ≠ test) the p value Larger sample size $\Longrightarrow$ smaller p value If $x^2 calc = \sum \frac{(D-E)^2}{2}$ . Reject: $x^2 calc > x^2 critical$ HypothesisH <sub>i</sub> are independent/in the ratio distributed <i>H</i> <sub>1</sub> : are not independent and n = 1 for others after n = 2 if approximating <i>p</i> or <i>H</i> / <i>p</i> but this newer comes up <b>Definitions</b> : <b>Type 1:</b> <i>H</i> <sub>0</sub> true, but we say it is false i.e. reject it <b>Calculations</b> : <b>Type 2:</b> <i>H</i> <sub>0</sub> Frob of NOT being in the critical region-see blue on left for the method of ow to find this Note: There is an alternative method which is harder. Only use this method if you're forced to. Re-arrange $z = \frac{T_{x}}{T_{x}}$ to get $\overline{x} = \mu + 2e_{x}\frac{\sigma}{m}$ $> P(\overline{x} > \mu + 2e_{x}\frac{\sigma}{m})$ where $\mu$ -original mean, $z_{x}$ =CV Lower= $\mu + 2e_{x}\frac{\sigma}{m}$ upper=100, unenew $\mu, \sigma = \frac{\sigma}{m}$ .
STATISTIC/ SAMPLE values, as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use 2 if $\sigma$ known OR $n \ge$ 30 Use T if $\sigma$ unknown OR n < 30 AP Stats: $\sigma$ unknown OR $\sigma$ = $\sigma_{2}$ , but	Tests, 1 Prop Z interval Tests, 1 Prop Z interval 2 <b>prop</b> : $\vec{p}_1 - \vec{p}_2 \pm x_c \int \frac{\vec{p}_1 \vec{q}_1 + \vec{p}_2 \vec{q}_2}{\vec{n}_1 + \vec{p}_2 \vec{q}_2}$ where $\vec{p}_1 = \frac{x_1}{n_1}, \vec{p}_2 = \frac{x_2}{n_2}$ Tests, 2 Prop Z interval 1 <b>mean</b> : $\vec{X} \pm Z_c \frac{\sigma_1}{\sigma_1} \otimes \vec{X} \pm T_c \frac{\sigma_1}{\sigma_1}$ ( $\vec{\sigma}$ unknown/ small sample) Tests, 2 OR T Interval $z_c$ : Invnorm with area $\frac{1-s_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-s_0}{2}$ and $\vec{q} = n-1$ 2 <b>means</b> : $\vec{X}_1 - \vec{X}_2 \pm Z_c \int \frac{\sigma_1^2}{\sigma_1} + \frac{\sigma_2}{\sigma_2} / \vec{X}_1 - \vec{X}_2 \pm T_c \cdot \vec{s}_p \int \frac{1}{n_1} + \frac{1}{n_2}$ where $s_p = \int \frac{(\sigma_1-1)s_2+(\sigma_2-1)s_2}{n_1+\sigma_2-2}$ (pooled formula) We pool when $\sigma_1 \approx \sigma_2$ : $S_1 < 2S_2$ or $S_2 < 2S_1$ If don't pool use formula: $\vec{X}_1 - \vec{X}_2 \pm T_c \int \frac{s_1^2}{s_1^2} + \frac{s_2^2}{s_2}$ $z_c$ : Invnorm with area $\frac{1-s_0}{\sigma_2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-s_0}{\sigma_1}, \mu = 0, \sigma = 1$ Tests, 2 OR T Interval. We usually use T $T_c$ : InVT with area $\frac{1-s_0}{\sigma_1}, d= n-1$	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B"	$\label{eq:response} \begin{array}{l} \mbox{variation in y variable} \mbox{variation in y variable} \mbox{variation in y variable} \mbox{variation in y variable} \mbox{Variable} V$		Upper tail if > test and in Goudie either tail if $\neq$ test) the p value Larger sample size $\Longrightarrow$ smaller p value If $x^2_{callc} = \sum_{i=1}^{i_{i}} Reject: x^2_{callc} > x^2_{critical}$ HypothesixH <sub>i</sub> : are independent/in the ratio/ distributed $H_i$ : are not independent/in the network of the independent Note: There is an alternative method which is harder Only use this method if you're forced to. Re-arrange $z = \frac{\pi_{i}}{\mu}$ to get $T = \mu + z_e \frac{\pi_{i}}{\pi}$ $<: P(\overline{X} < \mu + z_e \frac{\pi_{i}}{\pi})$ where $\mu$ -original mean, $z_e$ : CV Lower = -100, upper- $\mu + z_e \frac{\pi_{i}}{\pi}$ upper-100, u-new $\mu, \sigma - \frac{\pi_{i}}{\pi}$ $: P(\overline{X} > \mu + z_e \frac{\pi_{i}}{\pi} vaper-100, u-new \mu, m - \frac{\pi_{i}}{\pi}$ $\neq: P(\mu - z_e \frac{\pi_{i}}{\pi} < \mu + z_e - \frac{\pi_{i}}{\pi})$ where $\mu$ -original mean $x = P(\mu - z_e \frac{\pi_{i}}{\pi} < \mu + z_e \frac{\pi_{i}}{\pi})$ where $\mu$ -original mean
STATISTIC/ SAMPLE values, as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use 2 if $\sigma$ known OR $n \ge$ 30 Use 7 if $\sigma$ unknown OR $n \ge$ 30 Use 7 if $\sigma$ unknown OR $n \ge$ 30 d AP Stats: $\sigma$ unknown always, so can always, so can al	Tests, 1 Prop Z interval Tests, 1 Prop Z interval 2 prop: $\overline{p_1} - \overline{p_2} \pm x_c \sqrt{\frac{\beta_1 q_1}{n_1} + \frac{\beta_2 q_2}{n_2}}$ where $\overline{p_1} = \frac{x_1}{n_1}, \overline{p_2} = \frac{x_2}{n_2}$ Tests, 2 Prop Z interval 1 mean: $\overline{X} \pm Z_c \frac{\sigma_1}{q_1} \otimes \overline{X} \pm \overline{\zeta_c} \frac{\sigma_1}{q_1}$ (or unknown/ small sample) Tests, 2 OR T Interval $z_c$ : invrom with area $\frac{1-q_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-q_0}{2}, dE = n-1$ 2 means: $\overline{X}_1 - \overline{X}_2 \pm Z_c \sqrt{\frac{\sigma_1^2}{q_1^2} + \frac{\sigma_2}{q_2^2}}, (\overline{X}_1 - \overline{X}_2 \pm T_1, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}})$ where $s_p = \sqrt{\frac{(n_1-1)q_1^2 + (n_2-1)g_2^2}{q_1 + q_2^2}}$ (pooled formula) We pool when $\sigma_1 \approx \sigma_2$ : $S_1 < S_2$ or $S_2 < 2S_1$ If don't pool use formula: $\overline{X}_1 - \overline{X}_2 \pm T_c \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ $z_c$ : invnorm with area $\frac{1-q_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-q_0}{2}, \mu = 0, \sigma = 1$ Tests, 2 Samp Z or T interval 2 means paired: $d \pm T_c : \frac{d_1}{q_1}$ Tests, 2 OR T Interval. We usually use T $T_c$ : InvT with area $\frac{1-q_0}{2}, dE = n-1$	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B"	variation in y variable for a given amount of x variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations y = ax <sup>3</sup> where x → logx, y → log y On calculator: PwrReg(L, L) or LinReg (log L, log L) y = ab <sup>2</sup> where x → logx, y → log y On calculator: PwrReg(L, L) or LinReg (log L, log L) y = ab <sup>2</sup> where x → x, y → log y On calculator: PwrReg(L, L) or Subleg (L, log L) Transformed scatter must look linear Residual plot of TRANSFORMED no pattern starry night <b>Probability</b> P(A) = $\frac{n(D)}{n(D)} = \frac{number of favourable outcomes}{number of possible outcomes}$ P(A') = 1-P(A) i.e. Probabilities add to 1 P(AUB)=P(A)+P(B)-P(A ∩ B) P(A ∩ B) = 0 addition rule becomes: P(AUB)=P(A)+P(B) P(A ∩ B) = P(A)P(B) addition rule becomes: P(AUB)=P(A)+P(B)-P(A)P(B) P(AB) = $\frac{P(AB)}{P(B)}$ If independent: P(AB) = P(A)	$\label{eq:constraints} \begin{split} & \mbox{with } \mu \mbox{ and } \frac{\sigma}{\sqrt{n}} \sigma \mbox{Tcdf with } df \\ & \mbox{(see test statistics section } \\ & \mbox{above for } for \mbox{Tcdf} ) \\ \hline & \mbox{Chi-Squared} \\ \hline \\ $	Upper tail if > test and in Goudie either tail if ≠ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2_{calc} = \sum_{acc} \frac{(g-g)^2}{2\pi}$ Reject: $x^2_{calc} > x^2_{critical}$ Hypothesis. $y_{acc}$ is a generative standard to be a single standard tobs a single standard to be single standard to be sin
STATISTIC/ SAMPLE values as their values, not population values, not values, not values, not values, not samples use lif or unknown OR m $\geq$ 30 Use lif or unknown OR m $\geq$ 30 AP Stats: or unknown OR always, so can always, so can always was can alwas can always was can always was can alwas	Tests, 1 Prop Z interval Tests, 1 Prop Z interval 2 <b>prop</b> : $\widehat{p_1} - \widehat{p_2} \pm z_c \int_{\overline{p_1}}^{\overline{p_1}} \frac{F_{\overline{p_2}}}{r_{\overline{p_2}}}$ where $\widehat{p_1} = \frac{x_1}{n_1}, \widehat{p_2} = \frac{x_2}{n_2}$ Tests, 2 Prop Z interval 1 <b>mean</b> : $\overline{X} \pm Z_c \frac{\pi}{n_1} OR \overline{X} \pm T_c \frac{\pi}{n_1} (\sigma \text{ unknown}/\text{ small sample})$ Tests, 2 OR T Interval $z_c: \text{Invorm with area} \frac{1-96}{r_1}, \mu = 0, \sigma = 1$ $T_c: \text{ InvT with area} \frac{1-96}{r_1}, \frac{\pi}{n_1} - \overline{x_2} \pm T_c S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ where $S_p = \sqrt{\frac{(n_1-1)n^2 + (n_2-1)n^2}{n_1 + n_2-2}}$ (pooled formula) We pool when $\sigma_1 \approx \sigma_2: S_1 < S_2$ or $S_2 < S_1$ If don't pool use formula: $\overline{X}_1 - \overline{X}_2 \pm T_c \frac{s_1^2 + s_2^2}{n_1} + \frac{1}{n_2}$ $z_c: \text{Invnorm with area} \frac{1-96}{n_2}, \mu = 0, \sigma = 1$ $T_c: \text{ InvT with area} \frac{1-96}{n_2}, \sigma(\pi_1 + n_2 - 2) \text{ (no pool)}$ Tests, 2 Samp 2 or T Interval. We usually use T $T_c: \text{ InvT with area} \frac{1-96}{n_2}, \sigma(\pi_1 - n_2)$ Slope of regression line (2 possible formulae) $\circ \text{ slope}\pm T_c(s_b)$ where $s_p$ is standard error (SE)	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B" Bayes Theorem	variation in y variable for a given amount of x variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations y = ab <sup>2</sup> where x - logx, y - log y On calculator: PumReg(L, L, ) or LinReg (log L, log L,) y = ab <sup>2</sup> where x - t, y - log y On calculator: PumReg(L, L, ) or LinReg (log L, log L,) y = ab <sup>2</sup> where x - x, y - log y On calculator: Exped(L, L, ) or tinReg (L, log L,) Transformed scatter must look linear Residual plot of TRANSFORED to pattern starry night Probability P(A) = $\frac{m(A)}{m(D)} = \frac{muther of favourable outcomes}{muther of possible outcomes}$ P(A) = $1-P(A)$ Le, probabilities add to 1 P(A \DB) = P(A)+P(B)-P(A \DB) P(A \cap B) = 0 addition rule becomes: P(AUB)=P(A)+P(B) P(A \cap B) = P(A)P(B) P(A \cap B) = P(A)P(B) P(A \cap B) = P(A)P(B) P(A \cap B) = P(A)P(B)-P(A)P(B) P(A (B) = $\frac{P(A)B}{P(B)}$	with $\mu$ and $\frac{\sigma}{\sqrt{n}} \sigma$ Tcdf with df (see test statistics section above for of for Tcdf) <b>Chi-Squared</b> <b>Errors</b> Type 2 Error Steps: Step 1: Find CV (using invorm) $< area = \alpha$ (elch), $u=0, 0, \sigma=1$ $+ : area = \alpha$ (lech), $u=0, 0, \sigma=1$ $+ : area = \alpha$ (lech), $u=0, \sigma=1$ $= \frac{\sigma}{2}$ (lech), $u=0, \sigma=1$ $= 1$ (lech), $u=0, \sigma=1$ $= 1$ (lech), $u=0, \sigma=1$ $u=new \mu, \sigma=\frac{\sigma}{\sqrt{n}}$	Upper tail if > test and in Goudie either tail if $\neq$ test) the p value Larger sample size $\Longrightarrow$ smaller p value If $x^2 calc = \sum_{i=1}^{i} \frac{(g-E)^2}{\pi}$ . Reject: $x^2 calc > x^2 critical$ Hypothesix, are independent/in the ratiodistributed $H_i$ are not independent and n-1 for the method which is never $Type 1 : H_i$ true, but we say it is false i.e. reject it Calculations: Type 1: $H_i$ true, but we say it is false i.e. reject it Calculations: Type 2: $F_i$ for bod for Dieng in the critical region-see blue on left for the method of ow to find this Note: There is an alternative method which is harder. Only use this method if you're forced to. Re-arrange $z = \frac{F_{i}}{\pi}$ to get $\overline{x} = \mu + z_c \frac{\pi}{\eta}$ $z = P(\overline{x} < \mu + z_c \frac{\pi}{\eta})$ where $\mu = original mean, z_eCV$ Lower= $\mu + z_c \frac{\pi}{\eta}$ where $\mu = original mean, z_eCV$ Lower= $\mu + z_c \frac{\pi}{\eta}$ where $\mu = z_c \frac{\pi}{\eta}$ where $\mu = original mean$ Lower= $\mu - z_c \frac{\pi}{\eta}$ where $\mu = z_c \frac{\pi}{\eta}$ where $\mu = original mean$ Lower= $\mu - z_c \frac{\pi}{\eta}$ where $\mu = z_c \frac{\pi}{\eta}$ where $\mu = original mean$ Lower $\mu - z_c \frac{\pi}{\eta}$ where $\mu = z_c \frac{\pi}{\eta}$ where $\mu = original mean$ Lower $z = z_{\pi}^{-1}$ where $\mu = z_c \frac{\pi}{\eta}$ where $\mu = z_c \frac{\pi}{\eta}$ mean $\mu, \sigma = \frac{\pi}{\eta}$ To Increase/Decase the Errors:
STATISTIC/ SAMPLE values, as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use Z if $\sigma$ known OR $n \ge$ 30 Use T if $\sigma$ unknown OR n < 30 AP Stats: $\sigma$ unknown OR always, use can always, use can always, use can always, use ta for mean tests. We pool if T Test and $\sigma_1 \approx \sigma_2$ , but you will also notice that the mark scheme normally	Tests, 1 Prop Z interval Tests, 1 Prop Z interval 2 <b>prop</b> : $\vec{p}_1 - \vec{p}_2 \pm x_c \int_{\overline{p} \leq 0}^{\overline{p} \leq 0} \frac{ \vec{p}_1 \leq 1 - \vec{p}_2 \leq 0}{ \vec{n}_1 + \vec{n}_2 \leq 0}$ where $\hat{p}_1 = \frac{x_1}{n_1}, \hat{p}_2 = \frac{x_2}{n_2}$ Tests, 2 Prop Z interval 1 <b>mean</b> : $\vec{X} \pm \zeta_c = \frac{\sigma_1}{m_1} \otimes \vec{R} \pm T_c = \frac{\sigma_1}{m_1} (\sigma \text{ unknown}/ small sample)$ Tests, 2 OR T Interval $z_c$ : Invnorm with area $\frac{1-\omega_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-\omega_0}{2}, df = n-1$ 2 <b>means</b> : $\vec{X}_1 - \vec{X}_2 \pm Z_c \int_{\overline{\alpha}} \frac{\sigma_1^2}{n_1 + \sigma_2} \frac{2\sigma_2}{n_1 + \sigma_2} f(n - 1)$ We pool when $\sigma_1 \approx \sigma_2$ : $S_1 < S_2$ or $S_2 < 2S_1$ . If fort pool use formula: $\vec{X}_1 - \vec{X}_2 \pm T_c \int_{\overline{\alpha}} \frac{1}{n_1} + \frac{1}{n_2}$ $z_c$ : Invnorm with area $\frac{1-\omega_0}{n_1 + n_2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-\omega_0}{n_1}, df = n_1 + n_2$ $z_c$ : Invnorm with area $\frac{1-\omega_0}{n_1}, df = n_1 + n_2$ $z_c$ : Invnorm with area $\frac{1-\omega_0}{n_1 + n_2}, df = n_1 + n_2$ $z_c$ : Invnorm with area $\frac{1-\omega_0}{n_1}, df = n_1 + n_2$ . (so pool) Tests, 2 Samp 2 or T interval 2 <b>means paired</b> : $d \pm T_c; \frac{2d}{m}$ Tests, 2 OR T Interval. We usually use T $T_c$ : InvT with area $\frac{1-\omega_0}{n_2}, df = n - 1$ Siope of regression line (2 possible formulae) $\circ$ slope $\pm T_c(S_0)$ , where $s_0$ is standard error (SE) $\circ$ slope $\pm T_c$ ( $\frac{1-\sigma^2}{n_2}, \frac{\sigma}{m_1}$ where $s_0$ , $s_0$ are the SL 's	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B" Bayes Theorem	$\label{eq:started} \begin{array}{l} \text{variable} (b \ the regression line i.e measure of variation in y variable for a given amount of x variable. \\ \\ Scatter plot looks like a straight line High r value High r^2 value Residual plot has no pattern with uniform variation across x (starry night) \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\$		Upper tail if > test and in Goudie either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2 call \in \sum \frac{(G-E)^2}{2}$ . Reject: $x^2 call > x^2 critical$ HypothesisH <sub>4</sub> : are independent/in the ratio/distributed $H_1$ : are not independent/in the ratio/distributed $H_2$ : are not independent/in the ratio/distributed $H_1$ : are not independent/in the ratio/distributed $H_2$ : and $H_2$ is a factor of the ratio of the ratio $H_2$ : the ratio of rejecting the test i.e. being in the critical region) Type 2: H_3 (Forb of NOT being in the critical region-see blue on left for the method of out to find this Note: There is an alternative method which is harder. Only use this method for your forced to. Re-arrange $z = \frac{\pi_T}{T}$ to get $\overline{x} = \mu + z_e \frac{\pi_0}{\pi}$ $< P(\overline{x} < \mu + z_e \frac{\pi_0}{\pi})$ where $\mu$ -original mean, $z_e CV$ Lower = $-100$ . upper = $\mu + z_e \frac{\pi_0}{\pi}$ unew $\mu, \sigma - \frac{\pi}{\pi}$ $< P(\overline{x} > \mu + z_e \frac{\pi}{\pi})$ upper = 100, u=new $\mu, \sigma - \frac{\pi}{\pi}$ $\neq : P(\mu - z_e \frac{\pi}{\pi} < \overline{z} \approx \mu + z_e \frac{\pi}{\pi})$ unewe $\mu, \sigma - \frac{\pi}{\pi}$ $\neq : P(\mu - z_e \frac{\pi}{\pi} < \overline{z} \approx \mu + z_e \frac{\pi}{\pi})$ unewe $\mu, \sigma - \frac{\pi}{\pi}$ $= \frac{1}{10 \operatorname{tracess}} (Decrease Herrors)$ Prover: in minus prob of type $z = 1 - \beta$ Interess type: I increase is give. Changing sample size does nothing.
STATISTIC/ SAMPLE values, as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use 2 if $\sigma$ unknown OR $n \ge$ 30 Use 7 if $\sigma$ unknown OR <i>n</i> < 30 AP Stats: $\sigma$ unknown always, so can always use T for mean tests. We pool if T Test and $\sigma_1 \approx \sigma_2$ , but you will also normally doesn't pool	Tests, 1 Prop Z interval Tests, 1 Prop Z interval 2 prop: $\widehat{p_1} - \widehat{p_2} \pm x_c \left( \frac{\widehat{p_1} \widehat{q_1}}{n_1} + \frac{\widehat{p_2} \widehat{q_2}}{n_1} \right)$ where $\widehat{p_1} = \frac{x_1}{n_1}, \widehat{p_2} = \frac{x_2}{n_2}$ Tests, 2 Prop Z interval 1 mean: $\widehat{X} \pm Z_c \frac{\sigma_1}{\sigma_1} \otimes \widehat{X} \pm \overline{C_c} \frac{\pi}{n_1} (\sigma \text{ unknown}/\text{ small sample})$ Tests, 2 OR 1 Interval 2 means: $\widehat{X}_1 = \overline{X}_2 \pm Z_c \left( \frac{\sigma_1 \widehat{x}}{n_1} + \frac{\sigma_2}{n_2} \right) (\sigma + 1)$ Tests, 2 OR 1 Interval 2 means: $\widehat{X}_1 = \overline{X}_2 \pm Z_c \left( \frac{\sigma_1 \widehat{x}}{n_1} + \frac{\sigma_2}{n_2} \right) (X_1 - \overline{X}_2 \pm T_c S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}})$ where $S_p = \sqrt{\frac{(n_1 - 1)n_2 + (n_2 - 1)n_2}{n_1 + n_2 - 2}} (\text{pooled formula})$ We pool when $n_1 \approx \sigma_2$ : $S_1 < S_2$ or $S_2 < 2S_1$ If don't pool use formula: $\overline{X}_1 - \overline{X}_2 \pm T_c \left( \frac{s_1 \widehat{x}}{n_1} + \frac{s_2}{n_2} \right)$ $Z_c$ : Invnorm with area $\frac{1 - \omega_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1 - \omega_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1 - \omega_0}{2}, \sigma = 1$ Tests, 2 Samp Z or T interval 2 means paired: $d \pm T_c : \frac{d}{\sigma_1}$ Tests, 2 OR 1 Interval. We usually use T $T_c$ : InvT with area $\frac{1 - \omega_0}{2}$ of $n = -1$ Slope of regression line (2 possible formulae) $\circ$ slope $\pm T_c \left( \frac{(n_2 - 2)n_2 \widehat{x}}{(n_2 - 2)n_2 \widehat{x}} \right)$ where $s_x$ , $s_y$ are the S.D.'s $m_1 = \sigma_1 (-\frac{1}{2}, \frac{1}{n_1}, \frac{1}{n_2}, \frac{1}{n_2} \right)$	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B" Bayes Theorem Exclusive Events	variation in y variable for a given amount of x variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations y = ax <sup>b</sup> where x → logx, y → log y On calculator: Pwrleg(L, L, ) or Line(g (og L, log L, ) y = ab <sup>+</sup> where x → logx, y → log y On calculator: Pwrleg(L, L, ) or Line(g (og L, log L, ) Transformed scatter must look linear Residual plot of TRANSFORMED no pattern starry night <b>PrOAbility</b> P(A) = $\frac{n(A)}{n(D)} = \frac{number of favourable outcomes}{mumber of possible outcomes}$ P(A) = $-P(A) = -P(A) = 0$ addition rule becomes: P(AUB)=P(A)+P(B) P(A ∩ B) = P(A)P(B) addition rule becomes: P(AUB)=P(A)+P(B)-P(A)P(B) P(A ∩ B) = P(A)P(B) f independent: P(AB) = P(A) P(A B) = $\frac{P(B A)P(A)}{P(B A)P(A)}$ P(A B) = $\frac{P(B A)P(A)}{P(B A)P(A)}$	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcdf with df (see test statistics section above for of for Tcdf) Chi- Squared Errors Type 2 Error Steps: Step 1: Find CV (using invnorm) < area a $\alpha$ (left), $\mu$ =0, $\sigma$ =1 >: area $\alpha$ (left), $\mu$ =0, $\sigma$ =1 Step 2: Find error (normcdf) < : lower=-100, upper=CV $u$ =new $\mu$ , $\sigma$ = $\frac{\sigma}{\alpha}$ >: Lower=CV, upper=CV $u$ =new $\mu$ , $\sigma$ = $\frac{\sigma}{\alpha}$	upper tail if > test and in Goudie either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value Iff $x_{calc}^2 = \sum_{\alpha, \alpha} \frac{(\alpha - E)^2}{2}$ . Reject: $x_{calc}^2 > x^2$ -critical Hypothesci H <sub>2</sub> : are in generated with the rais/_ distributed iff (x_{calc}) = \sum_{\alpha, \alpha} \frac{(\alpha - E)^2}{2}. Reject: $x_{calc}^2 > x^2$ -critical Hypothesci H <sub>2</sub> : are in generated with the rais/_ distributed dir(orw 1)(columns-1) for independence and $n-1$ for others dir $n - 21$ agrowinating p or $\mu/a$ but this never comes up Definitions: Type 1: H <sub>1</sub> true, but we say it is false i.e. reject it Type 2: H <sub>0</sub> folse, but we say it is false i.e. reject it Type 2: H <sub>0</sub> Food of NOT being in the critical region-see blue on left for the method of ow to find this Note: There is an alternative method which is harder. Only use this method if you're forced to. Re-arrange $z = \frac{E_{T_{calc}}^2}{\pi}$ to get $\overline{x} = \mu + z_{c} \frac{\pi}{2}$ $< P(\overline{x} < \mu + z_{c} \frac{\pi}{2})$ where $\mu$ -original mean, $z_{c}$ -CV Lower = $-100$ . upper = $\mu + z_{c} \frac{\pi}{2}$ turnew $\mu, \sigma - \frac{\pi}{2}$ $\Rightarrow P(\{\overline{x} > \mu + z_{c} \frac{\pi}{2})$ where $\mu$ -original mean, $z_{c}$ -CV Lower = $\mu + z_{c} \frac{\pi}{2}$ where $\mu - \sigma$ ignial mean $z_{c}$ . To Increase/Decrease The Errors To Increase/Decrease The Errors To Increase/Decrease The Trop $\mu = \beta$ sample due does nothing. Hencerse Type 2 deces [see] = $h = \beta$ Hencerse Type 2 deces [see] = $h = \beta$ sample due does nothing.
STATISTIC/ SAMPLE values, as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use 2 if or unknown OR $n \ge 30$ Use 1 if or unknown OR $n \ge 30$ Use 1 if or unknown OR $n \ge 30$ AP Stats: or unknown always, so can always, so can always	Tests, 1 Prop Z interval Tests, 1 Prop Z interval 2 <b>prop</b> : $\widehat{p_1} - \widehat{p_2} \pm x_c \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$ where $\widehat{p_1} = \frac{x_1}{n_1}, \widehat{p_2} = \frac{x_2}{n_2}$ Tests, 2 Prop Z interval 1 <b>mean</b> : $\overline{X} \pm Z_c \frac{\sigma_1}{\sigma_c} \otimes \overline{X} \pm \overline{\zeta}_c \frac{\sigma_1}{\sigma_c}$ (or unknown/small sample) Tests, Z OR T Interval $Z_c$ : Invrom with area $\frac{1-q_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-q_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-q_0}{2}, \overline{X} = -\overline{X} \pm \overline{Z}, \overline{S}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ where $S_p = \sqrt{\frac{(n_1-1)s_1^2 \pm (n_2-1)s_2^2}{n_1+n_2-2}}$ (pooled formula) We pool when $\sigma_1 \approx \sigma_2$ : $S_1 < S_2$ or $S_2 < 2S_1$ If don't pool use formula: $\overline{X} = \overline{X}_2 \pm \overline{T}, \frac{s_1^2}{n_1} + \frac{s_2}{n_2}$ $Z_c$ : Invnorm with area $\frac{1-q_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-q_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-q_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-q_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-q_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-q_0}{2}, \eta = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-q_0}{2}, \eta = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-q_0}{2}, \eta = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-q_0}{2}, \eta = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-q_0}{2}, \eta = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-q_0}{2}, \eta = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-q_0}{2}, \eta = 0, \sigma = 1$ $T_c$ : InvT (with area $\frac{1-q_0}{2}, \eta = 0, \sigma = 1$ $T_c$ : InvT (Where $S_2, S_2, S_2, S_2, S_2, S_2, S_2, S_2, $	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B" Bayes Theorem Events Completely Randomise Design (CRD):	$\label{eq:started} \begin{array}{l} \mbox{variation in y variable} \mbox{variation in y variable} \mbox{variable} $	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcdf with df (see test statistics section above for of for Tcdf) Chi- Squared Errors Type 2 Error Steps: Step 1: Find CV (using invorm) (: a rea = a (left), u=0, 0, \sigma=1 ): area = a (left), u=0, 0, \sigma=1 ): area = a (left), u=0, 0, \sigma=1 ): area = a (left), u=0, 0, \sigma=1 (left), u=0, 0, d=1 $\frac{\sigma}{2}$ (left), u=0, 0, d=1 $\frac{\sigma}{2}$ (left), u=0, 0, d=1 (left), u=0, 0, d=1 ): area = a (left), u=0, 0, d=1 (left), u=0, 0, d=1 ): area = a (left), u=0, 0, d=1 (left), u=0, 0, d=1 ): area = a (left), u=0, 0, d=1 (left), u=0, 0, d=1 ): area = a (left), u=0, 0, d=1 (left), u=0, 0, d=1 (left), u=0, 0, d=1 ): area = a (left), u=0, 0, d=1 (left), u=0, d=1 (left),	upper tail if > test and in Goudie either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value if $x^2_{calc} = \sum_{i=1}^{i} \frac{(Q-E)^2}{\pi}$ . Reject: $x^2_{calc} > x^2_{critical}$ Hypothesis, $t_{i=1}^{i}$ are independent in the ratio_i distributed dif $x^2_{calc} = \sum_{i=1}^{i} \frac{(Q-E)^2}{\pi}$ . Reject: $x^2_{calc} > x^2_{critical}$ Hypothesis, $t_{i=1}^{i}$ are independent in the ratio_i. distributed dif $x^2_{calc} = \sum_{i=1}^{i} \frac{(Q-E)^2}{\pi}$ . Reject: $x^2_{calc} > x^2_{critical}$ Hypothesis, $t_{i=1}^{i}$ are independent in the ratio_i. distributed dif $x^2_{calc} = \sum_{i=1}^{i} \frac{(Q-E)^2}{\pi}$ . The value is a coper to others dif $x = -2$ if approximating $p$ or $p/s$ but is never comes up <b>Definitions</b> : <b>Type 1:</b> $H_i$ true, but we say it is faile i.e. reject it <b>Calculations</b> : <b>Type 1:</b> $H_i$ true, but we say it is faile i.e. reject it <b>Calculations</b> : <b>Type 2:</b> $R_i$ Frob of NOT being in the critical region-see <b>blue on left for the method</b> of ow to find this Note: There is an alternative method which is harder. Only use this method if you're forced to. Re-arrange $z = \frac{T_{eac}}{\pi}$ to get $\overline{x} = \mu + z_e \frac{\pi}{m}$ $z = P(\overline{x} < \mu + z_e \frac{\pi}{m})$ where $\mu$ -original mean, $z_eCV$ Lower $= 1-100$ , upper $\mu + z_e \frac{\pi}{m}$ unew $\mu, \pi \frac{\pi}{m}$ $\Rightarrow P(\overline{x} > \mu + z_e \frac{\pi}{m})$ upper $-100$ , unence $\mu, \mu = \sigma \frac{\pi}{m}$ $\Rightarrow P(\overline{x} > \mu + z_e \frac{\pi}{m})$ upper $-100$ , unence $\mu, \sigma \frac{\pi}{m}$ <b>To Increase/Decrease The Errors</b> : <b>To Increase type 1:</b> increase is gived. Charging sample size does rothing Increase Type 2: increase g is evel. Charging sample size does rothing Increase Type 2: increase g is
STATISTIC/ SAMPLE values as their values, not population values, not askyourself first is proportion or mean and then is it 1 or 2 samples Use 2 if $\sigma$ known OR $n \ge 300$ Use T if $\sigma$ unknown OR $n \ge 300$ Use	Tests, 1 Prop Z interval Tests, 1 Prop Z interval 2 <b>prop</b> : $\widehat{p_1} - \widehat{p_2} \pm x_c \int_{\overline{p_1}}^{\overline{p_1} + \overline{p_2}} \frac{\overline{p_1}}{\overline{p_1}} \int_{\overline{p_2}}^{\overline{p_2}} \frac{\overline{p_1}}{\overline{p_2}}$ where $\widehat{p_1} = \frac{x_1}{n_1}, \widehat{p_2} = \frac{x_2}{n_2}$ Tests, 2 Prop Z interval 1 <b>mean</b> : $\overline{X} \pm Z_c \frac{\sigma_1}{m_1} \otimes \overline{R} \pm T_c \frac{\sigma_1}{m_1}$ (for unknown/ small sample) Tests, 2 OR 1 Interval $z_c$ : Invnorm with area $\frac{1-\omega_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-\omega_0}{2}, \frac{\sigma_1}{m_1} + \frac{\sigma_2}{m_2}$ ( $\overline{X}_1 - \overline{X}_2 \pm T_c \frac{\sigma_1}{m_1} + \frac{1}{n_2}$ ) where $s_p = \sqrt{\frac{(m_1 - 1)^2 + (m_1 - m_2)^2}{m_1 + m_2}}$ (pooled formula) We pool when $\sigma_1 \approx \sigma_2$ : $S_1 < 2S_2$ or $S_2 < 2S_1$ If don't pool use formula: $\overline{X}_1 - \overline{X}_2 \pm T_c \frac{(s_1^2 + s_2^2)}{m_1 + m_2}$ $z_c$ : Invnorm with area $\frac{1-\omega_0}{2}, d= 0, d= 1$ $T_c$ : InvT with area $\frac{1-\omega_0}{2}, d= 0, d= 1$ $T_c$ : InvT with area $\frac{1-\omega_0}{2}, d= 0, d= 1$ Tests, 2 Samp 2 or T interval 2 <b>means paired</b> : $d \pm T_c; \frac{z}{m_1}$ Tests, 2 C RT 1 Interval. We usually use T $T_c: InvT$ with area $\frac{1-\omega_0}{2}, d= n-1$ Slope of regression line (2 possible formulae) $\circ$ slope $\pm T_c(S_0)$ where $s_p$ is standard error (SE) $\circ$ slope $\pm T_c(S_0)$ $M = n - 2$ Tests, LinkegTint (can only use if have raw data)	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B" Bayes Theorem Example Design (CRD): Research and	variable) to the regression line i.e measure of variation in y variable for a given amount of x variable. Scatter plot looks like a straight line High r value High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations $x = x^2$ where $x \rightarrow logx, y \rightarrow log y$ On calculator: Explex[l., or Explex[l., or Logic [l., or Config (ll., log. L.) $y = ab^2$ where $x \rightarrow x, y \rightarrow log y$ On calculator: Explex[l., or Explex[l., or Logic [l., or Logic [l., log. L.] Transformed scatter must look linear Residual plot of TRANSFORMED no pattern starry night Plotability P(A) = $\frac{p(A)}{n(U)} = \frac{number of favourable outcomes}{n(U)} P(A \cap B) = 0$ $P(A \cap B) = 0$ $P(A \cap B) = 0$ $P(A \cap B) = P(A) + P(B) - P(A \cap B)$ $P(A \cap B) = P(A) + P(B) - P(A) = P(A) + P(B) - P(A) = P(A) + P(B) - P(A) = P(A) = \frac{P(A)B}{P(B)} =$	with $\mu$ and $\frac{\sigma}{\sqrt{\pi}} \sigma$ Tcdf with df (see test statistics section above for of for Tcdf) Chi-Squared Errors Type 2 Error Steps: Step 1: Find CV (using invnorm) c: area = a (eltr), u=0, of = 1 >: area = a (eltr), u=0, of = 1 >: area = a (eltr), u=0, of = 1 = $\frac{\sigma}{2}$ (right), u=0, \sigma = 1 = $\frac{\sigma}{2}$ (right), u=0, \sigma = 1 Step 2: Find error (normcdf) c: Lower=-100, upper=CV u=new $\mu$ , $\sigma_{\pi}^{-1}$ $\neq$ : Lower=CV, upper=100 u=new $\mu$ , $\sigma_{\pi}^{-1}$	upper tail if > test and in Goudie either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2 calc = \sum \frac{(G-E)^2}{2}$ . Reject: $x^2 calc > x^2 critical$ HypothesisH, are independent/in the ratio distributed $H_1$ ; are not independent/in the ratio distributed $H_2$ ; are not independent/in the ratio distributed $H_1$ ; are not independent/in the ratio distributed $H_2$ ; are not independent in the ratio distributed $H_2$ ; and $H_2$ ratio is a distribute we is a ratio $H_2$ ; and $H_2$ ratio we ratio of ove to find this Note: There is an alternative method which is harder. Only use this method if you're forced to. Re-arrange $z = \frac{F_2}{F_2}$ to get $\overline{x} = \mu + 2e_{\sqrt{n}}^{\alpha}$ $> P(\overline{x} > \mu + 2e_{\sqrt{n}}^{\alpha})$ where $\mu$ -original mean, $z_e CV$ Lower= $\mu - 2e_{\sqrt{n}}^{\alpha}$ where $\mu$ -original mean, $z_e CV$ Lower= $\mu - 2e_{\sqrt{n}}^{\alpha}$ where $\mu$ = $e_{\sqrt{n}}^{\alpha}$ unewe $\mu, \sigma = \frac{\sigma}{\sqrt{n}}$ $To Increase V_2 excess given the Trons: Power: I minus proof type Z = 1 - \betaIncrease type 1: dis (a, type 1 error; take smallersamples, increase e_{\sqrt{n}} uneaver \mu, \sigma = \frac{\sigma}{\sqrt{n}}To Increase V_2 excess \mu reacess \mu reacess \mu.$
STATISTIC/ SAMPLE values, as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use 2 if $\sigma$ unknown OR $n \ge 30$ 00 Use T if $\sigma$ unknown OR $n \ge 310$ Use T if $\sigma$ unknown OR $n \ge 310$ 00 Use T if $\sigma$ unknown OR $n \ge 310$ $\Lambda P$ Stats: $\sigma$ unknown always, so can always, use T for mean tests. We pool if T Test and $\sigma_1 \approx \sigma_2$ , but you will also notice that the mark scheme normally doesn't pool regardless of size of $\sigma$	Tests, 1 Prop Z interval Tests, 1 Prop Z interval 2 <b>prop</b> : $\widehat{p_1} = \widehat{p_2} \pm x_c \left( \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2} \text{ where } \widehat{p_1} = \frac{x_1}{n_1}, \widehat{p_2} = \frac{x_2}{n_2} \right)$ Tests, 2 Prop Z interval 1 <b>mean</b> : $\widehat{X} \pm Z_c \frac{\sigma_1}{\sigma_1} \otimes \widehat{X} \pm T_c \frac{\pi}{\sigma_1} (\sigma \text{ unknown}/ \text{ small sample})$ Tests, 2 OR 1 Interval $z_c : \text{Invnorm with area} \frac{1-q_0}{p_2}, \mu = 0, \sigma = 1$ $T_c : \text{InvT with area} \frac{1-q_0}{p_2}, \alpha = 1, \alpha = 1$ $Z_c : \text{Invnorm with area} \frac{1-q_0}{p_2}, \alpha = 1, \alpha = 1$ (a) <b>2 means</b> : $\widehat{X}_1 = \overline{X}_2 \pm Z_c \left( \frac{\sigma_1 x_1}{\sigma_1 + \sigma_2}, \frac{\sigma_1 x_1}{\sigma_2}, \frac{\sigma_1 x_1}{\sigma_2}, \frac{\sigma_1 x_1}{\sigma_2}, \frac{\sigma_1 x_1}{\sigma_2}, \frac{\sigma_1 x_1}{\sigma_2}, \frac{\sigma_1 x_1}{\sigma_1 + \sigma_2}, \frac{\sigma_1 x_1}{\sigma_2}, \frac{\sigma_1 x_1}{\sigma_1 + \sigma_1 + \sigma_1 + \sigma_1 + \sigma_1 + \sigma_1 + \sigma_1 +$	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B" Bayes Theorem Excompletely Randomise Design (CRD):	variation in y variable hot the regression line i.e measure of variation in y variable for a given amount of x variable. Scatter plot looks like a straight line High r value High r <sup>2</sup> value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations y = ax <sup>b</sup> where x - logx, y - logy Oncalculator: Pwreg(L, L, ) or Linkeg (log, L, log L, ) Oncalculator: Pwreg(L, L, ) or Enkeg (log, L, log L, ) Oncalculator: Pwreg(L, L, ) or Enkeg (log, L, log L, ) Transformed scatter must look linear Residual plot of TRANSFORMED no pattern starry night Probability P(A) = $\frac{n(A)}{n(B)} = \frac{number of favourable outcomes}{number of possible outcomes}$ P(A) = $1 - P(A)$ i.e. probabilites add to 1 P(AA) = $1 - P(A)$ i.e. probabilites add to 1 P(AA) = $1 - P(A)$ i.e. Probabilites add to 1 P(AA) = $1 - P(A)$ i.e. Probabilites add to 1 P(AA) = $1 - P(A)$ i.e. P(AA) = $P(A)P(B)$ addition rule becomes: $P(AUB) = P(A) + P(B) - P(A)P(B)$ P(A(B) = $\frac{P(A B)}{P(B)} = \frac{P(A B)}{P(B)} = P(A)$ Each B(B) = $\frac{P(A B)}{P(B A)P(A)} + P(B A')P(A')$ geriment Templates Each subject receives only one treatment	with $\mu$ and $\frac{\sigma}{\sqrt{n}} \sigma$ Tcdf with df (see test statistics section above for of for Tcdf) Chi- Squared Errors Type 2 Error Steps: Step 1: Find CV (using invorom) <. area = $\alpha$ (left), $u=0, 0, \sigma=1$ $\Rightarrow$ : area = $\alpha$ (left), $u=0, 0, \sigma=1$ $\Rightarrow$ : area = $\alpha$ [left), $u=0, \sigma=1$ $\Rightarrow$ : area = $\alpha$ [left), $u=0, \sigma=1$ $\Rightarrow$ : lower=CV (using invorom) <. tower=.100, upper=CV $u=new \mu, \sigma=\frac{\sigma}{\sqrt{n}}$ $\Rightarrow$ : lower=CV1, upper=CV $u=new \mu, \sigma=\frac{\sigma}{\sqrt{n}}$ $\Rightarrow$ : lower=CV1, upper=CV2 $u=new \mu, \sigma=\frac{\sigma}{\sqrt{n}}$	upper tail if > test and in Goudie either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2 call \in \sum \frac{(G-E)^2}{2}$ . Reject: $x^2 call > x^2 criticalHypothesisH2 are independent/in the ratio/distributedH_1; are not independent/in the ratio/distributedH_1; are not independent/in the ratio/distributedH_2; are not independent/in the ratio/distributedH_1; are not independent/in the ratio/distributedH_2; are not independent/in the ratio/distributedDefinitions:Type 1: H0 true, but we say it is fable i.e. reject itCalculations:Type 2: This is (prob of rejecting the test i.e. being inthe critical region)Type 2 = B^2. For bod NOT being in the critical region-seeblue on left for the method of ow to fund thisNote: There is an alternative method which is harder.Only use this method if you're forced to.Re-arrange z = \frac{T_{TT}}{2} to get \overline{z} = \mu + z_{e_{TT}}^2(\overline{z} \ll \mu + z_{e_{TT}}^2) upper=100, u=new \mu, \sigma = \frac{\sigma}{\sigma_{TT}}>: P(\overline{x} > \mu + z_{e_{TT}}^2 upper=100, u=new \mu, \sigma = \frac{\sigma}{\sigma_{TT}}=: p(\mu - z_{e_{TT}}^2 \ll z) + z_{e_{TT}}^2 upper=100, u=new \mu, \sigma = \frac{\sigma}{\sigma_{TT}}=: p(\mu - z_{e_{TT}}^2 \propto z) + z_{e_{TT}}^2, upper=100, u=new \mu, \sigma = \frac{\sigma}{\sigma_{TT}}=: p(\mu - z_{e_{TT}}^2 \propto z) + z_{e_{TT}}^2, unew \mu, \sigma = \frac{\sigma}{\sigma_{TT}}=: p(\mu - z_{e_{TT}}^2 \propto z) + z_{e_{TT}}^2, unew \mu, \sigma = \frac{\sigma}{\sigma_{TT}}=: p(\mu - z_{e_{TT}}^2 \ll z) + z_{e_{TT}}^2, unew \mu, \sigma = \frac{\sigma}{\sigma_{TT}}=: p(\mu - z_{e_{TT}}^2 \approx z) + z_{e_{TT}}^2, unew \mu, \sigma = \frac{\sigma}{\sigma_{TT}}=: p(\mu - z_{e_{TT}}^2 \approx z) + z_{e_{TT}}^2, unew \mu, \sigma = \frac{\sigma}{\sigma_{TT}}=: p(\mu - z_{e_{TT}}^2 \approx z) + z_{e_{TT}}^2, if frequency, \overline{z} = \frac{2f_{TT}}{2}=: p(\mu - z_{e_{TT}}^2 = z) p(\mu - z_{e_{TT}}^2$
STATISTIC/ SAMPLE values, as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use 2 if $\sigma$ known OR $n \ge$ 30 Use 7 if $\sigma$ unknown OR n < 30 AP Stats: $\sigma$ unknown always, so can always,	Tests, 1 Prop Z interval Tests, 1 Prop Z interval 2 <b>prop</b> : $\widehat{p_1} - \widehat{p_2} \pm x_c \int_{\overline{p_1}}^{\overline{p_1}} \frac{\overline{p_1} - \overline{p_2}}{\overline{p_2}} \frac{\overline{x_1}}{\overline{p_1}} + \frac{\overline{p_2} - \overline{x_2}}{\overline{p_2}}$ Tests, 2 Prop Z interval 1 <b>i</b> mean: $\overline{X} \pm Z_c \frac{\sigma_1}{\sigma_1} \otimes \overline{X} \pm \overline{\zeta_c} \frac{\sigma_1}{\overline{p_1}} (\sigma \operatorname{unknown})$ small sample) Tests, 2 OR 1 Interval 2 <b>c</b> : invnorm with area $\frac{1-\omega_0}{\gamma_1} = 0, \sigma = 1$ $T_c$ : invT with area $\frac{1-\omega_0}{\gamma_1} = \frac{1}{2}$ ( $\overline{X}_1 - \overline{X}_2 \pm T_c \overline{sp} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ ) where $\overline{s_p} = \sqrt{\frac{(n_1-y)p_2^2 + (n_2-1)p_2^2}{\overline{p_1}}}$ (polded formula) We pool when $n_1 \approx \sigma_2$ : $S_1 < S_2$ , $\sigma_2 < 2S_1$ If don't pool use formula: $\overline{X}_1 - \overline{X}_2 \pm T_c \frac{s_1^2}{n_1^2} + \frac{s_2^2}{n_2}$ $z_c$ : invnorm with area $\frac{1-\omega_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-\omega_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-\omega_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-\omega_0}{2}, \sigma = 1$ $T_c$ : InvT with area $\frac{1-\omega_0}{2}, \sigma = 1$ $T_c$ : InvT interval 2 <b>means paired</b> : $d \pm T_c : \frac{d}{\sqrt{n}}$ Tests, 2 ORT 1 Interval 3 <b>Slope of regression</b> line (2 possible formulae) $\circ \operatorname{slope} \pm T_c (\frac{(n_1-y)p_2^2}{(n_1-y)p_2^2})$ where $s_x, s_y$ are the S.D.'s $T_c$ : InvT $(\frac{1-\sigma_1}{2}, df), df = n - 2$ Tests, InkergTint (can only use if have raw data) <b>8</b> <b>8</b> Random: SRS (independent) $N \ge 10n$ or $n \le .10n$ (pop size is 10 times samples size or samples zize is 10 tercent of non $\sigma = 2$ )	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B" Bayes Theorem Experimentations Experimenta	variation in y variable for a given amount of x variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations y = ax <sup>2</sup> where x → logx, y → log y On calculator: PwrReg(L, L) or LinReg (tog, L, log L) y = ab <sup>2</sup> where x → logx, y → log y On calculator: PwrReg(L, L) or LinReg (tog, L, log L) y = ab <sup>2</sup> where x → x, y → log y On calculator: PwrReg(L, L) or ExpReg (L), Log L) Transformed scatter must look linear Residual plot of TRANSFORMED no pattern starry night <b>Probability</b> P(A) = $\frac{m(A)}{m(B)} = \frac{minther of favourable outcomes}{minther of passible outcomes}$ P(A') = 1-P(A) i.e. probabilities add to 1 P(AUB)=P(A)+P(B)-P(A ∩ B) P(A ∩ B) = 0 addition rule becomes: P(AUB)=P(A)+P(B) P(A ∩ B) = P(A)P(B) if independent: P(AB) = P(A)P(B) P(A B) = $\frac{P(A B)}{P(B A)P(A)} + P(B A')P(A')$ periment Templates Each subject receives only one treatment	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcdf with df (see test statistics section above for of for Tcdf) Chi- Squared Errors Type 2 Error Steps: Step 1: Find CV(using invnorm) c: area = a (left), u=0, 0, ar1 >: area = a (left), u=0, 0, ar1 >: area = a (left), u=0, 0, ar1 : area = a (left), u=0, ar1 : area = a (left), u=0, ar1 : area	upper tail if > test and in double either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2_{calc} = \sum_{i=1}^{i} \frac{(Q-E)^2}{\pi}$ . Reject: $x^2_{calc} > x^2_{critical}$ Hypothesis, $u_{i=1}^{i}$ and explored in the table of the double of t
STATISTIC/ SAMPLE values, as their values, as their values, not population values Ask yourself first: Is proportion or mean and then is it l or 2 samples Use I if or unknown OR $n \ge 30$ Use I if or unknown OR $n \ge 30$ AP Stats: or unknown OR n < 30 AP Stats: or	Tests, 1 Prop Z Interval Tests, 1 Prop Z Interval 2 <b>prop</b> : $\vec{p}_1 - \vec{p}_2 \pm x_c \int_{\overline{p} + \overline{p}_1}^{\overline{p} + \overline{p}_2} \text{ where } \vec{p}_1 = \frac{x_1}{n_1} \cdot \vec{p}_2 = \frac{x_2}{n_2}$ Tests, 2 Prop Z Interval 1 <b>mean</b> : $\overline{X} \pm Z_c \frac{d}{m_1} \otimes R_c \overline{X} \pm T_c \frac{d}{m_1} (\sigma \text{ unknown}/\text{ small sample})$ Tests, 2 OR T Interval $z_c: \text{Invnorm with area} \frac{1-a_c}{p_c}, \mu = 0, \sigma = 1$ $T_c: \text{ InvT}$ with area $\frac{1-a_c}{p_c}, df = n - 1$ 2 <b>means</b> : $\overline{X}_1 - \overline{X}_2 \pm Z_c \int_{\overline{m} + \frac{1}{m_2}}^{2} (\overline{X}_1 - \overline{X}_2 \pm T, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}})$ where $s_p = \sqrt{\frac{(n_1 - 1)^2}{n_1 + n_2}}$ (pooled formula) We pool when $\sigma_1 \approx \sigma_2: S_1 < 2S_2 \text{ or } S_2 < 2S_1$ If don't pool use formula: $\overline{X}_1 - \overline{X}_2 \pm T, \frac{s_2}{p_1} + \frac{s_2}{n_2}$ $z_c: \text{Invnorm with area} \frac{1-a_c}{2}, df = n - 1$ 7 $T_c: \text{ InvT}$ with area $\frac{1-a_c}{2}, df = n - 2$ Tests, 2 Samp Z or T interval 2 <b>means</b> paired: $d \pm T_c: \frac{d}{2}$ Tests, 2 Samp Z or T interval 2 <b>means</b> paired: $d \pm T_c: \frac{d}{2}$ Tests, 2 C RT Interval 2 <b>means</b> paired: $d \pm T_c: \frac{d}{2}$ Tests, 2 C RT Interval 2 <b>means</b> paired: $d \pm T_c: \frac{d}{2}$ Tests, 2 Samp Z or T interval 3 <b>Sope</b> of regression line (2 possible formulae) 0 <b>slope</b> $T_c(S_0)$ where $s_p$ is standard error (SE) 0 <b>slope</b> $T_c(S_0)$ where $s_p$ is standard error (SE) 1 <b>state</b> $\frac{1-a_c}{2}, df + n - 2$ Tests, LinkegTint (can only use if have raw data) 8 <b>Random</b> : SRS <b>Independent</b> : $N \ge 10n \text{ or } n \le 1.0n$ (pop size is 10 times sample size or sample size is 10 percent of pop size) Normality:	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B" Bayes Theorem Experiment units Completely Randomise Design (CRD): Texperiment units	variation in y variable for a given amount of x variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations y a ax <sup>3</sup> where x - logx, y - log y On calculator: Purkeg(L, L) or Linkeg (log L, log L, y = ab <sup>2</sup> where x - logx, y - log y On calculator: Expleg(L, L) or Linkeg (log L, log L, y = ab <sup>2</sup> where x - x, y - log y On calculator: Expleg(L, L) or Linkeg (log L, log L, y = ab <sup>2</sup> where x - x, y - log y On calculator: Expleg(L, L) or Linkeg (log L, log L, y = ab <sup>2</sup> , where x - x, y - log y On calculator: Expleg(L, L) or Linkeg (log L, log L, y = ab <sup>2</sup> , where x - x, y - log y On calculator: Expleg(L, L) or Linkeg (log L, log L, y = ab <sup>2</sup> , where x - x, y - log y On calculator: Expleg(L, L) or Linkeg (log L, log L, y = ab <sup>2</sup> , where x - x, y - log y On calculator: Expleg(L, L) or Linkeg (log L, log L, l	$\label{eq:constraints} \begin{aligned} & \text{with } \mu \text{ and } \frac{\pi}{\sqrt{n}} \Theta^{-1} \text{Cdf with df} \\ & (\text{see test statistics section} \\ & \frac{1}{n} \text{ bore for } \text{for } \text{Tcdf}) \end{aligned}$	upper tail if > test and in double either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2 calc = \sum_{i=1}^{i} \frac{(g-E)^2}{\pi}$ . Reject: $x^2 calc > x^2 critical HypothesizHi are independent/in the ratio/ distributed H_i are not independent/in the ratio/ distributedH_i are not independent of H_i and H_i are notH_i are not independent of H_i and H_i are notH_i are not independent of H_i and H_i are notH_i are not independent of H_i are notH_i are not independent in H_i are notH_i increase H_i are not independent on H_i are notH_i increase H_i are not H_i are H_i are H_i are H_i are H_iH_i frequency: a^2 = \frac{\Sigma_i H_i}{\Sigma_i} are H_i are H_i are notH_i frequency: a^2 = \frac{\Sigma_i H_i}{\Sigma_i} are notH_$
STATISTIC/ SAMPLE values, as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use Z if $\sigma$ known OR $n \ge 30$ Use T if $\sigma$ unknown OR $n \ge 30$ Use T if $\sigma$ unknown OR $n \ge 30$ Use T if $\sigma$ unknown OR $n \ge 30$ Is a constant of the theory of theory of the theory of theory of the theory of the theory of the theory of theor	Tests, 1 Prop Z interval Tests, 1 Prop Z interval 2 <b>2 prop:</b> $\widehat{p}_1 - \widehat{p}_2 \pm x_c \int_{\overline{p}_1}^{\overline{p}_2 + \overline{p}_2} \frac{\overline{p}_3}{\overline{n}_1} + \frac{\overline{p}_2}{\overline{p}_2}$ where $\widehat{p}_1 = \frac{x_1}{n_1}, \widehat{p}_2 = \frac{x_2}{n_2}$ Tests, 2 Prop Z interval 1 <b>mean:</b> $\overline{X} \pm Z_c \frac{\sigma_1}{\sigma_1} \otimes \overline{X} \pm T_c \frac{\sigma_1}{\sigma_1}$ ( $\sigma$ unknown/ small sample) Tests, 2 OR 1 Interval $z_c$ : invnorm with area $\frac{1-\sigma_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-\sigma_0}{2}, \frac{\sigma_1}{\sigma_1} + \frac{\sigma_2}{\sigma_2}$ ( $\overline{x}_1 - \overline{X}_2 \pm T_c \overline{s}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ where $s_p = \sqrt{\frac{(n_1-1)s_2}{n_1+\alpha_2}}$ (pooled formula) We pool when $\sigma_1 \approx \sigma_2$ : $S_1 < 2S_2$ or $S_2 < 2S_1$ . If fort pool use formula: $\overline{X}_1 - \overline{X}_2 \pm T_c \sqrt{\frac{s_1^2}{n_1} + \frac{s_2}{n_2}}$ $z_c$ : invnorm with area $\frac{1-\sigma_0}{n_1+\alpha_2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-\sigma_0}{n_1}, dt = n - 2$ (no pool) Tests, 2 Samp 2 or T interval 2 <b>2 means paired</b> : $\frac{d}{d} \pm T_c \frac{d}{m}$ Tests, 2 OR T Interval. 2 <b>2 means paired</b> : $\frac{d}{d} \pm T_c \frac{d}{m}$ Tests, 2 OR T Interval 2 <b>3 isope 5 regression line</b> (2 possible formulae) $\circ$ slope $\pm T_c (S_0)$ . Where $s_1$ is standard error (SE) $\circ$ slope $\pm T_c (S_0)$ . Where $s_n$ sy are the S.D.'s $T_c$ : InvT ( $\frac{1-\sigma_0}{1+\sigma_1}, \frac{d}{d} = n - 2$ Tests, LinkegTint (can only use if have raw data) <b>Bandom</b> : SR <b>I indegendent</b> : $N \ge 10$ nor $n \le .10n$ (pop size is 10 times sample size or sample size is 10 percent of pop size) <b>Normality:</b> <b>Proportion: Mean (Z) Mean (T)</b>	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B" Bayes Theorem Events Completely Randomise Design (CRD):	variation in y variable for a given amount of x variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value High r <sup>2</sup> value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations y = $ax^{b}$ where $x \rightarrow logx, y \rightarrow logy$ On calculator: PwrReg(L, L) or LinReg (log L, log L) y = $ab^{t}$ where $x \rightarrow x, y \rightarrow log y$ On calculator: PwrReg(L, L) or ExpReg (L, log L) Transformed scatter must look linear Residual plot of TRANSFORMED no pattern starry night <b>Probability</b> $P(A) = \frac{n(A)}{n(D)} = \frac{number of favourable automes}{number of favourable automes}$ $P(A') = 1 - P(A) L = Drobabilite is add to 1P(AA D) = P(A) P(B) = P(A) P(B)P(A A D) = P(A)P(B) = P(A)P(B)P(A A D) = P(A)P(B) = P(A)P(B)P(A B) = P(A)P(B) = P(A)P(B)P(A B) = \frac{P(A B)}{P(B)} P(A) + P(B A')P(A')$ geriment Templates Each subject receives only one treatment	with $\mu$ and $\frac{\sigma}{\sqrt{n}} = 0^{-1}$ Cdf with df (see test statistics section above for of for Tcdf) Chi- Squared Errors Type 2 Error Steps: Step 1: Find CY (using innorm) $\exists$ : area = $\alpha$ (left), $u=0, 0$ = 1 $\Rightarrow$ : area = $\alpha$ (left), $u=0, 0$ = 1 $\Rightarrow$ : area = $\alpha$ (left), $u=0, 0$ = 1 $\Rightarrow$ : area = $\alpha$ (left), $u=0, 0$ = 1 $\Rightarrow$ : area = $\alpha$ (left), $u=0, 0$ = 1 $\Rightarrow$ : area = $\alpha$ (left), $u=0, 0$ = 1 $\Rightarrow$ : area = $\alpha$ (left), $u=0, 0$ = 1 $\Rightarrow$ : area = $\alpha$ (left), $u=0, 0$ = 1 $\Rightarrow$ : area = $\alpha$ (left), $u=0, 0$ = 1 $\Rightarrow$ : area = $\alpha$ (left), $u=0, 0$ = 1 $\Rightarrow$ : lower=CV.upper=CV $u=new \mu, \sigma = \frac{\sigma}{\sqrt{n}}$ $\Rightarrow$ : lower=CV1, upper=CV $u=new \mu, \sigma = \frac{\sigma}{\sqrt{n}}$ Mean Variance Note: can also use formula $\frac{s_{ex}}{n}$ Standard Dev	upper tail if > test and in Goudie either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2 calc = \sum \frac{(G-E)^2}{2\pi}$ . Reject: $x^2 calc > x^2 critical$ HypothesisH, are independent/in the ratio/distributed $H_i$ : are not independent/in the ratio/distributed $H_i$ : bis is c (prob of rejecting the test i.e. being in the critical region) $H_i$ : Do to: Here is an alternative method which is harder. Only use this method if you're forced to. Re-arrange $z = \frac{T_{in}}{2\pi}$ to get $\overline{x} = \mu + z_{in} \frac{\sigma}{\sigma_i}$ $> P(\overline{x} > \mu + z_i \frac{\sigma}{\sigma_i})$ where $\mu$ =original mean $z_i$ =CV Lower= $\mu + z_i \frac{\sigma}{\sigma_i}$ where $\mu$ =original mean $z_i$ =CV Lower= $\mu + z_i \frac{\sigma}{\sigma_i}$ where $\mu = z_i \frac{\sigma}{\sigma_i}$ unew $\mu, \sigma = \frac{\sigma}{\sigma_i}$ $T increase fyee: z_i decige let z_i et ye 1 = n\sigmaIncrease ye z: z_i decige let z_i et ye 1 = n\sigma,H_i no frequency: \overline{x} = \frac{Y_i \pi}{\sigma_i} if frequency: \overline{x} = \frac{Y_i \pi}{Y_i}If no frequency: \overline{x} = \frac{Y_i \pi}{\sigma_i} if frequency: \overline{x} = \frac{Y_i \pi}{Y_i}If no frequency: \overline{x} = \frac{Y_i \pi}{\sigma_i} if frequency: \overline{x} = \frac{Y_i \pi}{Y_i}f = \sqrt{variance}$
STATISTIC/ SAMPLE values, as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use 2 if $\sigma$ unknown OR $n \ge$ 30 Use 7 if $\sigma$ unknown OR $n \ge$ 30 Use 7 if $\sigma$ unknown OR $n \ge$ 30 AP 5 fats: $\sigma$ unknown always, so can always use T for mean tests. We pool if T Testand $\sigma_1 \approx \sigma_2$ , but you will also normally doesn't pool regardless of size of $\sigma$ Assumptions	Tests, 1 Prop Z interval Tests, 2 Prop Z interval 2 <b>prop</b> : $\vec{p}_1 - \vec{p}_2 \pm x_c \left( \frac{\vec{p}_1 \vec{q}_1}{n_1} + \frac{\vec{p}_2}{n_2} \text{ where } \vec{p}_1 = \frac{x_1}{n_1}, \vec{p}_2 = \frac{x_2}{n_2} \right)$ Tests, 2 Prop Z interval 1 <b>mean</b> : $\vec{X} \pm Z_c \frac{\sigma_1}{\sigma_1} \otimes \vec{X} \pm C_c \frac{\pi}{n_1} (\sigma \text{ unknown}/ small sample)$ Tests, 2 OR 1 Interval 2 <b>means</b> : $\vec{X}_1 = \overline{X}_2 \pm Z_c \left( \frac{\pi x_1}{n_1} + \frac{\pi x_2}{n_2} \right), \vec{x}_1 = \overline{X}_2 \pm T_c \cdot S_p \left( \frac{1}{n_1} + \frac{1}{n_2} \right)$ 2 <b>means</b> : $\vec{X}_1 = \overline{X}_2 \pm Z_c \left( \frac{\pi x_1}{n_1} + \frac{\pi x_2}{n_2} \right), \vec{X}_1 = \overline{X}_2 \pm T_c \cdot S_p \left( \frac{1}{n_1} + \frac{1}{n_2} \right)$ where $S_p = \sqrt{\frac{(n_1 + 1)n_2 + (n_2 - 1)n_2}{n_1 + n_2 - 2}}$ (pooled formula) We pool when $\sigma_1 \approx \sigma_2$ : $S_1 < S_2$ or $S_2 < 2S_1$ If don't pool use formula: $\vec{X}_1 = \overline{X}_2 \pm T_c \left( \frac{x_1}{n_1} + \frac{x_2}{n_2} \right)$ $Z_c$ : invorm with area $\frac{1 - m_2}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1 - m_2}{2}, \sigma(\pi_1 + n_2 - 2)$ (no pool) Tests, 2 Samp 2 or 1 Interval 2 <b>means paired</b> : $\vec{d} \pm T_c \cdot \frac{\pi}{\sigma_1}$ Tests, 2 OR 1 Interval 2 <b>means paired</b> : $\vec{d} \pm T_c \cdot \frac{\pi}{\sigma_1}$ Tests, 2 C R 1 Interval. We usually use T $T_c$ : InvT with area $\frac{1 - m_2}{2}, df = n - 1$ 5 <b>Slope of regression line</b> (2 possible formulae) $\circ$ <b>slope</b> $T_c ((\frac{1 - n^2 - 1)n_2}{(\frac{1 - n^2 - 1}{2}n_2})$ Tests, $T_c$ : InvT ( $\vec{t} = n - 2$ ] Tests, LinkegTint (con uniy use if have raw data) 8 <b>Random</b> : SR 1 <b>Independent</b> : $N \ge 10n \text{ on } n \le .10n$ (pop size is 10 times sample size or sample size is 10 percent of pop size) <b>Normality:</b> <b>Proprotion: Mean (2) Mean (1)</b> n < 30	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B" Bayes Theorem Exponential Completely Randomise Design (CRD):	variation in y variable for a given amount of x variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations y = ax <sup>3</sup> where x → logx, y → log y On calculator: Pwrleg(L, L, ) or Line(g (og L, log L, ) y = ab <sup>2</sup> where x → logx, y → log y On calculator: Pwrleg(L, L, ) or Line(g (og L, log L, ) y = ab <sup>2</sup> where x → x, y → log y On calculator: Pwrleg(L, L, ) or Line(log L, log L, ) Transformed scatter must look linear Residual plot of TRANSFORMED no pattern starry night <b>P</b> (A) = $\frac{m(M)}{n(M)} = \frac{mmher of favourable outcomes}{mmher of possible outcomes}$ <b>P</b> (A') = 1-P(A) i.e. probabilities add to 1 <b>P</b> (AUB)=P(A)+P(B)-P(A \cap B) <b>P</b> (A ∩ B) = 0 addition rule becomes: P(AUB)=P(A)+P(B) <b>P</b> (A(B)) = P(A)P(B) <b>P</b> (B A)P(A) <b>P</b> (B A)P(A') <b>P</b> (B	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcdf with df (see test statistics section above for of for Tcdf) Chi- Squared Errors Type 2 Error Steps: Step 1: Find CV (using invorm) < area = ar (left), $\mu$ =0, $\sigma$ =1 > area = ar (left), $\mu$ =0, $\sigma$ =1 + area = $\frac{\sigma}{2}$ (left), $\mu$ =0, $\sigma$ =1 Step 2: Find error (normcdf) < lowers-100, upper=CV $u=mew \mu$ , $\sigma = \frac{\sigma}{\sqrt{n}}$ >: Lower-CV, upper=ToV $u=mew \mu$ , $\sigma = \frac{\sigma}{\sqrt{n}}$ +: Lower-CV, upper=CV $u=mew \mu$ , $\sigma = \frac{\sigma}{\sqrt{n}}$ +: Lower-CV, upper=CV $u=mew \mu$ , $\sigma = \frac{\sigma}{\sqrt{n}}$	upper tail if > test and in double either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2_{calc} = \sum \frac{(0-E)^2}{2}$ . Reject: $x^2_{calc} > x^2_{critical}$ Hypothesis. If we is generated with the tail of distributed dif(cov)[(columns-1) for independence and $n-1$ for other diff $= n > 1$ approximating $p \ or \ \mu/a$ but this newer comes up <b>Definitions:</b> Type 1: $H_1$ true, but we say it is false i.e. reject it Type 2: $H_1$ false, but we say true i.e. accept it <b>Calculations:</b> Type 2: $H_2$ from 0 for length the test i.e. being in the critical region) Type 2: $H_2$ from 0 for Diening in the critical region-see blue on left for the method of ow to find this Note: There is an alternative method which is harder. Only use this method if you're forced to. Re-arrange $z = \frac{T_{eff}}{\pi}$ to get $\overline{x} = \mu + x_c \frac{\sigma}{\pi}$ $\approx P(\overline{X} > \mu + x_c \frac{\sigma}{\pi})$ where $\mu$ -original mean, $x_c = CV$ Lower = $\mu = 100$ , upper = $\mu + x_c \frac{\sigma}{\pi}$ unew $\mu, \sigma - \frac{\sigma}{\pi}$ $\approx P(\overline{X} > \mu + x_c \frac{\sigma}{\pi})$ where $\mu$ -original mean, $x_c = CV$ Lower = $\mu - x_c \frac{\sigma}{\pi}$ upper = 100, unewe $\mu, \sigma - \frac{\sigma}{\pi}$ $\approx P(\mu - x_c - \frac{\sigma}{\pi} < T > \mu + x_c \frac{\sigma}{\pi})$ unewe $\mu, \sigma - \frac{\sigma}{\pi}$ To Increase/Decrease the Errors: Power: in line tool for $y \ge 1 - \pi \beta$ , the smaller samples, increase $a_i$ line case $x_i$ if frequency: $x = \frac{Y_i x_i}{X^2}$ If no frequency: $\sigma^2 = \frac{Y_i x_i}{X^2} - x^2 = \frac{Y_i x_i - \pi^2}{X^2}$ If no frequency: $\sigma^2 = \frac{Y_i x_i}{X^2} - x^2 = \frac{Y_i x_i - \pi^2}{X^2}$ If no frequency: $\sigma^2 = \frac{Y_i x_i}{X^2} - x^2 = \frac{Y_i x_i - \pi^2}{X^2}$ If no frequency: $\sigma^2 = \frac{Y_i x_i}{X^2} - x^2 = \frac{Y_i x_i - \pi^2}{X^2}$
STATISTIC/ SAMPLE values, as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use 2 if o unknown OR $n \ge$ 30 Use 1 if o unknown OR $n \ge$ 30 Use 1 if o unknown OR $n \ge$ 30 AP Stats: o unknown always, so can always use T for mean tests. We pool if Y Test and $\sigma_1 \approx \sigma_2$ , but you will also notice that the mark scheme normally desn't pool regardless of size of o Assumptions	Tests, 1 Prop Z Interval Tests, 1 Prop Z Interval 2 <b>2</b> prop: $\vec{p}_1 - \vec{p}_2 \pm x_c \int_{\overline{p}_1}^{\overline{p}_2 + \overline{p}_2} where \hat{p}_1 = \frac{x_1}{n_1}, \hat{p}_2 = \frac{x_2}{n_2}Tests, 2 Prop Z Interval1 1 mean: \vec{X} \pm \zeta_c \frac{d}{m_1} \otimes \vec{R} + \vec{T}_c \frac{d}{m_1} (\sigma unknown/ small sample)Tests, 2 OR T Intervalz_c: Invnorm with area \frac{1-a_0}{p}, \mu = 0, \sigma = 1T_c: InvT with area \frac{1-a_0}{p}, \frac{d}{dt} = n - 12 means: \vec{X}_1 - \vec{X}_2 \pm \zeta_c \frac{d}{m_1} + \frac{a_0}{p_2} / \vec{X}_1 - \vec{X}_2 \pm T, s_0 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}where s_p = \sqrt{\frac{(n_1 - 1)^2 + (n_2 - 1)^2 + 2}{n_1 + n_2 + 2}} (pooled formula)We pool when \sigma_1 \approx \sigma_2: S_1 < 2S_2 or S_2 < 2S_1If don't pool use formula: \vec{X}_1 - \vec{X}_2 \pm T, \frac{s_0 + \frac{1}{n_1} + \frac{1}{n_2}}{n_2}z_c: Invnorm with area \frac{1-a_0}{2}, d=n + n_2 - 2 (no pool)\vec{dt} = \min(n_1 - 1, n_2 - 1) (no pool)Tests, 2 Samp Z or T interval2 means paired: \vec{d} \pm T_c, \frac{id}{m}Tests, 2 OR T Interval. We usually use TT_c: InvT with area \frac{1-a_0}{2}, d=n - 15 Slope of regression line (2 possible formulae)\circ slope \pm T_c (S_0) where s_n is standard error (SE)\circ slope \pm T_c (\frac{1-a^2}{p_0}), where s_n sy are the S.D.'sT_c: InvT (\frac{1-a_0}{m_0}, \frac{1}{m} - 2Tests, LinkegTInt (can only use if have raw data)1 Random: SR1 Independent: N \ge 10n \text{ or } n \le 10n (pop size is 10 timessample size or sample size is 10 times or n < 30pot polythouldn \ge 301 Normality:Nermality:Nermality:Nermality:Nermality$	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B" Bayes Theorem Completely Randomise Design (CRD):	variation in y variable for a given amount of x variation in y variable for a given amount of x variation in y variable for a given amount of x wariable Scatter plot looks like a straight line High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations y a car's where x - logx, y - log y On calculator: Purkleg(L, L) or inset (og L, log L) On calculator: Purkleg(L, L) or inset (og L, log L) Transformed scatter must look linear Residual plot of TRANSFORMED no pattern starry night P(A) = $\frac{n(A)}{n(B)} = \frac{number of favourable outcomes}{number of possible outcomes}$ P(A) = 1-P(A) i.e. probabilities add to 1 $P(AUB)=P(A)+P(B)-P(A \cap B)$ $P(A \cap B) = 0$ addition rule becomes: P(AUB)=P(A)+P(B)-P(A)P(B) P(A \cap B) = P(A)(B) = P(A)(B)/P(A)) $P(A B) = \frac{P(A B)}{P(B A)P(A)} + P(B A')P(A')$ If independent: $P(A B) = P(A)$ $P(A B) = \frac{P(A B A)P(A)}{P(B A)P(A)} + P(B A')P(A')$ periment Templates Each subject receives only one treatment $\frac{P(A B B)}{P(A B A)} = \frac{P(A B A)P(A)}{P(A B)} = \frac{P(A B A)P(A)}{P(A B)} = \frac{P(A B)}{P(A B)} = \frac{P(A B A)P(A)}{P(A B)} = \frac{P(A B)}{P(B A)P(A)} + P(B A')P(A')}$ periment Templates Each subject receives only one treatment $\frac{P(A B B)}{P(A B)} = \frac{P(A B A)}{P(A B)} = \frac{P(A B)}{P(A B)} = \frac{P(A B A)}{P(A B)} =$	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcdf with df (see test statistics section above for of for Tcdf) Chi- Squared Errors Type 2 Error Steps: Step 1: Find CV (usp (n, 0, 0)) $\approx$ area $\approx$ (left), $u=0, 0, 0=1$ $\approx$ area $\approx$ (left), $u=0, 0=1$ $\approx$ area (left), $u=0, 0=1$ $\approx$ area (left), $u=0, 0=1$ $\approx$ area (left), $u=0, 0=1$ $\approx$ area (left), $u=0, 0=1$ $\approx$	upper tail if > test and in double either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value if $x^2 calc = \sum_{i=1}^{i} \frac{(Q-E)^2}{n}$ . Reject: $x^2 calc > x^2 criticalHypothesis, U_i are independent in the ratio$
STATISTIC/ SAMPLE values, as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use Z if $\sigma$ known OR $n \ge 30$ Use T if $\sigma$ unknown OR $n \ge 30$ Use T if $\sigma$ unknown OR $n \ge 30$ Use T if $\sigma$ unknown OR $n \ge 30$ Is a T if $\sigma$ unknown OR $n \ge 30$ Is a C if $\sigma$ A Stats: $\sigma$ unknown OR $n \ge 30$ Is a C if $\sigma$ unknown OR $n \ge 30$ Is a C if $\sigma$ A Stats: $\sigma$ unknown OR $n \ge 30$ If $\sigma$ A Stats: $\sigma$ unknown OR $n \ge 30$ If $\sigma$ A Stats: $\sigma$ unknown OR $n \ge 30$ Is a C if $\sigma$ A Stats: $\sigma$ unknown OR $n \ge 30$ Is a C if $\sigma$ A Stats: $\sigma$ unknown OR $n \ge 30$ If $\sigma$ unknown O	Tests, 1 Prop Z interval Tests, 1 Prop Z interval 2 <b>prop</b> : $\vec{p}_1 - \vec{p}_2 \pm x_c \int_{\overline{p} \leq 0}^{\overline{p} \leq 0} \frac{ \vec{p}_1 \leq 1 - \vec{p}_2 \leq 0}{ \vec{n}_1 + \vec{n}_2 \leq 0}$ where $\vec{p}_1 = \frac{x_1}{n_1}, \vec{p}_2 = \frac{x_2}{n_2}$ Tests, 2 Prop Z interval 1 <b>mean</b> : $\vec{k} \pm \zeta_c = \frac{\sigma_1}{m_1} \otimes \vec{k} \pm \gamma_c = \frac{\sigma_1}{m_1} (\sigma \text{ unknown}/\text{ small sample})$ Tests, 2 OR 1 Interval $z_c$ : Invnorm with area $\frac{1-\omega_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-\omega_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-\omega_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-\omega_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with $\alpha = \frac{1-\omega_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with $\alpha = \frac{1-\omega_0}{n_1 + m_2}$ (pooled formula) We pool when $\sigma_1 \approx \sigma_2$ : $S_1 < 2S_2$ or $S_2 < 2S_1$ If fort pool use formula: $\overline{X}_1 - \overline{X}_2 \pm T, (\frac{s_1 + s_2}{n_1} + \frac{s_2}{n_2})$ $z_c$ : Invnorm with $\alpha = \frac{1-\omega_0}{2}, df = n, -1$ $T_c$ : InvT with $\alpha = \frac{1-\omega_0}{2}, df = n, -2$ (no pool) Tests, 2 Samp 2 or T interval 2 <b>Imeans paired</b> : $d \pm T_c; \frac{z_1}{m_1}$ Tests, 2 CRT Interval. We usually use T $T_c$ : InvT with $\alpha = \frac{1-\omega_0}{2}, df = n, -1$ 5 <b>Slope of regression line</b> (2 possible formulae) $\circ$ <b>slope</b> $\pm T_c(S_0)$ where $s_n$ is standard error (SE) $\circ$ <b>slope</b> $\pm T_c(S_0)$ where $s_n$ is standard error (SE) $\circ$ <b>slope</b> $\pm T_0$ ( $\sigma_1$ ) where $s_n$ is standard error (SE) $\circ$ <b>slope</b> $\pm T_0$ ( $\sigma_1$ ) where $s_n$ is 10 m r s = 10n r n < 10n (pop size is 10 times sample size or sample size is 10 percent of pop size) <b>Normality:</b> <b>Proprotion: Mean(Z)</b> $n \geq 30$ Note: Do for	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B" Bayes Theorem Events Completely Randomise Design (CRD):	variation in y variable for a given amount of x variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value High r <sup>2</sup> value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations y = ax <sup>8</sup> where x → logx, y → logy On calculator: PwrReg(L, L, ) or LinReg (log, L, log L, ) y = ab <sup>8</sup> where x → L, y → logy y On calculator: PwrReg(L, L, ) or LinReg (log, L, log L, ) Transformed scatter must look linear Residual plot of TRANSFORMED no pattern starry night Probability Probability Pr(A) = $\frac{n(L)}{n(L)} = \frac{number of favourable automes}{number of possible automes}$ P(A')=1-P(A) Le, Drobabilite s add to 1 P(AUB)=P(A)+P(B)-P(A \cap B) P(A \cap B) = 0 addition rule becomes: P(AUB)=P(A)+P(B) P(A ∩ B) = P(A)P(B) P(A ∩ B) = P(A)P(B) P(A ∩ B) = P(A)P(B) P(A B) = $\frac{P(A B)}{P(B)}$ If independent: P(A B) = P(A) P(A B) = $\frac{P(A B)}{P(B)}$ Each subject receives only one treatment if transmet) Measurement if t	with $\mu$ and $\frac{\sigma}{\sqrt{n}} = 0^{-1}$ Cdf with df (see test statistics section above for of for Tcdf) Chi- Squared Errors Type 2 Error Steps: Step 1: Find CY (using innorm) $\Rightarrow$ area $= \alpha$ (right), $u=0, 0=1$ $\Rightarrow$ : tower=CV, upper=CV $u=new \mu, \sigma = \frac{\pi}{\pi}$ = Mean	upper tail if > test and in Goudie either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2$ calls $\equiv \sum \frac{(G-E)^2}{2}$ . Reject: $x^2$ calls $\geq x^2$ critical HypothesisH, are independent/in the ratio/
STATISTIC/ SAMPLE values, as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use 2 if $\sigma$ unknown OR $n \ge 30$ Use 7 if $\sigma$ unknown OR $n \ge 30$ Use 7 if $\sigma$ unknown OR $n \ge 30$ a 30 AP Stats: $\sigma$ unknown always, so can always use T for mean tests. We pool if T Test and $\sigma_1 \approx \sigma_{2,b}$ but you will also notice that the mark scheme normally deen't pool regardless of size of $\sigma$ Assumptions	$\begin{split} \mathbf{v}_{1} & \mathbf{v}_{1} & \mathbf{v}_{1} & \mathbf{v}_{1} \\ \text{Tests, 1 Prop Z Interval} \\ \mathbf{2 \ prop:} & \widehat{p_{1}} = \widehat{p_{2}} \pm \chi_{c} \left( \frac{\beta_{1} q_{1}}{m_{1}} + \frac{\beta_{2} q_{2}}{m_{1}} \right) \\ \text{Tests, 2 Prop Z Interval} \\ \mathbf{1 \ mean:} & \widehat{X} \pm Z_{c} & \frac{\sigma_{1}}{m_{0}} \otimes \widehat{X} \pm T_{c} & \frac{\sigma_{1}}{m_{1}} (\sigma unknown/small sample) \\ \text{Tests, 2 OR 1 Interval} \\ \mathbf{2 \ c: lnvnorm with area} \frac{1-q_{0}}{m_{1}}, \mu = 0, \sigma = 1 \\ & T_{c}: \ln V \text{ with area} \frac{1-q_{0}}{m_{1}}, \frac{\sigma_{1}}{m_{2}} = X_{2} \pm T_{c} \left( \frac{\sigma_{1}}{m_{1}} + \frac{\sigma_{2}}{m_{2}} \right) \\ Means in the same same same same same same same sam$	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B" Bayes Theorem Experimental Experi	variation in y variable for a given amount of x variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations y = ax <sup>3</sup> where x + logx, y + logy On calculator: Pwring(L, L, ) or Line(tog) y = ab <sup>2</sup> where x + logx, y + logy On calculator: Pwring(L, L, ) or Line(tog) y = ab <sup>2</sup> where x + x, y - log y On calculator: Pwring(L, L, ) or Line(tog) y = ab <sup>2</sup> where x + x, y - log y On calculator: Pwring(L, L, ) or Line(tog) y = ab <sup>2</sup> where x + x, y - log y On calculator: Pwring(L, L, ) or Line(tog) P(A) = <sup>m(A)</sup> Transformed scatter must took linear Residual plot of TRANSFORMED no pattern starry night <b>Probability</b> P(A) = <sup>m(A)</sup> P(A) = <sup>m(A)</sup> P(A) = <sup>m(A)</sup> P(A) = 0 addition rule becomes: P(AUB)=P(A)+P(B)= P(A) = P(A) P(B) = 0 addition rule becomes: P(AUB)=P(A)+P(B)=P(A)+P(B) P(A) B) = P(A)P(B) P(A) B) = P(A)P(B) P(A B) = <sup>m(A)P(A)</sup> P(B A)P(A) + P(B A)P(A) P(A B) = <sup>m(A)P(A)</sup> P(B A)P(A) + P(B A)P(A') promisent Templates Each subject receives only one treatment <sup>m(A)</sup> Finance (Manus ruppen) Finance (Manus ruppen)	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcdf with df (see test statistics section above for of for Tcdf) Chi- Squared Errors Type 2 Error Steps: Step 1: Find CV (using invorm) < area at (left), $\mu$ =0, $\sigma$ =1 >: area at (left), $\mu$ =0, $\sigma$ =1 >: area at (left), $\mu$ =0, $\sigma$ =1 =: area = $\frac{\sigma}{2}$ (left), $\mu$ =0, $\sigma$ =1 Step 2: Find error (normcdf) < lowers-100, upper=CV uenew $\mu$ , $\sigma$ = $\frac{\sigma}{\sqrt{n}}$ >: Lower-CV, upper=T00 uenew $\mu$ , $\sigma$ = $\frac{\sigma}{\sqrt{n}}$ #: Lower-CV, upper=CV uenew $\mu$ , $\sigma$ = $\frac{\sigma}{\sqrt{n}}$ Mean Mean Mean Mariance Note:can also use formula $\frac{5\pi}{n}$	upper tail if > test and in Goudie either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2_{calc} = \sum_{\alpha} \frac{(0-x)^2}{2}$ . Reject: $x^2_{calc} > x^2_{critical}$ Hypothesis. <i>J<sub>a</sub></i> we importantly the status of
STATISTIC/ SAMPLE values, as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use 2 if $\sigma$ known OR $n \ge$ 30 Use 7 if $\sigma$ unknown OR $n \ge$ 30 Use 7 if $\sigma$ unknown OR $n \ge$ 30 Use 7 if $\sigma$ unknown CR n < 30 AP Stats: $\sigma$ unknown always, so can always, so can alw	$\begin{split} \mathbf{v}_{1} & \mathbf{v}_{1} & \mathbf{v}_{1} & \mathbf{v}_{1} \\ \text{Tests, 1 Prop Z interval} \\ \mathbf{2 prop:} & \widehat{p_{1}} = \widehat{p_{2}} \pm x_{c} \left( \frac{\beta_{1} q_{1}}{n_{1}} + \frac{\beta_{2} q_{2}}{n_{2}} \text{ where } \widehat{p_{1}} = \frac{x_{1}}{n_{1}} + \widehat{p_{2}} = \frac{x_{2}}{n_{2}} \\ \text{Tests, 2 Prop Z interval} \\ \mathbf{1 mean:} & \widehat{x} \pm Z_{c} \frac{q_{1}}{q_{1}} \otimes \widehat{x} \pm \overline{\zeta}_{c} \frac{\pi}{n_{1}} (\sigma \text{ unknown/ small sample}) \\ \text{Tests, 2 OR 1 Interval} \\ \mathbf{2 c}: \text{Invnorm with area} \frac{1-q_{0}}{2}, \mu = 0, \sigma = 1 \\ T_{c}: \text{ inv T with area} \frac{1-q_{0}}{2}, \frac{q_{1}}{n_{1}} + \frac{q_{2}}{n_{2}} / \widehat{X_{1}} - \widehat{X_{2}} \pm T_{1} S_{p} \sqrt{\frac{1}{n_{1}} + \frac{1}{n_{2}}} \\ \text{where } S_{p} = \sqrt{\frac{(n_{1}-1)s_{2}^{2} + (n_{2}-1)s_{2}^{2}}} (\text{pooled formula}) \\ \text{We pool when } n_{1} \approx \sigma_{2}: S_{1} < S_{2}, \sigma_{2} < S_{2} \\ \text{where } S_{p} = \sqrt{\frac{(n_{1}-1)s_{2}^{2} + (n_{2}-1)s_{2}^{2}}} (\text{pooled formula}) \\ \text{We pool when } n_{1} \approx \sigma_{2}: S_{1} < S_{2} \\ \text{z}: \text{Invnorm with area} \frac{1-q_{0}}{2}, \mu = 0, \sigma = 1 \\ T_{c}: \text{InvT with area} \frac{1-q_{0}}{2}, \mu = 0, \sigma = 1 \\ T_{c}: \text{InvT with area} \frac{1-q_{0}}{2}, \mu = 0, \sigma = 1 \\ T_{c}: \text{InvT with area} \frac{1-q_{0}}{2}, \mu = 0, \sigma = 1 \\ T_{c}: \text{InvT with area} \frac{1-q_{0}}{2}, \mu = 0, \sigma = 1 \\ T_{c}: \text{InvT with area} \frac{1-q_{0}}{2}, \mu = 0, \sigma = 1 \\ T_{c}: \text{InvT with area} \frac{1-q_{0}}{2}, \mu = 0, \sigma = 1 \\ T_{c}: \text{InvT with area} \frac{1-q_{0}}{2}, \mu = 0, \sigma = 1 \\ T_{c}: \text{InvT with area} \frac{1-q_{0}}{2}, \mu = 0, \sigma = 1 \\ T_{c}: \text{InvT with area} \frac{1-q_{0}}{2}, \mu = 0, \sigma = 1 \\ T_{c}: \text{InvT with area} \frac{1-q_{0}}{2}, \mu = 0, \sigma = 1 \\ T_{c}: \text{InvT with area} \frac{1-q_{0}}{2}, \mu = 0, \sigma = 1 \\ T_{c}: \text{InvT with area} \frac{1-q_{0}}{2}, \mu = 0, \sigma = 1 \\ T_{c}: \text{InvT with area} \frac{1-q_{0}}{2}, \mu = 0, \sigma = 1 \\ T_{c}: \text{InvT with area} \frac{1-q_{0}}{2}, \frac{1}{2}, \frac{1}{2},$	To Check Whether A Good Model  Linear Re Power Exponential Interpretations  Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B" Bayes Theorem  Completely Randomise Design (CRD):  Financy and Financy an	variation in y variable for a given amount of x variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations y = ax <sup>2</sup> where x → logx, y → log y On calculator: Pwr8g(L, L,) or timBeg (Dg, L, Dg, L) Y = ab <sup>2</sup> where x → L, y → log y On calculator: ExpReg(L, L,) or timBeg (Dg, L, Dg, L) Y = ab <sup>2</sup> where x → X, y → log y On calculator: ExpReg(L, L, ) or timBeg (Dg, L, Dg, L) Y = ab <sup>2</sup> where x → X, y → log y On calculator: ExpReg(L, L, ) or timBeg (Dg, L, Dg, L) Transformed scatter must look linear Residual plot of TRANSFORMED on pattern stary night Probability P(A) = <sup>m(D)</sup> / <sub>m(D</sub> = <sup>mumber</sup> of favourable outcomes mumber of possible outcomes P(A') = 1-P(A) i.e. probabilities add to 1 P(AUB)=P(A)+P(B)-P(A ∩ B) P(A ∩ B) = 0 addition rule becomes: P(AUB)=P(A)+P(B) P(A ∩ B) = P(A)P(B) P(A ∩ B) = P(A)P(B) P(A B) = <sup>P(ADB)</sup> / <sub>P(B)</sub> If independent: P(AB) = P(A) P(A B) = <sup>P(B A)P(A)</sup> P(A B) = <sup>P(B A)P(A)</sup> P(B A)P(A') + P(B A')P(A') epriment Templates Each subject receives only one treatment <sup>Ba</sup> (transmer) (temes represent the treatment) (temes represent) (temes	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcdf with df (see test statistics section above for of for Tcdf)         Chi- Squared         Type 2 Error Steps: Step 1: Find CV (above, 0, of 1)         :: area = ar (left), u=0, 0, of 1 $\neq$ : area = ar (left), u=0, 0, of 1 $\neq$ : area = ar (left), u=0, of 1 $\neq$ : area = ar (left), u=0, of 1 $\neq$ : area = ar (left), u=0, of 1 $\Rightarrow$ : area = ar (left), u=0, area $\Rightarrow$ : (lower-100, upper-CV u=new $\mu$ , $\sigma = \frac{\sigma}{2}$ $\Rightarrow$ : lower-CV, upper-CV u=new $\mu$ , $\sigma = \frac{\sigma}{\sqrt{n}}$ Mean         Variance Note:: an also use formula $\frac{\pi u}{n}$ Standard Dev $S_{xx}$ Unbiased Estimator ( $\overline{x}$ , S <sub>x</sub> )	upper tail if <i>x</i> less tail in Goudie either tail if <i>x</i> test) the p value Larger sample size $\Rightarrow$ smaller p value if $x^2_{calc} = \sum_{i=1}^{i} \frac{(g-E)^2}{\pi}$ . Reject: $x^2_{calc} > x^2_{critical}$ Hypothesis, <i>it</i> , are independent in the rabul-, distributed diff $x^2_{calc} = \sum_{i=1}^{i} \frac{(g-E)^2}{\pi}$ . Reject: $x^2_{calc} > x^2_{critical}$ Hypothesis, <i>it</i> , are independent in the rabul-, distributed diff $x^2_{calc} = \sum_{i=1}^{i} \frac{(g-E)^2}{\pi}$ . Reject: $x^2_{calc} > x^2_{critical}$ Hypothesis, <i>it</i> , are independent in the rabul-, distributed diff $x^2_{calc} = \sum_{i=1}^{i} \frac{(g-E)^2}{\pi}$ . Reject: $x^2_{calc} > x^2_{critical}$ Hypothesis, <i>it</i> , are independent in the rabul-, distributed diff $x^2_{calc} = \sum_{i=1}^{i} \frac{(g-E)^2}{\pi}$ . This is a (prob of rejecting the test i.e. being in the critical region) Type 2: <i>H</i> , fixels, but we say it is false i.e. reject it <b>Calculations:</b> Type 2: <i>F</i> , Forb of NOT being in the critical region-see blue on left for the method of ow to find this Note: There is an alternative method which is harder. Only use this method if you're forced to Re-arrange $z = \frac{x-x}{\pi}$ by mere $\mu$ -original mean, $z_eCV$ Lower: $\mu + z_e \frac{\pi}{\pi}$ by there $\mu$ -original mean, $z_eCV$ Lower: $\mu + z_e \frac{\pi}{\pi}$ by hyper $\mu + z_e \frac{\pi}{\pi}$ unew $\mu, \sigma \frac{\pi}{\pi}$ $\Rightarrow P(\overline{X} > \mu + z_e \frac{\pi}{\pi})$ where $\mu$ -original mean, $z_eCV$ Lower: $\mu - z_e \frac{\pi}{\pi}$ by hyper $\mu + z_e \frac{\pi}{\pi}$ unew $\mu, \sigma \frac{\pi}{\pi}$ To <i>Increase Obserose</i> The Errors: <b>To</b> Increase type 1: increase is level. Changing sample size does nothing Increase type 2: increase is $z^2 = \frac{F^2}{\pi}$ . $x^2 = \frac{2(x-x)^2}{\pi}$ If no frequency: $a^2 = \frac{E^2}{\pi}$ . $x^2 = \frac{2(x-x)^2}{\pi}$ $g_{xy} = \sum (x_1 - x_2) = \sum x_1^2 - \frac{(x_1)^2}{\pi}$ $s_{xy} = \sum (x_1 - x_2) = \sum x_1^2 - \frac{(x_1)^2}{\pi}$ $s_{xy} = \sum (x_1 - x_2) = \sum x_1^2 - \frac{(x_1)^2}{\pi}$
STATISTIC/ SAMPLE values as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use Z if $\sigma$ known OR $n \ge 30$ Use T if $\sigma$ unknown OR $n \ge 300$ Use T if $\sigma$ unknown OR $n \ge 300$ AP Stats: $\sigma$ unknown OR $\sigma_2$ Subt you will also notice that the mark scheme normally desen't pool regardless of size of $\sigma$ Assumptions	$\begin{split} \mathbf{v}_{1} & \mathbf{v}_{1} & \mathbf{v}_{1} & \mathbf{v}_{1} \\ \text{Tests, 1 Prop Z Interval} \\ \mathbf{2 \ prop:} \ \widehat{p_{1}} - \widehat{p_{2}} \pm x_{c} \left( \frac{\beta_{1} q_{1}}{n_{1}} + \frac{\beta_{2} q_{2}}{n_{1}} \right) \\ \text{where } \widehat{p_{1}} = \frac{x_{1}}{n_{1}} + \frac{\beta_{2}}{n_{2}} \right) \\ \mathbf{x}_{2} = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{j=1}^{N} \sum_{j=1}^{N} \sum_{i=1}^{N} \sum_{j=1}^{N} $	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B" Bayes Theorem Completely Randomise Design (CRR):	variation in y variable for a given amount of x variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value High r <sup>2</sup> value Residual plot has no pattern with uniform variation across x (starry night) <b>sgression - Transformations</b> $y = ax^{b}$ where $x \to logx, y \to logy$ On calculator: Pwreg(L, L, ) or Linkeg (log L, log L, ) On calculator: Pwreg(L, L, ) or Linkeg (log L, log L, ) On calculator: Pwreg(L, L, ) or Linkeg (log L, log L, ) Transformed scatter must look linear Residual plot of TRANSFORMED no pattern starry night <b>Probability</b> $P(A) = \frac{n(L)}{n(L)} = \frac{number of favourable outcomes}{number of possible outcomes}$ $P(A') = \frac{n(L)}{n(L)} = \frac{number of favourable outcomes}{number of possible outcomes}$ P(A = n(L) = n(A) = n(A) = n(A) = 0 addition rule becomes: $P(A D B) = P(A) = 0$ P(A = B) = P(A) + P(B) = P(A) = 0 P(A = B) = P(A) P(B) = P(A) + P(B) = P(A) = 0 P(A = B) = P(A) = P(A) = P(A) = 0 P(A = B) = P(A) = P(A) = P(A) = 0 P(A = B) = P(A) = P(B) = P(A) = P(A) = P(B) = P(A) = P(B) = P(A) = P(B) = P(A) = P(B) = P(A) = P(A) = P(B) = P(B) = P(B) = P(A) = P(B) = P(A) = P(B) = P(A) = P(B) = P(A) = P(B)	with $\mu$ and $\frac{\sigma}{\sqrt{n}} \sigma$ Tcd with df (see test statistics section above for of for Tcdf) Chi- Squared Errors Type 2 Error Steps: Step 1: Find CV (using innorm) $c; rare = \sigma$ (left), $u=0, \sigma=1$ $\Rightarrow: area = \sigma$ (left), $u=0, \sigma=1$ $\Rightarrow: area = \sigma$ (left), $u=0, \sigma=1$ $\Rightarrow: area = \frac{\sigma}{2}$ (left), $u=0, \sigma=1$ $f: area = \frac{\sigma}{2}$ (left), $u=0, \sigma=1$ Step 2: Find error (normcdf) c: lower=-CV, upper=100 $u=new \mu, \sigma=\frac{\sigma}{\sqrt{n}}$ $\Rightarrow: lower=CV, upper=CV$ $u=new \mu, \sigma=\frac{\sigma}{\sqrt{n}}$ $\Rightarrow: lower=CV, upper=CV$ $u=new \mu, \sigma=\frac{\sigma}{\sqrt{n}}$ f: lower=CV, upper=CV $u=new \mu, \sigma=\frac{\sigma}{\sqrt{n}}$ Standard Dev $S_{XX}$ Unbiased Estimator ( $\overline{x}, S_X$ )	upper tail if > test and in Goudie either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2$ calc $\equiv \sum_{i=1}^{G(2-E)^2}$ . Reject: $x^2$ cartical HypothesizH, are independent/in the ratio/_ distributed $H_i$ are not independent/in the ratio/_ distributed $H_i$ by $E_i$ $H_i$ for the method of out to find this Note: There is an alternative method which is harder. Only use this method if you're forced to. Re-arrange $z = \frac{F_i + 1}{F_i}$ to get $\overline{x} = \mu + z_e \frac{\sigma_i}{\sigma_i}$ $> P(\overline{x} > \mu + z_e \frac{\sigma_i}{\sigma_i})$ Where $\mu$ -original mean, $z_e CV$ Lower= $\mu + z_e \frac{\sigma_i}{\sigma_i}$ Where $\mu$ -original mean, $z_e CV$ Lower= $\mu + z_e \frac{\sigma_i}{\sigma_i}$ Where $\mu$ -original mean, $z_e CV$ Lower= $\mu + z_e \frac{\sigma_i}{\sigma_i}$ Where $\mu$ -original mean, $z_e CV$ Lower= $\mu + z_e \frac{\sigma_i}{\sigma_i}$ Where $\mu$ -original mean, $z_e CV$ Lower= $\mu + z_e \frac{\sigma_i}{\sigma_i}$ Where $\mu$ -original mean, $z_e CV$ Lower= $\mu + z_e \frac{\sigma_i}{\sigma_i}$ Where $\mu$ -original mean, $z_e CV$ Lower= $\mu + z_e \frac{\sigma_i}{\sigma_i}$ Where $\mu$ -original mean, $z_e CV$ Lower= $\mu + z_e \frac{\sigma_i}{\sigma_i}$ Where $\mu$ -original mean, $z_e CV$ Lower= $\mu + z_e \frac{\sigma_i}{\sigma_i}$ where $\mu$ are $\mu$ are $\sigma_i \frac{\sigma_i}{\sigma_i}$ To Increase/Decrease The Errors: Prover: Initians prob of type $\Xi = 1 - \beta_{in}$ samples increase $\sigma_i$ If no frequency: $\overline{a} = \frac{\Sigma A^2}{\lambda} = \overline{x}^2 = \frac{\Sigma A^2}{\lambda}$ If no frequency: $\overline{a} = \frac{\Sigma A^2}{\lambda} = \overline{x}^2 = \frac{\Sigma A^2}{\lambda}$ $\overline{a} = \sqrt{variance}$ $S_{xx} = \sum (x_i - \overline{x}) = \sum (x_i - $
STATISTIC/ SAMPLE values, as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use 2 if $\sigma$ unknown OR $n \ge 30$ Use T if $\sigma$ unknown OR $n \ge 30$ Use T if $\sigma$ unknown OR $n \ge 30$ and AP Stats: $\sigma$ unknown always, so can always use T for mean tests. We gool if T Testand $\sigma_1 = \sigma_2$ , but you will also notice that the mark scheme normally desen't pool regardless of size of $\sigma$ Assumptions	$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	To Check Whether A Good Model  Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional ''A given B'' Bayes Theorem Excompletely Randomise Design (CRD):	variation in y variable for a given amount of x variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations y = x <sup>b</sup> where x - 1 (agx, y - 1 (agy - 1)) or calculator: PwrReg(L, L) or LinReg (tog L, log L)) y = ab <sup>2</sup> where x - 1 (agx, y - 1 (agy - 1)) On calculator: PwrReg(L, L) or LinReg (tog L, log L)) y = ab <sup>2</sup> where x - 1 (agx, y - 1 (agy - 1)) On calculator: PerReg(L, L) or LinReg (tog L, log L)) y = ab <sup>2</sup> where x - x (x - 1 (agy - 1)) On calculator: PerReg(L, L) or LinReg (tog L, log L)) Transformed scatter must took linear Residual plot of TRANSFORMED no pattern starry night <b>P</b> (A) = $\frac{n(A)}{n(D)} = \frac{number of favourable outcomes}{number of possible outcomes}$ P(A) = $-P(A) = \frac{n(A)}{n(B)} = P(A)P(B) - P(A \cap B) = P(A)P(B) - P(A \cap B) = 0$ addition rule becomes: P(AUB)=P(A)+P(B)-P(A)P(B) P(A) B) = P(A)P(B) = P(A)P(B) P(A) B) = $P(A)B = P(A)P(A)$ P(A B) = $\frac{P(A)B}{P(B A)P(A)} + P(B A)P(A)$ P(A B) = $\frac{P(B A)P(A)}{P(B A)P(A)} + P(B A)P(A')$ geriment Templates Each subject receives only one treatment <b>P</b> (torman) + Kouw represent Biblock first and then CRD on each block	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcd with df (see test statistics section above for of for Tcdf)       Chi-Squared       Errors       Type 2 Error Steps: Step 1: Find CV (using innorm) (: area = a: (left), u=0, 0, \sigma=1 >: area = a: (left), u=0, 0, \sigma=1 >: area = a: (left), u=0, 0, \sigma=1 = $\frac{\sigma}{2}$ (left), u=0, 0, \sigma=1 (left), u=0, 0, d=1 := $\frac{\sigma}{2}$ (left), u=0, 0, \sigma=1 Step 2: Find error (normcff) <: lower=-700, upper=CV u=new $\mu$ , $\sigma^{-}_{\pi}$ >: lower=-700, upper=CV u=new $\mu$ , $\sigma^{-}_{\pi}$ >: lower=-700, upper=CV u=new $\mu$ , $\sigma^{-}_{\pi}$ >: lower=CV, upper=CV u=new $\mu$ , $\sigma^{-}_{\pi}$ >: lower=CV, upper=CV u=new $\mu$ , $\sigma^{-}_{\pi}$ Standard Dev       Sxx       Unbiased Estimator ( $\overline{x}$ , Sx)	upper tail in 2 less tail in Gouble either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2_{calc} = \sum_{\alpha, \alpha} \frac{(\alpha - E)^2}{R}$ Reject: $x^2_{calc} > x^2_{critical}$ Hypothesis. If $x$ is the important in the risk of distributed thypothesis. If $x$ is the important in the risk of distributed diff (movel) (columns. 1) for independence and $n-1$ for other diff $n = 2.1$ approximating p or $\mu/\alpha$ but this newer comes up <b>Definitions:</b> Type 1: H <sub>0</sub> true, but we say it is false i.e. reject it Type 2: H <sub>0</sub> from 0 of rejecting the test i.e. being in the critical region) Type 2 = $\beta$ . Frob of NOT being in the critical region-see blue on left for the method of ow to find this Note: There is an alternative method which is harder. Only use this method if you're forced to. Re-arrange $z = \frac{E_{T}}{\pi}$ to get $\overline{x} = \mu + z_{c}\frac{\pi}{\pi}$ $< P(\overline{x} > (\mu + z_{c}\frac{\pi}{\pi}))$ where $\mu$ -original mean, $z_{c}$ -CV Lower = $-100$ , upper = $\mu + z_{c}\frac{\pi}{\pi}$ unence $\mu, \sigma - \frac{\pi}{\pi}$ $\Rightarrow P(\overline{x} > \mu + z_{c}\frac{\pi}{\pi})$ where $\mu$ -original mean, $z_{c}$ -CV Lower = $\mu + z_{c}\frac{\pi}{\pi}$ upper = 100, u-new $\mu, \sigma - \frac{\pi}{\pi}$ $\Rightarrow P(\mu - z_{c}\frac{\pi}{\pi})$ where $\mu$ -original mean, $z_{c}$ -CV Lower = $\mu + z_{c}\frac{\pi}{\pi}$ upper = 100, u-new $\mu, \sigma - \frac{\pi}{\pi}$ To increase/Decrease The Errors: Power: 1 minos to fby $\overline{z} = 1 + z_{c}\frac{\pi}{\pi}$ upper a lower, is take smaller samples, increase $z_{1}$ increases $\overline{z}$ Data If no frequency: $\sigma^{2} = \frac{\Sigma x^{2}}{\Sigma - x^{2}} = \frac{\Sigma (z - \rho^{2})}{K - 1}$ If no frequency: $\sigma^{2} = \frac{\Sigma x^{2}}{\Sigma - x^{2}} = \frac{\Sigma (z - \rho^{2})}{K - 1}$ If no frequency: $\sigma^{2} = \frac{\Sigma x^{2}}{\Sigma - x^{2}} = \frac{\Sigma (z - \rho^{2})}{K - 1}$ $s_{xy} = \sum xy - \frac{\Sigma x(2)}{\pi - 1} = \sum xy - \pi x\overline{y}$ $s_{xy} = \sum xy - \frac{\Sigma x(2)}{\pi - 1} = \sum xy - \pi x\overline{y}$ $s_{xy} = \sum xy - \frac{\Sigma x(2)}{\pi - 1} = \sum xy - \pi x\overline{y}$ $s_{xy} = \sum xy - \frac{\Sigma x(2)}{\pi - 1} = \sum xy - \pi x\overline{y}$ $s_{xy} = \sum xy - \frac{\Sigma x(2)}{\pi - 1} = \sum xy - \pi x\overline{y}$ $s_{xy} = \sum xy - \frac{\Sigma x(2)}{\pi - 1} = \sum xy - \pi x\overline{y}$ $x_{xy} = \sum x + x - \frac{\Sigma x}{\pi - 1} = \frac{\Sigma x}{\pi - 1} = \sum x - \frac{\Sigma x}{\pi - 1} = \frac{\Sigma x}{\pi - 1} = \sum x - \Sigma$
STATISTIC/ SAMPLE values, as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use 2 if $\sigma$ unknown OR $n \ge$ 30 Use 7 if $\sigma$ unknown OR $n \ge$ 30 Use 7 if $\sigma$ unknown OR $n \ge$ 30 Use 7 if $\sigma$ unknown always, so can always use T for mean tests. We pool if T Test and $\sigma_1 \approx \sigma_2$ , but you will also normally doesn't pool regardless of size of $\sigma$ Assumptions	$\begin{split} \mathbf{v}_{1} & \mathbf{v}_{1} & \mathbf{v}_{1} & \mathbf{v}_{1} \\ \text{Tests, 1 Prop Z Interval} \\ \mathbf{2 prop:} & \widehat{p_{1}} = \widehat{p_{2}} \pm x_{c} \left( \frac{\beta_{1} q_{1}}{m_{1}} + \frac{\beta_{2} q_{2}}{m_{1}} \right) \\ \text{Tests, 2 Prop Z Interval} \\ \mathbf{1 mean:} & \widehat{x} \pm Z_{c} & \frac{q_{1}}{q_{1}} \otimes \widehat{x} \pm Z_{c} & \frac{q_{1}}{m_{1}} (\sigma unknown/small sample) \\ \text{Tests, 2 OR T Interval} \\ \mathbf{2 c: invnorm with area} \frac{1-q_{0}}{p_{1}}, \mu = 0, \sigma = 1 \\ & T_{c}: invT with area \frac{1-q_{0}}{p_{1}} & \frac{1}{m_{1}} = \overline{x_{2}} \pm \overline{x_{1}} - \overline{x_{2}} \pm \overline{x_{1}} + \frac{q_{2}}{q_{2}} + \overline{x_{2}} + \overline{x_{1}} + \frac{q_{2}}{q_{2}} + \overline{x_{2}} + x_$	To Check Whether A Good Model  Linear Re Power  Exponential Interpretations  Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B" Bayes Theorem  Completely Randomise Design (CRD):  Randomized (complete) Block Design: Randomized (complete) Block Design:	variation in y variable for a given amount of x variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations y = ax <sup>2</sup> where x → logx, y → logy On calculator: Perkeg(L, L) or Linkeg (tog, L, log L) y = ab <sup>2</sup> where x → logx, y → logy On calculator: Perkeg(L, L) or Linkeg (tog, L, log L) y = ab <sup>2</sup> where x → x, y → log y On calculator: Perkeg(L, L, L) or Linkeg (tog, L, log L) y = ab <sup>2</sup> where x → x, y → log y On calculator: Perkeg(L, L) or Linkeg (tog, L, log L) Y = ab <sup>2</sup> where x → x, y → log y On calculator: Perkeg(L, L) or Linkeg (tog, L, log L) Y = ab <sup>2</sup> where x → x, y → log y On calculator: Perkeg(L, L) or Explore (L, log L) Y = ab <sup>2</sup> where y = togstbe autcomes P(A) = <sup>m(L)</sup> P(A) = <sup>m(L)</sup> P(A) = <sup>n(L)</sup> P(A ∩ B) = 0 addition rule becomes: P(AUB)=P(A)+P(B) P(A ∩ B) = P(A)(B) P(A ∩ B) = P(A)(B) P(A B) = <sup>P(ADB)</sup> P(B A)P(A) P(B A)P(A) + P(B A')P(A') periment Templates Each subject receives only one treatment <sup>ma</sup> treaments) → tenare reasons = treaments) → tenare treaments = treaments) → tenare treaments = trea	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcd with df (see test statistics section above for of for Tcdf)         Chi- Squared         Type 2 Error Steps: Step 1: Find C (using invnorm) c: area $= \alpha$ (left), $u=0, 0, \sigma=1$ $\Rightarrow$ : area $= \alpha$ (left), $u=0, \sigma=1$ $\Rightarrow$ : area $= \alpha$ (left), $u=0, \sigma=1$ Step 2: Find error (normcdf) c: lower=-100, upper-CV $urnew \mu, \sigma=\frac{\sigma}{2}$ >: Lower=C-100, upper=CV $urnew \mu, \sigma=\frac{\sigma}{2}$ $urnew \mu, \sigma=\frac{\sigma}{2}$ Unbiased Estimator ( $\overline{x}, S_x$ )         Quartiles Being in the $p^n$ percentile means you did better than $p^n$ is of execute.	upper tail if > test and in Goudie either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value if $x^2_{calc} = \sum_{i=1}^{i} \frac{(g-E)^2}{n}$ . Reject: $x^2_{calc} > x^2_{critical}$ Hypothesis, $u_i$ are independent/in the ratio/
STATISTIC/ SAMPLE values as their values, not population values, not askyourself first Is proportion or mean and then is it 1 or 2 samples Use Z if $\sigma$ known OR $n \ge 300$ Use Z if $\sigma$ unknown OR $n \ge 300$ Use Z if $\sigma$ samples unknown OR $n \ge 300$ Use Z if $\sigma$ samples and $\sigma \ge 300$ AP Stats: $\sigma$ unknown OR $n \ge 300$ Use T if $\sigma$ unknown OR $n \ge 300$ Use T if $\sigma$ Use T if $\sigma$ unknown OR $n \ge 300$ Use T if $\sigma$ Use T if $\sigma$ Use T if $\sigma$ unknown OR $n \ge 300$ Use T if $\sigma$ Use T if $\sigma$ unknown OR $n \ge 300$ Use T if $\sigma$ Use T if	$\begin{array}{ c c c c c } \hline V_{1} & V_{$	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Completely Randomise Design (CRR): Bayes Theorem Completely Randomise Design (CRR): Experiment units Completely Block Design: Randomized (complete) Block Design:	variation in y variable for a given amount of x variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations y = ax <sup>b</sup> where x → logx, y → logy On calculator: Pwreg(L, L, ) or Linkeg (og L, log L, ) On calculator: Pwreg(L, L, ) or Linkeg (og L, log L, ) On calculator: Pwreg(L, L, ) or Linkeg (og L, log L, ) On calculator: Pwreg(L, L, ) or Explore (L, log L, ) Transformed scatter must look linear Residual plot of TRANSFORMED no pattern starry night Probability Probability P(A) = $\frac{n(A)}{n(L)} = \frac{number of favourable outcomes}{number of patourable outcomes}$ $P(A) = \frac{n(A)}{n(L)} = \frac{number of favourable outcomes}{number of patourable outcomes}$ $P(A) = \frac{n(A)}{n(L)} = \frac{n(A)BP(A)}{n(B)} = P(A) P(B)$ P(A n B) = 0 addition rule becomes: P(AUB)=P(A)+P(B) P(A n B) = P(A)P(B) P(A n B) = P(A)P(B) P(A n B) = P(A)P(B) P(A B) = $\frac{P(A B)}{P(B)}$ If independent: P(A B) = P(A) P(A B) = $\frac{P(A B)}{P(B A)P(A)} + P(B A')P(A')$ geriment Templates Each subject receives only one treatment ** (formation) + (formation) ** (formation) + (forma	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcd with df (see test statistics section above for of for Tcdf) Chi- Squared Errors Type 2 Error Steps: Step 1: Find CY (using innorm) $c; area = \sigma$ (left), $u=0, \sigma=1$ $\Rightarrow: area = \sigma$ (left), $u=0, \sigma=1$ $\Rightarrow: area = \sigma$ (left), $u=0, \sigma=1$ $\Rightarrow: area = \frac{\sigma}{2}$ (left), $u=0, \sigma=1$ $f: area = \frac{\sigma}{2}$ (left), $u=0, \sigma=1$ $f: tower=-CV, upper=TO u=new \mu, \sigma=\frac{\sigma}{\sqrt{n}}\Rightarrow: tower=-CV, upper=TO u=new \mu, \sigma=\frac{\sigma}{\sqrt{n}}f: tower=-CV, upper=TO u=new \mu, \sigma=\frac{\sigma}{\sqrt{n}}MeanVarianceNote: can also use formula \frac{k_{ax}}{n}Standard DevS_{xx}Unbiased Estimator (\bar{x}, S_x)QuartilesBeing in the pn percentile meansyou did better than p X of people.$	upper tail if > test and in Goldble either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2$ calc $\equiv \sum \frac{(G-E)^2}{2\pi}$ . Reject: $x^2$ and $x^2$ critical HypothesisH, are independent/in the ratio/ distributed $H_1$ are not independent/in the ratio/ distributed $H_2$ are not independent/in the ratio/ distributed $H_1$ are not independent/in the ratio/ distributed $H_2$ are not independent of not the inserver comes up $H_2$ e 2 if $F_2$ for bod NOT being in the critical region-see blue on left for the method of ow to find this Note: There is an alternative method which is harder. Only use this method if you're forced to. Re-arrange $z = \frac{F_2H}{\pi}$ to get $\overline{x} = \mu + z_c \frac{\pi}{\pi}$ upper = 100, upper $\mu + z_c \frac{\pi}{\pi}$ upper $\mu = 0$ night mean, $z_c = CV$ Lower = $\mu + z_c \frac{\pi}{\pi}$ upper $\mu + \mu = 0$ night and mean Lower = $\mu + z_c \frac{\pi}{\pi}$ upper $\mu + z_c \frac{\pi}{\pi}$ unew $\mu, \sigma - \frac{\pi}{\pi}$ To Increase (Decaase The Errors) Phoreses Type 1: increase gives! Changing angles size does nother, increase $\pi_1$ be registered. The night are not path of $\pi_1$ H for frequency: $\sigma^2 = \frac{E_2H}{2T} - x^2 = \frac{E_2(x-y)^2}{\pi}$ If no frequency: $\sigma^2 = \frac{E_2H}{2T} - x^2 = \frac{E_2(x-y)^2}{\pi}$ $F_{XY} = \sum x_X - \sum (X_X - \sqrt{2}) = \sum x_Y - x_X - \overline{x_Y}$ $\overline{x_X} = \sum (X_X - \sqrt{2}) = \sum x_Y - \frac{(-\Sigma_X)}{\pi} - \frac{x_Y}{\pi} = \frac{x_Y}{\pi} = \frac{\pi^{-1}}{\pi} = \frac{\pi^{-1}}{\pi$
STATISTIC/ SAMPLE values, as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use Z if $\sigma$ unknown OR $n \ge 30$ 00 Use T if $\sigma$ unknown OR $n \ge 310$ Use Z if $\sigma$ unknown OR $n \ge 310$ 00 Use T if $\sigma$ unknown OR $n \ge 310$ $\sigma$ samples use Z if $\sigma$ samples use Z if $\sigma$ use T if $\sigma$ unknown always, so can always, so can alway	$\begin{aligned} & \operatorname{Tests}, 1 \operatorname{Pop } \mathbb{Z} \operatorname{interval}^{\operatorname{Interval}} \\ & 2 \operatorname{prop}: \widehat{p_1} = \widehat{p_2} \pm \chi_c \left( \frac{\beta_1 q_1}{m_1} + \frac{\beta_2 q_2}{m_2} \operatorname{where } \widehat{p_1} = \frac{x_1}{n_1}, \widehat{p_2} = \frac{x_2}{n_2} \\ & \operatorname{Tests}, 2 \operatorname{Prop } \operatorname{Interval} \\ & 1 \operatorname{mean}: \overline{X} \pm \zeta, \frac{\sigma_1}{\sigma_1} \otimes \overline{X} \pm \zeta, \frac{\sigma_1}{\sigma_1} (\sigma \operatorname{unknown}) \operatorname{small sample} \right) \\ & \operatorname{Tests}, 2 \operatorname{Col } \operatorname{Interval} \\ & z_c : \operatorname{invnorm with area} \frac{1-q_0}{p_2}, \mu = 0, \sigma = 1 \\ & \mathcal{T}_c : \operatorname{Inv} \operatorname{Twi with area} \frac{1-q_0}{p_2}, \alpha = 0, \sigma = 1 \\ & \mathcal{T}_c : \operatorname{Inv} \operatorname{Twi with area} \frac{1-q_0}{p_2}, \alpha = 1, \alpha = 1 \\ & 2 \operatorname{means}: \overline{X}_1 = \overline{X}_2 \pm Z_c, \left( \frac{\sigma_1 x_1}{\sigma_1 + \sigma_2}, \frac{\sigma_1 x_1}{\sigma_2} + \overline{X}_2 \pm T_c, \overline{sp}, \frac{1}{n_1} + \frac{1}{n_2} \\ & \text{where } s_p = \sqrt{\frac{q_1 + 1 \sigma_2 + 2 \sigma_1 \sigma_2 + 1 \sigma_2}{n_1 + m_2 + 2 \sigma_2}} (\text{pooled formula}) \\ & \text{We pool when } \sigma_1 \approx \sigma_2 : S_1 < 2S_2 \text{ or } S_2 < 2S_1 \\ & \text{if don't pool use formula: } \overline{X}_1 = \overline{X}_2 \pm T_c, \frac{s_1 x_1}{\sigma_1 + s_2} + \frac{s_2 \sigma_2}{\sigma_1 + s_2} \\ & z_c : \operatorname{Invnorm with area} \frac{1-q_0}{\sigma_1 + m_1 \sigma_1 - m_2}, \mu = 0, \sigma = 1 \\ & \mathcal{T}_c : \operatorname{Inv} \operatorname{Interval} & \frac{1-q_0}{\sigma_1 + m_1 \sigma_1 - m_2}, \mu = 0, \sigma = 1 \\ & \mathcal{T}_c : \operatorname{InvT with area} \frac{1-q_0}{\sigma_1 + m_1 \sigma_1 - m_2}, \mu = 0, \sigma = 1 \\ & \mathcal{T}_c : \operatorname{InVT with area} \frac{1-q_0}{\sigma_1 + m_1 \sigma_1 - m_2}, \mu = 0, \sigma = 1 \\ & \mathcal{T}_c : \operatorname{InVT with area} \frac{1-q_0}{\sigma_1 + m_1 \sigma_1 - m_2}, \mu = 0, \sigma = 0 \\ & \operatorname{Interval} & \operatorname{Ittraval} \\ & \operatorname{Ittravel} \mathcal{I}_c \in \mathcal{I}_c \otimes \mathcal{I}_c $	To Check Whether A Good Model  Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B" Bayes Theorem Excompletely Randomise Design (CRD): Completely Randomise Design (CRD): Randomized (complete) Block Design: Randomized (complete) Block Design:	variation in y variable for a given amount of x variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations y = x <sup>b</sup> where x - 1 (agx, y - 10gy On calculator: Purkeg(L, L) or Linkeg (tog L, log L) y = ab <sup>2</sup> where x - 1 (agx, y - 10gy On calculator: Purkeg(L, L) or Linkeg (tog L, log L) y = ab <sup>2</sup> where x - x (x - 10gx, y) On calculator: Purkeg(L, L) or Linkeg (tog L, log L) y = ab <sup>2</sup> where x - x (x - 10gx, y) On calculator: Purkeg(L, L) or Linkeg (tog L) y = ab <sup>2</sup> where x - x (x - 10gx, y) On calculator: Purkeg(L, L) or Linkeg (tog L) y = ab <sup>2</sup> where x - x (x - 10gx, y) On calculator: Purkeg (L, L) or Linkeg (tog L) y = ab <sup>2</sup> where y - y (L) or Displet (L) (and L) Transformed scatter must look linear Residual plot of TRANSFORMED no pattern starry night P(A) = $\frac{n(A)}{n(B)} = \frac{n(A)BP of J + P(B) - P(A \cap B)}{P(A \cap B) = P(A) P(B) = P(A) = P(A)}$ P(A \B) = $P(A \cap B) = P(A)P(B)$ P(A \B) = $P(A \cap B) = P(A)P(B)$ P(A B) = $P(A B) = P(A)P(A)$ P(A B) = $\frac{P(B A)P(A)}{P(B A)P(A)} + P(B A')P(A')$ geriment Templates Each subject receives only one treatment # f tronsmat} + Kouwa represent Block first and then CRD on each block	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcd with df (see test statistics section above for of for Tcdf) Chi- Squared Errors Type 2 Error Steps: Step 1: Find CV (using invorm) <, area = a: (left), u=0, 0, o=1 >: area = a: (left), u=0, 0, o=1 >: area = a: (left), u=0, 0, o=1 :: area = a: (left), u=0, o=1 :: area = a: area =	upper tail in 2 less tail in Gouble either tail in 2 less test) the p value Larger sample size $\Rightarrow$ smaller p value Integr sample size $\Rightarrow$ smaller p value If $x^2_{-calc} = \sum_{i=1}^{i} \frac{(0-E)^2}{i}$ . Reject: $x^2_{-calc} > x^2_{-critical}$ Hypothesis, we is genoment/in the nixip, distributed dir(nov 1)(columns.1) for independence and $n-1$ for other dir $n > 21$ approximating p or $\mu/a$ but this newer comes up Definitions: Type 1: $H_0$ true, but we say it is false i.e. reject it Type 2: $H_0$ from 0 for federation the critical region-see bue on left for the method of ow to find this is harder. Only use this method if you're forced to. Re-arrange $x = \frac{E_{T}}{\pi t}$ to get $\overline{x} = \mu + x_c \frac{\pi}{\pi}$ $< P(\overline{X} > (\mu + x_c \frac{\pi}{\pi}))$ where $\mu$ -original mean, $x_c = CV$ Lower = $-100$ , upper = $\mu + z_c \frac{\pi}{\pi}$ uneaw $\mu, \sigma = \frac{\pi}{\pi}$ $x = P(\overline{X} > (\mu + x_c \frac{\pi}{\pi}))$ where $\mu$ -original mean, $x_c = CV$ Lower = $\mu + z_c \frac{\pi}{\pi}$ upper = 100, u-neaw $\mu, \sigma = \frac{\pi}{\pi}$ $x = P(\mu - x_c \frac{\pi}{\pi} (x > \mu + x_c \frac{\pi}{\pi}))$ where $\mu$ -original mean, $x_c = CV$ Lower = $\mu + z_c \frac{\pi}{\pi}$ upper = 100, u-neaw $\mu, \sigma = \frac{\pi}{\pi}$ $x = P(\mu - x_c \frac{\pi}{\pi})$ upper = 100, u-neaw $\mu, \sigma = \frac{\pi}{\pi}$ To increase/Decrease The Errors: Power: 1 minos to of type 2 = 1 + gr - x^2 = \frac{2E_{T}-P^2}{K} If no frequency: $\sigma^2 = \frac{E_{T}^2 - x^2 = \frac{E_{T}-P^2}{K}$ If no frequency: $\sigma^2 = \frac{E_{T}^2 - x^2 = \frac{E_{T}-P^2}{K}$ If no frequency: $\sigma^2 = \frac{E_{T}^2 - x^2 = \frac{E_{T}-P^2}{K}$ $x_T = \frac{E_{T}}{x_T}$ , $x_T = (\sum_{T=1}^{2} - \sum_{T=1}^{2} - \pi \overline{x}) = \frac{E_{T}-P^2}{K}$ $\overline{x_T} = \frac{E_{T}}{x_T}$ , $x_T = (\sum_{T=1}^{2} - \frac{E_{T}-P^2}{\pi}) = \frac{E_{T}-P^2}{\pi}$ $x_T = \frac{E_{T}}{x_T}$ , $x_T = (\sum_{T=1}^{2} - \frac{E_{T}-P^2}{\pi}) = \frac{E_{T}-P^2}{\pi}$ $x_T = \frac{E_{T}}{x_T}$ , $x_T = (\sum_{T=1}^{2} - \frac{E_{T}-P^2}{\pi}) = \frac{E_{T}-P^2}{\pi}$ $x_T = \frac{E_{T}}{x_T}$ , $x_T = (\sum_{T=1}^{2} - \frac{E_{T}-P^2}{\pi}) = \frac{E_{T}-P^2}{\pi}$ $x_T = \frac{E_{T}}{x_T}$ , $x_T = (\sum_{T=1}^{2} - \frac{E_{T}-P^2}{\pi}) = \frac{E_{T}-P^2}{\pi}$ $x_T = \frac{E_{T}}{x_T}$ , $x_T = (\sum_{T=1}^{2} - \frac{E_{T}-P^2}$
STATISTIC/ SAMPLE values, as their values, not population values Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use 2 if $\sigma$ known OR $n \ge 30$ Use 1 if $\sigma$ unknown OR $n \ge 30$ Use 1 if $\sigma$ unknown OR $n \ge 30$ Use 3 if $\sigma$ unknown OR $n \ge 30$ Use 5 if $\sigma$ unknown CR $n \ge 30$ To rmean tests. We pool if T rest and $\sigma_1 \approx \sigma_2$ , but you will also normally doesn't pool regardless of regardless of size of $\sigma$ Assumptions	Tests, 1 Prop Z interval Tests, 2 Prop Z interval 2 <b>2</b> prop: $\vec{p}_1 = \vec{p}_2 \pm x_c \left( \frac{\vec{p}_1 \vec{q}_1}{n_1} + \frac{\vec{p}_2 \vec{q}_2}{n_2} \text{ where } \vec{p}_1 = \frac{x_1}{n_1}, \vec{p}_2 = \frac{x_2}{n_2} \right]$ Tests, 2 Prop Z interval 1 <b>1</b> mean: $\vec{X} \pm Z_c \frac{d}{m_1} \otimes \vec{X} \pm T_c \frac{d}{m_1}$ ( $\sigma$ unknown/small sample) Tests, 2 CR 1 Interval 2 <b>c</b> : invnorm with area $\frac{1-q_0}{p_1}, \mu = 0, \sigma = 1$ $T_c$ : invT with area $\frac{1-q_0}{p_1}$ df = <b>n</b> -1 2 <b>2</b> means: $\vec{X}_1 = \vec{X}_2 \pm Z_c \left( \frac{q_1 \vec{X}}{n_1} + \frac{q_2}{n_2} \right) \vec{X}_1 = \vec{X}_2 \pm T_c s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ where $s_p = \sqrt{\frac{(n_1-1)n_2 + (n_2-1)n_2}{p_1}}$ (poled formula) We pool when $n_1 \approx \sigma_2$ : $S_1 < S_2$ or $S_2 < 2S_1$ If don't pool use formula: $\vec{X}_1 = \vec{X}_2 \pm T_c \left( \frac{n_1 \vec{X}}{n_1} + \frac{x_2}{n_2} \right)$ $x_c$ : invnorm with area $\frac{1-q_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-q_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-q_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-q_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-q_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-q_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT with area $\frac{1-q_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : InvT (where $s_1, s_2, \sigma \in 1$ Interval. We usually use T $T_c$ : InvT (where $s_1, s_2, \sigma \in 1$ Interval. We usually use T $T_c$ : InvT ( $\frac{1}{n_1}, \frac{1}{n_2}, \frac{1}{n_1}, \frac{1}{n_2}, \frac{1}{n_2}, \frac{1}{n_1}, \frac{1}{n_2}, \frac{1}{n_2}, \frac{1}{n_2}, \frac{1}{n_2}, \frac{1}{n_2}, \frac{1}{n_2}, \frac{1}{n_1}, \frac{1}{n_2}, \frac{1}{n_1}, \frac{1}{n_2}, \frac{1}{n$	To Check Whether A Good Model  Linear Re Power  Exponential Interpretations  Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B" Bayes Theorem  Completely Randomise Design (CRD):  Final Completely Block Design: Randomized (complete) Block Design:	variation in y variable for a given amount of x variation in y variable for a given amount of x variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations y = ac <sup>2</sup> where x - logx, y - logy On calculator: Parking(L, L) or Linkeg (Og, L, Og, L) y = ab <sup>2</sup> where x - x - x y - log y On calculator: Parking(L, L) or Linkeg (Og, L, Og, L) y = ab <sup>2</sup> where x - x - x y - log y On calculator: Parking(L, L) or Linkeg (Og, L, Og, L) y = ab <sup>2</sup> where x - x y - log y On calculator: Parking(L, L) or Linkeg (Og, L, Og, L) y = ab <sup>2</sup> where x - x y - log y On calculator: Parking(L, L) or Linkeg (Og, L, Og, L) y = ab <sup>2</sup> where x - x y - log y On calculator: Parking (L, L) or Linkeg (Og, L) og L) y = ab <sup>2</sup> where y - probabilities and to a linker P(A) = <sup>m</sup> (A) = <sup>miniber</sup> of favourable outcomes mumber of passible outcomes P(A) = 0 P(A) = P(A)P(B) = P(A)P(B) P(A) = P(A)B) = P(A)P(B) P(A) = P(A)B) = P(A)P(B) P(A) = P(A)B) = P(A)P(B) P(A) = P(B) = P(A)P(A) P(B) = P(A)B) = P(A)P(A) P(B) = P(B)P(A)P(A) + P(B)A')P(A') periment Templates Each subject receives only one treatment # # frommat + Missin repro- set format + Miss	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcd with df (see test statistics section above for of for Tcdf)         Chi- Squared         Type 2 Error Steps: Step 1: Find C (using invnorm) c: area $= \alpha$ (left), $u=0, \sigma=1$ $*$ area $= \alpha$ (left), $u=0, \sigma=1$ $*$ area $= \alpha$ (left), $u=0, \sigma=1$ $*$ area $= \alpha$ (left), $u=0, \sigma=1$ $s$ tower=C/10, upper=C/2 $u=new \mu, \sigma=\frac{\sigma}{\sqrt{n}}$ $u=new \mu, \sigma=\frac{\sigma}{\sqrt{n}}$ $w = low = -C/1$ , upper=C/2 $u=new \mu, \sigma=\frac{\sigma}{\sqrt{n}}$ Unbiased Estimator ( $\overline{x}, S_x$ )         Quartiles       Being in the $p^n$ percentile means you did better than $p X$ of people. e.g. (lower quartile ( $S^n$ percentile) $P(X < \alpha) = 0.25$ $258$ were below the value a	upper tail if > test and in Goudie either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value if $x^2_{calc} = \sum_{i=1}^{i} \frac{(g-E)^2}{n}$ . Reject: $x^2_{calc} > x^2_{critical}$ Hypothesis, $u_i$ are independent/in the ratio/_identified of the other state of submitted of the other state other state of the other state of th
STATISTIC/ SAMPLE values as their values, not population values. Not first is proportion or mean and then is it 1 or 2 samples Use 2 if $\sigma$ known OR $n \ge 00$ Use T if $\sigma$ unknown OR $n \ge 00$ Use T if $\sigma$ samples unknown OR $n \ge 00$ Use T if $\sigma$ samples unknown OR $n \ge 00$ Use T if $\sigma$ unknown OR $n \ge 00$ Use T if $\sigma$ and the thermal samples of size of $\sigma$ Assumptions Interpretation of 2 sample C: (+, -) cannot say whether a diff $(-, -): 2^{rod} > 1^{rd}$ (+, +): 1 $r^{td} > 2^{rod}$	Tests, 1 Prop Z Interval Tests, 1 Prop Z Interval 2 <b>2 prop:</b> $\vec{p}_1 - \vec{p}_2 \pm x_c \left[ \frac{\beta_1 q_1}{m_1} + \frac{\beta_1 q_2}{m_2} \text{ where } \vec{p}_1 = \frac{x_1}{m_1}, \vec{p}_2 = \frac{x_2}{m_2} \text{ Tests, 2 Prop Z Interval}$ 1 <b>1 mean:</b> $\vec{X} \pm \chi_c \frac{q_1}{m_1} \otimes \vec{Q}_1 = \vec{Q}_1 \text{ (on known/ small sample)} \text{ Tests, 2 OR 1 Interval}$ 2 <b>2 means:</b> $\vec{X}_1 - \vec{X}_1 \pm \chi_c \frac{q_1}{m_1} + \frac{q_2}{m_2}$ ( $\vec{x}_1 - \vec{x}_2 \pm T, \vec{s}_p \sqrt{\frac{1}{m_1} + \frac{1}{m_2}}$ ) where $\vec{s}_p = \sqrt{\frac{1}{m_1 + \frac{1}{m_2}}} \frac{d(n - n - 1)}{m_1 + \frac{1}{m_2}}$ 2 <b>2 means:</b> $\vec{X}_1 - \vec{X}_1 \pm \chi_c \frac{q_1}{m_1} + \frac{1}{m_2}$ ( $\vec{X}_1 - \vec{X}_2 \pm T, \vec{s}_p \sqrt{\frac{1}{m_1} + \frac{1}{m_2}}$ ) where $\vec{s}_p = \sqrt{\frac{1}{m_1 + \frac{1}{m_2}} \frac{d(n - 1)}{m_1 + m_2}}$ (pooled formula) We pool when $\sigma_1 \approx \sigma_2$ : $S_1 < 2S_2$ or $S_2 < 2S_3$ If don't pool use formula: $\vec{X}_1 - \vec{X}_2 \pm T_c \sqrt{\frac{1}{m_1} + \frac{1}{m_2}}$ $\vec{x}_c$ : Invnorm with area $\frac{1-q_1}{m_1}$ , $d(n - 1, m_2 - 1)$ (no pool) Tests, 2 Samp Z or T Interval 2 <b>2 means paired</b> : $d \pm \frac{1}{T_c}$ ; $\frac{d}{m_1}$ Tests, 2 OR T Interval 2 <b>2 means paired</b> : $d \pm \frac{1}{T_c}$ ; $\frac{d}{m_1}$ Tests, 2 C R T Interval 3 <b>2 means paired</b> : $d \pm \frac{1}{T_c}$ ; $\frac{d}{m_1}$ Tests, 2 C R T Interval 4 <b>2 means paired</b> : $d \pm \frac{1}{T_c}$ ; $\frac{d}{m_1}$ Tests, 2 C R T Interval 5 <b>3 lope of regression</b> line (2 possible formula) 0 $\leq \text{ solpe}\pm T_c(\vec{S}_n)$ , Where $s_n$ is standard error (SE) 0 $\leq \text{ slope}\pm T_c \left(\frac{1}{m_1} + \frac{m_2}{m_2}\right)$ <b>1 Indegendent</b> : $N \ge 10$ for $n \le 1.0n$ (pop size is 10 times sample) size is 10 percent of pop size) <b>1 Normality:</b> <b>1 Proportion:</b> $\frac{n}{n} \ge 30$ Note: No for $n \ge 30$ <b>2 Normality:</b> <b>1 ONE INTERVAL In context/ lying BETWEEN 2 values: We are. Sc onfident that the time roug polyines is difference intervals the we compute would contain the true polying. We rest is the spectrum may this, about, S d h the confidence intervals the we compute would contain the true polying. The confidence interval that we compute would contain the true polying. The confidence interval the vecompute would contain the true pol</b>	To Check Whether A Good Model Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Completely Randomise Design (CRD): Bayes Theorem Completely Randomise Design (CRD): Linear Re Randomized (complete) Block Design: Randomized (complete) Block Design:	variation in y variable for a given amount of x variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations y = ax <sup>8</sup> where x → logx, y → logy On calculator: Pwreg(L, L, L) or Linkeg (og L, log L) y = ab <sup>x</sup> where x → x, y → log y On calculator: ExpReg(L, L, L) or Linkeg (og L, log L) Transformed scatter must look linear Residual plot of TRANSFORMED no pattern starry night Probability P(A) = $\frac{n(L)}{n(L)} = \frac{number of favourable automes}{number of favourable automes}$ P(A) = $\frac{n(L)}{n(L)} = \frac{number of favourable automes}{number of patometable automes}$ P(A) = $\frac{n(L)}{n(L)} = \frac{number of favourable automes}{number of patometable automes}$ P(A) = $\frac{n(L)}{n(L)} = \frac{number of favourable automes}{number of pattern starry night}$ P(A) = $\frac{n(L)}{n(L)} = \frac{number of favourable automes}{number of patometable automes}$ P(A) = $\frac{n(L)}{n(L)} = \frac{n(L)P(A) = 0}{n(L)P(B)}$ addition rule becomes: P(AUB)=P(A) + P(B) P(A) = $\frac{P(A B) = P(A)P(B)}{P(B A)P(A)}$ P(A B) = $\frac{P(B A)P(A)}{P(B A)P(A)}$ periment Templates Each subject receives only one treatment ** {Transat} + Maxare represe ** {Transat} + Maxar	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcd with df (see test statistics section above for of for Tcdf) Chi-Squared	upper tail if > test and in Goudie either tail if $\neq$ test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2$ calc: $= \sum_{i=1}^{i} \frac{(Q-E)^2}{n}$ . Reject: $x^2$ and $x^2$ critical Hypothesis, $H_i$ are independent in the ratio
STATISTIC/ SAMPLE values as their values, not population values. Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use Z if $\sigma$ known OR $n \ge 30$ Use Z if $\sigma$ unknown OR $n \ge 30$ Use Z if $\sigma$ <b>a</b> samples unknown OR $n \ge 30$ Use Z if $\sigma$ <b>b</b> Stats: $\sigma$ unknown OR $n \ge 30$ Use Z if $\sigma$ <b>b</b> Stats: <b>b</b> Stats: $\sigma$ unknown OR $n \ge 30$ Use T if $\sigma$ <b>b</b> Stats: <b>b</b> Sta	Tests, 1 Prop Z interval Tests, 2 Prop Z interval 2 <b>2 prop:</b> $\vec{p}_1 = \vec{p}_2 \pm x_c \left( \frac{\vec{p}_1 \cdot \vec{q}_1 + \vec{p}_2 \cdot \vec{q}_3}{n_1 + n_2 \cdot \vec{q}_3} \text{ where } \vec{p}_1 = \frac{x_1}{n_1}, \vec{p}_2 = \frac{x_1}{n_2} \right)$ Tests, 2 Prop Z interval 1 <b>mean:</b> $\vec{X} \pm \zeta_c = \frac{\sigma_1}{\sigma_1} (\sigma \text{ unknown}/ small sample)$ Tests, 2 OR 1 Interval $z_c: \text{Invnorm with area} \frac{1-\sigma_0}{2}, \mu = 0, \sigma = 1$ $T_c: \text{ InvT with area} \frac{1-\sigma_0}{2}, \frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}, \vec{X}_1 - \vec{X}_2 \pm T_c \cdot \vec{p}_0 \frac{1}{n_1} + \frac{1}{n_2}$ where $s_p = \sqrt{\frac{(\alpha_1 - 1)s_1^2 + (\alpha_2 - 1)s_2^2}}$ (pooled formula) We pool when $\sigma_1 \approx \sigma_2: S_1 < 2S_2$ or $S_2 < 2S_1$ If fort pool use formula: $\vec{X}_1 - \vec{X}_2 \pm T_c \frac{s_1^2}{n_1^2} + \frac{s_2^2}{n_2}$ $z_c: \text{Invnorm with area} \frac{1-\sigma_0}{n_1 n_1 + n_2}, \mu = 0, \sigma = 1$ $T_c: \text{InvT with area} \frac{1-\sigma_0}{n_1 n_1 + n_2}, \mu = 0, \sigma = 1$ $T_c: \text{InvT with area} \frac{1-\sigma_0}{n_1 n_1 - n_2}, \mu = 0, \sigma = 1$ $T_c: \text{InvT with area} \frac{1-\sigma_0}{n_1 n_1 - n_2}, \mu = 0, \sigma = 1$ $T_c: \text{InvT with area} \frac{1-\sigma_0}{n_1 n_1 - n_2}, \mu = 0, \sigma = 1$ $T_c: \text{InvT with area} \frac{1-\sigma_0}{n_1 n_2}, \frac{\sigma_1 - 1}{n_2} = 0$ (no pool) Tests, 2 Samp 2 or T interval 2 <b>Imeans paired:</b> $d \pm T_c; \frac{d}{m_1}$ Tests, 2 C RT Interval. We usually use T $T_c: \text{InvT with area} \frac{1-\sigma_0}{n_1 - n_2}, \frac{\sigma_1 - n_1}{n_2}$ <b>Slope of regression line</b> (2 possible formulae) $\circ \text{ slope} T_c(s_0), \text{ where } s_1$ is standard error (SE) $\circ \text{ slope} T_c(s_0), \text{ where } s_1$ is standard error (SE) $\circ \text{ slope} T_c(s_0), \text{ where } s_1$ is standard error (SE) $\sigma  sample size or sample size is 10 percent of pop size)$ <b>Proportion:</b> $\frac{n \geq 30}{n c a 30}$ $n (1 - \vec{p}) \geq 10$ $n \geq 30$ Note: Do for $n \geq 30$ Note: Do	To Check Whether A Good Model  Linear Re Power Exponential Interpretations Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B" Bayes Theorem Excompletely Randomise Design (CRD): Completely Randomise Design (CRD):	variation in y variable for a given amount of x variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value High r <sup>2</sup> value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations y = x <sup>5</sup> where x - 1 (agx, y - 1 (agy) On calculator: Pwreg(L, L) or Linkeg (log, L) (agy) On calculator: Pwreg(L, L) or Explore (L) (agy) On calculator: Pwreg(L, L) or Explore (L) (agy) On calculator: Pwreg(L, L) or Explore (L) (bas) Pr(A = 1.4(A) = number of favourable outcomes number of possible autcomes number of possible autcomes (P(A) = 1.4(A) = number of favourable outcomes number of possible autcomes (P(A) = 1.4(A) = number of favourable outcomes number of possible autcomes (P(A) = 1.4(A) = number of favourable outcomes number of possible autcomes (P(A) = number of favourable outcomes (P(A) = number of favourable outcomes (P(A) = 0.4(B) = number of favourable outcomes (P(A) = num	with $\mu$ and $\frac{\sigma}{\sqrt{\pi}}$ or Tcd with df (see test statistics section above for of for Tcdf) Chi- Squared Errors Type 2 Error Steps: Step 1: Find CV (using invorm) <: area = a (left), u=0, 0, or 1 :: area = a [left), u=0, 0, or 1 :: area = a [left], u=0, or 1 :: bower-CV, upper-CV u=new $\mu, \sigma_{eff}^{-}$ :: bower-CV, upper-CV :: bower-CV,	upper tail in 2 less tand in 6000ie either tail in 2 less tand in 6000ie either tail in 2 less tand in 6000ie either tail in 2 less test test) the p value Larger sample size $\Longrightarrow$ smaller p value If $x^2$ cuite $\sum \frac{(G-E)^2}{2}$ . Reject: $x^2$ cuits $z^2$ critical HypothesisH, are independent/in the ratiodistributed $H_1$ ; are not independent/in the ratiodistributed $H_2$ ; are not independent/in the ratiodistributed $H_1$ ; are not independent/in the ratiodistributed $H_2$ ; are not independent and n-1 to others of an -2 if approximating por $\mu/\nu$ but this newer comes up Definitions: Joi on dependence and n-1 to others of the ratio_distributed in the critical region (Type 2: H, fishe, but we say it is fable i.e. reject it Calculations: Type 1: Ho, this is a (prob of rejecting the test i.e. being in the critical region) Type 2 = $p^2$ rob of NOT being in the critical region see blue on left for the method of ow to find this. Note: There is an alternative method which is harder. Only use this method if you're forced to. Re-arrange $z = \frac{F_{xx}}{F_{xx}}$ to get $\overline{z} = \mu + z_{x} \frac{\sigma}{\sigma_{x}}$ $: P(\overline{x} > \mu + z_{x} \frac{\sigma}{\sigma_{x}})$ where $\mu$ -original mean $z_{x}$ =CV Lowers $\mu + z_{x} \frac{\sigma}{\sigma_{x}}$ upper $= \mu + z_{x} \frac{\sigma}{\sigma_{x}}$ unewe $\mu, \sigma \frac{\sigma}{\sigma_{x}}$ $: P(\overline{x} > \mu + z_{x} \frac{\sigma}{\sigma_{x}})$ where $\mu$ -original mean $z_{x}$ =CV Lowers $\mu + z_{x} \frac{\sigma}{\sigma_{x}}$ upper $= \mu - z_{x} \frac{\sigma}{\sigma_{x}}$ unewe $\mu, \sigma \frac{\sigma}{\sigma_{x}}$ $To increase/Decrease the there. To increase T_{x} = \frac{F_{x}}{\sigma_{x}} upper = \mu - z_{x} \frac{\sigma}{\sigma_{x}} unewe \mu, \sigma \frac{\sigma}{\sigma_{x}} To increase/Decrease the there T_{x} H in o frequency: \overline{x} = \frac{\Sigma x^2}{2T} - \overline{x}^2 = \frac{\Sigma (x - \overline{x})^2}{2T} H in o frequency: \overline{x} = \frac{\Sigma (x - \overline{x})^2}{2T} - \overline{x} = \frac{T_{x} - \overline{x}}{2T} \sigma = \sqrt{variance} S_{xx} = \sum (x_{x} - \overline{x}) = \frac{\Sigma (x - \overline{x})^2}{2T} - \overline{x} = \frac{T_{x} - \overline{x}}{2T} \overline{x} = \frac{T_{x} - \overline{x}}{2T} - \frac{T_{x} - \overline{x}}{2T} = \frac{T_{x} - \overline{x}}{2T} \overline{x} = \frac{T_{x} - \overline{x}}{2T} - \frac{T_{x} - \overline{x}}{2T} S_{xy} = \sum xy - (\frac{\Sigma x})^2 - \frac{\Sigma x^2}{2T} - \frac{\Sigma x^2}{2T} - \frac{T_{x} - \overline{x}$
STATISTIC/ SAMPLE values as their values, not population values. Ask yourself first: Is proportion or mean and then is it 1 or 2 samples Use 2 if $\sigma$ unknown OR $n \ge 30$ Use 7 if $\sigma$ unknown OR $n \ge 30$ Use 7 if $\sigma$ unknown OR $n \ge 30$ Use 7 if $\sigma$ unknown OR $n \ge 30$ O use 7 if $\sigma$ unknown OR $n \ge 30$ O use 7 if $\sigma$ unknown always, so can always use T for mean tests. We pool if T Testand $\sigma_1 \approx \sigma_2$ , but you will also notice that the mark scheme normally deen't pool regardless of size of $\sigma$ Assumptions Interpretation of 2 sample C1: (+, -) cannot say whether a diff (-, -): 2 <sup>-rd</sup> > 1 <sup>st</sup> (+, +): 1 <sup>st</sup> > 2 <sup>rd</sup> Decrease C1 Hints for	<b>V u</b> Tests, 1 Prop Z interval <b>2</b> prop: $\vec{p}_1 = \vec{p}_2 \pm x_c$ $\left[\frac{\vec{p}_1 \vec{q}_1}{n_1} + \frac{\vec{p}_2}{n_2} \text{ where } \vec{p}_1 = \frac{x_1}{n_1}, \vec{p}_2 = \frac{x_2}{n_2}$ Tests, 2 Prop Z interval <b>1</b> mean: $\vec{X} \pm Z_c = \frac{\sigma_1}{\sigma_1} (\sigma \text{ unknown/small sample)$ Tests, 2 CR T Interval $z_c$ : invrorm with area $\frac{1-\sigma_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : invT with area $\frac{1-\sigma_0}{2}, \mu = 0, \sigma = 1$ $T_c$ : invT with area $\frac{1-\sigma_0}{2}, fx_1 - \vec{X}_2 \pm T_c \cdot \vec{s}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ where $s_p = \sqrt{\frac{\alpha_1 + \beta_2 + \alpha_2}{n_1 + \sigma_2 - 2}} (x_1 - \vec{X}_2 \pm T_c \cdot \vec{s}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ where $s_p = \sqrt{\frac{\alpha_1 + \beta_2 + \alpha_2}{n_1 + \sigma_2 - 2}} (polded formula)$ We pool when $\sigma_1 \approx \sigma_2$ : $S_1 < 2S_2$ or $S_2 < 2S_1$ If don't pool use formula: $\vec{X}_1 - \vec{X}_2 \pm T_c \sqrt{\frac{1}{n_1^2} + \frac{\alpha_2^2}{n_2}}$ $z_c$ : invorm with area $\frac{1-\sigma_0}{2}, fx_1 - n_2 - 2$ (no pool)         Tests, 2 Samp 2 or T interval <b>2 means paired</b> : $d \pm T_c \cdot \frac{\sigma_0}{2\pi}$ Tests, 2 CR T Interval <b>2 means paired</b> : $d \pm T_c \cdot \frac{\sigma_0}{2\pi}$ Tests, 2 CR T Interval <b>5 slope of regression line</b> (2 possible formulae) $\sigma$ slope $\pm T_c ((-\sigma^{-2})s_2^{-2})$ where $s_n$ so are the S.D.'s $T_c$ : InvT ( $\frac{1-\sigma^2}{2}, \frac{\sigma_1}{(n-2)s_2^{-2}}$ where $s_n$ so are the S.D.'s $T_c$ : InvT ( $\frac{1-\sigma^2}{2}, \frac{\sigma_1}{(n-2)s_2^{-2}$	To Check Whether A Good Model  Linear Re Power  Exponential Interpretations  Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B" Bayes Theorem  Completely Randomise Design (CRD):  Figurementations  Randomized (complete) Block Design: Randomized (complete) Block Design: Matched Pairs Design: Each subject ge	variation in y variable for a given amount of x variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations y = ac <sup>3</sup> where x → logx, y → logy On calculator: Payleg(L, L) or Linkeg (og L, log L) y = ab <sup>2</sup> where x → logx, y → logy On calculator: Payleg(L, L) or Linkeg (og L, log L) y = ab <sup>2</sup> where x → logx, y → logy On calculator: Payleg(L, L) or Linkeg (og L, log L) y = ab <sup>2</sup> where x → logx, y → logy On calculator: Payleg(L, L) or Linkeg (og L, log L) Transformed scatter must look linear Residual plot of TRANSFORMED no pattern starry night P(A) = <sup>m(L)</sup> P(A) = <sup>m(L)</sup> P(A) = <sup>m(L)</sup> P(A) = <sup>m(L)</sup> P(A ∩ B) = 0 addition rule becomes: P(AUB)=P(A)+P(B) P(A ∩ B) = P(A)P(B) P(A ∩ B) = P(A)P(B) Addition rule becomes: P(AUB)=P(A)+P(B)=P(A)P(B) P(A B) = <sup>m(L)</sup> P(B A)P(A) + P(B A')P(A') pr(B A)P(A) + P(B A')P(A') prepriment Templates Each subject receives only one treatment <sup>max</sup> frommat} for more some flows resume flows	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcdf with df (see test statistics section above for of for Tcdf) Chi-Squared Errors Type 2 Error Steps: Step 1: Find CV (using knorm) c. area $\approx \alpha$ (left), $\mu$ =0, $\sigma$ =1 $\approx 2 = \alpha \alpha$ (left), $\mu$ =0, $\sigma$ =1 $\approx 2 = \alpha \alpha$ (left), $\mu$ =0, $\sigma$ =1 $\approx 2 = \alpha \alpha$ (left), $\mu$ =0, $\sigma$ =1 Step 2: Find error (normcdf) c. lower=-100, upper=CV $u = \alpha \omega \mu$ , $\sigma^{-}_{\sqrt{\alpha}}$ $\approx 1 = 10 = 10^{-1}$ (left), $u$ =0, $\sigma$ =1 Step 2: Find error (normcdf) c. lower=-100, upper=CV $u = \alpha \omega \mu$ , $\sigma^{-}_{\sqrt{\alpha}}$ $\approx 1 = 10 = 10^{-1}$ (left), $u$ =0, $\sigma$ =1 Mean Mean Mean Variance Note: can also use formula $\frac{s_{xx}}{n}$ Standard Dev $s_{xx}$ Unbiased Estimator ( $\overline{x}$ , $S_x$ ) Quartiles Being in the $\beta^n$ percentile means you dia better than $\beta$ is of people. e.g. lower quartile ( $\overline{S}$ percentile) $P(X < \alpha) = 0.25^{-2}$ Z5% were below the value a Shape Centre	upper tail in 2 less tand in Bouble either tail in 2 less test) the p value Larger sample size $\Rightarrow$ smaller p value If $x^2_{calc} = \sum_{i=1}^{i} \frac{(0-E)^2}{\pi}$ . Reject: $x^2_{calc} > x^2_{critical}$ Hypothesis, at method modent/in the ratio/_distributed dir(movil)(columns-1) for independence and n=1 for other dir = n < 21 approximating p or $\mu/c$ but this newer comes up <b>Definitions:</b> Type 1: $H_0$ true, but we say it is false i.e. reject it Type 2: $H_0$ true, but we say it is false i.e. reject it Type 2: $H_0$ true, but we say it is false i.e. reject it <b>Calculations:</b> Type 1: $H_0$ true, but we say it is false i.e. reject it <b>Calculations:</b> Type 2: $P(\nabla cold NOT being in the critical region-see blue on left for the method of ow to find this Note: There is an alternative method which is harder. Only use this method if you're forced to. Re-arrange z = \frac{T_{eac}}{\pi} to get \overline{x} = \mu + z_c \frac{\sigma}{\pi}> P(\overline{X} > \mu + z_c \frac{\sigma}{\pi}) where \mu-original mean, z_e CVLower: \mu + z_c \frac{\sigma}{\pi} upper 100, \mu-new \mu, \sigma - \frac{\sigma}{\pi}\Rightarrow P(\overline{X} > \mu + z_c \frac{\sigma}{\pi}) upper 100, \mu-new \mu, \sigma - \frac{\sigma}{\pi}To increase/Decrease The Errors:Power: initions bod fyre 2 = 1 - \betaIncrease type 3 increase 5.DatIf no frequency: a^2 = \frac{\Sigma_c}{2T} - a^2 = \frac{\Sigma(x-x)^2}{\pi}If frequency: a^2 = \frac{\Sigma_c}{2T} - a^2 = \frac{\Sigma(x-x)^2}{\pi}If no frequency: a^2 = \frac{\Sigma_c}{2T} - a^2 = \frac{\Sigma(x-x)^2}{\pi}a = \sqrt{variance}S_{xx} = \sum (x_1 - x_2)^2 = \sum x_1^2 - \frac{(\Sigma_c x)^2}{\pi} - \frac{\pi}{\pi}S_{xy} = \sum x_y - \frac{(\Sigma_x)^2}{\pi} - \frac{(\Sigma_x)^2}{\pi} - \frac{\Sigma_c}{\pi}E = \frac{\Sigma_x}{\pi} \cdot S_x = \sqrt{\frac{\Sigma_x}{2T} - \frac{2}{\pi}} = \frac{\Sigma_c}{\pi} + \frac{\pi}{\pi}Upper Quartile a = \frac{\pi}{\pi} or \frac{25m}{\pi}Lower Quartile a = \frac{\pi}{100} or \frac{25m}{\pi}Lower Amedian-Mode or 0, -0, -0, -0,$
STATISTIC/ SAMPLE values as their values, not population values. Not first is proportion or mean and then is it 1 or 2 samples Use 2 if $\sigma$ known OR $n \ge 30$ Use 2 if $\sigma$ unknown OR $n \ge 30$ Use 2 if $\sigma$ samples unknown OR $n \ge 30$ Use 2 if $\sigma$ <b>a</b> samples of $n < 30$ AP Stats: $\sigma$ unknown OR n < 30 AP Stats:	$\begin{aligned} \mathbf{y}_{1}^{\mathbf{y}_{1}} & \mathbf{x}_{1}^{\mathbf{y}_{1}} & \mathbf{y}_{1}^{\mathbf{x}_{1}} & \mathbf{y}_{1}^{\mathbf{x}_{2}} & \mathbf{y}_{1}^{\mathbf{x}_{1}} & \mathbf{y}_{1}^{\mathbf{x}_{2}} & \mathbf{y}_{1}^{\mathbf{x}_{1}} & \mathbf{y}_{2}^{\mathbf{x}_{2}} & \mathbf{y}_{1}^{\mathbf{x}_{2}} & \mathbf{y}_{1}^{\mathbf{x}_{2}$	To Check Whether A Good Model  Linear Re Power  Exponential Interpretations  Probability of event A Complementary Events Combined Events (Addition Rule) Mutually Exclusive Events Independent Events Conditional "A given B" Bayes Theorem  Completely Randomise Design (CRD):  Feyerimental unity  Matched Pairs Design: Each subject ge  Feyerimental unity  Feyerimental u	variation in y variable for a given amount of x variation in y variable for a given amount of x variable Scatter plot looks like a straight line High r value Residual plot has no pattern with uniform variation across x (starry night) gression - Transformations y = ax <sup>2</sup> where x → logx, y → logy On calculator: Pwreg(L, L, ) or Lindeg (log, L, log L, log y) = ab <sup>2</sup> where x → x, y → logy On calculator: Pwreg(L, L, ) or Lindeg (log, L, log L, log L, log L, log L, log L, log L, log L, log L, log y) = ab <sup>2</sup> where x → x, y → logy On calculator: Pwreg(L, L, ) or Lindeg (log, L, log L, log x) = ab <sup>2</sup> where x → x, y → logy On calculator: Pwreg(L, L, ) or Lindeg (log, L, log L, log L, log L, log y) = ab <sup>2</sup> where x → x, y → logy On calculator: Pwreg(L, L, ) or Lindeg (log, L, log L,	with $\mu$ and $\frac{\sigma}{\sqrt{n}}$ or Tcd with df (see test statistics section above for of for Tcdf) Chi-Squared Frors Type 2 Error Steps: Step 1: Find CV(apple, 0, $\sigma$ =1 $\Rightarrow$ : area = $\alpha$ (left), $u$ =0, $\sigma$ =1 $\Rightarrow$ : area = $\alpha$ (left), $u$ =0, $\sigma$ =1 $\Rightarrow$ : area = $\alpha$ (left), $u$ =0, $\sigma$ =1 $\Rightarrow$ : area = $\alpha$ (left), $u$ =0, $\sigma$ =1 $\Rightarrow$ : area = $\alpha$ (left), $u$ =0, $\sigma$ =1 Step 2: Find error (normcdf) <li>clower=-100, upper=CV <math>u=new \mu, \sigma = \frac{\sigma}{\sqrt{n}}</math> <math>\Rightarrow</math>: Lower=CV, upper=CV <math>u=new \mu, \sigma = \frac{\sigma}{\sqrt{n}}</math> Mean Variance Note: can also use formula <math>\frac{s_{xx}}{n}</math> Standard Dev <math>S_{xx}</math> Unbiased Estimator (<math>\overline{x}, S_x</math>) Quartiles Being in the <math>\beta^n</math> percentile means you did better than <math>\beta</math> x of people. e.g. lower quite (<math>\overline{S}^n</math> percentile) P(X &lt; a) = 0.25 25% were below the value a Shape Centre Spread</li>	upper tail in 2 lets 1 and in Bouble either tail in 2 lets the p value Larger sample size $\Rightarrow$ smaller p value If $x^2_{calc} = \sum_{i=1}^{i} \frac{(Q-E)^2}{\pi}$ . Reject: $x^2_{calc} > x^2_{critical}$ Hypothesis, it, are independent in the ratio_i distributed of (m) H <sub>i</sub> , are not independent in the ratio_i. distributed of (m) H <sub>i</sub> , are not independent in the ratio_i. distributed of (m) H <sub>i</sub> , are not independent in the ratio_i. distributed of (m) H <sub>i</sub> , are not independent in the ratio_i. distributed of (m) H <sub>i</sub> , are not independent in the ratio_i. distributed of (m) H <sub>i</sub> , are not independent in the ratio_i. distributed of (m) H <sub>i</sub> , are not independent in the ratio_i. distributed of (m) H <sub>i</sub> , are not independent in the ratio_i. distributed of (m) H <sub>i</sub> are not independent in the ratio_i. distributed of (m) H <sub>i</sub> are not independent in the ratio_i. distributed of (m) H <sub>i</sub> are not independent in the ratio_i. distributed of (m) H <sub>i</sub> are not independent in the ratio_i. distributed of (m) H <sub>i</sub> are not independent in the ratio_i. distributed of (m) H <sub>i</sub> are not independent in the ratio_i. A state <b>State</b> (m) H <sub>i</sub> the same independent in the ratio independent (m) use this method if you're forced to. Re-arrange z = $\frac{T_{ex}}{\pi}$ to get $\overline{x} = \mu + z_e \frac{\sigma}{\pi}$ $\approx P(\overline{x} > \mu + z_e \frac{\sigma}{\pi})$ where $\mu$ -original mean, $z_e CV$ Lower: $\mu + z_e \frac{\sigma}{\pi}$ upper 100, unenew $\mu, \sigma \frac{\sigma}{\pi}$ $\Rightarrow P(\overline{x} > \mu + z_e \frac{\sigma}{\pi})$ where $\mu$ -original mean, $z_e CV$ Lower: $\mu = z_e \frac{\sigma}{\pi}$ by the re user signal mean, $z_e CV$ Lower: $\mu = z_e \frac{\sigma}{\pi}$ by the re user signal mean, $z_e CV$ Lower: $\mu = z_e \frac{\sigma}{\pi}$ by the re user signal mean, $z_e CV$ Lower: $\mu = z_e \frac{\sigma}{\pi}$ by the re $\mu = original mean, z_e CV$ Lower: $\mu = z_e \frac{\sigma}{\pi}$ by the re $\mu = original mean, z_e CV$ Lower: $\mu = z_e \frac{\sigma}{\pi}$ if frequency: $\overline{x} = \frac{\Sigma T}{\Sigma T}$ If no frequency: $\overline{x} = \frac{\Sigma T}{T} = \overline{x} = \frac{\Sigma T}{T}$ $\overline{x} = \frac{\Sigma L}{\Sigma T} = \frac{\Sigma L}{T} = \frac{\Sigma T}{T}$ $\overline{x} = \frac{\Sigma L}{\Sigma T} = \frac{\Sigma L}{T} = \frac{\Sigma L}{T}$ $\overline{x} = \frac{\Sigma L}{\Sigma T} = \frac{\Sigma L}{T} = \frac{\Sigma L}{T} = \frac$