S3 June 2013(R) MA

Q1) List all males 1-300 and females 1-100 then use random numbers to pick 15 females and 45 males.

(No. of each group = $\frac{\text{group size}}{\text{population size}}$ x sample size)

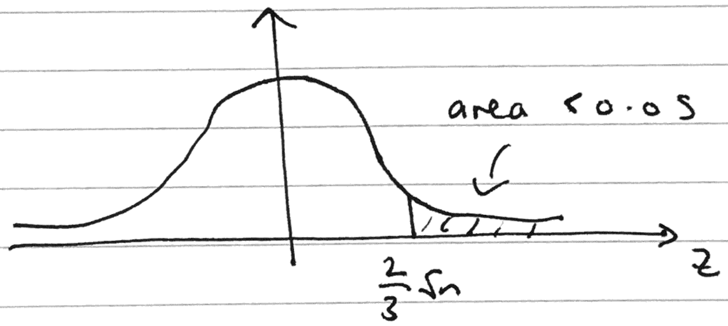
Q2) $X \sim N(40, 3^2)$ then by C.L.T, $\bar{X} \sim N(40, \frac{3^2}{n})$

What we want is...

$$P(\bar{X} > 42) < 0.05$$

$$P\left(Z > \frac{42-40}{\frac{3}{\sqrt{n}}}\right) < 0.05$$

$$P\left(Z > \frac{2}{3}\sqrt{n}\right) < 0.05$$



$P(Z > z) = 0.05$ when $z = 1.6449$

the area < 0.05 which means $\frac{2}{3}\sqrt{n} > 1.6449$

$$\sqrt{n} > \frac{3(1.6449)}{2}$$

$\therefore n > 6.087... \rightarrow \boxed{n_{min} = 7}$

Q3a)

Town	Population	No. of...	R_p	R_N	d	d^2
A	211	10	1	2	1	1
B	356	2	2	1	1	1
C	1047	12	3	3	0	0
D	2463	21	4	5	1	1
E	4892	16	5	4	1	1
F	6479	25	6	6	0	0
G	6571	67	7	10	3	9
H	6573	45	8	8	0	0
I	9845	48	9	9	0	0
J	14784	34	10	7	3	9
						22

$\sum d^2 = 22$

$$\therefore r_s = 1 - \frac{6(22)}{10(99)} = \boxed{0.867}$$

b) $H_0: \rho = 0$
$H_1: \rho > 0$

critical value: ± 0.6485
(1-tail, 2.5%)

$0.867 > 0.6485$
Result is significant.
Reject H_0.
Evidence of +ve correlation between population and no. of council employees.

c) $H_0: \rho = 0$
$H_1: \rho > 0$

critical value: ± 0.6319

$0.627 < 0.6319$
$\therefore$ Result is insignificant.
Accept H_0 - No evidence of a positive correlation

d) Data is likely not to be jointly normally distributed (required for PMCC) so Spearman's is more suitable. So use conclusion from (b).

Q4a) $\frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}} = \frac{282 \times 100}{600} = \boxed{47}$

b) $\nu = (\text{rows} - 1)(\text{columns} - 1) = (4-1)(4-1) = \boxed{9}$

c) 2.5%

$$\left[\begin{array}{l} \chi^2_9(0.010) = 21.666 \\ \chi^2_9(0.025) = 19.023 \end{array}\right]$$

d) $\underline{H_0}$: hair colour of population is in the ratio 2:6:1:3

$\underline{H_1}$: hair colour of population is <u>not</u> in the ratio 2:6:1:3.

<u>BLACK</u>: Expected no. $= \frac{2}{12} \times 600 = 100$

<u>BROWN</u>: Expected no. $= \frac{6}{12} \times 600 = 300$

<u>RED</u>: Expected no. $= \frac{1}{12} \times 600 = 50$

<u>BLONDE</u>: Expected no. $= \frac{3}{12} \times 600 = 150$

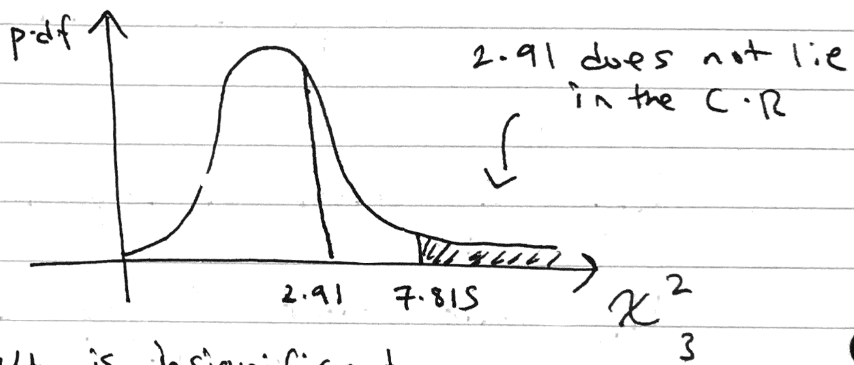
Colour	Black	Brown	Red	Blonde
O_i	105	282	48	165
E_i	100	300	50	150
$\frac{(O-E)^2}{E}$	0.2500	1.0800	0.0800	1.5000

$$X^2 = \sum \frac{(O-E)^2}{E} = 2.91$$

$$\nu = 4 - 1 = 3$$

$\therefore$ critical value $= \chi^2_3(5\%) = 7.815$

$$2.91 < 7.815$$



$\therefore$ Result is insignificant
Evidence suggests hair colour does in fact occur in the given ratio (Accept H_0).

Q5a) 90% C.I = $\left[\bar{x} \pm 1.6449 \frac{\sigma}{\sqrt{n}}\right]$

$\therefore$ width of given interval $= \frac{2(1.6449)\sigma}{\sqrt{n}}$

$$\frac{121.2 - 118.8}{2(1.6449)} = \frac{\sigma}{\sqrt{n}} = \frac{4000}{5483}$$

$$\bar{x} = \frac{1}{2}(121.2 + 118.8) = 120$$

so 98% C.I : $\left[\bar{x} \pm 2.3263 \frac{\sigma}{\sqrt{n}}\right]$

$$\left[120 \pm 2.3263\left(\frac{4000}{5483}\right)\right]$$

$$\left[118.3,\ 121.7\right]$$

b) circumference $= 2\pi r = \underline{\underline{\pi d}} \Rightarrow [118.3\pi,\ 121.7\pi]$

c) let X = no. of confidence intervals containing μ,

$X \sim B[3, 0.98]$

$P(X=3) = \binom{3}{3}(0.98^3) = \boxed{0.941}$

Q6a) $X \sim U[a-1, a+5]$

$E(X) = \frac{a-1+a+5}{2} = \frac{2a+4}{2} = a+2$

$Var(X) = \frac{(b-a)^2}{12} = \frac{(a+5-a+1)^2}{12} = \frac{6^2}{12} = 3$

$\therefore \boxed{\bar{X} \sim N\left(a+2, \frac{3}{50}\right)}$

(by C.L.T)
as n is large

b) sample mean is normally distributed so . . .

$\left[\bar{x} \pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right)\right]$

use σ of actual distribution not sample

$\left[17.2 \pm 1.96\left(\frac{\sqrt{3}}{\sqrt{50}}\right)\right]$

$[17.2 \pm 0.48]$

$[16.72, 17.68]$

this is a C.I for $(a+2)$ since $\mu = a+2$!

so subtract 2 from both limits to give C.I for a . . .

$\boxed{[14.72, 17.68]}$

Q7a) $H_0: \mu_{new} = \mu_{old}$ critical value : ± 1.6449 (5%, 1-tail)

$H_1: \mu_{new} < \mu_{old}$

$$\text{Test Statistic} = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} = \frac{67 - 60}{\sqrt{\frac{25^2}{100} + \frac{10^2}{150}}}$$

$$= 2.66$$

$2.66 > 1.6449$

$\therefore$ Result is significant

Reject H_0.

Evidence suggests mean has in fact decreased after the factory closed.

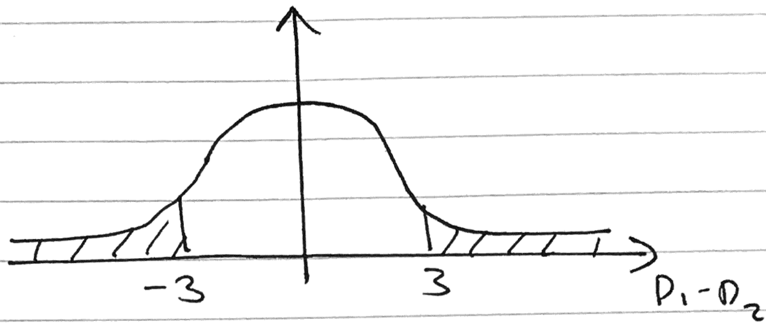
b) Allows us to assume sample means are normally distributed as n is large.

c) sample variance = population variance.

(Test Statistic by definition uses population variance)

(Q8a) P(required) = $P(|D_1 - D_2| > 3)$

$(D_1 - D_2) \sim N(0, 2(1.2^2))$



$P(|D_1 - D_2| > 3) = 2P(D_1 - D_2 > 3)$

$= 2\left[P\left(Z > \frac{3-0}{\sqrt{2(1.2^2)}}\right)\right] = 2P(Z > 1.77)$

$= 2\left[1 - P(Z < 1.77)\right] = \boxed{0.0768}$

b) P(required) = $P(C < \frac{4}{5}D) = P(C - \frac{4}{5}D < 0)$

let $A = C - \frac{4}{5}D$, $E(A) = 44 - \frac{4}{5}(54) = 0.8$

$Var(A) = 0.8^2 + \left(\frac{4}{5}\right)^2(1.2^2)$
$= 1.5616$

$\therefore A \sim N(0.8, 1.5616)$

P(required) $= P(A < 0) = P\left(Z < \frac{-0.8}{\sqrt{1.5616}}\right)$

$= P(Z < -0.64) = 1 - P(Z < 0.64)$

$= 1 - 0.7389 = \boxed{0.2611}$

E = empty box.

c) $P(\text{required}) = P(D_1+D_2+D_3+D_4+D_5+D_6+E-(C_1+C_2+C_3+C_4+C_5+C_6+E)>50)$

let $X = D_1+D_2+D_3+D_4+D_5+D_6+E$,

$X \sim N(352, 13.64)$ $\left[\begin{array}{l} E(X) = 6(E(D)) + E(E) \\ Var(X) = 6Var(D) + Var(E) \end{array}\right]$

$(D_1+\dots+D_6 \neq 6D)$

let $Y = C_1+C_2+C_3+C_4+C_5+C_6+E$,

$\left[\begin{array}{l} E(Y) = 6E(C)+E(E) \\ Var(Y) = 6Var(Y)+Var(E) \end{array}\right]$ $Y \sim N(292, 8.84)$

$P(\text{required}) = P(X-Y>50)$

subtract means, add variances.

$(X-Y) \sim N(60, 22.48)$

$\therefore P(X-Y>50) = P\left(Z > \frac{50-60}{\sqrt{22.48}}\right)$

$= P(Z > -2.10)$

$= P(Z < 2.10)$

$= \boxed{0.9821}$