

Please indicate whether each statement is true or false

	True	False	Statement
1A	<input type="checkbox"/>	<input checked="" type="checkbox"/>	The Central Limit Theorem (CLT) suggests that an unbiased sample will be normally distributed if it is the result of a reasonably large number of draws from a population
1B	<input checked="" type="checkbox"/>	<input type="checkbox"/>	The mean of the sampling distribution for a proportion is expected to have the same value even when the sample size increases.
1C	<input checked="" type="checkbox"/>	<input type="checkbox"/>	The Central Limit Theorem (CLT) will apply in situations where the number of random draws (sample size of a random sample) is reasonably large and the population is normal.
1E	<input type="checkbox"/>	<input checked="" type="checkbox"/>	For a population that is not normally distributed, the distribution of sample percentages will have the same shape as the population when the sample (randomly drawn) is reasonably large.
1G	<input checked="" type="checkbox"/>	<input type="checkbox"/>	The sampling distribution for unbiased random samples of reasonable size will be centered on the expected value of the mean.
1H	<input checked="" type="checkbox"/>	<input type="checkbox"/>	The Central Limit Theorem (CLT) implies that when its conditions are met, a sampling distribution can be correctly summarized by only an expected value and a standard deviation

2. In 2001, a survey organization takes a simple random sample of 1,600 adults in Los Angeles, California, a large American city. Among this sample of adults, it was found that 975 support the death penalty, 525 support life imprisonment with no parole and the rest did not believe in penalties for homicide. It was noted that support for the death penalty had changed from a survey taken in 1991 when approximately 80% of adults in Los Angeles supported the death penalty.

- a. Is it possible to construct a 95% confidence interval for the population percentage of Los Angeles adults who support the death penalty in 2001. (circle one)

YES

NO

If you circled YES, please construct a 95% confidence interval in the space below. If you circled NO, please use the space to explain why it is not possible to construct a 95% confidence interval.

$$\frac{975}{1600} = .6094 \quad 60.94\% \pm 1.96 \sqrt{\frac{(.6094)(.3906)}{1600}}$$

- b. If the sample size were 400 instead of 1600 it would (circle one to fill in the blank) the width of any confidence interval constructed from the sample information

Increase

Decrease

Not Affect

- c. If the level of confidence were 99% instead of 95% it would (circle one to fill in the blank) the width of any confidence interval from the sample information

Increase

Decrease

Not Affect

- d. Suppose it was known that actually 60% of all adults in Los Angeles support the death penalty. So if a simple random sample of 625 adults were to be taken, the SE for the sample percentage of death penalty supporters is calculated to be about 2%. You should assume these numbers are correct.

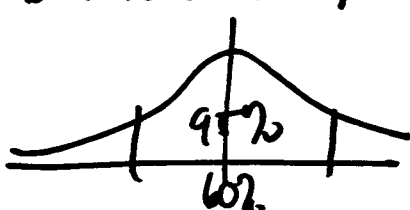
A student, looking at the numbers in part d, interprets them as follows: this means that there is about a 95% chance for the percentage of death penalty supporters in the sample to be in the range $60\% \pm 4\%$. (circle one)

The student is correct

The student is not correct

Please explain your choice below:

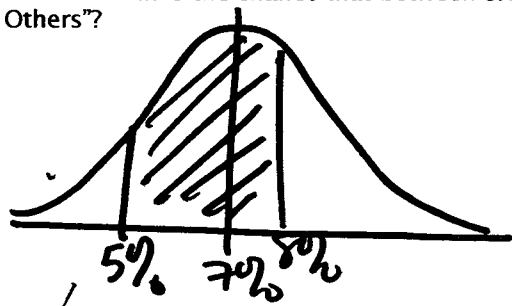
625 is a large enough sample and by the CLT 95% percent of all samples should fall w/i 2 SE of the pop. parameter



$$100 - 71 - 10 - 12 = 7$$

3. The most recent enrollment statistics for the entire Los Angeles Unified School District revealed that 71% were identified as "Hispanic", 10% as "Non-Hispanic White", 12% as "Black" or "African American" and the remainder were called "All Others". A recent survey of excellent quality conducted by UCLA on 225 students in the Los Angeles Unified School District revealed that 6% of the enrolled students were "Asian"

A. What is the chance that between 5% and 8% of students in a sample of 225 will be identified as "All Others"?



$$z_{.08} = \frac{.08 - .07}{\sqrt{\frac{(.07)(.93)}{225}}} = .59 \quad \text{area is } .7224$$

$$z_{.05} = \frac{.05 - .07}{\sqrt{\frac{(.07)(.93)}{225}}} = -1.18 \quad \text{area is } .1190$$

$$.7224 - .1190 = .6034$$

B. If the sample size were 400 instead of 225 it would (circle one to fill in the blank) the standard error for the sample count of "All Others" students

☒ Increase

☐ Decrease

☐ Not be enough information to calculate

C. Can you construct a 90% confidence interval for the percentage of "Asian" students in the Los Angeles Unified School District using information from the original sample of 225? (circle one)

☒ YES

☐ NO

If yes, please construct it in the space below, if no, please explain why this is not possible.

$$6\% \pm 1.65 \left(\sqrt{\frac{(.06)(.94)}{225}} \right)$$

(or 1.64)

$$6\% \pm 2.61 \text{ or } [3.39\%, 8.61\%]$$

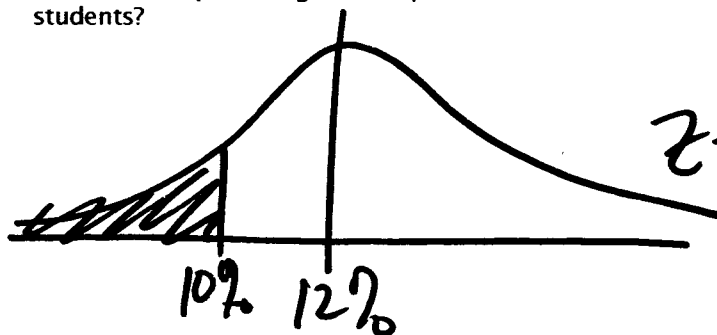
D. (continued from part C) Suppose it is possible to calculate a confidence interval. If the level of confidence were changed to 80% instead of 90% it would (circle one to fill in the blank) the width of any confidence interval from the sample information

☒ Increase

☐ Decrease

☐ Not Affect

E. What percentage of samples of size 225 will have fewer than 10% "Black or African American" students?



$$z = \frac{.10 - .12}{\sqrt{\frac{(.12)(.88)}{225}}} = -.92$$

$$\text{about } .1788 \sim [17.88\%]$$

4. Some friends take you to a casino and you are confronted with two games.

GAME A works like this: you can bet \$8 on a number, and if your number comes up, you win \$11, if not, you lose your \$8. Your number comes up 35% of the time.

GAME B works like this: you can bet \$3 on a number, and if your number comes up, you win \$2, if not, you lose your \$3. Your number comes up 55% of the time.

A) Please calculate the Expected Value and the Standard Deviation for each of these games

A

+11	-8
.35	.65

Expected Value
 $(+11)(.35) + (-8)(.65)$
 $= -1.35$

SD

$$\sqrt{(.35)(11 - (-1.35))^2 + (.65)(-8 - (-1.35))^2}$$

$$= 9.0624$$

B

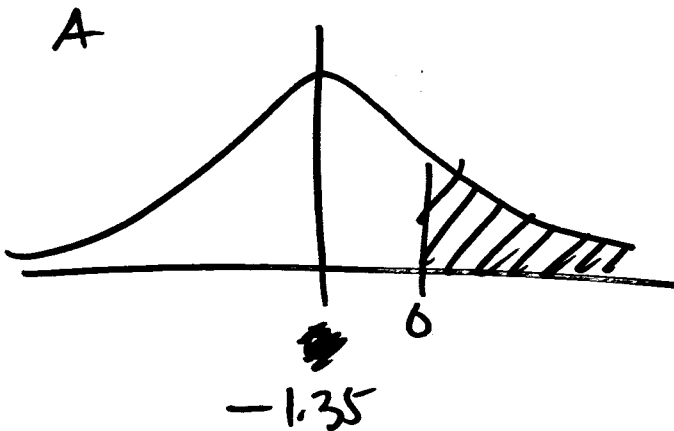
+2	-3
.55	.45

$(2)(.55) + (-3)(.45)$
 $= -.25$

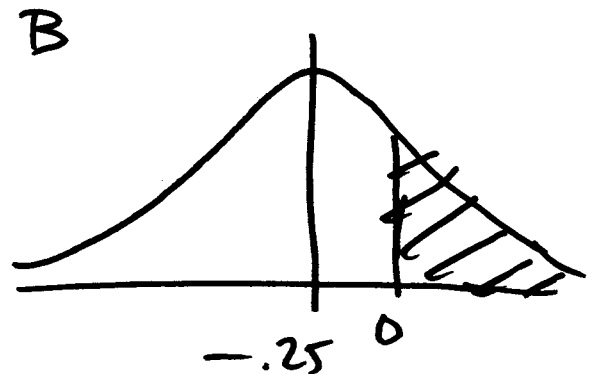
$$\sqrt{(.55)(2 - (-.25))^2 + (.45)(-3 - (-.25))^2}$$

$$= 2.4875$$

B) Based on your calculations in part A, which game is more likely to give you winnings (as opposed to losing)? Justify your selection in the space below.



$$z = \frac{0 - (-1.35)}{9.0624} = .15$$



$$z = \frac{0 - (-.25)}{2.4875} = .10$$

GAME B, the area > 0 is larger

5. Indicate whether each statement is true or false concerning the following 68% confidence interval that a student generated from a sample of size 144. Suppose the population was not normal and suppose the confidence interval was correctly calculated:

$$21\% \pm 1 * \left(\frac{\sqrt{144} * \sqrt{.21 * .79}}{144} * 100 \right) = 21\% \pm 3.39\%$$

WRONG
FORMULA

	True	False	
6A			Assume for this statement, the confidence interval was correctly calculated, then in 68% of all samples, the true population percentage would lie within Sample percentage \pm Standard Error for a Percentage.
6B			The value $\sqrt{.21 * .79}$ is also called the standard deviation for a percentage
6C			Suppose the sample size were increased to 169, the width of the confidence interval would also increase
6D			Suppose the confidence level were increased from 68% to 90%, the width of the confidence interval would also increase

7. So you went to college, but your best friend in high school decided to skip college and work for a fast food chain called "Down and Out Burgers". After 2 years, you are accumulating a lot of bills but your best friend has now been promoted to manager and is earning \$90,000 per year. You however know a lot about statistics and decide to pay your friend a visit, hoping to get a free burger perhaps. Your friend tells you that managing a restaurant is not an easy job. His "Down and Out Burgers" was designed to accommodate 6,000 customers per day. If the restaurant has 6,000 customers it will earn \$11,000 on that day. If instead there are 2,000 or fewer customers, the restaurant will lose \$7,000 on that day. If there are 10,000 or more customers, it will lose \$5,000 on that day. The history of the restaurant reveals that on 45% of the days the restaurant has 6,000 customers and on 35% of the days it has 2,000 or fewer customers.

- e. What is the expected value of the earnings for this particular "Down and Out"?

outcome	11,000	-7000	-5000
probab	.45	.35	.20

$$E(x) = \$1,500$$

- b. What is the standard deviation of those earnings?

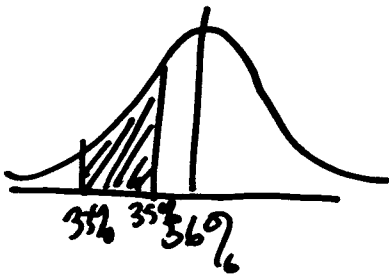
$$\sqrt{.45(11,000 - 1,500)^2 + .35(-7000 - 1,500)^2 + (.20)(-5000 - 1,500)^2}$$

$$= \$8,623$$

8. The Super Bowl is the number one party event of the year for Americans, exceeding even New Year's Eve celebrations. Suppose it is known that the typical party has an average of 17 partygoers with a standard deviation of 3.3. On an ordinary Sunday afternoon, the average number of calories consumed in America is 600 but 36% will consume more than 2,000 calories. And only 5% will get drunk. Please assume that calories are normally distributed

The Harvard School of Public Health decided to study the effects of attending Super Bowl Sunday parties on the caloric consumption of Americans. 850 Americans were selected by random-digit dialing and interviewed by telephone. 490 Americans reported that they had attended a Super Bowl party, 110 did not attend a party but watched the Super Bowl on television at home. The remainder did not attend a Super Bowl party or watch the game. The calories consumed by the partygoers had a mean 1,330 with a standard deviation of 600. The calories consumed by the non-party goers had a mean of 560 with a standard deviation of 100. Among the party goers, 77% reported getting "drunk", among non-party goers who watched the game 15% reported getting "drunk" and only 7% of the non-party goers/non Bowl watchers reported getting "drunk" on Super Bowl Sunday. The average party had 19 partygoers. Please assume that all of the sample sizes are reasonably large and no biases exist.

A. What is the chance that a sample of size 850 will have between 33% and 35% of those surveyed consuming more than 2,000 calories? (5 points)



$$z = \frac{.35 - .36}{\sqrt{\frac{(.36)(.64)}{850}}} = .61 \quad \text{area is } .2709$$

$$z = \frac{.33 - .36}{\sqrt{\frac{(.36)(.64)}{850}}} = -1.82 \quad \text{area is } .0344$$

→ chance

$$\boxed{.2709} - .0344 = .2365$$

B. Please construct an approximate 75% confidence interval for the population percentage of partygoers who reported getting "drunk". (4 points)

$$77\% \pm 1.15 \left(\sqrt{\frac{(.77)(.23)}{490}} \right)$$

$$77\% \pm 2.19\%$$

$$\approx (74.81\%, 79.19\%)$$

(continued from above)

C. According to statistical theory, do we meet the conditions necessary to allow us to construct an 80% confidence interval for the population percentage of non-party going/non-Super Bowl watching Americans who got drunk on Super Bowl Sunday? (circle one)

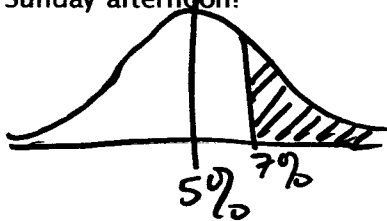
YES

NO

If yes, please construct it in the space below, if no, please explain why this is not allowable.

$$7\% \pm 1.28 \sqrt{\frac{(.07)(.93)}{250}} \Rightarrow 7\% \pm 2.07\%$$

D. If 5% of all Americans are drunk on a typical Sunday afternoon, what is the chance that a simple random sample of 256 Americans will show that at least 7% are drunk on a typical Sunday afternoon?



$$z = \frac{.07 - .05}{\sqrt{\frac{(.05)(.95)}{256}}} = \cancel{1.47} + 1.47$$

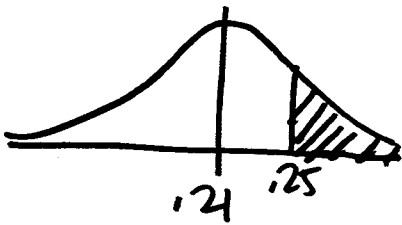
area is .0708 or 7.08%

9. Indicate whether the statement is true or false

	True	False	
A		X	In hypothesis testing, an $\alpha = .10$ means that there is a 10% chance that the null hypothesis is wrong
B	X		In hypothesis testing, a null rejected at $\alpha = .05$ will always be rejected at $\alpha = .10$
C	X		90% confidence means that there is a 90% chance that the sample you select will generate an interval that contains the true parameter value.
D	X		The proper use of the Z statistic requires that the population be normally distributed
E		X	For confidence intervals, decreasing the sample size will have the same effect as decreasing the confidence level.

10. A marketing company wishes to determine the extent to which people who currently own personal computers would be willing to "upgrade" to handheld "communication devices" such as the Palm Pilot (manufactured by 3Com). Because of logistical considerations, the survey focused on households that purchased new personal computers within the last year. A list of all such households was obtained from manufacturers through warranty registration records, and from this list the marketing company determined that amount paid for a computer was normally distributed with an average of \$3,100 and a standard deviation of \$1355 and 21% of the households reported that they planned to purchase a handheld "communication device" in the next five years. From this list, the company selected a simple random sample (SRS) of size 64 and conducted in-person household interviews. Analysis of the sample revealed that the 64 households paid an average of \$2,850 for their personal computer (with a standard deviation of \$1200) and 25% of the households said they will purchase a handheld "communication device" sometime in the next five years.

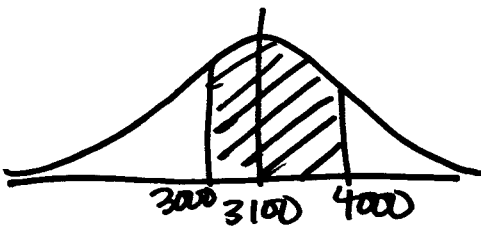
- a. What is the chance of getting a sample proportion of 25% or more from a random sample of size 64 from the population of households?



$$z = \frac{.25 - .21}{\sqrt{\frac{(.21)(.79)}{64}}} = .79$$

$$\sim 21.5\%$$

- b. What percentage of all computers ^{cost} between \$3,000 and \$4,000?



$$z_{4000} = \frac{4000 - 3100}{1355} = .66$$

$$z_{3000} = \frac{3000 - 3100}{1355} = -.07 \quad [27.33\%]$$

$$.7454 - .4721 = .2733$$

- c. Please test the hypothesis that the percentage of households who plan to purchase a handheld "communication device" has increased.

$$Use \alpha = .05$$

$$H_0: p = .21$$

$$H_1: p > .21$$

$$z = \frac{.25 - .21}{\sqrt{\frac{(.21)(.79)}{64}}} = .79$$

$$p\text{-value is } .2148$$

$$.2148 > .05$$

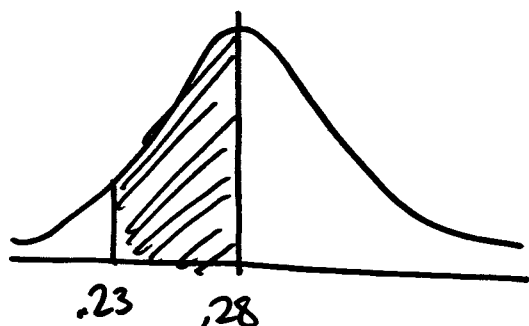
DO NOT REJECT THE NULL. DIFFERENCE IS NOT STATISTICALLY SIG.
No evidence of an increase in plans to purchase

11. The Dull Computer Company manufactures 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 of Dull has stated publicly that if customers make unassisted online purchases of their computers, they will have a mean delivery time of 36 hours from time of purchase with a standard deviation of 11 hours and a mean cost of \$1,603 with a standard deviation of \$400. He also went on to state that 28% of those computers are delivered in less than 24 hours. Assume that both delivery time and cost are normally distributed variables.

A consumer research organization decided to test the CEO's claim by purchasing 100 computers from Dull at randomly selected times and days. The 100 purchases were randomly divided into two groups: 51 were purchased by telephone and involved talking to a live salesperson, the remaining 49 were unassisted online purchases. The delivery time of the 49 had a mean of 35 hours with a standard deviation of 16 hours and they also had a mean cost of \$1,588 with a standard deviation of \$676. 11 of the 49 computers were delivered in less than 24 hours.

Assume 49 is large enough.

(a) What is the chance of selecting a sample of size 49 that has between 23% and 28% of the computers delivered in less than 24 hours?

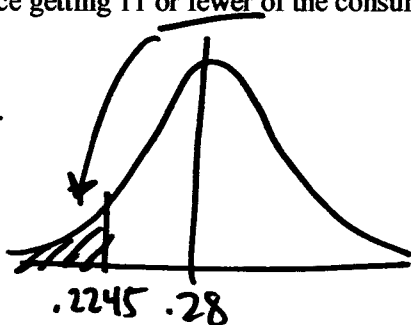


$$z = \frac{.23 - .28}{\sqrt{\frac{(.28)(.72)}{49}}} = -.78$$

area is $(.50 - .2177) = .2823$
or 28.23%

(b) What is the chance getting 11 or fewer of the consumer research group's purchases delivered in less than 24 hours?

$$\frac{11}{49} = .2245$$



$$z = \frac{.2245 - .28}{\sqrt{\frac{(.28)(.72)}{49}}} = -.87$$

area is $.1922$ or 19.22%

(c) Please test the hypothesis that fewer than 28% of computers are being delivered in less than 24 hours. Use $\alpha = .05$

$$H_0: p = .28$$

$$H_1: p < .28$$

$$z = \frac{.2245 - .28}{\sqrt{\frac{(.28)(.72)}{49}}} = -.87$$

p-value is .1922

.1922 > .05
do not reject null
evidence suggests
that 28% are being
delivered in < 24 hr

12. Congratulations, you have finished college. The bad news is this: you never learned a thing about the stock market, but find yourself applying for a job as a broker on Wall Street. You discover that the field has been narrowed down to you and one other applicant, "The Oracle". Both of you were told to select 36 different stocks for a portfolio (a collection of stocks) and the winner (the one with the higher return) gets the job. Since you know nothing about stocks, you selected 49 at random. Here are the results:

Portfolio	Number of Stocks Selected	Proportion that Made Money
Yours	36	53%
Oracle's	36	55%

You can assume that a sample of 36 is large enough and that Oracle chose his randomly too.

The director says to you "I like you better because you are a nice person, but unfortunately, you did not do as well as the Oracle". You respond that you can demonstrate that your return is, statistically speaking, equal to and not less than Oracle's return. Surprised, the director says, "Can you? Well, if you can prove it, I'll hire you instead!"

A. Please state a null and alternative hypothesis consistent with the claim.

$$H_0: p = .55$$

$$H_1: p < .55$$

B. Please provide the appropriate statistical test to test this claim. Give a numerical result

$$Z = \frac{.53 - .55}{\sqrt{\frac{(.55)(.45)}{36}}} = -.2412$$

C. Please state the p-value for this test

.4052

D. Use an $\alpha = .05$ as your decision rule. On the basis of the test results did you get the job or not?

(circle one)

YES

NO

Please briefly explain your choice.

p-value > .05 so we did not reject the null. In other words, ~~to~~ your proportion is not different, statistically speaking, than Oracle's. ~~At least random sample~~
~~was given~~

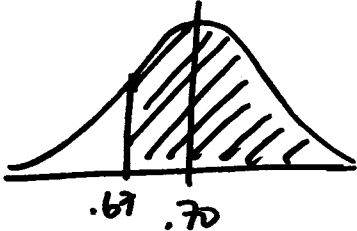
13. You got a job working for a marketing company and your supervisor is planning a sample survey of households in Los Angeles. Your supervisor instructs you to contact households by random-digit dialing phone numbers. Your supervisor knows from past experience that about 70% of the households you contact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

- (a) If you randomly sample and then dial 1500 telephone numbers, based on your supervisor's knowledge of household response, what are the mean and standard error of the percentage of households who respond?

$$\text{mean} = 70\%$$

$$SE = \sqrt{\frac{(.70)(.30)}{1500}} = .0118$$

- (b) In a random sample of size 1500 households, based on your supervisor's knowledge of household response, please find the chance that you will get at least 1000 responses.



$$\frac{1000}{1500} = .67 \text{ or } 67\% \text{ we are expecting } 70\%$$

$$\frac{.67 - .70}{\sqrt{\frac{(.70)(.30)}{1500}}} = -2.54 \quad 1 - .0053 = \boxed{99.45\%}$$

- (c) Suppose you dialed 1500 phone numbers at random and got 1001 responses, please test the hypothesis that the your supervisor's percentage of 70% has decreased. USE $\alpha = .05$

$$H_0: p = .70$$

$$H_1: p < .70$$

$$Z = \frac{.6673 - .70}{\sqrt{\frac{(.70)(.30)}{1500}}} = -2.76$$

reject the null $.0029 < .05$
 Stat. significant
 Looks like the %
 is less than 70%.

p-value is .0029

- (d) Assume your supervisor is stupid and does not know the business. Suppose you dialed 1500 phone numbers at random and got 1001 responses. Please construct an 80% confidence interval for the percentage of households who respond.

$$\frac{1001}{1500} = .6673 \text{ or } 66.73\%$$

$$66.73\% \pm 1.28 \left(\sqrt{\frac{(.6673)(.3327)}{1500}} \right)$$

$$66.73\% \pm 1.56\%$$

$$(65.17\%, 68.29\%)$$

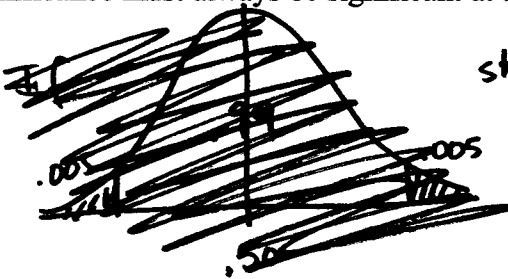
14. You are planning to perform a significance test of

H_0 : proportion = .50

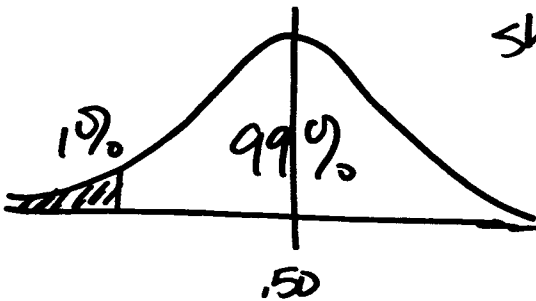
Versus

H_1 : proportion < .50

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.



~~shaded areas are rejection re~~



shaded area is rejection region

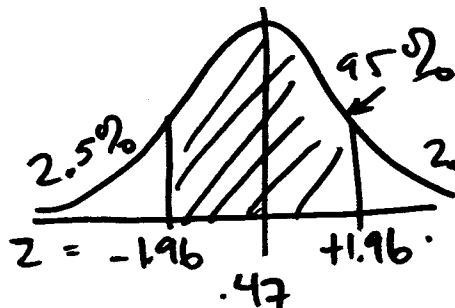
if $z = -2.38$ this gives 1% in the "tail"

TRUE if a test is significant at the 5% level ($z = 1.96$ or $z = -1.96$) then any z greater than it falls in the rejection region

15. 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.

SUPPOSE

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?



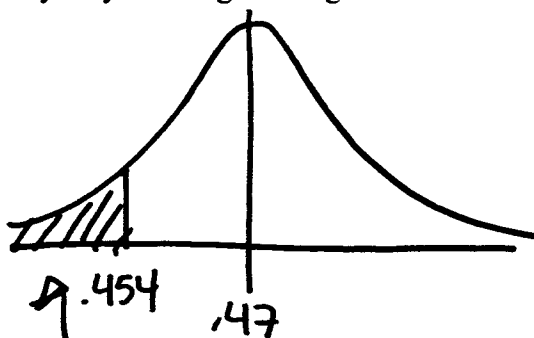
tails are 2.5% each

$$.47 \pm (1.96) \left(\sqrt{\frac{(.47)(.53)}{1025}} \right)$$

$$.47 \pm .0306 \text{ or}$$

$$.4394 \text{ to } .5006$$

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?



$$z = \frac{.454 - .47}{\sqrt{\frac{(.47)(.53)}{1025}}} = -1.0263 \approx -1.03$$

area is .1515 or
probability is .1515

c) Please construct a 99% confidence interval for the percentage of women who do not get enough time for themselves.

$$.47 \pm 2.58 \left(\sqrt{\frac{(.47)(.53)}{1025}} \right) \Rightarrow .47 \pm .0402$$

$$\left(\text{or } 2.57 \right. \\ \left. \text{or } 2.575 \right)$$

$$(42.98\%, 51.02\%)$$

16. 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% of X-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?

$$H_0: p = .45$$

$$H_1: p < .45$$

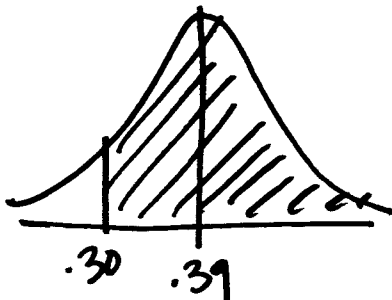
$$Z = \frac{.39 - .45}{\sqrt{\frac{(.45)(.55)}{121}}} = -1.33$$

p-value is .092 or 9.2%
since 9.2% > 5%

DO NOT REJECT THE NULL. THIS IS NOT STATISTICALLY SIGNIFICANT
Gen X-ers do not appear to be different in their beliefs from americans in general

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 64 will have at least 30% of the surveyed Generation Y-ers believing in Social Security?



$$Z = \frac{.30 - .39}{\sqrt{\frac{(.39)(.61)}{64}}} = -1.48$$

AREA TO THE RIGHT OF -1.48

$$1 - .0694 = .9306 \approx 93.1\%$$