Stat 10 Lew FINAL IS 3/20/03 11:30am-2:30pm

FINAL WILL BE HELD IN THE LECTURE HALL

1. Mr. Joe Potato works very hard during the workweek (Monday through Friday), but likes to watch a lot of television on weekends. The number of minutes of television viewing for Joe, on each of 60 consecutive days, was recorded. For this data set of 60 values, which of the following would be true? (Choose one)

a. The mean of this data set would smaller than the median.
b. The data set would be skewed left.
c. The data set would be skewed right.
d. Both a and b are true
e. Both a and c are true

2.The GPA (on a scale of 1-4) of a sample of students at UCLA has sample mean 2.9106 and sample standard deviation 0.256, with 95% confidence interval (2.488,3.333). Which of the following is correct?

a. If we sample many times, the proportion of intervals computed that cover the true mean is about 95%.
b. There's a 95% chance that the true mean falls between 2.488 and 3.333.
c. The estimate is within 95% of the true mean.
d. The population has a normal distribution with mean 2.9106 and standard deviation 0.256.
e. None of the above are correct

3.Which of the following is always true?

a. 95% of the data are within 2 standard deviations of the mean.
b. The distribution of a variable is always bell-shaped.
c. Histograms always have means and medians.
d. Histograms whose means are greater than their medians are always left skewed
e. None of the above are correct

4. At least 68% of the values in a data set fall within 1 standard deviation of the mean. TRUE or FALSE. must be normal

5. If the smallest value in a data set is removed, it would cause the standard deviation to decrease. TRUE or FALSE. decreases variability

6. Which of the random variables listed below are continuous?

a) The time it takes for a tow truck to arrive.
b) The number of buttons on a shirt.
c) The distance a long jumper jumps in a competition.
d) All of the above.
e) Only (a) and (b)
f) Only (b) and (c)
g) Only (a) and (c)

7. If the correlation between body weight and annual income were high and positive, we could conclude that:

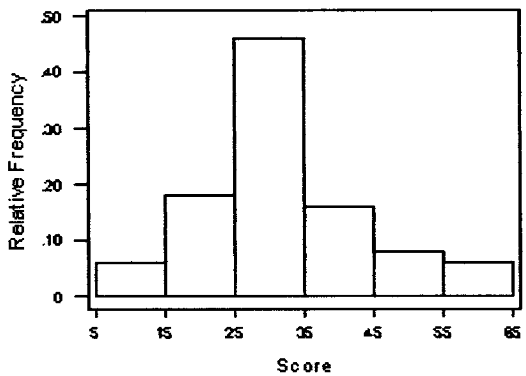
a. high incomes cause people to eat more food.
b. low incomes cause people to eat less food.
c. high income people tend to spend more of their income on food than low income people, on average.
d. high income people tend to be heavier than low income people, on average.
e. high incomes cause people to gain weight.

8. The correlation coefficient provides:

a. a measure of the extent to which changes in one variable cause changes in another variable.
b. a measure of the strength of the linear association between two categorical variables.
c. a measure of the strength of the association (not necessarily linear) between two categorical variables.
d. a measure of the strength of the linear association between two quantitative variables.
e. a measure of the strength of the linear association between a quantitative variable and a categorical variable.

FINAL WILL BE HELD IN THE LECTURE HALL

9. In the histogram below, which of the following is true? (choose 1)
a) The proportion of scores greater than 15 is 0.46.
b) The proportion of scores between 15 and 35 is 0.65.
c) The proportion of scores less than 55 is 0.05
d) Both a and b are true.
e) Both b and c are true.
f) Both a and c are true.



10. A correlation coefficient of -.7 is a negative and weaker correlation than +.50. TRUE OR FALSE

11. The standard deviation is a common measure of variability that displays the average distance of scores from the mean. TRUE OR FALSE?

12. Two confidence intervals are calculated for a proportion p: a 90% and a 99% confidence interval. Each confidence interval is based on the same random sample. Which one of the following statements is true?

A. The 99% confidence interval would be narrower.
B. The 90% confidence interval would be wider.
C. The 99% confidence interval would be wider.
D. It is not possible to determine which is wider and which is narrower, based on the information given.

13. Which of the following statements is most appropriate for a hypothesis test?

A. When the p-value is small, say less than .05, we can reject the null hypothesis or equivalently, accept the alternative hypothesis.
B. When the p-value is not small, for example if the p-value is greater than .10, we cannot reject the null hypothesis.
C. It is almost always correct to say "I accept the alternative hypothesis".
D. It is almost always correct to say "I accept the null hypothesis".

14. Mark ONE of the columns

True	False	Question
X		The sample standard deviation of a data set must be zero or larger.
	X	If X is a continuous random variable which is normally distributed with a mean of 100 and a standard deviation of 15 then the probability that X>115 is 0.5.
X		In order to have a valid probability distribution, the sum of the probabilities must equal to 1 or 100% and the probabilities themselves cannot be negative or greater than 100%.
X		The height of a randomly selected UCLA football player is a quantitative variable.
	X	If the mean of a variable is less than the median of that variable, it is correct to say that the distribution is right skewed.
X		The idea behind statistical inference is to find distribution(s) of statistic(s).

1. The IQ scores of adult humans (age 18 and over) is approximately normal with a mean of 100 and a standard deviation of 15.

(a) How low is the lowest 5% of all IQ scores (that is, at or below what IQ score is the lowest 5%) How high is the highest 10% of IQ scores (that is, at or above what IQ Score is the highest 10%)?

Use $z = -1.65$ for lowest 5% $\Rightarrow -1.65 = \frac{X-100}{15}$

Use $z = +1.30$ for highest 10% $\Rightarrow +1.30 = \frac{X-100}{15}$

solve for X in each case so

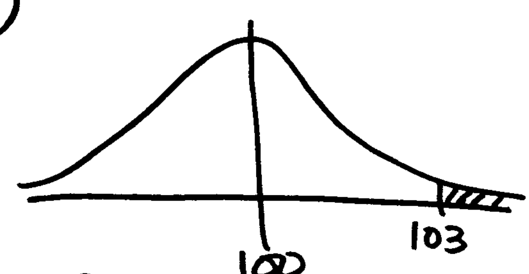
$X = 75.25$ and 119.50

(b) A simple random sample of 225 college students is drawn from the adult human population. The sample average is 103 and the sample standard deviation is 30. Please test the hypothesis that college students have higher IQ scores than the average human. State a null and an alternative hypothesis, perform a test, state a p-value and explain your result (do you reject or not reject the null and why). Use a 5% level of significance as your decision rule.

Null: Any difference between the statistic and the parameter is due to chance error, the true mean = 100

Alt: Any difference is NOT due to chance error and the true mean is > 100

Test: $$z = \frac{103-100}{\left(\frac{\sqrt{225} * 15}{225}\right)} = \frac{3}{1} = +3.0$$



P-value is

$$\frac{100-99.73}{2} \cong .14\%$$ much less than 5%

Conclusion: REJECT THE NULL

College students have HIGHER IQs than Average

2. Investors ask about the relationship between returns on investments (the money you make by investing your money) in the United States and on investments overseas. Below is a table of total returns on investments on U.S. and overseas stocks over a 10 year period.

	Year	Overseas % Return	U.S. % Return
	1987	24.6	5.1
	1988	28.5	16.8
	1989	10.6	31.5
	1990	-23	-3.1
	1991	12.8	30.4
	1992	-12.1	7.6
	1993	32.9	10.1
	1994	6.2	1.3
	1995	11.2	37.6
	1996	6.4	23
Average	1991.5000	9.8100	16.0300
Standard Deviation	2.7386	15.6493	12.6810

(a) Suppose the correlation, r, of the U.S. and overseas returns is .32. Using this information describe the relationship between U.S. and overseas returns in plain English.

the relationship is weakly positive
It suggests that as US returns
increase, so do overseas returns

(b) Find the regression line of overseas returns on U.S. returns. Please interpret the values of the slope and of the intercept of this line.

$$b = \text{slope} = (.3239)\left(\frac{15.6493}{12.6810}\right) = .3997$$

$$a = \text{Intercept} = (9.81) - [(.3997)(16.03)] = 3.403$$

$$\underset{(y)}{\text{overseas returns}} = 3.403 + .3997\left(\text{US returns}\right)$$

interpretation if US Returns = 0 the overseas returns are expected to be 3.403. For every one unit increase in US returns, overseas goes up by .3997

(continued from above)

(c) In 1993, the return on U.S. stocks was 10.1%, what was the predicted return on overseas stocks. Is the predicted return the same as the actual return? If it is the same, please explain why this is so. If it is different, please explain why they are different.

If we "plug" 10.1 into the regression equation (part b) we get $y = 7.44$. It's not the same as the true 1993 y of 32.9. This happens b/c the line is a "best linear fit" and it is NOT expected to go through all of the points

3. You got a job working for a marketing company and your supervisor is planning a sample survey of households in Los Angeles. Your supervisor instructs you to contact households by random-digit dialing phone numbers. Your supervisor knows from past experience that about 70% of the households you contact in this manner will respond.

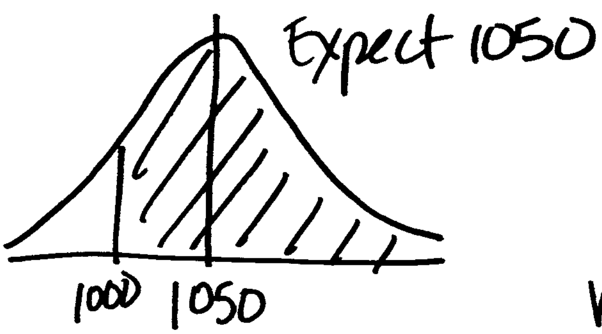
(a) If you randomly dial 1500 telephone numbers, what are the mean and standard error of the number of households who respond?

$\boxed{1}\ \boxed{0}$
.7 .3

Expected mean = 1,050 b/c $1500 \times .70 = 1{,}050$
or (box avg)*(# draws)

$SE_{sum} = \sqrt{1500} \times (1-0)\sqrt{.7 \times .3} = 17.75$

(b) Find the probability that you will get at least 1000 responses.



$z = \dfrac{1000 - 1050}{17.75} = -2.82$

round to 2.80

prob. is $99.49 + \left(\dfrac{100 - 99.49}{2}\right) = 99.75\%$

4. You are planning to perform a significance test of

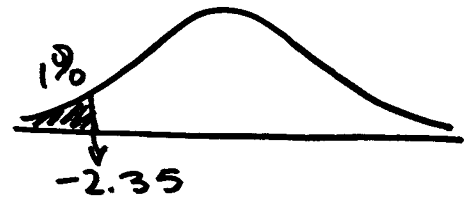
H_0: mean = 0

Versus

H_1: mean < 0

What values of Z would lead you to reject H_0 at the 1% level of significance? Then answer this question: True or False and explain why. A significance test that is significant at the 1% level of significance must always be significant at the 5% level of significance.

1) Any Z less than -2.35 b/c



2) TRUE. 1% is always less than 5%, see table A105

5. An investigator looks up the rainfall in a certain city on January 15 for the past 70 years. She finds the average rainfall on that day to be 0.30 inches and the SD to be about 0.14 inches. She then concludes that the interval from 0.25 to 0.35 inches is a 99.7% confidence interval for the average rainfall on January 15 in the city. Is this conclusion justified? Why or why not?

No. This is NOT a random sample of rainfall and it is not a repeatable process (it's history). She should not calculate C.I.s

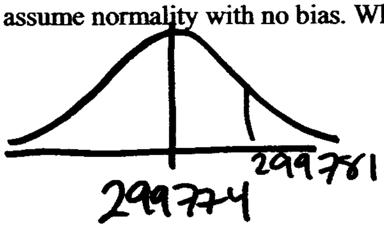
6. The speed of light is measured 2,500 times by a new process. The average of these 2,500 measurements is 299,774 kilometers per second, with an SD of 14 kilometers per second.

a. Find an approximate 95% confidence interval for the speed of light. (You may assume normality, with no bias.)

$$299{,}774 \pm 2\left(\frac{\sqrt{2500} \times 14}{2500}\right)$$

$$299{,}774 \pm .56$$

b. Now the investigators determine the speed of light again by taking the average of another 2,500 measurements. The average is now 299,781 kilometers per second. Is this a surprising result? Again, you may assume normality with no bias. Why or why not?



$$z = \frac{299781 - 299774}{\frac{\sqrt{2500} * 14}{2500}} = \frac{7}{.56} = 12.5$$

z is "off the chart". THIS IS SURPRISING. A 2nd sample has almost NO CHANCE of being so far from the E.V.

7. In government data, a household consists of all occupants of a dwelling unit. Choose an American household at random and count the number of people it contains. Here is the assignment of proportions for your outcome:

Number of persons	1	2	3	4	5	6	7
Box Proportions	0.25	0.32	???	???	0.07	0.03	0.01

The proportions of 3 people in a household is the same as that of 4 people. These proportions are marked with a ???.

a) Find the proportions of households that contain 3 people.

$1.0 - (.25 + .32 + .07 + .03 + .01) = 1.0 - .68$
$= .32$

if 3=4 then $\frac{.32}{2} = \boxed{.16}$

b) Pretend the table above is a box model. What is the box average?

$(1 \times .25) + (2 \times .32) + (3 \times .16) + (4 \times .16)$
$+ (5 \times .07) + (6 \times .03) + (7 \times .01) = \boxed{2.61}$

c) 100 families are going to be drawn at random from the "box" and will become a part of a new study on poverty. What is the expected number of people in the study?

$\underbrace{100}_{\text{\# of draws}} \times \underbrace{2.61}_{\text{box avg.}} = 261$ people

8. Suppose that 47% of all adult women think they did not get enough time for themselves. An opinion poll interviews 1025 randomly chosen women and records the sample proportion that doesn't feel they get enough time for themselves. This statistic will vary from sample to sample if the poll is repeated. The sampling distribution is approximately normal with mean 0.47 and standard error about 0.016.

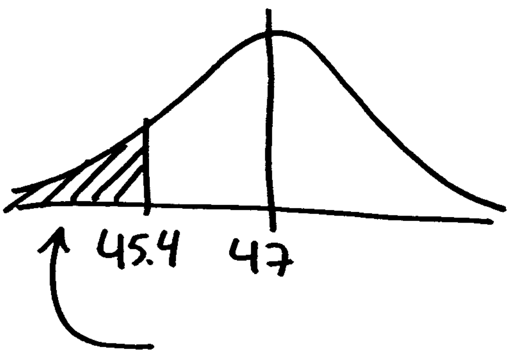
a) The truth about the population is 0.47. In what range will the middle 95% of all sample results fall for samples of size 1,025?

$$47\% \pm 2\left(\frac{\sqrt{1025} * \sqrt{.47 \times .53}}{1025} \times 100\right)$$

$\hookrightarrow 1.5589$

$47\% \pm 3.1178\%$

b) What is the probability that a new poll of size 1,025 gets a sample in which fewer than 45.4% say they do not get enough time for themselves?



$$z = \frac{45.4-47}{1.5589} \hookrightarrow -1.03$$

use $z = 1.05$

$$\frac{100-70.63}{2} = 14.685\%$$

9.A study of many families gave the following results:

average height of father=68 inches, SD=3 inches
average height of daughter=63 inches, SD=2.5 inches
r=0.6

Using the regression method, estimate the height of a daughter whose father is 62 inches tall

1. $z = \frac{62-68}{3} = -\frac{6}{3} = -2$ so he's 2 SD below average
2. $-2 \times \underset{r}{0.6} = -1.2$
3. $-1.2 * 2.5 = -3$ inches so our estimate is $63 - 3 = 60$ inches

10. Does salt cause high blood pressure? One large study was done at 52 centers in 32 counties. Each center recruited 200 subjects in 8 age- and sex- groups. Salt intake was measured, as well as blood pressure and several possible confounding variables. After adjusting for age, sex, and the confounding variables, 25 of the centers found a positive association between diastolic pressure and salt intake; 27 found a negative association. Do the data support the theory that salt causes high blood pressure? Answer yes or no, and explain briefly.

No, Not conclusively, this is an observational study
Not a randomized controlled experiment,
Association is NOT causation so we should
avoid saying SALT "causes" high blood pressure

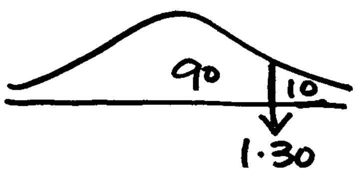
11. A study on pre-meds, selected at random, gives the following results for the medical college admissions test (MCAT) and undergraduate GPA (grade point average):

Average GPA: 3.3; Standard deviation = 0.4
Average MCAT: 10; Standard deviation = 1.1
Correlation coefficient = 0.65

Suppose the percentile rank of one student's GPA is 90%. Predict the student's percentile rank on the MCAT. The scatter diagram is football shaped and the MCAT and the GPA are normal.

If GPA is the 90th percentile then $z = +1.3$

b/c



so $(1.3 \times .65) = .85 = z_{MCAT}$

correlation

area is $\left(\frac{100-60.47}{2}\right) + 60.47 \cong 80\%$ so around the 80th percentile

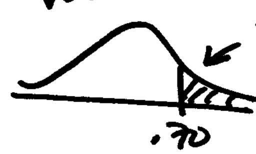
12. The pregnancy duration of human females (age 18 and over) is approximately normal with a mean of 266 days and a standard deviation of 16 days. It is believed that older pregnant women have longer pregnancy durations. A simple random sample of 121 older pregnant women is drawn from the population of all pregnant women. The average pregnancy duration for the sample is 267 days and the sample standard deviation is 35. Please test the hypothesis that older women have longer pregnancy durations than the average woman. State a null and an alternative hypothesis, perform a test, state a p-value and explain your result (do you reject or not reject the null and why). Use a 5% level of significance as your decision rule.

16 = BOX SD   266 = BOX AVG

Null: older = younger
AH: older > younger (longer)

test $z = \frac{267-266}{\left(\frac{\sqrt{121} * 16}{121}\right)} = \frac{1}{1.4545} = .69 \cong .70$

p-value is



DO NOT REJECT NULL B/c 24% > 5%

No evidence to support the alternative

FINAL WILL BE HELD IN THE LECTURE HALL

13. The following comes from a recent article in *The Wall Street Journal* :

Generation X-ers Aren't Relying On the Survival of Social Security

BY JOHN SIMONS

According to the most recent Wall Street Journal/NBC News poll, only 39% ofX-ers believe that Social Security will still be able to provide benefits when they retire. That compares to recent surveys of all Americans which show that 45% think so.

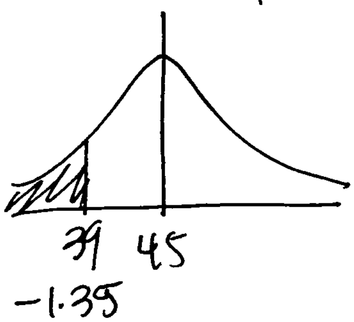
Assume that 45% figure is a stable, long-run, historical fact about American beliefs about Social Security benefits. Also assume that the survey of Generation X-ers had 121 respondents.

A. Test the hypothesis that the belief that Social Security will still be able to provide retirement benefits has decreased over time. State the null and the alternative, perform a test, and state a p-value. Please use a 5% level of significance as your decision rule. On the basis of your test results, do you think that Generation X-ers are like other Americans in their beliefs about social security or are they different?

Null: The percentage is equal to 45%

Alt: the percentage is < 45%

$$z = \frac{39-45}{\left(\frac{\sqrt{121} * \sqrt{.45 \times .55}}{121} \times 100\right)} \approx -1.35$$

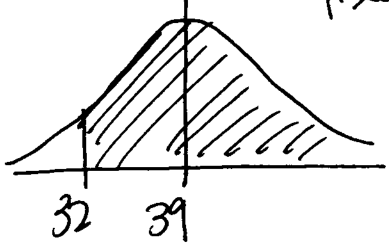


$$\frac{100-82.30}{2} = 8.9\%$$

8.9% > 5% so do not reject the null. The sample does not provide evidence that these people are different from older americans in their beliefs

B. Suppose in a few years the Wall Street Journal decided to replicate this study (i.e.draw a new sample) on Generation Y-ers (that's you all, I think...). Let's assume that the 39% figure is now the stable, long run fact about belief in Social Security benefits by Americans.

What is the chance that a sample of 100 will have at least 32% of the surveyed Generation Y-ers believing in Social Security? fixed. Sorry.



$$z = \frac{32-39}{\left(\frac{\sqrt{100} * \sqrt{.39 \times .61}}{100} \times 100\right)} \approx -1.45$$

area is $85.29\% + \left(\frac{100\% - 85.29\%}{2}\right) = 92.65\%$

or

$$\frac{85.29\%}{2} + 50\% = 92.65\%$$

14. Here are two statistics on all persons who consider themselves medical doctors in 1999:

\$820,000 dollars per year
\$141,000 dollars per year

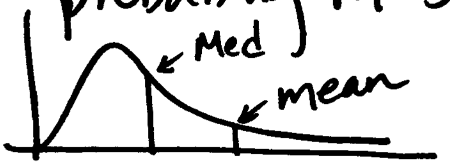
Which one of these numbers is the mean salary and which one is the median?

The mean is 820,000

The median is 141,000

Explain your choice in the space below. Be brief. This is not a long answer.

the minimum salary is 0, but the max can be very large therefore the distribution is probably RT SKEWED and so $\bar{x}$ > Median



15. High Bias and High Variance are both considered undesirable features of certain sample statistics (such as a sample mean for example). You are working with a team on a marketing study, a sample of size 100 is drawn. One of the variables you are interested in is the average time spent on the internet on any day. You plan to construct confidence intervals and perform some unspecified hypothesis tests. Studies always have problems, and today you have your choice: High Bias or High Variance. Which one would you rather deal with and why?

High Variance.

Your estimates may be more trustworthy and usable IF the differences we're examining (i.e. Chance Error) are also large. At least the variance issue could be addressed with a large sample.

This will not be so for high bias. First, your estimate will be OFF and second, getting a larger biased sample will not help.

16. A marketing survey interviewed 1000 adults selected at random from the population of all U.S. adults. Of the adults, 529 said they currently own a personal computer. When asked about the manufacturer of their computer, 144 of them said "Dull", 115 of them said "Compact", 275 of them said "some other company" and the rest of them said "I don't know". The mean time of ownership (in months) for the 529 was 12.9 with a standard deviation of 8.7.

(a) A Compact executive saw the survey and is now upset, he believes that the survey was poorly done and argues that Compact's true market share is 25% (i.e. he thinks that 25% of all adults who own computers own a Compact) and cannot be nearly as low as the survey suggests.

Let's help the executive out. Please test the hypothesis that Compact's market share is actually 25%. Use a 5% level of significance as your decision rule. State the null hypothesis, the alternative hypothesis, perform a test, give a p-value, and state your conclusion in plain English: would you reject the null and on the basis of your test result do you also think the survey was poorly done?

Null: proportion is = .25 or 25%

Alt: proportion is < .25 or < 25% $\qquad \frac{115}{529} = .2174$

or %

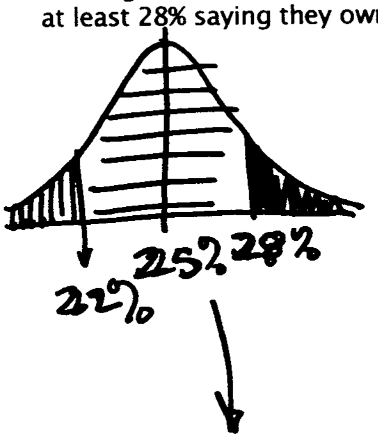
$$z = \frac{21.74 - 25}{\left(\frac{\sqrt{529} \cdot \sqrt{.25 \times .75}}{529} \times 100\right)} = \frac{-3.26}{1.88} \approx -1.75$$

← SE%

p-value is < .05

so we REJECT THE NULL, we don't think the executive is correct. We think our survey is right

(b) Suppose Compact's market share REALLY IS 25%. What is the chance that among 529 computer owners you would get less than 22% of them saying they owned a Compact? What is the chance that you would get between 22% and 28% saying they owned a Compact? What is the chance that you would get at least 28% saying they owned a Compact?



AREA $= z = \frac{22-25}{\left(\frac{\sqrt{529} \times \sqrt{.25 \times .75}}{529} \times 100\right)} = -1.60$

AREA $= z = \frac{28-25}{\left(\frac{\sqrt{529} * \sqrt{.25 \times .75}}{529} \times 100\right)} = +1.60$

So = area between ± 1.60 z is 89.04%

and the other two areas are $\frac{100-89.04}{2} = \boxed{5.48\%}$

17 A study of 250 first year college students, selected at random after their first full year of college, gives the following results for the course units (NUNITS) first year GPA (grade point average) and high school SAT I score:

Average GPA:	3.3;	Standard deviation GPA = 0.4
Average NUNITS:	38;	Standard deviation NUNITS: 6.1
Average SAT I:	1010;	Standard deviation SAT I = 201

Correlation coefficient for GPA and NUNITS= .35
Correlation coefficient for GPA and SAT= .65
Correlation coefficient for NUNITS and SAT= -.55

Assume SAT scores have a minimum of 400 and a maximum of 1600 and are normally distributed. NUNITS has a minimum of 12 and a maximum of 60 and is not normally distributed. GPA has a minimum of 0.0 and a maximum of 4.3 and it is not normally distributed.

(a) Please interpret the value of the correlation coefficient for SAT and NUNITS in plain English. Discuss the direction and magnitude its value implies.

$r = -.55$ negative, therefore as SAT goes up NUNITS goes down and vice versa. It is not particularly strong or weak.

(b) A student is interested in regressing GPA on SAT I. Using the information at the top of the page, please find the regression equation. Clear identify the slope, intercept, x and y variables.

Y=GPA X=SATI

$$\text{slope} = (+.65)\left(\frac{0.4}{201}\right) = .0013$$

$$\text{Intercept} = 3.3 - (.0013)(1010) = 1.9935$$

$$GPA = .0013(\text{SAT I}) + 1.9935$$

(c) Please interpret the values of slope and intercept you calculated in part (b) in plain English.

A slope of .0013 suggests that for every one unit increase in SAT I score we expect to see a .0013 increase in GPA (or you could say that you would need a 100 point increase in SAT to see a .13 increase in GPA)
An intercept of 1.9935 suggests that students with a SAT I = 0 will have a GPA of 1.99

18. The amount of money all college students earn in the year after graduation is right skewed with a mean of $26,600 and a standard deviation of $5000. Apparently, 20% of all students earn no money in the year after graduation. Administrators at UCLA believe that UCLA students earn more money in the year after graduation than the average college student. A simple random sample of 250 UCLA students is drawn from the population of all recent UCLA graduates. The average for the sample was $27,300 and the sample standard deviation is $9,500. The sample also revealed that 15% of UCLA students earned no money in the year after graduation.

(a) Is it possible to test the hypothesis that UCLA students earn more than other college students? If you think it is, please state a null and an alternative hypothesis, perform a test, state a p-value and explain your result (do you reject or not reject the null and tell us how to interpret the result -- in plain English) and use a 5% level of significance as your decision rule.

If you think it is not possible to test the hypothesis, please use the space below to explain why this is not possible.

YES - the sample is of reasonable quality and size.

NULL: The average is $26,000 for UCLA students

ALT: The average is > $26,000 for UCLA students

$$Z = \frac{27,300 - 26,600}{\left(\frac{\sqrt{250} * 5000}{250}\right)} = \frac{700}{316.23} = 2.21 \sim 2.20$$

$$\text{p-value is } \left(\frac{100 - 97.22}{2}\right) = 1.39\%$$

Reject the null (1.39% < 5%)

UCLA students appear to earn more money that the average

FINAL WILL BE HELD IN THE LECTURE HALL

19.A simple lottery gamed called "The Daily 3" is played in California. You pay $1 to play and you get to choose a 3-digit number. Each digit can have any value from 0 to 9 and you can have repeated values (e.g. 111, 222 etc.) The State of California randomly selects a 3 digit number at the end of each day and pays you $500 if your number is selected. There are 1,000 3 digit combinations (different 3 digit numbers).

If you were to play every day for one year (365 plays). What is the chance that you will win money? You can treat your plays as if they were a random sample of size 365.

500	-1
$\frac{1}{1000}$	$\frac{999}{1000}$

$$\text{BOX AVG} = (500 * 1/1000) + (-1 * 999/1000) \cong -.50\ (-.499)$$

$$EV = 365 \times -.499 = -182.135$$

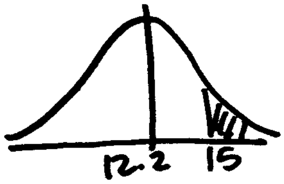
$$SE = \sqrt{365} * (500 - -1)\sqrt{\tfrac{1}{1000} + \tfrac{999}{1000}} = 302.529$$

$$Z = \frac{0 - -182.135}{302.529} = +.60$$

$\boxed{27.425\%}$

20.Suppose it known that Americans spend an average of 12.2 minutes in the shower each day. If the standard deviation of showering times is 2.3 minutes and times are normally distributed:

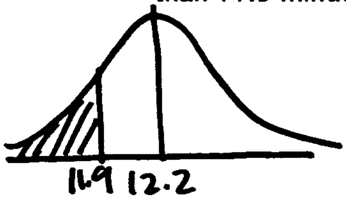
a. Find the percentage of Americans who spend at least 15 minutes in the shower



$$z = \frac{15-12.2}{2.3} = 1.22 \approx +1.20$$

$$\frac{100-76.99}{2} = \boxed{11.51\%}$$

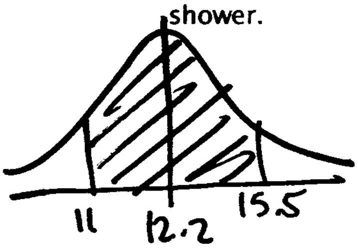
b. Calculate the chance that the average time of a random sample of 144 Americans will be less than 11.9 minutes



$$z = \frac{11.9-12.2}{\left(\frac{\sqrt{144}\cdot(2.3)}{144}\right)} = \frac{-.3}{.192} = -1.57 \approx -1.55$$

$$\frac{100-87.89}{2} = \boxed{6.055\%}$$

c. Find the percentage of Americans who spend between 11 minutes and 15.5 minutes in the shower.

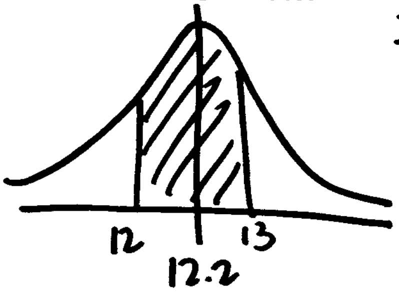


$$z_{15.5} = \frac{15.5-12.2}{2.3} = 1.43 \approx +1.45$$

$$z_{11} = \frac{11-12.2}{2.3} = -.52 \approx -.50$$

$$\frac{85.29+38.29}{2} = \boxed{61.79\%}$$

d. What percentage of random samples of size 100 will have a sample average between 12 and 13 minutes?



$$z_{13} = \frac{13-12.2}{\left(\frac{\sqrt{100} * 2.3}{100}\right)} = 3.48 \approx 3.50$$

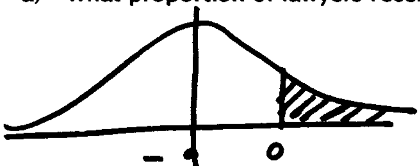
$$\frac{99.953+60.47}{2} = \boxed{80.21\%}$$

$$z_{12} = \frac{12-12.2}{\left(\frac{\sqrt{100} * 2.3}{100}\right)} = -.87 \approx -.85$$

FINAL WILL BE HELD IN THE LECTURE HALL

21. Lawyers frequently receive a "year-end bonus" because law firms are partnerships and the money earned is shared among partners. There are approximately 1,000,000 lawyers in the U.S. and year 2002's "year-end bonus", when calculated as a percentage change of the 2001 "year-end bonus", is normally distributed with a mean year-end bonus of -9% (a decrease, 2002 was a bad year compared to 2001) and a standard deviation of 16%. SHOW YOUR WORK FOR FULL CREDIT.

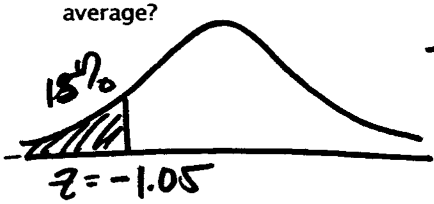
a) What proportion of lawyers received year-end 2002 bonuses that were larger than their year-end 2001?



$z = \frac{0-(-9)}{16} = \frac{9}{16} = .5625 \approx .55$

$\frac{100-41.77}{2} = \boxed{29.12\%}$

b) A sample of 100 lawyers has an average percentage change at the 15^{th} percentile, what is the value of its average?



$-1.05 = \frac{X-(-9)}{\frac{\sqrt{100}*16}{100}}$. solve for X $\boxed{X = -10.68\%}$

c) How many lawyers have bonuses between -1% and +5%?

$z_5 = \frac{5-(-9)}{16} = \frac{14}{16} = .875 \approx .90$

$z_{-1} = \frac{-1-(-9)}{16} = \frac{8}{16} = .50$

$\frac{63.19-38.29}{2} = \boxed{24.90\%}$

d) If it is possible to do this, please calculate the median and interquartile range for the probability histogram of samples of lawyers of size 225. If any of these are not possible to calculate, please identify the ones that are not calculable and please give values for the ones that are calculable.

Median = mean = -9%

75th $z = +.65 \longrightarrow +.65 = \frac{X-(-9)}{\frac{\sqrt{225}\cdot 16}{225}}$ solve for X = -8.31

25th $z = -.65 \longrightarrow -.65 = \frac{X-(-9)}{\frac{\sqrt{225}\cdot 16}{225}}$ solve for X = -9.69

IQR is -8.31 - -9.69 = 1.38

22. Mark ONE of the columns

True	False	Statement
	X	You can tell if a distribution is right-skewed if you know its standard deviation and its mean
X		Standard Errors are either zero or positive values
	X	For normal distributions, Chance Error and the Standard Error are equivalent
X		Selection Bias is a result of mistakes made by the person or persons who design a study which involves sampling
	X	Non-response Bias is caused by researchers who fail to write a survey questionnaire properly
	X	Chance error is the source of difference between statistics and parameters
X		Observational studies differ from randomized controlled experiments in that the researchers do not assign the subjects to treatment or control
X		Association may point to causation, but it is not the same as causation
	X	An extremely large biased sample (e.g. > 10,000) generates better estimates population parameters than a small random sample (e.g. a little over size 100)
X		A quantitative variable can be discrete or continuous

FINAL WILL BE HELD IN THE LECTURE HALL

23.The Dull Computer Company manufacturers its own computers and delivers them directly to customers who order them via the Internet. Dull's market dominance has arisen from its quick delivery and competitive pricing. The CEO (Chief Executive Officer) of Dull has stated publicly that if customers make unassisted online purchases of their computers (i.e. no interaction with a salesperson), the computers will have an average delivery time of 56 hours from the time of purchase (with a standard deviation of 18 hours). He also noted that about 20% of these computers cost less than $1300. The CEO refused to give a mean cost of these computers but said the standard deviation was $600. The cost of these computers is not normally distributed nor is the delivery time.

A consumer research organization decided to test the CEO's delivery time claim and cost claims by purchasing 200 computers from Dull at randomly selected times and days. The 200 purchases were randomly divided into two groups: 72 were purchased by telephone and involved talking to a live salesperson, the remaining 128 were unassisted online purchases. 32 of the 128 computers were delivered in less than 45 hours. The mean cost of the 128 purchases was $1,661 Please assume that the purchases (i.e. 200, 72, 128) constitute reasonably large samples of good quality.

(a) Using the information above and from what you have learned in this course, can you construct a 80% confidence interval for the population percentage of computers that will be delivered in less than 45 hours? If yes, circle yes below, justify your response and then construct the interval. If no, circle no below and justify your response. (5 points)

YES NO

conditions of CLT have been met

$$\frac{32}{128} = .25 \qquad 25\% \pm (1.30)\left(\frac{\sqrt{128} * \sqrt{.25 \times .75}}{128} \times 100\right)$$

$$25\% \pm 4.98\%$$

(b) Using the information above and from what you have learned in this course, can you construct a 80% confidence interval for the mean cost of all computers involved in online unassisted purchases? If yes, circle yes below, justify your response and then construct the interval. If no, circle no below and justify your response. (5 points)

YES NO

CLT has been met

$$1{,}661 \pm (1.30)\left(\frac{\sqrt{128} * 600}{128}\right)$$

$$1{,}661 \pm 68.94$$