Math 21-01 | Introductory Statistics | Fall 2015
Midterm 1 | October 7th, 2015

NAME: Midterm 1 Solutions

HONESTY STATEMENT: My signature below indicates that I neither gave nor received assistance from any person or source, other than those specified in the instructions given below. All violations will result in an F in the course and you will be reported to the appropriate Dean.

Signature: xxxxxxxxxxxxxxxxxxxx

INSTRUCTIONS: Please mark your answers on the exam in the space provided. You are not allowed to use any calculators, books, or notes. Answers given on the exam can be left as unsimplified formulas when appropriate. If your answer space is messy/difficult to read, please circle your final answer.

SCORE:

Problem 1:	10	(out of 10)
Problem 2:	15	(out of 15)
Problem 3:	10	(out of 10)
Problem 4:	15	(out of 15)
Bonus:	5	(Maximum 5 points)
Total:	55	(out of 50)

PROBLEM 1:

Please circle ONE for each question.

(a) Which of the following is a parameter?

$\bar{x}$ | σ | $\tilde{x}$ | n

(b) The number of ways to pick r objects from a set of n objects without replacement is:

$n!$ | ${}_nC_r$ | ${}_nP_r$ | n^r

(c) Which of the following is not a measure of relative standing?

Z-score | Mode | Percentile | Quantile

(d) If we take multiple samples from the same population, we expect our samples to have similar ____?

Ranges | Medians | Z-Scores | Means

(e) The number of ways to pick r objects from a set of n objects with replacement is:

$n!$ | ${}_nC_r$ | ${}_nP_r$ | n^r

(f) Which of the following is not necessarily unique?

Variance | Mode | Median | Range

(g) The number of ways to order r objects from a set of n objects without replacement is:

$n!$ | ${}_nC_r$ | ${}_nP_r$ | n^r

(h) Which of the following is a statistic?

μ | σ | N | $\bar{x}$

(i) The number of ways to order n objects out of without replacement is:

$n!$ | ${}_nC_r$ | ${}_nP_r$ | n^r

(j) You are given a list of letter grades from an assignment. What kind of data is this?

Nominal | Ordinal | Interval | Ratio

PROBLEM 2:
Consider the following sample:

1 3 3 3 5 6 8 8 8 15

Find the range, mean, median, and mode of the sample.

Range = 15 - 1 = 14
Mean = $\frac{1+3+3+3+5+6+8+8+8+15}{10} = \frac{60}{10} = 6$
Median = 5.5
Modes = 3 and 8

The standard deviation of the sample is 4.0. How would one determine that? What is the variance? What is the coefficient of variation?

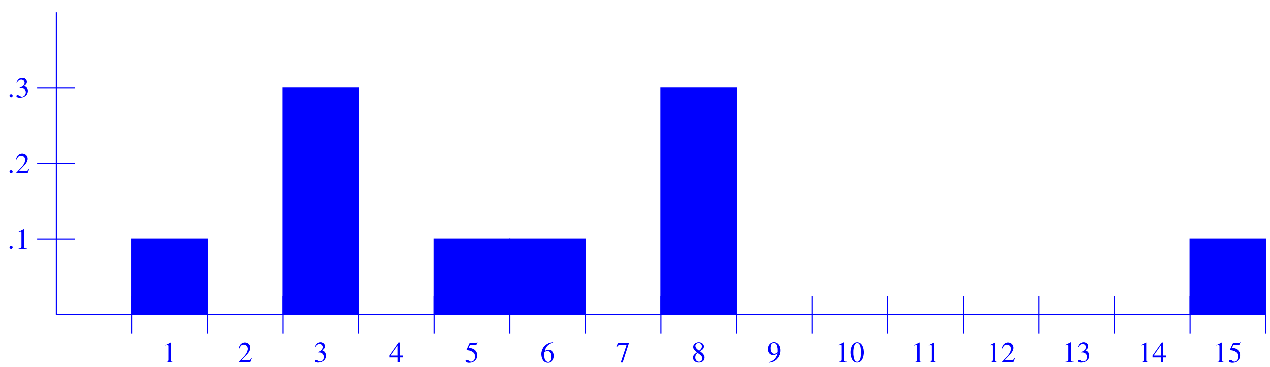
Standard deviation is $s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$, or, square root of the variance. Variance can be calculated as such:

$$s^2 = \frac{(-5)^2+(-3)^2+(-3)^2+(-3)^2+(-1)^2+0^2+2^2+2^2+2^2+9^2}{9}$$

$$\frac{25+9+9+9+1+0+4+4+4+81}{9} = \frac{146}{9}, (\text{which is } \approx 16)$$

And then we see that $s = \sqrt{16} = 4$ as we were given.
Coefficient of variation is $\frac{s}{\bar{x}} = 4/6 = .66$.

Draw a histogram. Does the data appear to be normally distributed? Why or Why not?



The data does not look normally distributed. Normally distributed data should look roughly like a bell curve, and have the two main properties: symmetric and rises to a maximum then falls. This data appears to do neither.

What is the Z-score of the value 1? What is the Z-score of the value 15? Which one is more unusual?

Z-score of $1 = \frac{1-6}{4} = -1.25$

Z-score of $15 = \frac{15-6}{4} = 2.25$

15 is more unusual.

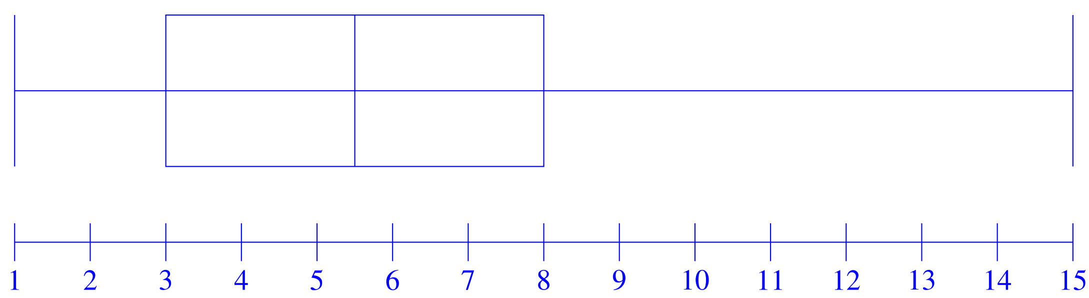
What is the percentile of the value 3? What is the percentile of the value 5?

Percentile of $3 = 10$

Percentile of $5 = 40$

Give a five number summary of the data, and draw a box plot.

Min $= 1$
$Q1 = 3$
$Q2 = 5.5$
$Q3 = 8$
Max $= 15$



Judging by relative standing, are there any outliers? If so, draw a modified box plot to reflect this.

$IQR = 8 - 3 = 5$
Maximum usual value = $Q3 + (1.5) * (IQR) = 8 + 7.5 = 15.5$
Minimum usual value = $Q1 - (1.5) * (IQR) = 3 - 7.5 = -4.5$

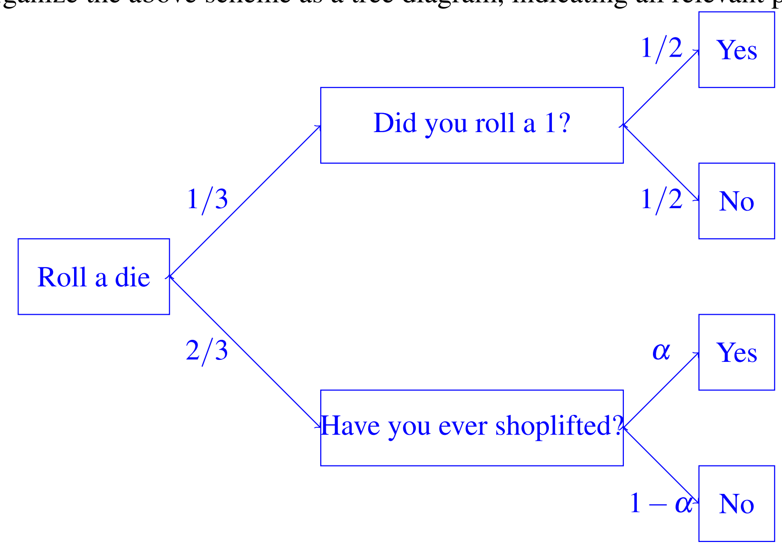
Thus, there are no outliers. One thing to note is that even though 15 is more than two standard deviations from the mean, in this context, it is NOT an outlier.

PROBLEM 3:

In social sciences, often times studies will use what is called a *randomized response sample*. The setup is as follows: You want to ask some number of people a sensitive question (ex. 'have you ever shoplifted?'), but are aware that people will be hesitant to tell the truth. To remedy this, you insert some randomness.

A test subject first rolls a six-sided die in private. If they roll a one or a two, then they answer the question "Did you roll a 1?". If they roll a three, four, five, or six, then they answer the question "have you ever shoplifted?". Under this scheme, researchers cannot know for sure that someone who answered "Yes" is a shoplifter, and so assume now that people always answer truthfully. Let the probability of someone has shoplifted be denoted by α.

Organize the above scheme as a tree diagram, indicating all relevant probabilities.



Write an expression for the probability that someone has shoplifted given they answered yes.

By Bayes' theorem,

$$P(\text{Shoplifted} \mid \text{Answered Yes}) = \frac{\left(\frac{2}{3}\right)(\alpha)}{\left(\frac{2}{3}\right)(\alpha) + \left(\frac{1}{3}\right)\left(\frac{1}{2}\right)}$$

Write an expression for the probability of a Yes response.

By the law of total probability,

$$P(\text{Answered Yes}) = \left(\frac{2}{3}\right)(\alpha) + \left(\frac{1}{3}\right)\left(\frac{1}{2}\right)$$

Say in a sample of 300 subjects, you receive 70 'Yes' responses. Use the relative frequency to approximate the probability of a 'Yes' response, and solve for α in the above expression.

$$P(\text{Answered Yes}) \approx \frac{70}{300} = \frac{2\alpha}{3} + \frac{1}{6}$$

$$\left(\frac{7}{30} - \frac{1}{6}\right)\left(\frac{3}{2}\right) = \alpha$$

$$\alpha = \frac{1}{10}$$

PROBLEM 4:

Problem 1 (a) is a multiple choice question with four options, one of which is correct. If you were to randomly choose among the four options, what is the probability of guessing the correct answer?

$$\frac{1}{4}$$

Problem 1 is made up of ten similar multiple choice questions Let X be the random variable that counts the number of correctly answered questions. What distribution does X have?

The Binomial Distribution

With X as above, what are the following:

$p = 1/4$

$n = 10$

$P(X = r) = {}_{10}C_r \left(\frac{1}{4}\right)^r \left(\frac{3}{4}\right)^{n-r}$

$\mu = (10)\left(\frac{1}{4}\right) = 2.5$

$\sigma^2 = (10)\left(\frac{1}{4}\right)\left(\frac{3}{4}\right) = \frac{30}{16} = 1.875(\approx 2)$

$\sigma = \sqrt{\frac{30}{16}} \approx \sqrt{2} \approx 1.4$

Maximum usual value = $\mu + 2\sigma = 2.5 + 2\left(\sqrt{\frac{30}{16}}\right) \approx 5.3$

Minimum usual value = $\mu - 2\sigma = 2.5 - 2\left(\sqrt{\frac{30}{16}}\right) \approx -0.3$

Say there is a student that I think has randomly chosen their answers as described, and they correctly answered 9 out of 10 questions. Write an expression for the probability that X = 9 (under these assumptions). Should I reject my assumptions?

Right off the bat, 9 is higher than the maximum usual value, but we need to do a bit more than that.

$$P(X=9) = {}_{10}C_9 \left(\frac{1}{4}\right)^9 \left(\frac{3}{4}\right)^1 = \frac{30}{4^{10}}$$

To find out if I should reject my assumption, I need to see if $P(X \geq 9)$ is small since this is a probability distribution.

$$P(X=10) = {}_{10}C_1 \left(\frac{1}{4}\right)^{10} \left(\frac{3}{4}\right)^0 = \frac{1}{4^{10}}$$

$$P(X \geq 9) = \frac{31}{4^{10}}$$

Since $P(X \geq 9)$ is really small (even without calculating exactly, is clearly less than 0.05 or 0.01 for that matter), getting 9 correct is a rare event, thus we should reject our assumption.

In other words, this student probably did not guess randomly.

BONUS:

We learned of a probability distribution that, under appropriate conditions, can approximate the probability distribution in Problem 4. What is this mysterious probability distribution? Under our setup from Problem 4, is it appropriate to use it for approximations? Why or why not? If so, approximate the probability that at least two questions are answered correctly.

The mystery distribution is the Poisson distribution, $P(X=r) = \frac{\mu^r}{r!e^{\mu}}$. However, the conditions are not satisfied!

$$n = 10 < 100$$

$$\mu = 2.5 < 10$$

Therefore, NO, it is not appropriate to approximate using Poisson.

If we decied to try anyways, the probability would be:

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - \frac{2.5^0}{0!e^{2.5}} - \frac{2.5^1}{1!e^{2.5}} = 1 - \frac{3.5}{e^{2.5}}$$