

1) a) Axioms are $\underline{P(E) \geq 0}$ for any event E
 (in an event space \mathcal{F} - not required)

$$\underline{P(\Omega) = 1}$$

$$\underline{\text{If } E \cap F = \emptyset, P(E \cup F) = P(E) + P(F)}$$

b) i) A random variable is a function from the sample space to the real numbers: $X: \Omega \rightarrow \mathbb{R}$

ii) Let $F(x) = P(X \leq x)$ be the distribution function of a random variable X . X is continuous iff $F(x)$ is continuous & differentiable $\forall x$.

c) ~~Problem~~ $P(X \leq b) = P((X \leq a) \cup (a < X \leq b))$ when $b > a$
 $= P(X \leq a) + P(a < X \leq b)$ (axiom 3)

But $P(X \leq b) = F(b)$, $P(X \leq a) = F(a)$ by definition.
 Hence $\underline{P(a < X \leq b) = F(b) - F(a)}$ as required

From above, we have $F(b) = F(a) + P(a < X \leq b)$ ($b > a$)
 and $P(a < X \leq b) \geq 0$ (axiom 1)

Hence, when $b > a$ we have $F(b) \geq F(a)$, & $\underline{F(\cdot)}$ is nondecreasing as required.

d)
$$P(X=x) = \lim_{\varepsilon \rightarrow 0} P(x-\varepsilon < X \leq x)$$

$$= \lim_{\varepsilon \rightarrow 0} (F(x) - F(x-\varepsilon))$$

$$= \underline{0}$$
 For a continuous random variable, since by definition F is continuous in this case.

2) a) Density is $f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$

Obⁿ fⁿ is $F(x) = \int_{-\infty}^x f(u) du$
 $= \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$

3

b) Obⁿ fⁿ of X is $F(x) = P(X \leq x) = P(-\log U^2 \leq x)$
 $= P(\log U \geq -x/2)$
 $= P(U \geq e^{-x/2})$
 $= 1 - F_U(e^{-x/2})$
 $= 1 - e^{-x/2} \quad (x > 0)$

This is the obⁿ fⁿ of an exponential obⁿ with parameter 1/2.

Density of X is $f(x) = \frac{1}{2} e^{-x/2} \quad (x > 0)$

Hence MGF is $M_X(t) = E(e^{tx}) = \int_0^{\infty} e^{tx} \cdot \frac{1}{2} e^{-x/2} dx$
 $= \frac{1}{2} \left[\frac{e^{(t-1/2)x}}{t-1/2} \right]_0^{\infty} = \frac{(1-2t)^{-1}}{\uparrow}$ (not required) 5

c) $S = -\sum_{i=1}^n \log U_i^2 = \sum_{i=1}^n X_i$, where $X_i \sim \text{Exp}(1/2)$ independently for each i

Hence $M_S(t) = \prod_{i=1}^n M_{X_i}(t) = (1-2t)^{-n}$ 2

d) Comparing $M_S(t)$ with the expression given, we have equality if $n = m/2$. Hence $m = 2n$ & $S \sim \chi_{2n}^2$. 2

e) $P(\prod_{i=1}^6 U_i > 0.1) = P(\prod_{i=1}^6 U_i^2 > 0.01)$
 $= P(-\sum_{i=1}^6 \log U_i^2 < -\log 0.01)$ $\log_{10} 10^{-2}$
 $= P(Y < 4.61)$ where $Y \sim \chi_{12}^2$
 $\approx (0.8 \times 0.0274) + (0.2 \times 0.0420)$
 (linear interpolation in Table 7)

χ^2
 11 $\left\{ \begin{array}{l} 0.95 \rightarrow 4.2 \\ 0.975 \rightarrow 5.2 \end{array} \right.$

$= \frac{0.030}{\text{TOTAL}}$
 0.95 0.975
 4.2 5.2

7
 75

$$32) a) i) P(X < 2) = P(X=0) + P(X=1) = \underline{e^{-\lambda}(1+\lambda)}$$

$$\begin{aligned} ii) P(Y < X | X < 2) &= P(Y < X \cap X < 2) / P(X < 2) \\ &= P(Y=0 \cap X=1) / P(X < 2) \\ &= P(Y=0)P(X=1) / P(X < 2) \text{ (independence)} \\ &= e^{-\lambda} \cdot \lambda e^{-\lambda} / (e^{-\lambda}(1+\lambda)) = \underline{\frac{\lambda e^{-\lambda}}{1+\lambda}} \text{ as req'd} \end{aligned}$$

iii) ~~But~~ X & Y are independent with the same distribution; therefore the event " $X=r, Y=k$ " has the same probability as " $X=k, Y=r$ ". Thus

$$P(Y < X) = \sum_{k=0}^{\infty} \sum_{r=k+1}^{\infty} P(X=r, Y=k) = P(X < Y)$$

(NB full argument not required - just some awareness of symmetry)

$$\text{Now } P(Y < X) = 1 - P(X \leq Y) = 1 - P(X=Y) - P(X < Y)$$

$$\Rightarrow 2P(Y < X) = 1 - P(X=Y), \quad P(Y < X) = \frac{1}{2}[1 - P(X=Y)] \text{ as req'd}$$

$$ii) P(X=Y) = \sum_{k=0}^{\infty} P(X=Y | Y=k) P(Y=k) \quad (\text{L \& T P})$$

$$= \sum_{k=0}^{\infty} P(X=k | Y=k) P(Y=k)$$

$$= \sum_{k=0}^{\infty} P(X=k) P(Y=k) \text{ (independence)}$$

$$= \sum_{k=0}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!} \cdot \frac{e^{-\lambda} \lambda^k}{k!} = \sum_{k=0}^{\infty} \frac{e^{-2\lambda} \lambda^{2k}}{(k!)^2} \text{ as req'd}$$

$$\text{When } \lambda=1, \quad P(Y < X) = \frac{1}{2}(1 - 0.309) = \underline{0.346}$$

$$v) P(Y < X) = P(Y < X | X < 2) P(X < 2) + P(Y < X | X \geq 2) P(X \geq 2) \quad (\text{L \& T P})$$

Rearranging gives the result.

$$\text{When } \lambda=1 \text{ we have } P(Y < X | X < 2) = \frac{1}{2} e^{-1} = 0.184 \text{ from (ii)}$$

$$P(X < 2) = 2e^{-1} = 0.736 \text{ from (i)}$$

$$\begin{aligned} \text{Hence } P(Y < X | X \geq 2) &= \frac{[0.346 - (0.184 \times 0.736)]}{[1 - 0.736]} \\ &= \underline{0.798} \end{aligned}$$

- b) Let X be the # of accidents at some location in December; then $X \sim \text{Poi}(1)$. Let Y be the # of accidents at the same location in January, and suppose the speed cameras have no effect. Then $Y \sim \text{Poi}(1)$ regardless of whether a camera has been installed. However, cameras have only been installed at locations where $X \geq 2$. Therefore the probability of fewer accidents at these locations, if speed cameras have no effect, is

$$P(Y < X | X \geq 2) = 0.798$$

So if speed cameras have no effect, you'd expect around ~~80%~~ 80% of camera sites to show a reduction — which is exactly what has been observed. The government's conclusion is not justified.

Other intelligent / aware comments will be rewarded.
Fatuous waffle will be ignored.

4 a) i) Let X be the GA-3 concentration; then $X \sim N(525, 15^2)$

$$\begin{aligned}
 P(500 < X \leq 550) &= P\left(\frac{500-525}{15} < Z \leq \frac{550-525}{15}\right) \\
 &= P(-1.667 < Z \leq 1.667) = 2\Phi(1.667) - 1 \\
 &= (2 \times 0.9525) - 1 \quad (\text{Table 4}) \\
 &= \underline{0.905}
 \end{aligned}$$

5

ii) Let Y be the # of batches allowing germination; then $Y \sim \text{Bin}(4, 0.905)$.

$$\begin{aligned}
 P(Y=4) &= 0.905^4 = \underline{0.671} \\
 P(Y=3) &= \binom{4}{3} 0.905^3 (1-0.905) = \underline{0.282}
 \end{aligned}$$

4

b) Require b : $P(-25/b < Z \leq 25/b) = 0.99$

$$2\Phi(25/b) - 1 = 0.99 \Rightarrow \Phi(25/b) = 0.995$$

Hence $25/b = 2.5758$ (Table 5) & $b = \underline{9.71 \text{ ppm}}$

5

c) In this case, $X \sim N(500, 20^2)$ & $P(500 < X \leq 550) = P(0 < Z \leq 2.5)$

$$= \Phi(2.5) - \Phi(0) = 0.99379 - 0.5 = 0.49379$$

$$P(Y=3) = \binom{4}{3} (0.49379)^3 (1-0.49379) = \underline{0.244}$$

5

d) Let E be the event "half selected contains new solution". Then, unconditionally, $P(E) = 1/2$.

$$\begin{aligned}
 \text{We require } P(E|Y=3) &= \frac{P(Y=3|E)P(E)}{P(Y=3|E)P(E) + P(Y=3|E^c)P(E^c)} \\
 &\quad (\text{Bayes})
 \end{aligned}$$

$$= \frac{(0.282 \times 1/2)}{(0.282 \times 1/2) + (0.244 \times 1/2)} = \underline{0.536}$$

6

TOTAL

25

5) a) The LS estimates minimize $S(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$

$$\frac{\partial S}{\partial \beta_0} = -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)$$

$$= 0 \text{ when } \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) = 0 \Rightarrow n\bar{y} - n\beta_0 - n\bar{x} = 0$$

Hence $\hat{\beta}_0$ satisfies $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$, where $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$\frac{\partial S}{\partial \beta_1} = -2 \sum_{i=1}^n x_i (y_i - \beta_0 - \beta_1 x_i)$$

$$= 0 \text{ when } \sum_{i=1}^n x_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

At the minimum we have $\sum_{i=1}^n x_i (y_i - \bar{y} - \hat{\beta}_1 (x_i - \bar{x})) = 0$

$$\Rightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^n x_i (y_i - \bar{y})}{\sum_{i=1}^n x_i (x_i - \bar{x})}$$

Now $\sum_{i=1}^n x_i (y_i - \bar{y}) = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$, since

$$\sum_{i=1}^n \bar{x} (y_i - \bar{y}) = \bar{x} \sum_{i=1}^n (y_i - \bar{y}) = 0 \text{ by definition of } \bar{y}.$$

Similarly, $\sum_{i=1}^n x_i (x_i - \bar{x}) = \sum_{i=1}^n (x_i - \bar{x})^2$.

Hence $\hat{\beta}_1 = C_{xy} / C_{xx}$, where

$$C_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$C_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$$

6) i) For these data, $\bar{x} = 65$ (trivial!), $\bar{y} = 9.127$,

$$C_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - 11\bar{x}^2 = 110$$

$$C_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - 11\bar{x}\bar{y} = 20.1$$

Hence $\hat{\beta}_1 = 0.183$, $\hat{\beta}_0 = -2.747$ (rounding errors tolerated!)

ii) Estimated s.e. is $\sqrt{0.4/110} = 0.060$, hence 95% CI for β_1 is $0.183 \pm 0.060 t_9(0.025) = (0.046, 0.319)$. This excludes zero, so there is evidence that temperature affects sugar conversion.

TOTAL

5
5

6) a) Given observations x_1, \dots, x_n , the sample mean is $\frac{1}{n} \sum_{i=1}^n x_i$
 & the sample variance is $\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ or $\frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} \right]$
 (either is acceptable)

b) Lower 2½% point of $F_{15,43} = 1 / (\text{upper } 2\frac{1}{2}\% \text{ point of } F_{43,15})$

Table 12(c) doesn't give this; it does, however, give upper % points for $F_{24,15} (= 2.701)$ & $F_{40,15} (= 2.395)$. Hence the required value is between $1/2.701 = 0.370$ and $1/2.395 = 0.418$, as required.

c) i) To test $H_0: \delta_1^2 / \delta_2^2 = 1$ against $H_1: \delta_1^2 / \delta_2^2 \neq 1$, the test statistic is $F = S_1^2 / S_2^2$.

Under H_0 , $F \sim F_{15,43}$. The upper 2½% point of this distribution is around 2.2 (Table 12c) and the lower 2½% point is between 0.370 & 0.418 (part (b) above). Hence we accept H_0 if F is between ≈ 0.418 & ≈ 2.2 , reject otherwise.

The observed value of F is $332.2 / 714.8 = 0.465$, which lies within the acceptance region regardless of the exact value of $F_{15,43} (0.025) \Rightarrow$ accept $H_0: \delta_1^2 = \delta_2^2$.

ii) CI is $\bar{x}_1 - \bar{x}_2 \pm t_{0.025} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

where $t_{0.025}$ is upper 2½% point of $t_{15+43-2} \approx 2.00$ (Table 10)

$$\begin{aligned} \& S_p^2 &= \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{(15 \times 332.2) + (43 \times 714.8)}{58} \\ &= 615.85 \end{aligned}$$

$$\begin{aligned} \text{Hence CI is } & (63.25 - 62.18) \pm (2 \times 24.81 \times \sqrt{\frac{1}{16} + \frac{1}{44}}) \\ &= \underline{(-11.42, 17.56)} \end{aligned}$$

d) Assumptions are that each sample is drawn from a normal db^{\wedge} . If satisfied, the fact that the CI includes zero, & the hypothesis of equal variances is not rejected, suggests that student performance is comparable in each year.

e) Histograms don't look that normal - NB "bump" in 70-80 range in both years (or 50-60 range in 2002, although this is based on many fewer students) - assumptions a bit dubious.

