S3 June 2008 (MA)

Q1a) mean $= \frac{\sum x}{n} = \frac{6046}{36} = \boxed{167.94}$

$s^2 = \frac{1}{n-1}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) = \frac{1}{35}\left(1016338 - \frac{(6046)^2}{36}\right)$

$= \boxed{27.03}$

b) 99% C.I: $\left[\bar{x} \pm 2.5758\left(\frac{\sigma}{\sqrt{n}}\right)\right]$

$\Rightarrow \left[167.94 \pm 2.5758\left(\frac{5.1}{\sqrt{36}}\right)\right]$

$\Rightarrow [165.8,\ 170.1]$

Q2) H_0: There is no association between gender and type of course taken.

H_1: There is an association between gender and type of course taken

$$\text{Expected no.} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$$

EXPECTED :

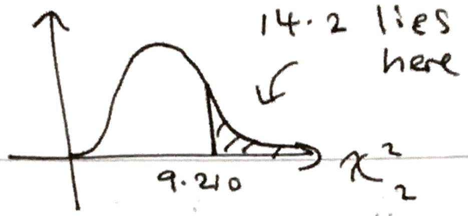
	Arts	Sciences	Humanities	
Boy	37.1	37.1	40.8	115
Girl	32.9	32.9	36.2	102
	70	70	77	217

O_i	E_i	$\frac{(O-E)^2}{E}$
30	37.1	1.3588
50	37.1	4.4854
35	40.8	0.8245
40	32.9	1.5322
20	32.9	5.0581
42	36.2	0.9293
		$\simeq 14.2$

$$X^2 = \sum \frac{(O-E)^2}{E} = 14.2$$

$$\nu = (\text{rows}-1)(\text{columns}-1) = (2-1)(3-1) = 2$$

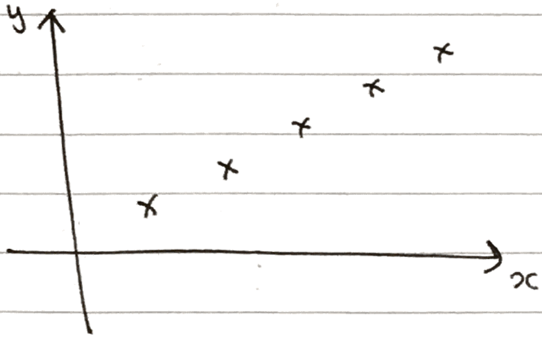
$$\text{critical value} = \chi^2_2(1\%) = 9.210$$



$14.2 > 9.210$

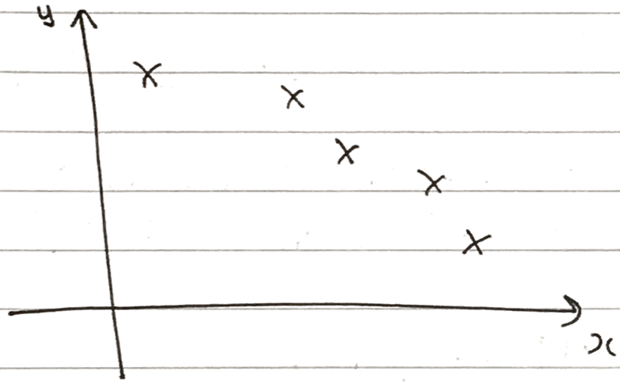
$\therefore$ Result is significant. (reject H_0).
Evidence suggests that there are is in fact an association between gender & type of course taken.

Q3ai)



$r = 1$ means a perfect straight line

ii)



$r_s = -1$ means a perfect monotonic negative correlation. This does not necessarily mean a straight line. $r > -1$ tells us its not a perfect straight line.

bi)

Dogs	Judge 1	Judge 2	d	d^2
A	1	1	0	0
B	4	2	2	4
C	2	4	2	4
D	3	3	0	0
E	5	5	0	0
F	6	7	1	1
G	7	6	1	1
				10

$\sum d^2 = 10. \quad \therefore r_s = 1 - \dfrac{6(10)}{7(48)} = \boxed{0.821}$

ii) $H_0: p = 0$
$H_1: p > 0$

judges in agreement indicates a positive correlation

critical value: 0.7143
(5% 1tail)

$0.821 > 0.7143$
$\therefore$ result is significant
evidence suggests that the judges are in agreement (reject H_0).

Q4a) $M \sim N(84, 11^2)$

not 4M as we are taking a sample of four men.

$P(\text{required}) = P(M_1 + M_2 + M_3 + M_4 < 350)$

LET $A = M_1 + M_2 + M_3 + M_4$,

$E(A) = 4E(M) = 4(84) = \boxed{336}$
$Var(A) = 4Var(M) = \boxed{484}$

$\therefore A \sim N(336, 484)$

$P(M_1 + M_2 + M_3 + M_4 < 350)$

$= P\left(Z < \frac{350 - 336}{\sqrt{484}}\right) = P(Z < 0.64)$

$= \boxed{0.7389}$

b) $W \sim N(62, 10^2)$

$P(\text{required}) = P(M < \frac{3}{2}W)$

$= P(M - \frac{3}{2}W < 0)$

let $B = M - 1.5W$,

$E(B) = 84 - 1.5(62) = -9$
$Var(B) = Var(M) + 1.5^2 Var(W)$
$= 11^2 + 1.5^2(10^2) = 346$.

$\therefore B \sim N(-9, 346)$

$P(\text{required}) = P(B < 0) = P(Z < \frac{9}{\sqrt{346}})$

$= P(Z < 0.48) = \boxed{0.6844}$

Q5a) - Only cleaners are chosen - what about managers?
- The first 50 cleaners to leave are all part of the same shift so their opinions may vary considerably when compared with other days due to the nature of the job → leads to bias.

bi) Gather an ordered list of all employees 1-550. Then using random numbers, select the first employee from the range (1-11) then pick every 11^{th} employee afterwards.

Why 11 : $k = \frac{\text{total employees}}{\text{sample size}} = \frac{550}{50} = 11$

(so every 11^{th} person is to be chosen)

ii) Label cleaners 1-495 & managers 1-55...

no. of cleaners needed = $\frac{495}{550} \times 50 = 45$

no. of managers needed = $\frac{55}{550} \times 50 = 5$

... & use random numbers to select 45 cleaners and 5 managers.

c) 390, 372.

Keep going across the top row. Numbers do not need to be right next to each other to count. → eg 39$\boxed{0, 67}$2

067 would count here.

Q6a) $p = \frac{0(11) + 1(21) + 2(30) + 3(20) + 4(12) + 5(3) + 6(2) + 7}{10 \times 100}$

$= \boxed{0.223}$

$p = \left[\frac{\text{no. of cuttings that do not grow}}{\text{total no. of cuttings}}\right]$

b) let $X \sim B[10, 0.2]$ where X = no. of cuttings that don't grow

$E_0 = r = 100 \times P(X=0) = 100\left[1-0.2\right]^{10} = \boxed{10.74}$

$E_2 = r = 100 \times P(X=2) = 100\left[\binom{10}{2}(0.2^2)(0.8^8)\right]$

$= \boxed{30.20}$

$t = 100 - \sum E_i = \boxed{3.28}$

c) H_0 : Binomial $(n=10, p=0.2)$ is a suitable model for these data.

H_1 : Binomial $(n=10, p=0.2)$ is not a suitable model for these data.

d) $E<5$ so last cell needs to be pooled with the second last. ν = no. of cells after pooling -1
$\therefore \nu = 5-1=4$.

NOTE : the value of p calculated in (a) is not the value we used so don't subtract a degree of freedom.

e) critical value $= \chi^2_4(5\%) = 9.488$.

$4.17 < 9.488$

$\Rightarrow$ 4.17 is not within the critical region so accept H_0. Evidence suggests $B[10, 0.2]$ is a suitable model for these data.

Q7a) H_0: $\mu_{male} = \mu_{female}$ critical value $= \pm 1.96$
(5%, two-tail)

H_1: $\mu_{male} \neq \mu_{female}$

$$\text{Test Stat} = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

$$= \frac{6.86 - 5.48 - (0)}{\sqrt{\frac{3.62^2}{100} + \frac{4.51^2}{200}}}$$

$$= 2.86$$

$2.86 > 1.96$

$\therefore$ Result is significant (reject H_0).
Evidence suggests that there is a difference in the mean amounts spent on junk food by male/female teenagers.

b) Allows us to assume sample means are normally distributed. (ie $\bar{F}$ and $\bar{M}$).