① ⓐ Stratified v/s Simple: • It reduces the variability as you can estimate statistics for each strata instead of the entire sample as with

• Stratified sampling is more representative than simple random sampling

ⓑ Stratified v/s Quota: • Stratified is a random process compared to quota sampling so less bias with stratified sampling. [2]

• In Quota sampling, non-responses are not recorded. So stratified sampling might be a more suitable alternative if non-responses are important statistics.

② H_0 : There is no association between catching influenza and concentration of the new drug. [10]

H_1 : There is an association between catching influenza and concentration of the new drug.

O_i	E_i	E_i	$(O_i - E_i)^2/E_i$
12	$135/11$	12.27	0.006061
15	$162/11$	14.27	0.005051
29	$260/11$	23.63	1.2713
23	$312/11$	28.36	1.01428
9	$155/11$	14.09	1.8393
22	$186/11$	14.09	1.53275

So $X^2 = 5.61456...$

$= 5.61$ (3 s.f)

$V = 2$

C.V X^2_2 (10%) $= 4.605$.

So, since $5.61 > 4.605$

i.e: in critical region. Reject H_0 and accept H_1. There is significant evidence to suggest that there is an association between catching influenza and concentration of the new drug.

③ ⓐ When the data isn't ~~bivo~~ distributed bivariately normal. [1]

ⓑ [3]

	6	4	7	5	3	2	1
	6	5	7	2	3	4	1
d	0	-1	0	3	0	-2	0
d^2	0	1	0	9	0	4	0

$\therefore \sum d^2 = 14$

$\Rightarrow r_s = 1 - \frac{6 \times 14}{7 \times 48}$

$= 3/4 = 0.75$

(c) $H_0: \rho_s = 0$, $H_1: \rho_s > 0$ C.V: 0.7143. Since $0.75 > 0.7143$ then reject H_0 and accept H_1. There is significant evidence to suggest that there is a positive correlation between sales of ice-cream and sales of sunglasses.

(d) $H_0: \rho = 0$, $H_1: \rho > 0$ C.V: 0.6649. Since $0.65 < 0.6649$ then accept H_0. There is insignificant evidence to suggest that there is a positive correlation between sales of ice-cream and sunglasses. [2]

(e) Non-linear positive correlation. [1]

(4) $W \sim N(60, 5^2)$ [5]

(a) $W_1 - W_2 \sim N(0, 50) \Rightarrow P(|W_1 - W_2| > 2) = P(W_1 - W_2 > 2) + P(W_1 - W_2 < -2)$

$= P(Z > \frac{1}{5}\sqrt{2}) + P(Z < -\frac{1}{5}\sqrt{2}) = P(Z > 0.28) + P(Z < -0.28)$

$= 0.3897 + 0.3897 = 0.7794.$

(b) $C \sim N(40, 1.5^2) \Rightarrow Y = W_1 + \dots + W_{12} + C \sim N(760, 302.25)$ [3]

$E(W_1 + \dots + W_{12} + C) = 12 \times 60 + 40 = 760$; $Var(W_1 + \dots + W_{12} + C) = 12 \times 5^2 + 1.5^2$

(c) $P(Y > 800) \Rightarrow P(Z > (800 - 760)/\sqrt{302.25}) = P(Z > 2.30)$ [2]

$= 1 - 0.9893 = 0.0107.$

(5) (a) $H_0: \mu_E = \mu_{NE}$ Test-Statistic $Z = \dfrac{26.3 - 24.8}{\sqrt{0.589047619}} = 1.95$ [6]

$H_1: \mu_E > \mu_{NE}$

Since critical region is $Z \geq 1.6449$ then $1.95 > 1.6449$ and reject H_0. Accept H_1. There is significant evidence to suggest that the mean capacity of those who exercise is greater than those who don't.

(b) Assumed that $s^2 = \sigma^2$. Assume sample sizes are large enough [2]

(c) $\sum x = 35 \times 26.3 + 31.7 = 952.2$

$\sum x^2 = 34 \times 12.2 + 35 \times 26.3^2 + 31.7^2 = 25628.84$

So $s^2 = \dfrac{25628.84 - \frac{1}{36} \times 952.2^2}{35} = 12.7$ (3 s.f)

[8]

(6) (a) *

Number of suitcases	0	1	2, 3, 4
observed Frequency	6	25	19
Expected frequency	12.01	20.58	17.41
$(O_i - E_i)^2/E_i$	3.0075	0.94929	0.14521

$\therefore X^2 = 4.10$ (3 s.f) ; $V = 2$; C.V : $X_2^2(5\%) = 5.991$

* H_0 : $B(4, 0.3)$ is a good model

H_1 : $B(4, 0.3)$ is not a good model

Since $4.10 < 5.991$ accept H_0. $B(4, 0.3)$ is a good model.

[3]

(b) $r = e^{-1.8} \times 100 \frac{1.8^2}{2} = 26.78$

$s = 16.07$

	0	1	2	3	≥ 4
O_i	5	40	31	18	6
E_i	16.53	29.75	26.78	16.07	10.87
$(O_i - E_i)^2/E_i$	8.0424	3.5315	0.66499	0.2318	2.1819

$\therefore X^2 = 14.65$

[6]

(c) H_0 : Poisson is a good model. $V = 4 - 1 - 1 = 2$

H_1 : Poisson is not a good model. C.V : $X_2^2(1\%) = 9.21$

Since $14.65 > 9.21$ reject H_0. significant. Poisson not good model.

[4]

(7) (a) $19.5 \pm 1.6449 \times \frac{1.5}{\sqrt{50}} = [19.15\%, 19.85\%]$

(b) Paul should recommend they lower their stated fat content as 20% is above the interval.

[2]

(c) $P(|\bar{X} - \mu| > 0.5) \leq 0.05 \Leftrightarrow P(Z < \frac{0.5}{\sqrt{\frac{4}{n}}}) \leq$ 0.05 and $\Rightarrow \frac{0.5}{\sqrt{\frac{4}{n}}} \geq 1.6449$

[5]

$\Rightarrow \sqrt{n} \times \frac{0.5}{2} \geq 1.6449 \Rightarrow n \geq 43.3 \Rightarrow n_{min} = 44$