Sample Midterm Questions
Stat111 – Summer 2008

NOTE: This sample midterm is shorter than the actual midterm will be.

Question 1
Do IQ scores of Self-concept affect GPA? Based on 102 eighth-grade students in a rural Midwestern school with their IQ score, Self-concept scores and GPA we have the following information.

\( r = 0.542 \) for the regression of GPA on Self-concept

\( r = 0.632 \) for the regression of GPA on IQ

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>STD Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPA</td>
<td>102</td>
<td>8.3</td>
<td>3.5</td>
</tr>
<tr>
<td>SELF-CONCEPT</td>
<td>102</td>
<td>52.3</td>
<td>13.4</td>
</tr>
</tbody>
</table>

(a) Suppose that we are interested in predicting GPA by Self-concept scores. Calculate the best-fit line for \( Y = \text{GPA} \) and \( X = \text{Self-concept scores} \).

\[
\hat{y} = b_x - a
\]

\[
b = r \cdot \frac{s_y}{s_x} = 0.542 \left( \frac{3.5}{13.4} \right) \approx 0.1416
\]

\[
a = \bar{y} - b \bar{x} = 8.3 - 0.1416 \cdot 52.3 \approx 0.896
\]

\[
\hat{y} = 0.1416x + 0.896
\]

(b) Julia's Self-concept score is 31. Predict her GPA using the best-fit line.

\[
\hat{y} = 0.1416 \cdot (31) + 0.896 \approx 5.28
\]
(c) If an eighth-grade student decreases his Self-concept score by 12, what change do you expect in his GPA?

Can't do! Not a causal formula because the data are observational. Can't know if self-concept is the only variable that influences GPA (probably isn't!!).

(d) Which variable, Self-concept or IQ has a stronger association with GPA and how do you know this?

IQ - it has a larger correlation with GPA than self-concept.

Question 2

A population of students contains 32% freshmen, 50% sophomores and 18% juniors. Three students are independently selected at random from the population to take a particular Psychology class.

(a) Find the probability that at least one sophomore is selected.

This is a binomial because we're interested in sophomore vs. non-sophomore.

\[ n = 3, \ p = .50 \]

\[ P(X \geq 1) = 1 - P(X=0) = 1 - \frac{1}{8} = \frac{7}{8} \]
(b) Find the probability that at least one sophomore is selected given that you are told no juniors are selected.

\[ P = P(\text{sophomore} \mid \text{no juniors}) = \frac{P(\text{sophomore})}{P(\text{no juniors})} = \frac{50\%}{82\%} \approx 0.61 \]

\[ n = 3 \]

\[ P(X \geq 1) = 1 - P(X = 0) = 1 - {3 \choose 0} (0.61)^0 (1 - 0.61)^3 \]

\[ = 1 - 0.06 \approx 94\% \]

(c) Let \( X = \) number of freshmen selected. Write down the probability distribution of \( X \).

\[
\begin{array}{c|cccc}
X & 0 & 1 & 2 & 3 \\
p(X) & 0.31 & 0.44 & 0.21 & 0.03
\end{array}
\]

\[ P(X = k) = {3 \choose k} (0.32)^k (1 - 0.32)^{3-k} \]

(d) What is the mean and standard deviation of \( X \)?

\[
M_x = \sum_{i=1}^{n} x_i \cdot p(x_i) = 0 \cdot 0.31 + 1 \cdot 0.44 + 2 \cdot 0.21 + 3 \cdot 0.03 \]

\[ = 0.95 \]

\[
\sigma_x = \sqrt{\sum (x - M_x)^2 \cdot p(x)} = \sqrt{0.638475} \approx 0.799
\]
Question 3

On the math SAT men have traditionally averaged 500 with a standard deviation of 110. Women have traditionally averaged about 460, again with a standard deviation of about 110. The distribution of scores for men and women are both normally distributed.

(a) What proportion of men score below 460?

\[ M_X = 500 \]
\[ \sigma_X = 110 \]
\[ Z = \frac{460 - 500}{110} = -0.3636 \]
\[ p\text{-value} = 0.3594 \]

(b) In a random sample of 250 men, what is the probability that their mean SAT score is less than 490?

\[ n = 250 \]
\[ \bar{X} = \frac{\sum X}{250} \]
\[ M_{\bar{X}} = M_X = 500 \]
\[ \sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = 6.957 \]
\[ Z = \frac{\bar{X} - M_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{490 - 500}{6.957} \]
\[ = -1.437 \]
\[ p\text{-value} = 0.0749 \]

(c) In that same sample of 250 men, what is the approximate probability that at least 75 of them have an SAT score below 460?

Binomial! \[ n = 250 \]
\[ p = 0.36 \ (\text{from part (a)} \]
\[ n \cdot p = 90 > 10 \]
\[ n \cdot (1-p) = 160 > 10 \]

So we can use Normal approximation.

\[ M_X = n \cdot p = 90 \]
\[ \sigma_X = \sqrt{n \cdot p \cdot (1-p)} = 7.59 \]
\[ Z = \frac{75 - 90}{7.59} = -1.98 \]
\[ p\text{-value} = 1 - P(Z < -1.98) \]
\[ = 1 - 0.0244 \]
\[ = 0.9756 \]
Question 4

Shown below are the scores for the final exam for 50 students in Physics 101.

\[
\begin{array}{ll}
\text{Mean} & 43.18 \\
\text{Std Dev} & 29.8 \\
N & 50
\end{array}
\]

(a) Approximately how many students received a score of 70 or greater?

Total number of students: \( N = 50 \)

Upper quartile (Q3) is at about 65, so a bit fewer than a quarter of students scored 70 or more. \( \frac{50}{4} = 12.5 \), so I’d approximate about 10 students.

(b) The professor decided to grade on a curve. He multiplied all of the scores by 1.1 and then subtracted 5 points. What is the mean and what is the standard deviation for the rescaled exams?

\[
\begin{align*}
X &= \text{unscaled scores} \\
M_X &= 43.18 \\
\sigma_X &= 29.8 \\
Y &= \text{scaled scores} \\
Y &= 1.1X + (-5) \\
M_Y &= 1.1M_X - 5 = 42.5 \\
\sigma_Y &= 1.1 \cdot \sigma_X = 32.78
\end{align*}
\]
(c) The test scores were clearly not normally distributed, as you can see from the histogram. If the test scores (the original test scores, NOT the rescaled) were normally distributed and the mean and the standard deviation were the same, how many students would you anticipate scoring 20 or below?

\[
Z = \frac{20 - 43.18}{29.8} = -0.564
\]

\[ p-value \approx 0.2877 \]

\[ n \cdot p \approx 50 (0.2877) \approx 14 \]

This is close to boxplot.

**Question 5**

A group of Anthropology students want to know if graduate students are more likely to be alcoholics than other people. The anthropology students created a survey that helped them diagnose alcoholism. They then went to the Irish Pub on Monday night during Quizzo to find graduate students to survey. They performed a simple random sample of the graduate students in the Irish Pub.

The Anthropology students found that 64% of the people surveyed showed signs of alcoholism. The population rate of alcoholism is 23%.

(a) What can you conclude from their study?

Can't conclude anything about graduate students in general because sample was taken only at a bar. It's not representative of entire population of grad students. Also may be response bias because people will answer differently in front of friends.

(b) Why are anthropologists always messing up statistics? Seriously.

I wish I knew.