Spring 2015 Model Solutions. (31 total)

1 a). Binomial Probability

Assumptions: N known in advance
trials independent
p fixed. (1)

(1)
$N = 4$
$k = \cancel{4}\ 3$
p = prob of scoring 8500 $= 7/39$

$$P(\text{score 8500 three times out of 4}) = \frac{4!}{3!\,1!}\left(\frac{7}{39}\right)^3\left(1-\frac{7}{39}\right)^1$$

$$= 4 \times \left(\frac{7}{39}\right)^3 \times \frac{32}{39}$$

$$= 0.019$$ (1)

b) [Hard].

Total bonus of 27,780 in 3 plays means that I scored 9250 once, 8500 once and 10000 once.

OR that I scored 9250 3 times

call these A, B, C.

However I could have got them in any order

ABC
ACB
BAC
BCA
CAB
CBA

6 options. (1) each have prob.

$$\frac{4}{39} \times \frac{7}{39} \times \frac{28}{39}$$ (1)

$$+ \left(\frac{4}{39}\right)^3$$

$\Rightarrow$ prob. of bonus of 27850 in 3 plays is

$$6 \times \frac{4}{39} \times \frac{7}{39} \times \frac{28}{39} + \left(\frac{4}{39}\right)^3 = 0.08 + 0.001$$

c)

observed frequency	expected freq.
8	4
10	7
21	28.

(1)

H_0 : the 3 options occur with the stated probabilities

H_1 : the 3 options do not occur with the stated probabilities

(1).

$$\chi^2 = \frac{(8-4)^2}{4} + \frac{(10-7)^2}{7} + \frac{(21-28)^2}{28}$$

$$= 7.04$$ (1)

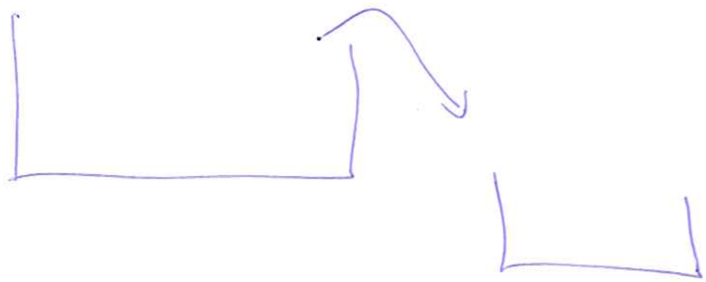
dof = # terms in χ^2 sum -1 $= 3-1$ $= 2.$ (1).

χ^2, 2 dof. from the table 7.04 is between $p=5\%$ and $p=1\%$

(1) $\Rightarrow$ reject H_0

conclude that the two positions are <u>not</u> chosen at random.

d)



$N = 39$

mean = ~~9653.8~~ 9461.5

What's in the box?

4 tickets with 9250 on them

7 tickets with 8500 on them

28 tickets with 10,000 on them

average / mean of box is $\dfrac{4 \times 9250 + 7 \times 8500 + 28 \times 10000}{39}$ | 1

$= 9653.8$

$$SD_{box} = \sqrt{\left(\frac{4 \times (9250 - 9653.8)^2 + 7 \times (8500 - 9653.8)^2 + 28 \times (10000 - 9653.8)^2}{39} \right)}$$ | 1

$= 584.56$

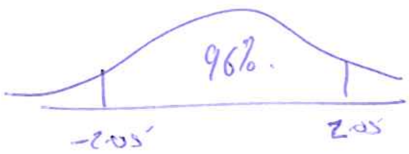
H_0 : mean of box is 9653.8.

H_1 : mean of box is not 9653.8. | 1

$$z = \frac{\text{observed} - \text{expected}}{SE}$$

$$SE_{mean} = \frac{SD_{box}}{\sqrt{\#\text{draws}}} = \frac{584.56}{\sqrt{39}} = 93.605$$

$$z = \frac{9461.5 - 9653.8}{93.605} = -2.05$$ (1)



from table, central area 95.96.

H_1 is such that variations in either direction cast doubt on H_0.

$\Rightarrow$ ~~t~~two-tailed test

$p \simeq 4\%$ (1)

Reject H_0

Conclude that it does not appear that the two positions were chosen at random. (1)

2/ a) This was an observational study. 1

No intervention was made.

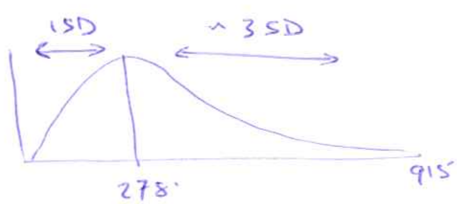
b) that cell phone use and texting is related to 1

GPA, anxiety and satisfaction with life

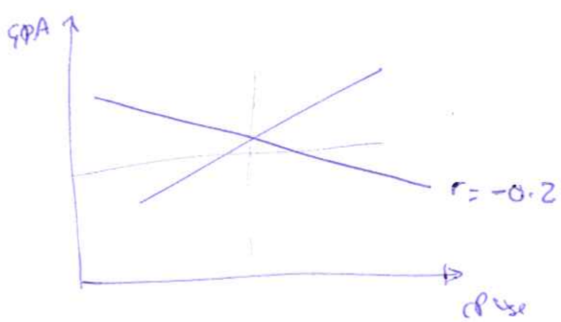
c) Cell phone use/texting was negatively related to GPA 1

and positively related to anxiety.

d) No - the distribution has a long right tail 1



e)



for each $1\,SD_{cpu}$ increase in CPuse, we associate a $r \times 1\,SD_{GPA}$ decrease in GPA.

1 hour = 60 mins $\hat{=}$ $\frac{60}{218}\,SD_{CPuse}$

corresponding decrease in GPA is $0.2 \times \frac{60}{218} \times 0.59$

$= \underline{0.032}$ 1

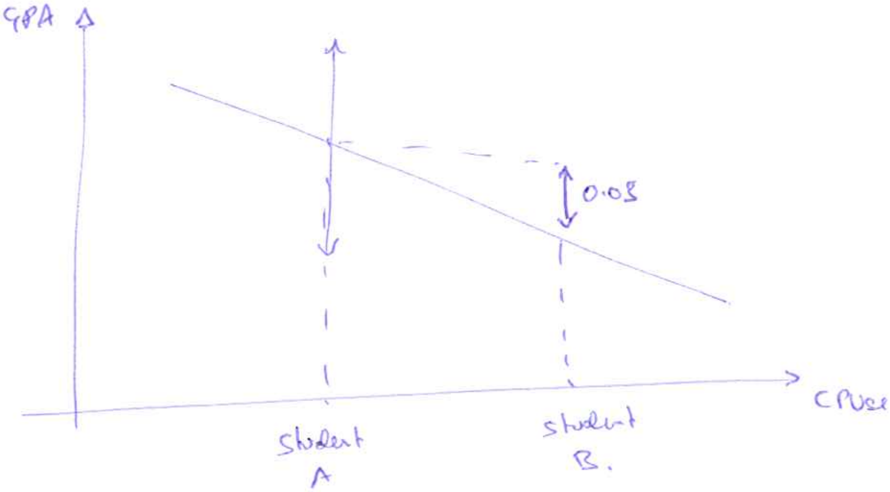
f) Because the distribution of CPUse does not follow the normal curve, the scatter diagram will not be "football shaped", and so one of the assumptions for a valid regression does not hold. (1)

g) the results show an association between increased CPUse and reduced GPA, but do not show causation. (1)

(there may be a causal link - time spent on the phone is not spent studying - but the results only show association).

BONUS.

g)(~~f~~).



RMS error is $\sqrt{1-r^2} \times SD_{GPA}$

$= \sqrt{1-0.2^2} \times 0.59$

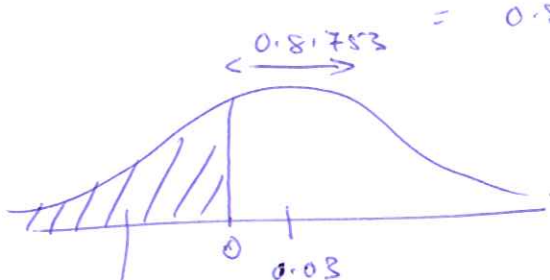
$= 0.578.$

consider the difference in GPA.

By analogy of the ~~SD of~~ SE of difference studied in class

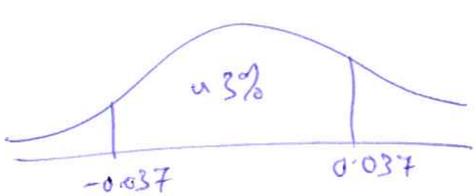
$SD_{\text{diff in GPA}} = \sqrt{0.578^2 + 0.578^2}$

$= 0.81753.$



prob. that student A scores lower than student B.

0 in standard units is $\frac{0 - 0.03}{0.81753} = -0.037$



from table, central area is ~3%

$\Rightarrow$ left tail is $0.5 \times (100-3)$

$= 48.5$

$\Rightarrow$ student A has prob 0.485 of scoring less than student B.

3

8

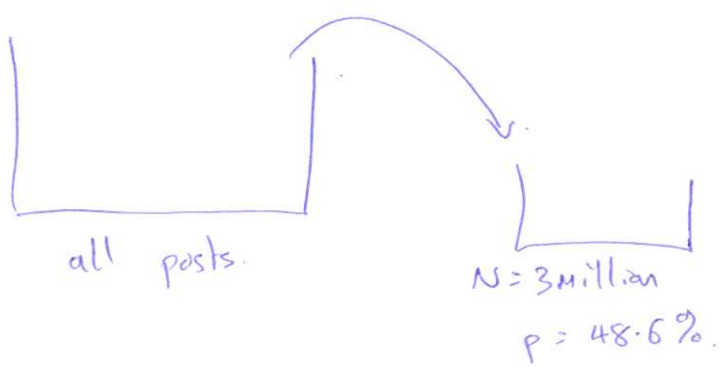
3/

a). H_0 : Reducing the number of posts with positive content in a user's news feed does not affect how much positive content that user posts.

H_1 : Reducing the number of posts with positive content in a user's news feed reduces the positive content that that user posts.

1

b)



Expected value for the population is 48.6%.

$$SE_{\%} = \frac{SD_{box}}{\sqrt{\#draws}} \times 100$$

$$SD_{box} = (1-0)\sqrt{0.486 \times 0.514}$$

1

$$= 0.4998$$

$$\simeq 0.5$$

$$SE_{\%} = \frac{0.5}{\sqrt{3000000}} \times 100 = 0.029\%$$

1

95% CI for % of all posts with positive sentiment is

$$48.6\% \pm 2 \times 0.029\%$$

1

$$48.54 \rightarrow 48.66.$$

c) we do not know whether the CI covers the true (1)
population % or not.

d) No. The range is extremely narrow, and so even slight changes over time are likely to mean that the % of positive posts 3 ½ years later will almost certainly have (1)
changed.

e). Yes - this was a controlled experiment, so we can (1)
conclude causation.

f) The observed difference is very small. For any individual (1)
it is unlikely to be important.

4, [Bonus].

there is also the effect of the counter. - people who count carefully get higher #'s, and count carefully for both brands.

Conversely, those who do not count carefully will have lower numbers for both brands.

2

5, [Bonus]

this will happen 1 time out of 100.

He knows that; she doesn't.

1