1. The next four questions use information from this statement, but each question is separate (i.e. you can get the first one wrong and it won't affect the others): The Medical College Admissions Test (MCAT) is constructed to be normally distributed with a mean of 9 and a standard deviation of 3. SHOW YOUR WORK FOR FULL CREDIT.

A. A UCLA Study selected a simple random sample (SRS) of 49 MCAT test takers. Please assume the sample is of good quality. What is the chance that the sample will have an average that is greater than 10.1?

$$Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{10.1-9}{\frac{3}{\sqrt{49}}} = \frac{1.1}{.4286} = 2.57$$ the area under the curve greater than 2.57 is .0051 this is the chance.

B. A poorly constructed sample might yield a sample average that is in the lowest 5% of all possible sample averages. Suppose UCLA plans to select a simple random sample (SRS) of 81 MCAT test takers. At or below what score is the lowest 5% of sample averages for samples of size 81?

The Z score associated with the lowest 5% is a –1.65 or a –1.64 or a –1.645
We need to find a sample average from this information (I'll use -1.65 here)

$$Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \Rightarrow -1.65 = \frac{\bar{x}-9}{\frac{3}{\sqrt{81}}} \Rightarrow \bar{x} = \left[-1.65*\left(\frac{3}{\sqrt{81}}\right)\right]) + 9$$ solving for x-bar gives 8.45

C. UCLA selected a simple random sample of size 100 MCAT test takers. The sample revealed that 15% of the test takers were Hispanic. Please construct a 90% confidence interval for the population percentage of Hispanic test takers.

$$\hat{p} \pm Z^{*}\left(\sqrt{\frac{(\hat{p})(1-\hat{p})}{n}}\right) \Rightarrow .15 \pm 1.645\left(\sqrt{\frac{.15*.85}{100}}\right) \Rightarrow .15 \pm .059$$

D. UCLA would like to estimate the average age of an MCAT test taker with 95% confidence with a margin of error of no more than ±3 years. What is the smallest sample size necessary to accomplish this task?

TYPO TYPO TYPO! To do this one, I needed to give you a value for sigma (σ), the population standard deviation. Let us suppose that sigma is equal to 10 years. Then the solution is

$$n = \left(\frac{Z*\sigma}{MoE}\right)^2 = \left(\frac{1.96*10}{3}\right)^2 = 42.68 \approx 43$$

2. 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 as large as or larger than their year-end 2001 bonuses?

$$Z = \frac{x-\mu}{\sigma} = \frac{0--9}{16} = \frac{9}{16} = .56$$ The area to the right of a Z=.56 is .2877

b) A simple random sample of 100 lawyers has an average year-end bonus at the 25th percentile, what is the actual value of that average?

The Z score associated with the lowest 25% is a –.67
We need to find a sample average from this information

$$Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \Rightarrow -.67 = \frac{\bar{x}--9}{\frac{16}{\sqrt{100}}} \Rightarrow \bar{x} = \left[-.67*\left(\frac{16}{\sqrt{100}}\right)\right]) - 9$$ solving for x-bar gives –10.07

c) A simple random sample of 100 lawyers has an average year-end bonus of –7.5%, what is the chance of getting a sample average of –7.5% or higher?

$$Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{-7.5--9}{\frac{16}{\sqrt{100}}} = \frac{1.5}{1.6} = .94$$ the area under the curve greater than .97 is .1660 this is the chance.

d) What is the chance that a lawyer, selected at random, will have a year-end bonus of -7.5% or higher?

$$Z = \frac{x-\mu}{\sigma} = \frac{-7.5--9}{16} = \frac{1.5}{16} = .09$$ The area to the right of a Z=.09 is .4641 this is the chance

e) What percentage of lawyers have year end bonuses between –1% and +5%?

Need 2 Z calculations here:

$$Z = \frac{x-\mu}{\sigma} = \frac{-1--9}{16} = \frac{8}{16} = 0.50$$ area to the left is .6915

$$Z = \frac{x-\mu}{\sigma} = \frac{5--9}{16} = \frac{14}{16} = 0.88$$ area to the left is .8106

The percentage is .8106-.6915 =.1191 that is 11.91%

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

(a) A simple random sample of size 256 is drawn from the adult human population. What is the chance that the sample average will exceed 101?

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{101 - 100}{\frac{15}{\sqrt{256}}} = \frac{1}{\frac{15}{16}} = 1.07$$ the area under the curve greater than 1.07 is .1423 this is the chance.

(b) How large of a sample would a researcher need to select to insure that he or she is within plus or minus 1 IQ point of the population mean IQ with 99% confidence?

$$n = \left(\frac{Z * \sigma}{MoE}\right)^2 = \left(\frac{2.576 * 15}{1}\right)^2 = 1493.0496 \approx 1494$$

(c) A simple random sample of 256 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 a research 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.

H0: μ=100 (college student IQs are the same as the average human)
H1: μ > 100 (college student IQs are higher than the average human's)

Test: $$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{103 - 100}{\frac{15}{\sqrt{256}}} = \frac{3}{\frac{15}{16}} = 3.20$$

p-value: the area to the right (greater than) Z=3.20 is .0007

Conclusion: We reject the null in favor of the alternative. This is statistically significant because .0007 is less than .05. This means it appears that college students have higher IQs than the average human.

4. Los Angeles International Airport handles an average of 6,000 international passengers an hour. Suppose 80% can pass through primary security, but the rest are detained for interrogation by the FBI. And suppose the FBI can handle 1,500 passengers an hour without unreasonable delays for travelers and extra costs to the airlines (due to missed flights and connections).

a. Suppose the FBI decides to randomly sample passengers in order to speed up the screening process. What is the chance that a simple random sample of 100 will have between 21 and 24 passengers detained for interrogation by the FBI?

Need two Z scores for this one

$$Z = \frac{n\hat{p} - np}{\sqrt{np(1-p)}} = \frac{21-20}{\sqrt{100(.2)(.8)}} = .25$$

$$Z = \frac{n\hat{p} - np}{\sqrt{np(1-p)}} = \frac{24-20}{\sqrt{100(.2)(.8)}} = 1$$

The area between a Z=1 and a Z=.25 is .8413-.5987 = .2426 and this is the chance

We choose np=20 because 80% pass through but 20% don’t and since n=100, np=20

b. Certain ethnic/racial groups appear to be detained at much higher rates than others. Suppose a human rights organization sends 64 persons who appear to be of middle eastern origin through the airport and 21 are detained for interrogation. Please test the hypothesis that persons of middle eastern origin are detained in higher proportions than the typical traveler. 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. You may treat the 64 as if it were a simple random sample and it is of reasonable size.

21/64= .328

H0: p=.20 (middle eastern origin are detained at the same rate as others)
H1: p > .20 (middle eastern origin are detained at a higher rate than others)

Test: $$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{.328 - .20}{\sqrt{\frac{.20 * .80}{64}}} = \frac{.128}{.05} = 2.56$$

p-value: the area to the right (greater than) Z=2.56 is .0052

Conclusion: We reject the null in favor of the alternative. This is statistically significant because .0052 is less than .05. This means it appears that middle eastern origin persons are detained at a higher rate than others.

5. Here is a correlation table:

	insurance	unemploy	poverty	income	incarceratio	violence
insurance	1.0000					
unemploy	0.4872	1.0000				
poverty	0.1106	0.5222	1.0000			
income	0.4115	-0.1345	-0.7099	1.0000		
incarceratio	0.4473	0.4389	0.4756	-0.1276	1.0000	
violence	0.5284	0.4381	0.3883	-0.0014	0.8144	1.0000

Please identify the strongest positive correlation, the strongest negative correlation and the weakest correlation in the table.

Strongest positive is .8144
Strongest negative is -.7099
Weakest in table is -.0014

6. Here is a correlation table: **THIS IS A HUGE TYPO**

	insurance	unemploy	poverty	income	incarceratio	violence
insurance	1.0000					
unemploy	0.4872	1.0000				
poverty	0.1106	0.5222	1.0000			
income	0.4115	-0.1345	-0.7099	1.0000		
incarceratio	0.4473	0.4389	0.4756	-0.1276	1.0000	
violence	0.5284	0.4381	0.3883	-0.0014	0.8144	1.0000

Please identify the strongest positive correlation, the strongest negative correlation and the weakest correlation in the table.

Try the same question on a different table

	mpg	fuelcap	length	weight	luggage	rearsh~r	doortop
mpg	1.0000						
fuelcap	-0.7798	1.0000					
length	-0.3524	0.4724	1.0000				
weight	-0.8743	0.8831	0.5096	1.0000			
luggage	-0.6450	0.7484	0.2765	0.7002	1.0000		
rearshoulder	-0.3574	0.2746	0.2958	0.3998	0.2099	1.0000	
doortop	-0.7641	0.7389	0.0218	0.7586	0.8656	0.2555	1.0000

e-mail me if you want the answer.

7. Below is a cross-tabulation or the result of a tabulate command for two variables: INCOME and CORRECTIONS.

```
tabulate income corrections, row column
           |      corrections
    income |      LOW         HIGH |     Total
-----------+----------------------+----------
     LOW   |         6           7 |        13
           |     46.15       53.85 |    100.00
           |     20.69       30.43 |     25.00
-----------+----------------------+----------
     MED   |        14          12 |        26
           |     53.85       46.15 |    100.00
           |     48.28       52.17 |     50.00
-----------+----------------------+----------
     HIGH  |         9           4 |        13
           |     69.23       30.77 |    100.00
           |     31.03       17.39 |     25.00
-----------+----------------------+----------
     Total |        29          23 |        52
           |     55.77       44.23 |    100.00
           |    100.00      100.00 |    100.00
```

(a) Please identify the row marginals

13, 26, 13 or 25%, 50%, 25%

(b) Please identify the column marginals

29, 23 or 55.77 44.23

(c) If you were sitting in front of the computer and were asked to modify the command “tabulate income corrections, row column” in order to have Stata compute the chi-sq statistic, please show the modified command in the space below:

tabulate income corrections, row column chi

(d) If you were told that the p-value for the Chi-Square statistics were less than .05, how would interpret the results (i.e. would you reject or not reject the null hypothesis and what does it mean, plainly speaking, to reject the null for a Chi-Square test of a two-way table?

Reject the null
The two variables are associated, they are related in some way.

8. Here is a cross-tabulation or results of a tabulate command on ceramic pieces found at a Mayan archaeological site. DIAM is the size of the piece. POLISH refers to the level of shine still left on the piece after a millennia where 1=Low, 2=Medium-Low, 3=Medium-High and 4=High.

```
tabulate diam polish, row column
           |                   POLISH
      DIAM |         1          2          3          4 |     Total
-----------+--------------------------------------------+----------
     SMALL |       533        252         13        204 |     1,002
           |     53.19      25.15       1.30      20.36 |    100.00
           |     58.25      78.26      15.66      39.53 |     54.58
-----------+--------------------------------------------+----------
     LARGE |       382         70         70        312 |       834
           |     45.80       8.39       8.39      37.41 |    100.00
           |     41.75      21.74      84.34      60.47 |     45.42
-----------+--------------------------------------------+----------
     Total |       915        322         83        516 |     1,836
           |     49.84      17.54       4.52      28.10 |    100.00
           |    100.00     100.00     100.00     100.00 |    100.00
```

(a) Please identify the row marginals

1002 and 834 OR 54.58 and 45.42

(b) Please identify the column marginals

915,322,83,516 OR 49.84,17.54,4.52,28.10

(c) If you were sitting in front of the computer and were asked to modify the command “tabulate diam polish, row column” in order to have Stata compute the chi-sq statistic, please show the modified command in the space below:

tabulate diam polish, row column chi

(d) If you were told that the p-value for the Chi-Square statistics were less than .05, how would interpret the results (i.e. would you reject or not reject the null hypothesis and what does it mean, plainly speaking, to reject the null for a Chi-Square test of a two-way table?

Reject the null
The two variables are associated, they are related in some way.

9. The UCLA housing office wants to estimate the mean monthly rent for one bedroom apartments in the 90024 zip code. A random sample of size 36 is selected from the zip code area and the sample mean is found to be $1,100. The sample standard deviation is $200. **TYPO: I NEED TO GIVE YOU THE POPULATION STANDARD DEVIATION SIGMA TO DO THIS PROBLEM. PRETEND IT IS $200.**

a. Please construct a 95% confidence interval for the mean monthly rent of all one bedroom apartments in the zip code.

$$\bar{x} \pm Z^*\left(\frac{\sigma}{\sqrt{n}}\right) \Rightarrow 1{,}100 \pm 1.96\left(\frac{200}{\sqrt{36}}\right) \Rightarrow 1{,}100 \pm 65.33$$

b. Please construct a 99% confidence interval for the mean monthly rent of all one bedroom apartments in the zip code.

$$\bar{x} \pm Z^*\left(\frac{\sigma}{\sqrt{n}}\right) \Rightarrow 1{,}100 \pm 2.576\left(\frac{200}{\sqrt{36}}\right) \Rightarrow 1{,}100 \pm 85.87$$

c. What is the minimum sample size needed to generate a 95% confidence interval which has a margin of error of ± $50?

$$n = \left(\frac{Z * \sigma}{MoE}\right)^2 = \left(\frac{1.96 * 200}{50}\right)^2 = 61.47 \approx 62$$

10. The Los Angeles Department of Transportation fines Taxicab companies that take too long to respond to customers. Each year, the Department of Transportation tests each taxicab company by making 100 independent calls to that company (you can treat this as a random sample of size 100) and uses this sample to check the proportion (or percentage) of calls that are answered in a timely manner where within 15 minutes is defined as timely.

Suppose that a taxi company's true proportion of calls answered in a timely manner is .81 (or 81%). That is, in the course of thousands and thousands of calls, the taxicab company would be able to successfully answer 81% of the calls within fifteen minutes.

(a) What is the chance that a random sample of 100 calls made to this taxi company will only show 77% **(TYPO: OR FEWER)** were answered in a timely manner?

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{.77 - .81}{\sqrt{\frac{.19 * .81}{100}}} = \frac{-.04}{.0392} = -1.02 \text{ the chance is .1539}$$

(b) The Los Angeles Department of Transportation doesn't know that this company has a true proportion of calls answered in a time manner of .81 or 81%. Instead, it makes 100 independent calls to the company (treat it as a random sample) and finds that only 75% were answered in a timely manner. Please construct a 98% confidence interval for the proportion of calls answered in a timely manner.

$$\hat{p} \pm Z^*\left(\sqrt{\frac{(\hat{p})(1-\hat{p})}{n}}\right) \Rightarrow .75 \pm 2.326\left(\sqrt{\frac{.75 * .25}{100}}\right) \Rightarrow .75 \pm .1007$$

11. A bank wonders whether omitting the annual credit card fee for customers who charge at least $2400 in a year would increase the amount charged on the credit card. The bank makes this offer to a simple random sample of 25 of its existing credit card customers. It then compares how much these customers charge this year with the amount they charged last year. The mean percentage change is +12% with a standard deviation of 6%. The amounts are nearly normally distributed. **TYPO: I NEED TO GIVE YOU THE POPULATION STANDARD DEVIATION SIGMA TO DO THIS PROBLEM. PRETEND IT IS 20%.**

(a) Is there a statistically significant increase in the mean amount charged? Please state a null and alternative hypothesis, perform the appropriate test, state a p-value and use an alpha=.05 as your decision rule. Give you conclusion in plain English please – do you reject or not reject the null and what does this mean?

H0: μ=0 (omitting annual fee makes no difference in the amount charged)
H1: μ > 0 (omitting annual fee increases the amount charged)

Test: $Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{12 - 0}{\frac{20}{\sqrt{25}}} = \frac{12}{\frac{20}{5}} = 3.00$

p-value: the area to the right (greater than) Z=3.00 is .0013

Conclusion: We reject the null in favor of the alternative. This is statistically significant because .0013 is less than .05. This means it appears that college students have higher IQs than the average human.

(b) Please construct a 99% confidence interval for the mean percentage change.

$$\bar{x} \pm Z^{*}\left(\frac{\sigma}{\sqrt{n}}\right) \Rightarrow 12 \pm 2.576\left(\frac{6}{\sqrt{25}}\right) \Rightarrow 12 \pm 3.091$$

12. The pregnancy duration of human females (age 18 and over) is approximately normal with a mean of 266 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.

(a) Suppose we actually knew $\sigma = 16$ days. Please test the hypothesis that older women have longer pregnancy durations than the average woman. State a null and research hypothesis, perform a test, state a p-value and explain your result (do you reject or not reject the null and why). Use a 1% level of significance or an alpha of .01 as your decision rule.

H0: μ=266 (older women do not appear to have longer durations)
H1: μ > 266 (older women have longer durations)

Test: $Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{267 - 266}{\frac{16}{\sqrt{121}}} = \frac{1}{\frac{16}{11}} = 1.45$

p-value: the area to the right (greater than) Z=1.45 is .0735

Conclusion: We do not reject the null in favor of the alternative. This is not statistically significant because .0735 is greater than .01. This means older women do not appear to have longer durations

13. 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.

975/1600=.6094

$$\hat{p} \pm Z^{*}\left(\sqrt{\frac{(\hat{p})(1-\hat{p})}{n}}\right) \Rightarrow .6094 \pm 1.96\left(\sqrt{\frac{.6094 * .3906}{1600}}\right) \Rightarrow .6094 \pm .0239$$

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

14. 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 Districted 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"?

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{.05 - .07}{\sqrt{\frac{.07 * .93}{225}}} = \frac{-.02}{.017} = -1.18$$ area to the left is .1190

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{.08 - .07}{\sqrt{\frac{.07 * .93}{225}}} = \frac{.01}{.017} = 0.59$$ area to the left is .7224

The area between is .7224-.1190 = .6034 and this is the chance that between 5% and 8% will be identified as "all others"

B. If the sample size were 400 instead of 225 it would (circle one to fill in the blank) the standard deviation for the possible sample counts 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.

$$\hat{p} \pm Z^*\left(\sqrt{\frac{(\hat{p})(1-\hat{p})}{n}}\right) \Rightarrow .06 \pm 1.645\left(\sqrt{\frac{.06 * .94}{225}}\right) \Rightarrow .06 \pm .026$$

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?

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{.10 - .12}{\sqrt{\frac{.12 * .88}{225}}} = \frac{-.02}{.0216} = -0.92$$ the area to the left is .1788 so this is the chance of having fewer than

15. 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 with a standard deviation of 1000, but 36% will consume more than 2,000 calories. And only 5% will get drunk. Please assume that calories are not normally distributed

The Hardvard 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 1,130 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 an average of 1,130 calories consumed if the true population average is 600 with a population standard deviation of 1,000 calories?

$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{1130 - 600}{\frac{1000}{\sqrt{850}}} = \frac{530}{34.299} = 15.45$, this is "off the chart" for Table A, so we interpret it as having no chance.

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

$$\hat{p} \pm Z^*\left(\sqrt{\frac{(\hat{p})(1-\hat{p})}{n}}\right) \Rightarrow .77 \pm 1.15\left(\sqrt{\frac{.77 * .23}{490}}\right) \Rightarrow .77 \pm .022$$

C. Please 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.

$$\hat{p} \pm Z^*\left(\sqrt{\frac{(\hat{p})(1-\hat{p})}{n}}\right) \Rightarrow .07 \pm 1.282\left(\sqrt{\frac{.07 * .93}{250}}\right) \Rightarrow .07 \pm .021$$

16.Indicate whether the statement is true or false

	True	False	
A		XX	In hypothesis testing, an alpha=.10 means that there is a 10% chance that the null hypothesis is wrong
B	XX		In hypothesis testing, a null rejected at alpha=.05 will always be rejected at alpha=.10
C	XX		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	XX		For confidence intervals, decreasing the sample size will have the same effect as decreasing the confidence level.

17. An opinion poll interviews 144 randomly chosen women and records the sample proportion that doesn't feel they get enough time for themselves among several other variables

a) Suppose the poll reveals that 47% of the women surveyed report that they do not get enough time for themselves. Please construct a 99% confidence interval for the percentage of women who do not get enough time for themselves.

$$\hat{p} \pm Z^* \left(\sqrt{\frac{(\hat{p})(1-\hat{p})}{n}} \right) \Rightarrow .47 \pm 2.576 \left(\sqrt{\frac{.47 * .53}{144}} \right) \Rightarrow .47 \pm .107$$

b) Suppose the researchers would like to reduce the width of the confidence interval by exactly one-half. If this is possible to do, please show, using calculations, how this is to be done. If it is impossible to do this, please indicate why this is impossible in the space below.

Quadruple the sample size to 576 OR reduce the confidence to about 80%, Z=1.288 rounded 1.29

18. Company SUCCESS-2000 advertises that by using its program, high school students can increase their Verbal SAT scores. A high school teacher is skeptical of this claim. She suspects the average score will be the same and there will be no increase. The teacher plans to examine data from a sample of students who use the program to see if the claim is true. If μ represents the **mean increase** in Verbal SAT score if all high school students used the method, the null and the alternative (research) hypotheses are:

A) H_o: $\mu < 0$ and H1: $\mu = 0$
B) H_o: $\mu = 0$ and H1: $\mu > 0$
C) H_o: $\mu = 0$ and H1: $\mu < 0$
D) H_o: $\mu = 0$ and H1: $\mu \neq 0$
E) H_o: $\bar{x} = 0$ and H_1: $\bar{x} < 0$

19. We carry out a hypothesis test of the hypotheses: H_0: $\mu = 0$ and H1: $\mu > 0$ and obtain a P-value of 0.002. Which is the correct interpretation of that value?

A) The chance that the null hypothesis is true is 0.002.
B) The chance that the alternative (research) hypothesis is true is 0.998.
C) The chance of obtaining a value of the test statistic as or larger than the value actually obtained if H_0 is true is 0.002.
D) The chance of obtaining a value of the test statistic > 0 is only 0.002
E) The chance that the alternative (research) hypothesis is true is 0.002

Questions 20 & 21 refer to the following case:
A particular model car advertises that it gets 30 miles per gallon in city driving**. A consumer group wishes to see if this is true or if the gas mileage is lower than 30 miles per gallon.** They try a sample of 4 cars of this model, and find the gas mileage of each car after 15,000 miles of city driving. They compute the mean gas mileage for the 4 cars and find it to be $\bar{x} = 27.5$ miles per gallon. Assume the distribution of city gas mileage of cars is normal with mean μ and standard deviation $\sigma = 2$ miles per gallon. Consider

H_0: $\mu = 30$
H_A: $\mu < 30$

20. Calculate the p-value.

$Z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{27.5 - 30}{\frac{2}{\sqrt{4}}} = \dfrac{-2.5}{1} = -2.5$ the p-value is the area to the left of -2.5 or .0062

21. Assume that we want to work with $\alpha = 0.05$. What would be the appropriate conclusion?
A) The advertisement is correct, that model has a mileage of 30 miles per gallon
B) The advertisement is wrong, the mileage is under 30 miles per gallon
C) We can not arrive at a conclusion because we have not tried all the cars
D) We reject the alternative (research) hypothesis because the p-value is negative
E) The mileage for the population of all cars of that type is 27.5 miles per gallon

22. In a test of hypothesis we decided to make α =0.05. The p-value was 0.136
Which would be the correct decision?
A) Reject the null hypothesis
B) Do not reject the null hypothesis
C) Change the α to 0.3
D) The results are inconclusive
E) The results are significant at the 0.05 level

23. A sample of 121 seniors from a large metropolitan area school district had a mean Math SAT score of $\bar{x}$ =450. Suppose we know that the standard deviation of the population of Math SAT scores for seniors in the district is σ =100. Estimate the mean Math SAT score μ for the population of seniors with a 90% confidence.

$$\bar{x} \pm Z^*\left(\frac{\sigma}{\sqrt{n}}\right) \Rightarrow 450 \pm 1.645\left(\frac{100}{\sqrt{121}}\right) \Rightarrow 450 \pm 14.9595$$

24. The mean batting average of 16 major-league baseball players selected at random is 0.263. Suppose batting averages are normally distributed with a population standard deviation of $\sigma = 0.031$.

a) Find a 98% confidence interval for the mean batting average of all major-league baseball players.

$$\bar{x} \pm Z^*\left(\frac{\sigma}{\sqrt{n}}\right) \Rightarrow .263 \pm 2.326\left(\frac{.031}{\sqrt{16}}\right) \Rightarrow .263 \pm .018$$

b) How large of a sample of baseball players would we need to estimate the mean batting average to within .008 with 98% confidence?

$$n = \left(\frac{Z * \sigma}{MoE}\right)^2 = \left(\frac{2.326 * .031}{.008}\right)^2 = 81.24 \approx 82$$

c) In 1999, the batting average of all major-league baseball players had a mean of .270. We wish to check if the mean batting average is **TYPO different this season. It should have read is "higher this season" or it should have read "is lower this season"**. State the hypotheses we would use in conducting a test of significance. DON'T ACTUALLY PERFORM THE TEST.

H0: μ=.270 (batting average this season is same as 1999)
H1: μ > .270 (batting average this season is higher)

H0: μ=.270 (batting average this season is same as 1999)
H1: μ < .270 (batting average this season is lower)

d) The information above would lead to a test value of $z = -.903$ if we were to conduct the test. What is the *P*-value?

P = .1841 if you were using an alpha of .05 or even .10, you wouldn't reject the null

25. 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", 175 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? (15 points)

115/529 = .2174 the sample proportion (p-hat) who said “Compact”

H0: p=.25 (what the executive believes, treat it as a parameter)
H1: p < .25 (what your survey data suggests is the truth)

Test: $$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{.2174 - .25}{\sqrt{\frac{.25 * .75}{529}}} = \frac{-.0326}{.0188} = -1.7316$$

p-value: the area to the left (less than) Z=-1.73 is .0418

Conclusion: We reject the null in favor of the alternative. This is statistically significant because .0418 is less than .05. This means it appears the executive is wrong in his thinking and that his company sells less computers than he thinks.

(b) Suppose computer ownership among adults is actually distributed binomial, that is B(1000, .56). What is the chance that you would have between 550 and 575 adults saying that they own a personal computer? (5 points)

If B(1000, .56) then we expect 1000*.56=560 computers owned in 1000 adults

Need 2 Z calculations here

$$Z = \frac{n\hat{p} - np}{\sqrt{np(1-p)}} = \frac{550 - 560}{\sqrt{1000(.56)(.44)}} = -.64$$

$$Z = \frac{n\hat{p} - np}{\sqrt{np(1-p)}} = \frac{575 - 560}{\sqrt{1000(.56)(.44)}} = .96$$

The area between a Z=.96 and a Z=-.64 is .8315-.2611= .5704 this is the chance

(continued from above)
(c) Suppose Compact's true market share is REALLY 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?

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{.22 - .25}{\sqrt{\frac{.25 * .75}{529}}} = \frac{-.03}{.0188} = -1.59$$ area to the left is .0559 (chance that less than 22% owned a compact)

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{.28 - .25}{\sqrt{\frac{.25 * .75}{529}}} = \frac{.03}{.0188} = 1.59$$ area to the right is .0559 (chance that more than 28% owned a compact)

the area between Z=-1.59 and Z=1.59 is 1.0 – (.0559 + .0559) = .8882

26. At UCLA, it is believed that USC students earn less money immediately after graduation than other college students. A simple random sample of 16 USC students is drawn from the population of all recent USC grads. The average for the sample was $34,300 and the sample standard deviation is $10,500. A national study of all college graduates had an average of $39,800 after graduation. Please test the hypothesis that USC students earn less than other college students. State a null and a research hypothesis, perform the appropriate test, state either an exact or an approximate p-value and explain your result (do you reject or not reject the null, is the result statistically significant or not and tell us how to interpret the result -- in plain English). Use a 5% level of significance for your decision rule. (10 points) **TYPO: I NEEDED TO GIVE YOU A VALUE FOR SIGMA. ASSUME SIGMA=$18,000.**

H0: μ=39800 (USC students earn as much as other students)
H1: μ < 39800 (USC students earn less)

Test: $Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{34300-39800}{\frac{18000}{\sqrt{16}}} = \frac{-5500}{4500} = -0.81$

p-value: the area to the left (less than) Z=-0.81 is .2090

Conclusion: We do not reject the null in favor of the alternative. This is not statistically significant because .2090 is not less than .05. This means there is no evidence to suggest that USC students earn less than other students.

27. The amount of money that college students earn immediately after graduation is approximately normal with a mean of $34,300 and a standard deviation of $12,400. At UCLA, it is believed that UCLA students earn more money immediately after graduation than other college students. A simple random sample of 16 UCLA students is drawn from the population of all recent UCLA grads. The average for the sample was $39,800 and the sample standard deviation is $10,500. Please test the hypothesis that UCLA students earn more than other college students. State a null and a research hypothesis, perform the appropriate test, state either an exact or approximate p-value and explain your result (do you reject or not reject the null, is the result statistically significant or not and tell us how to interpret the result -- in plain English). Use a 1% level of significance for your decision rule. (10 points)

H0: μ=34,300 (UCLA students earn as much as other students)
H1: μ > 34,300 (UCLA students earn more)

Test: $Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{39800-34300}{\frac{10500}{\sqrt{16}}} = \frac{5500}{2625} = 2.10$

p-value: the area to the right (greater than) Z=2.10 is .0179

Conclusion: We reject the null in favor of the alternative. This is statistically significant because .0179 is less than .05. This means there is evidence to suggest that UCLA students earn more than other students.