Spring 2013

Model Solution

31 total

1 a) Binomial Probability

\[
\begin{align*}
N &= 4 \\
R &= 4.3
\end{align*}
\]

\[
P = \text{prob of scoring 8500} = \frac{7}{39}
\]

\[
P(\text{8 score 8500 three times or one of 4}) = \frac{4!}{3!1!} \left( \frac{32}{39} \right)^3 \left( 1 - \frac{32}{39} \right)^1
\]

\[
= 4 \times \left( \frac{32}{39} \right)^2 \times \frac{32}{39}
\]

\[
= 0.019
\]

b) [Hard]

Total bonus if 27,780 in 3 plays means that I scored 9250 once, 8500 once an 10000 once. OR that I scored 9250 3 times.

call these A, B, C.

However I could have got them in any order.

\[
\begin{align*}
\text{ABC} & \quad \text{7} \\
\text{ACB} & \\
\text{BAC} & \quad \text{6 choices, each has prob.}
\end{align*}
\]

\[
\begin{align*}
\text{BCA} & \quad \frac{4 \times 7 \times 28}{39 \times 39 \times 39} \\
\text{CAB} & \\
\text{CBA} & \quad \left( \frac{4}{39} \right)^3
\end{align*}
\]

\[
\Rightarrow \text{prob of bonus is 27,850 in 3 plays is}
\]

\[
6 \times \frac{4}{39} \times \frac{7}{39} \times \frac{28}{39} + \left( \frac{4}{39} \right)^3 = 0.08 + 0.001
\]
<table>
<thead>
<tr>
<th></th>
<th>observed frequency</th>
<th>expected freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>21</td>
<td>28</td>
<td></td>
</tr>
</tbody>
</table>

$H_0$: the 3 options occur with the stated probabilities

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

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

$$= 7.04$$

$\# \text{dof} = \text{#terms in } \chi^2 \text{ sum} - 1 = 3 - 1 = 2.$

$\chi^2, 2 \text{dof. from the table } 7.04 \text{ is between } p = 0.05 \text{ and } p = 0.10$

$\Rightarrow \text{reject } H_0$

conclude that the 3 options are not chosen at random.
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 \( \bar{x} \) of box is:

\[
\frac{4 \times 9250 + 7 \times 8500 + 28 \times 10000}{39}
\]

\[= 9653.8\]

\[SD_{box} = \sqrt{\frac{4 \times (9250 - 9653.8)^2 + 7 \times (8500 - 9653.8)^2 + 28 \times (10000 - 9653.8)^2}{39}}\]

\[= 584.56\]

\(H_0: \) mean of box is 9653.8.

\(H_1: \) mean of box is not 9653.8.

\[
z = \frac{\text{observed} - \text{expected}}{\text{SE}_{mean}}
\]

\[\text{SE}_{mean} = \frac{SD_{box}}{\sqrt{\text{#draws}}} = \frac{584.56}{\sqrt{39}} = 93.605\]
\[ z = \frac{9461.5 - 9653.8}{93.605} = -2.05 \] (1)

From Table, central area 95.96.

It is such that variances of either direction can


doubt on Ho.

\[ \Rightarrow \text{two-tailed test} \]

\[ p = 4\% \] (1)

Reject Ho

Conclude that it does not appear that the two positions were chosen at random.
2. a) This was an observational study. No intervention was made.

b) That cell phone use and texting is related to GPA, anxiety and satisfaction with life.

c) Cell phone use/texting was negatively related to GPA and positively related to anxiety.

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

![Graph showing long right tail distribution with 1SD and 3SD marks.

GPA vs CRU]

For each 1SD increase in CRU, we associate a r = -0.2 decrease in GPA.

1 hour = 60 mins = \frac{60}{218} 5D CRU

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

= 0.032.
f) Because the distribution of CPU use 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)

9) The results show an association between increased CPU use and reduced GPA, but do not show causation.

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

By analogy of the same 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.

In standard units, \( Z = \frac{0 - 0.03}{0.81753} = -0.037 \)

From table, central area is 0.37%

\[ \Rightarrow \text{left tail is } 0.05 \times (100 - 3) \]

\[ = 48.5 \%
\]

\[ \Rightarrow \text{student A has prob } 0.485 \text{ of scoring less than student B.} \]
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.

b) 

\[
SE_\% = \frac{SD_{\text{box}}}{\sqrt{N}} \times 100
\]

\[
SD_{\text{box}} = (1-\hat{p}) \sqrt{\hat{p}(1-\hat{p})} = 0.0499 \\
\approx 0.05
\]

\[
SE_\% = \frac{0.5}{\sqrt{3,000,000}} \approx 0.029\%
\]

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

\[
48.6\% \pm 2 \times 0.029\% \Rightarrow 48.54 \approx 48.66.
\]
c) We do not know whether the CI covers the true 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 3½ years later will almost certainly have changed.

e) Yes - this was a controlled experiment, so we can conclude causality.

f) The observed difference is very small. For any individual, 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 carelessly for both brands.

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

5. [Bonus]
This will happen if they are at 100.
He knows that; she doesn't.