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

1. Suppose you know that the mean height of US males is 69 inches and that the standard deviation is 4 inches. Also suppose that the shape of the histogram for heights of US males is approximately normal. Which of the following follow from this information?
a) At least 75% of US males have heights between 61 and 77 inches.
b) Approximately 68% of US males have heights between 65 and 73 inches.
c) At most 25% of US males have heights above 77 inches.
d) All of the above.
e) None of the above
f) Only a and b
g) Only b and c
h) Only a and c

2. 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

3.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

4.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

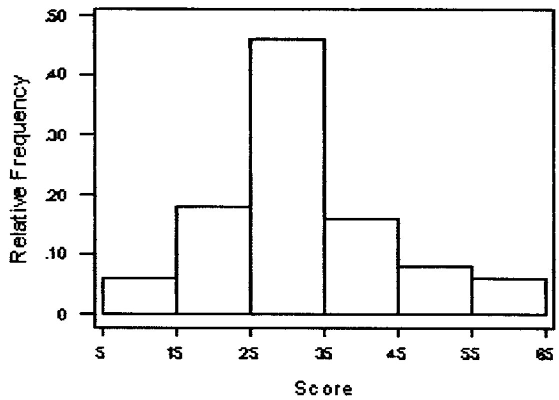
5. ~~At le~~ast 68% of the values in a data set fall within 1 standard deviation of the mean. TRUE or FALSE. must be normal.

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

7. 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)

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

8. 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.



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

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

11. 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.

TYPO

12. Which of the following statements is not appropriate for an 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 never correct to say "I accept the null hypothesis".
D. It is almost always correct to say "I accept the null hypothesis".

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).

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

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% $\Rightarrow -1.65 = \dfrac{X-100}{15}$

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

solve for X in each case so

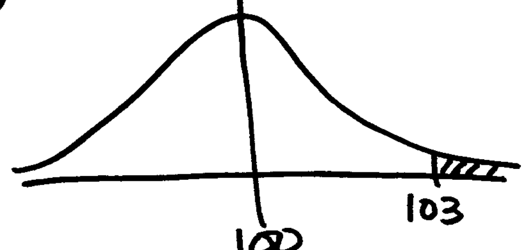
$X = 75.25$ and 119.50

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



p-value is $\dfrac{100-99.73}{2} \cong .14\%$ much less than 5%.

Conclusion: REJECT THE NULL
College students have HIGHER IQs than Average

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

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

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	X·Y
	1987	24.6	5.1	125.46
	1988	28.5	16.8	478.80
	1989	10.6	31.5	333.90
	1990	-23	-3.1	71.30
	1991	12.8	30.4	389.12
	1992	-12.1	7.6	-91.96
	1993	32.9	10.1	332.29
	1994	6.2	1.3	8.06
	1995	11.2	37.6	421.12
	1996	6.4	23	147.20
Average	1991.5000	9.8100	16.0300	221.529 (average of X·Y)
Standard Deviation	2.7386	15.6493	12.6810	

(a) Find the correlation, r, of the U.S. and overseas returns then describe the relationship between U.S. and overseas returns in words, using r to make your description more precise.

$$r = \frac{(221.529) - [(16.03) * (9.81)]}{(12.6810)(15.6493)} = .3239$$

weakly positive. Suggest US returns increase so do Overseas

(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.

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

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

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

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.

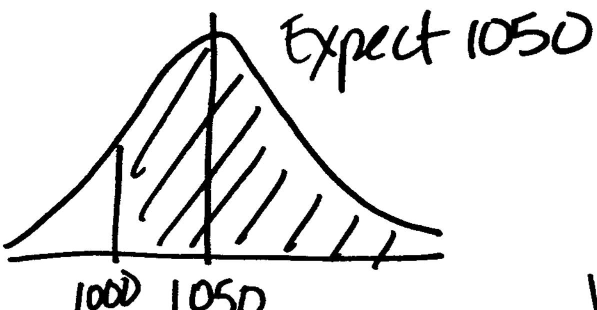
(a) If you randomly dial 1500 telephone numbers, what are the mean and standard error of <u>the number</u> of households who respond?

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

Expected mean = 1,050 b/c 1500×.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

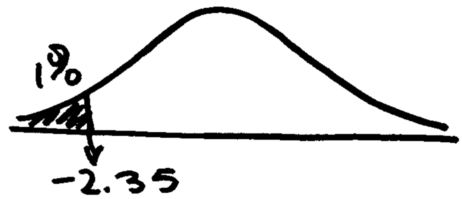
H_0: mean = 0

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.)

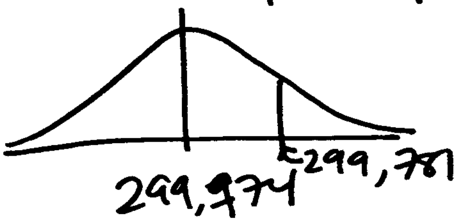
$$299{,}774 \pm 2\left(\frac{\sqrt{2500} \times 14}{2500}\right)$$

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

b. Now the investigators determine the speed of light once more by the same procedure and get 299,781 kilometers per second. Is this a surprising result? Why or why not?

$\rightarrow n = 1$

Not surprising



$$z = \frac{299781 - 299774}{14} = +.5$$

this is w/i 1 SD so it's reasonable

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 probabilities for your outcome:

Number of persons	1	2	3	4	5	6	7
Probability	0.25	0.32	??? .16	??? .16	0.07	0.03	0.01

The probability of finding 3 people in a household is the same as the probability of finding 4 people. These probabilities are marked ??? in the table of the distribution.

Find the probability that a household contains 3 people.

Box fractions MUST sum to 1.0 or 100% so

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

if 3 and 4 are the same then $\frac{.32}{2} = .16$

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

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

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?

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

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

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.

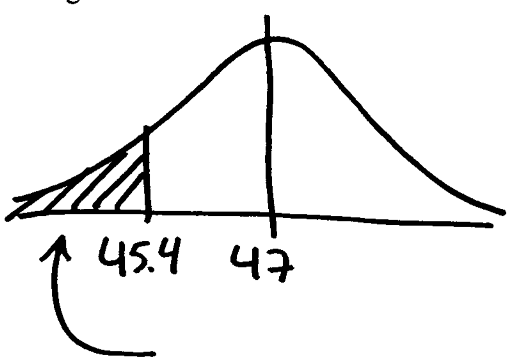
a) The truth about the population is 0.47. In what range will the middle 95% of all sample results fall for samples of size 1,025?

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

$\hookrightarrow 1.5589$

$$47\% \pm 3.1178\%$$

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 $-2 \times \underset{\downarrow\, r}{0.6} = -1.2$

③ $-1.2 * 2.5 = -3$ inches so our estimate is $63 - 3 = 60$ inches

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

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)

p-value is

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

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

No evidence to support the alternative