S3 June 2015 (MA)

Q1(a) Label all the books 1-160 then use random numbers to select 10 books

b)

Book	Page#	$R_{p\#}$	R_{Borrow}	d	d^2
A	50	1	1	0	0
B	212	6	2	4	16
C	115	4	3	1	1
D	80	2	4	2	4
E	301	8	5	3	9
F	90	3	6	3	9
G	356	10	7	3	9
H	283	7	8	1	1
I	152	5	9	4	16
J	317	9	10	1	1
					66

$\sum d^2 = 66$

$$\therefore r_s = 1 - \frac{6(66)}{10(99)} = \boxed{0.6}$$

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

critical value: ± 0.5636

$0.6 > 0.5636$

$\therefore$ Result is significant.

Reject H_0.

There is evidence to support the Librarian's belief.

Q2a) $H_0: \mu_g - \mu_s = 1.5$ [g = in a group, s = on their own]

$H_1: \mu_g - \mu_s > 1.5$

critical value $= \pm 2.3263$
(1%, 1-tail)

$$\text{Test Statistic} = \frac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

$$= \frac{8.7 - 6.6 - (1.5)}{\sqrt{\frac{2.1^2}{80} + \frac{1.4^2}{65}}}$$

$$= 2.05\ldots$$

$2.05 < 2.3263$

$\therefore$ Result is insignificant.
Accept H_0.
Evidence suggests that the researcher's belief is incorrect.

b) No - C.L.T allows us to assume sample means are normally distributed as n is large for both samples. The parent distribution(s) needn't follow a normal distribution.

Q3a) Label staff 1-16 and children 1-40...

no. of each group needed = $\frac{\text{group size}}{\text{population size}} \times$ sample size

no. of staff $= \frac{16}{56} \times 14 = 4$

no. of children $= \frac{40}{56} \times 14 = 10$

... and use random numbers to pick 4 staff and 10 children.

b) mean $= \frac{\sum x}{n} = \frac{437}{14} = \boxed{31.2}$

$(\hat{\sigma}^2) = s^2 = \frac{1}{n-1}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)$

$= \frac{1}{13}\left(26983 - \frac{(437)^2}{14}\right)$

$= \boxed{1026.3}$

c) $s.e = \frac{\sigma}{\sqrt{n}} = \frac{s}{\sqrt{n}} = \frac{\sqrt{1026.3}}{\sqrt{14}} \simeq \boxed{8.56}$

d) Standard error within each group is quite small relative to the answer in (c). This is understandable because when combining both groups the overall variation in weight will be bigger as we are combining staff and children — two groups with very different mean weights hence the ~~4m~~ average distance from the mean will be much greater (when combining).

In simpler terms, the standard error is just the standard deviation but for a sample.

Q4a) $H_0 : \mu = 0.5$
$H_1 : \mu \neq 0.5$

$0.4633 < 0.5 < 0.5127$

$\therefore$ Result is insignificant. Evidence suggests mean is equal to 0.5.

significance level : 10%
(100 − 90 = 10)

b) 95% C.I : $\left[\bar{x} \pm 1.96\left(\frac{\sigma}{\sqrt{150}}\right)\right]$

$$\left[0.479 \pm 1.96\left(\frac{\sigma}{\sqrt{150}}\right)\right]$$

find σ by using interval from (a):

$\Rightarrow 0.5127 = \bar{x} + 1.6449\left(\frac{\sigma}{\sqrt{100}}\right)$ ~ 1

$\Rightarrow 0.4633 = \bar{x} - 1.6449\left(\frac{\sigma}{\sqrt{100}}\right)$ ~ 2

1 − 2 : $0.0494 = \frac{2(1.6449)}{10}\sigma$

$\therefore \sigma = 0.15016\ldots$

hence 95% C.I : [0.455, 0.503] //

hence $D \sim N(0, \frac{6}{5}\sigma^2)$

b) $P(Y_1 > \bar{X} + \sigma) = P(Y_1 - \bar{X} > \sigma)$

$= P(D > \sigma) = P(Z > 0.91)$

$= 1 - P(Z < 0.91) = \boxed{0.181}$

c) U_1 and $\bar{U}$ are not independent as $\bar{U}$ depends on U_1 (i.e U_1 affects the value of $\bar{U}$).

In (iia), we used the variance formula to find $Var(Y_1 - \bar{X})$ and thus the distribution of D. That was ok as Y_1 and $\bar{X}$ are independent of each other. But U_1 and $\bar{U}$ aren't so we can't apply the distribution of D to calculate the stated probability.

Remember, the variance formula for two random variables requires both to be independent.

d) $P(U_1 - \bar{U} > \sigma)$

rewrite $\bar{U}$ in terms of $U_1, U_2, \ldots, U_5$ as we can't use variance formula.

$U_1 - \bar{U} = U_1 - \left(\frac{U_2 + U_3 + U_4 + U_5}{5}\right)$

$= \frac{4U_1 - (U_2 + U_3 + U_4 + U_5)}{5}$

let $\frac{4U_1 - (U_2 + \cdot\cdot + U_5)}{5} = F$

$E(F) = \frac{4}{5}E(U_1) - \frac{1}{5}E(U_2 + .. + U_5)$

$= \frac{4\mu}{5} - \frac{4}{5}\mu = 0$

$\text{Var}(F) = \frac{1}{25}\text{Var}(4U_1 - (U_2 + .. + U_5))$

here we can use variance formula as $U_1, U_2, \ldots, U_5$ are independent

$= \frac{16}{25}\text{Var}(U_1) + \frac{4}{25}\text{Var}(U_2)$

$= \left(\frac{16}{25} + \frac{4}{25}\right)\sigma^2 = 0.8\sigma^2$

$\therefore F \sim N(0, 0.8\sigma^2)$

$P(\text{required}) = P(F > \sigma) = P\left(Z > \frac{\sigma - 0}{(\sqrt{0.8})\sigma}\right)$

$= P\left(Z > \frac{1}{\sqrt{0.8}}\right)$

$= P(Z > 1.12) = 1 - P(Z < 1.12)$

$= \boxed{0.1314}$

Q6a) H_0: $U[0,10]$ is a suitable model

H_1: $U[0,10]$ is not a suitable model.

D	O_i	E_i	$\frac{(O-E)^2}{E}$
0-4	22	40	8.1
4-7	39	30	2.7
7-9	25	20	1.25
9-10	14	10	1.6

because theres 100 children.

Expected values $= \frac{b-a}{n} \times 100$, where

b = upper limit of interval

a = lower limit

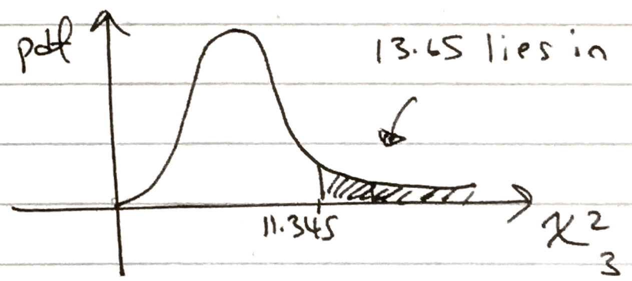
n = total range (10)

$X^2 = \sum \frac{(O-E)^2}{E} = 13.65$

$\nu = 4-1 = 3$

$\therefore$ critical value $= \chi^2_3(1\%) = 11.345$

$13.65 > 11.345$



13.65 lies in the C.R

$\therefore$ Result is significant.
Reject H_0.
Evidence suggests that
$U[0,10]$ is not a suitable model for these data.

b) Area $= \pi r^2 \propto$ Frequency $\therefore \pi r^2 k = f$.

In green region: Area $= 16\pi$
Freq $= 16$

hence $\frac{\text{Area}}{\pi} =$ frequency (expected).

$$\therefore r = \frac{\pi(9^2 - 7^2)}{\pi} = \boxed{32}$$

$$s = \frac{\pi(10^2 - 9^2)}{\pi} = \boxed{19}$$

c) Result is insignificant so Henry's model is perfectly fine.

d) H_0: No association between gender and colour chosen.

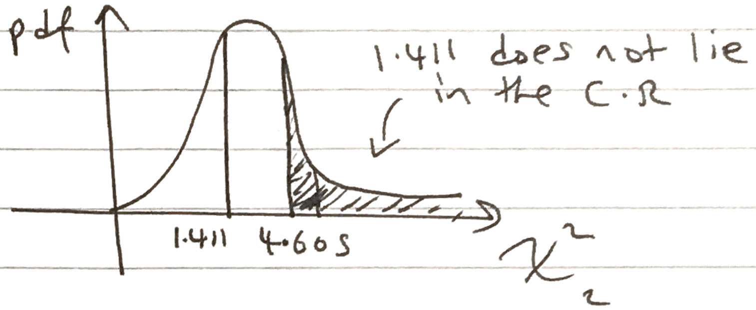
H_1: An association between gender and colour chosen exists.

e) $$\frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}} = \frac{39 \times 65}{100} = 25.35$$

f) Expected frequency < 5 in the cell (Boys : Yellow) so the column needs to be pooled with the one before it.
$\gamma = (\text{rows} - 1)(\text{columns} - 1) = (2-1)(3-1) = 2$

g) $\chi^2_2(10\%) = 4.605$

$1.411 < 4.605$



∴ Result is insignificant.
Phoebe's belief is not supported.