FINAL WILL BE HELD IN HAINES A2 THURSDAY 12/12/02 8am-11am

13. 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

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

Intercept $= 3.3 - (.0013)(1010) = 1.9935$

GPA = .0013(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

FINAL WILL BE HELD IN HAINES A2 THURSDAY 12/12/02 8am-11am

14.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

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

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

Reject the null $(1.39\% < 5\%)$

UCLA students appear to earn more money that the average

(continued from above)

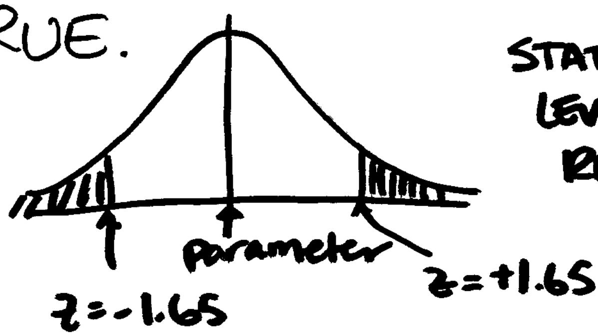
(b) Suppose a hypothesis test results in a p-value of .16 or 16%. Explain (using calculations or a verbal response or both) what this p-value means to your good friend who knows little about statistics and hasn't taken Stat 10 yet.

a p-value is an OBSERVED SIGNIFICANCE LEVEL (or you could call it a probability value).

This the chance of getting a test statistic, z as extreme or more extreme than the observed one (the one you got) This chance is computed based on the null being right. The smaller the p-value the stronger the evidence against the null. This one (.16) is large so we wouldn't reject the null

(c) Answer True or False and justify your response (using calculations are OK, but we are looking for a verbal response here for full credit). Achieving statistical significance at the 5% level (for a one-sided test) is like constructing a 90% confidence interval that doesn't contain the parameter.

TRUE.



STATISTICAL SIG. AT THE 5% LEVEL IMPLIES A SAMPLE RESULT THAT WAS AS FAR OR FARTHER AWAY THAN $z = 1.65$ or $z = -1.65$ S.E. FROM THE NULL

A 90% CONFIDENCE THAT DOESN'T CONTAIN THE PARAMETER MEANS THAT THE SAMPLE RESULT WAS FURTHER THAN ±1.65 S.E. FROM THE TRUE PARAMETER.

THEY ARE VERY SIMILAR IN THEIR CONSTRUCTION (interpretations are different though)

FINAL WILL BE HELD IN HAINES A2 THURSDAY 12/12/02 8am-11am

15. A simple lottery game called "The Daily 3" is played in California. You pay $1 to play and choose a 3-digit-number. The state of California chooses a 3 digit number at random and pays you $500 if your number is chosen. There are 1,000 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

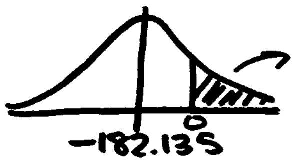
$\boxed{+500}$	$\boxed{-1}$
$\frac{1}{1{,}000}$	$\frac{999}{1000}$

$(500\times\frac{1}{1000})+(-1\times\frac{999}{1000}) = -.499$ u $= -.50$

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

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

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



$\boxed{27.425\%}$

16. Psychologist Amos Tversky, now deceased, studied public perceptions of probability. Here is an example of this work:

(a) Tversky asked randomly selected subjects to choose between 2 public health programs that affect 600 people. Program A has the probability of 50% of saving all 600 and probability 50% that all 600 will die. Program B always guarantee to save exactly 400 out of the 600 people. What is the expected number of people saved by Program A?

A

$\boxed{600}$	$\boxed{0}$
.5	.5

300 will be saved

B
400 saved

(b) Tversky then offered people a different choice. Again, Program A has the probability of 50% of saving all 600 and probability 50% that all 600 will die. Program C will always lose exactly 200 out of the 600 people. What is the difference between this choice and the choice given in part (a)?

$\boxed{600}$	$\boxed{0}$
.50	.5

300 will be saved

C
400 will be saved

the wording is different

(c) When surveyed, the majority of his randomly selected subjects choose Program B when given the choices in (a). When given the choices in (b), most people chose Program A. Do the subjects appear to use expected values in their choices? Answer yes or no and please explain your choice in plain English.

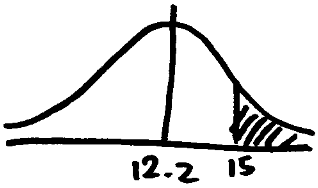
No. People do not always use logic/expected values in their choices. The calculations above suggest that both situations have the same outcomes and yet because of wording differences, people make unusual choices.

Stat 10 Lew FINAL IS 12/12/02 8am-11am

FINAL WILL BE HELD IN HAINES A2 THURSDAY 12/12/02 8am-11am

17. Americans spend an average of 12.2 minutes in the shower. If the standard deviation of the variable is 2.3 minutes and the variable is normally distributed:

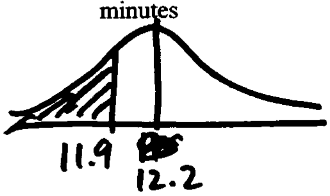
(a) Find the percentage of Americans who spend at least 15 minutes in the shower.



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

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

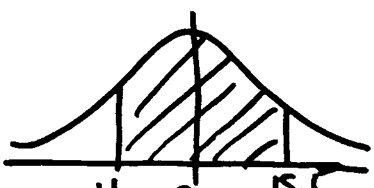
(b) Find the chance that the mean time of a random sample of 144 Americans who shower will be less than 11.9 minutes



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

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

(c) Find the percentage of Americans who spend between 11 minutes and 15.5 minutes in the showe

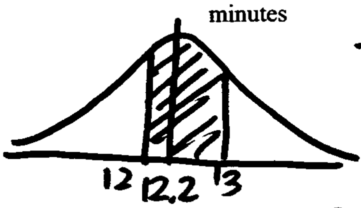


$$z_{15.5} = \frac{15.5-12.2}{2.3} = \frac{3.3}{2.3} = 1.43 \approx 1.45$$

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

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

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



$$z_{13} = \frac{13-12.2}{\left(\frac{\sqrt{100}*2.3}{100}\right)} = \frac{.8}{.23} = 3.48 \approx 3.50$$

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

$$z_{12} = \frac{12-12.2}{\left(\frac{\sqrt{100}*2.3}{100}\right)} = \frac{-.2}{.23} = -.87 \approx -.85$$

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 groups
X		Association may point to causation, but association is not the same as causation
	X	An extremely large biased sample (e.g. > 1000) generates better estimates of the population parameters than a small random sample (e.g. a little over size 100)
X		A quantitative variable can be discrete or continuous