MAT 223, Final Exam, Fall 2015 Name KEY

**Instructions**: Show all work. If you use your calculator to make computations, state the command you used, any values you entered into the function, and state the results of the calculation operation, then be sure to interpret it to answer the question in lieu of using by-hand computations. (I cannot grade work I do not have.) Problems that provide only answers and no work when work is required will not receive full credit. Round appropriately. Be sure to answer all parts of each question completely. For all hypothesis tests, be sure to state both hypotheses clearly and with correct notation even if not explicitly asked to do so.

1. When we are constructing confidence intervals or a hypothesis test, why do we use the sampling distribution standard error rather than the standard deviation of the population distribution? (4 points)

because we are calculating properties of a sample statistic not an individual observation

2. Find the sampling distribution standard deviation for a sample of 103 people, with 61 respondents answering in the affirmative. (5 points)

$\hat{p} = \frac{61}{103}$ $1-\hat{p} = \frac{42}{103}$

$$s_{\hat{p}} = \sqrt{\frac{\left(\frac{61}{103}\right)\left(\frac{42}{103}\right)}{103}} = \sqrt{\frac{.592(.408)}{103}} = .0484 \approx 4.8\%$$

3. A random sample of 1301 adult Americans was asked whether television was a luxury or a necessity. Of those surveyed, 521 indicated that television was a luxury they could do without. Construct and interpret at 95% confidence interval for the number of Americans who believe that television is a luxury. (8 points)

1PropZInt

$x = 521$
$n = 1301$
C-level: .95

(.37384, .42709)

or (37.4%, 42.7%)

We are 95% confident that between 37.4% and 42.7% of Americans believe TV is a luxury.

4. A recent Gallup poll asked American the number of book they read in the previous year. A preliminary survey found that the standard deviation of their sample was 16.6 books. Under that assumption, what is the minimum sample size needed to obtain a margin of error estimate of 4 books with 90% confidence? (8 points)

$$n = \left(\frac{z_{\alpha/2}\, s}{E}\right)^2 = \left(\frac{1.645 \cdot 16.6}{4}\right)^2 = 46.6045\ldots$$

47 or more.

5. In a random sample of 100 estate tax returns that was audited by the IRS, it was determined that the mean amount of additional tax owed was $3421 with a standard deviation of $2583. Construct and interpret a 99% confidence interval for the mean additional amount owed on estate tax returns. (8 points)

T-Interval (Stats)

(2742.6, 4099.4)

$\bar{x} = 3421$

$s_x = 2583$ (sample)

$n = 100$

C-level: .99

We are 99% confident that between $2,742.60 and $4099.40 is owed on all audited types returns by the IRS.

6. Consider the following scenario:
According to the CTIA-The Wireless Association, the mean monthly cell phone bill was $47.47 in 2010. A researcher suspects that the mean monthly cell phone bill is different today. (15 points)
   a. State the null and alternative hypotheses in this situation with appropriate notation.

$H_0: \mu = 47.47$

$H_a: \mu \neq 47.47$

   b. Is the hypothesis test one-tailed or two-tailed?

two-tailed

c. What kind of information would you need in addition to what was provided to determine if the test is a z-test or a t-test?

Sample size & Standard deviation

d. What is the relationship between the p-value and the chance of a Type I error in the context of this problem?

The p-value is the chance that the null hypothesis is true but we obtained an unusual result just by chance (i.e. our sample says ≠ $ 47.47, but it really is still $ 47.47)

e. What is a Type II error in the context of this problem?

It is the chance we obtain a result that says we are equal to (or sufficiently close to) $ 47.47 when in fact, the average bill has changed.

7. The mean score on the SAT Math Reasoning exam is 516. A test prep company claims that the mean score of students who take their course is higher than this. Conduct a test of this claim if the mean score of students who take the test prep course is 522 based on a sample of 87 students. It is known that the population standard deviation of test is 126. Is there sufficient evidence to think that the test prep company is correct using a significance level of $\alpha = 0.01$? (8 points)

$H_0: \mu = 516$
$H_a: \mu > 522$

Z-Test (Stats)
$\mu_0 = 516$
$\sigma = 126$
$\bar{x} = 522$
$n = 87$
$\mu > \mu_0$

$z = .444$
$p = .328 > .01$
fail to reject $H_0$

There is not sufficient evidence to think SAT scores have improved.

8. In August 2002, 47% of parents who had children in grades K-12 were satisfied with the quality of education the students received. In September 2010, the Gallup organization conducted a poll of 1013 parents who have children in K-12 and asked the same question. If those surveyed, 437 indicated that they were satisfied. Is this sufficient evidence to conclude that attitudes toward education have changed since 2002? (8 points)

$H_0: p = .47$
$H_a: p \neq .47$

1PropZTest
$p_0 = .47$
$X = 437$
$n = 1013$
$prop \neq p_0$

$z = -2.462$
$p = .0138 < .05$

reject $H_0$
There is sufficient evidence to think attitudes toward education have changed.

9. In 2002, the mean age of an inmate on death row was 40.7 years, according to data obtained from the U.S. Department of Justice. A sociologist wondered whether the mean age of a death-row inmate has changed since then. She randomly selected 32 death-row inmates and found their mean age is 38.9 with a standard deviation of 9.6. Is there sufficient evidence to conclude that the mean age of death row inmates is actually lower now than it was in 2002? (8 points)

$H_0: \mu = 40.7$
$H_a: \mu < 40.7$

T-Test (Stats)
$\mu_0 = 40.7$
$\bar{x} = 38.9$
$Sx = 9.6$
$n = 32$

$\mu < 40.7$

$t = -1.06$
$p = .1485 > .05$

fail to reject $H_0$
There is not sufficient evidence to think age of death row inmates is lower than before

10. List the events in the sample space of tossing a decahedral (10-sided) die with a fair coin. (4 points)

{1T, 1H, 2T, 2H, 3T, 3H, 4T, 4H, 5T, 5H, 6T, 6H, 7T, 7H, 8T, 8H, 9T, 9H, 10T, 10H}

11. Answer each of the following questions based on the contingency table shown. (12 points)

| | Dog | Cat | Total |
|---|---|---|---|
| Male | 42 | 10 | 52 |
| Female | 9 | 39 | 48 |
| Total | 51 | 49 | 100 |

a. What is the probability, according to this sample 100 people, that some is both male and owns a cat?

$$\frac{10}{100} = 10\%$$

b. What is the probability that someone is either male or owns a cat?

$$\frac{52}{100} + \frac{49}{100} - \frac{10}{100} = \frac{91}{100} = 91\%$$

c. What is the probability of being male, given that the person owns a cat?

$$\frac{10}{49} \approx 20.4\%$$

d. Is the probability of being male and the probability of owning a cat independent? Show calculations to justify your answer.

no, they are not independent.
The probability of being male is 52% in this sample, not 20.4%.

12. Suppose that there is a large bowl of marbles that contains 35 blue marbles, 11 white marbles, 41 black marbles and 17 green marbles. What is the probability of selecting a green marble followed by a blue marble (without replacement)? (4 points)

$35+11+41+17=104$

$\frac{17}{104} * \frac{35}{103} = .0555$ or $5.6\%$

13. What is the probability of get two pairs in a five card poker hand? (6 points)

$$\frac{13(4C2)\,12(4C2)\,44}{52C5} = .095078$$

or $9.5\%$

14. Suppose that there are 14 kids on a baseball team. The coach has to choose a line-up of nine batters to start the first inning. How many ways can that line-up be chosen? (4 points)

$14P9 = 726,485,760$

15. Suppose that a password must be 10 characters long and must start with a capital letter and end with a number, but all other characters can be any letter (capital or lower case), any number, or any of 32 special characters on the standard keyboard. How many possible passwords of this type are there? (5 points)

$52+10+32=94$

$26 \cdot 94^8 \cdot 10 = 1.58 \times 10^{18}$

16. Suppose that 231 tickets are sold at a raffle for $10 per ticket. The top prize given away is worth $500 in prize money, the two second prizes are worth $150 in prize money each, and the three third prizes are worth $50 each. What is the expected value of purchasing a ticket to this raffle? [Hint: It may help to construct a table.] (10 points)

| $x$ | 490 | 140 | 40 | $-10$ |
|---|---|---|---|---|
| $P(x)$ | $1/231$ | $2/231$ | $3/231$ | $225/231$ |

$-\$5.89$

$490(1/231) + 140(2/231) + 40(3/231) - 10(225/231) = -5.887$

17. Consider a 60-sided dice, with probability of 12/60=1/5 to roll a number divisible by 5. Call obtaining such a number divisible by 5 a success. (16 points)
    a. What is the probability of obtaining exactly 10 successes if one rolls 100 such dice?

$p = 1/5,\ n = 100,\ x = 10$

$\text{binomialpdf}(100, 1/5, 10) = .00336\ldots$

    b. What is the probability of obtaining no more than 2 successes if one rolls only 12 such dice?

$n = 12,\ x = 2 \quad p = 1/5$

$\text{binomialcdf}(12, 1/5, 2) = .5583\ldots$

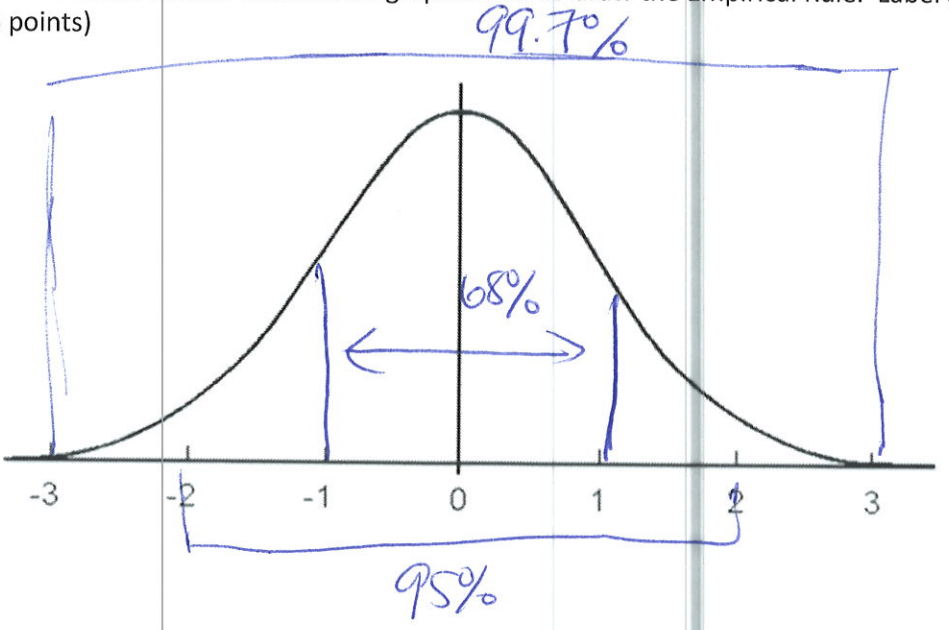
    c. What is the expected number of successes if one rolls 50 such dice repeatedly?

$np = 50(1/5) = 10$

    d. For the scenario in part c, what is the standard deviation of distribution on obtains from such samples?

$\sigma = \sqrt{np(1-p)} = \sqrt{50(1/5)(4/5)} = \sqrt{8} = 2.828\ldots$

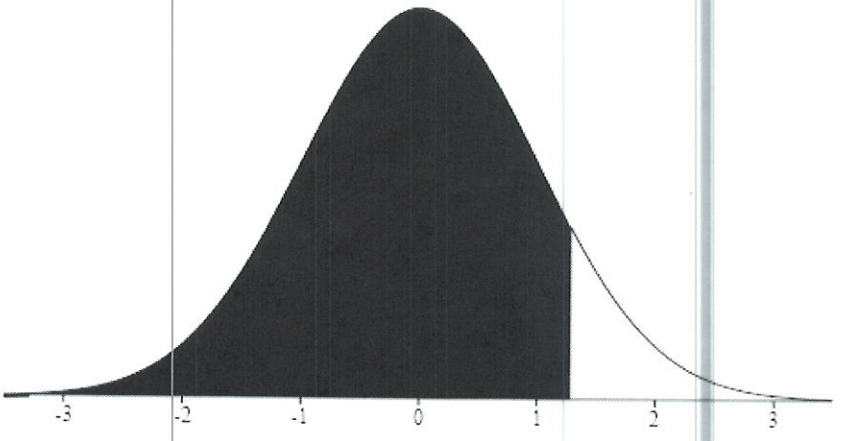
18. Use the blank normal distribution graph below to draw the Empirical Rule. Label appropriately. (5 points)




19. What are the qualitative differences between the standard normal distribution and the t-distribution? (4 points)

bigger tails, depends on sample size

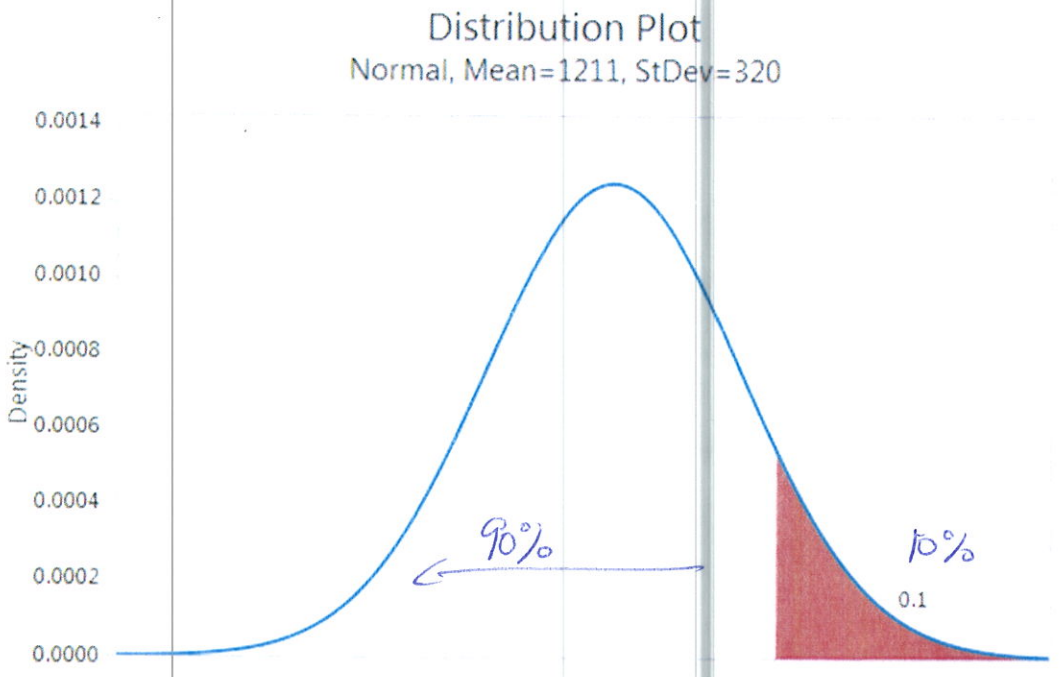
20. Find the probability represented by the shaded region bounded at $z = 1.35$. (4 points)




normalcdf(-E99, 1.35) = .91149...

91.1%

21. Find the x-value on the boundary of the region shown in the graph below that marks the bottom of the top 10% of the distribution, with the given mean and standard deviation shown. (6 points)




invNorm(.90, 1211, 320) = 1621.096501

22. For each of the variables below, determine the type of variable it is: a) qualitative or quantitative, b) level of measurement, c) if the variable is quantitative, is it discrete or continuous? (9 points)

a. Length of voicemail messages.

quantitative, ratio, continuous

b. Marital status

qualitative, nominal

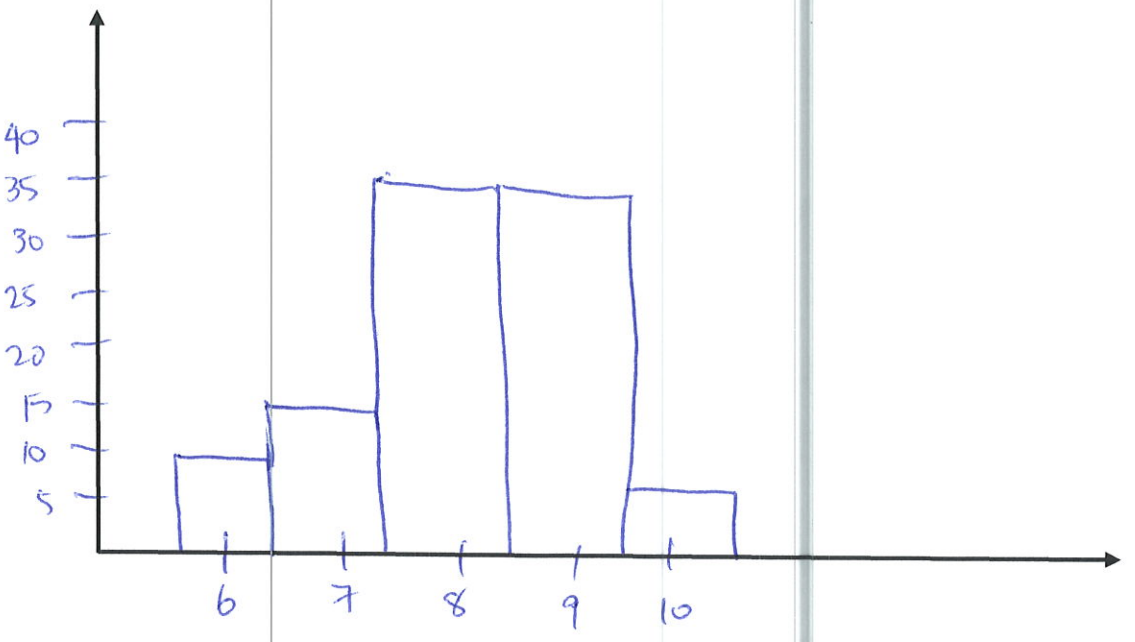
c. Runs scored (baseball)

quantitative, ratio, discrete

23. The frequency table below gives scores for the number of students in a statistics class who got a particular score on a 10-point quiz. Fill in the relative frequency values for each score in the last column. (5 points)

| Score | Frequency | **Relative Frequency** (round your answers to one decimal place) → after converting to % |
|---|---|---|
| 6 | 2 | 2/20 = 10% |
| 7 | 3 | 3/20 = 15% |
| 8 | 7 | 7/20 = 35% |
| 9 | 7 | 7/20 = 35% |
| 10 | 1 | 1/20 = 5% |
| | 20 | |

Sketch a bar graph of the relative frequencies. Be sure to label your graph appropriately. (5 points)



24. Below is data in a stem-and-leaf plot.

| Stem | Leaf |
|---|---|
| 4 | 0 |
| 5 | 4 8 |
| 6 | 0 0 0 3 3 5 5 7 7 9 |
| 7 | 0 0 1 1 1 2 2 3 4 4 4 6 7 7 8 8 9 |
| 8 | 0 0 0 1 2 3 4 8 |
| 9 | 0 1 2 3 |
| 10 | 9 |

Key: 6|0 = 60.

a. Find the mean and standard deviation of the data. (3 points)

$\bar{x} = 73.9767$ or $74.0$

$S_x = 12.078$ or $12.08$

b. Find the 5-number summary. (3 points)

Min = 40
$Q_1 = 67$
Med = 74
$Q_3 = 80$
Max = 109

c. Determine if there are any outliers in the data set. (3 points)

IQR = 80 − 67 = 13 × 1.5 = 19.5

67 − 19.5 = 47.5
80 + 19.5 = 99.5

yes, both 40 & 109

d. What is the percentile of the value 90 in this distribution? (3 points)

90 is 39th term

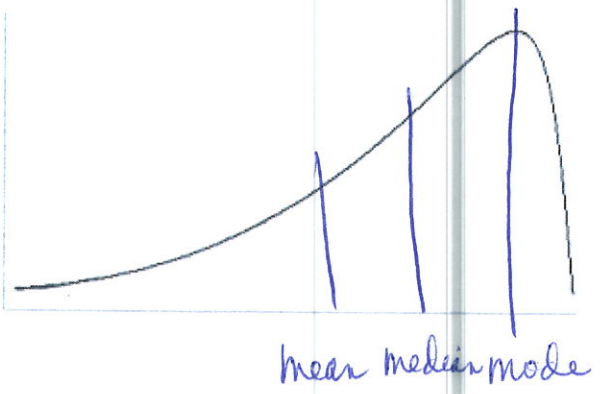
$\frac{39}{43} = 90.7\%$ile

Same result using mean and st. dev.

e. Use this information to graph (to scale) a box-and-whisker plot. (5 points)

25. For the distribution shown below, determine the shape of the distribution (symmetric, uniform, skewed right, skewed left, normal, etc.). Label the relative positions on the graph of the mean, median and mode. (3 points)

left-skewed




26. Below are a set of exam scores and revision time for those exams. Answer each of the questions based on the data in the table.

| ↓ | C1 | C2 |
|---|---|---|
| | Exam score | Revision time |
| 1 | 94 | 80.0 |
| 2 | 88 | 60.0 |
| 3 | 71 | 56.0 |
| 4 | 75 | 47.0 |
| 5 | 71 | 43.0 |
| 6 | 68 | 40.0 |
| 7 | 73 | 67.0 |
| 8 | 57 | 16.0 |
| 9 | 59 | 21.0 |
| 10 | 65 | 54.0 |
| 11 | 67 | 40.0 |
| 12 | 60 | 39.0 |
| 13 | 60 | 38.0 |
| 14 | 45 | 8.0 |
| 15 | 61 | 34.0 |
| 16 | 50 | 16.0 |

a. Plot the data on a scatter plot on your calculator and sketch the graph in the space next to the table. Does the data seem to be approximately linear? (6 points)

Yes, looks approximately linear

b. Find a linear regression equation for the data in the table. State the equation below. (4 points)

$y = 1.405x - 52.274$

c. What is the correlation for the regression equation? Does it agree or disagree with your answer in part a? Why or why not? (4 points)

$r = .89959...$

Yes, the correlation looks high

d. What percent of the variation in revision time is explained by the relationship to exam score? (6 points)

$r^2 = .80926...$ 80.9%

e. Is it appropriate to predict revision time for a score of 77? What about 35? Why or why not? What is your prediction in each case? (6 points)

77 is within the original range of data, 35 is outside it. 77 is pretty safe, 35 is sketchier.

$1.405(77) - 52.274 = 55.911$ (about 56)

$1.405(35) - 52.274 = -3.099$ Negative time for revisions is bad

27. Describe the relationship between slope and correlation (if one exists). (3 points)

they have the same sign but usually different magnitudes