MATH 117 - ELEMENTS OF STATISTICS
FINAL EXAM REVIEW FOR CHAPTERS 1 – 7 (REVISED-FALL 2018)
MONTGOMERY COLLEGE

Selected problems from various sources included the textbook:
Statistics: Unlocking the Power of Data, Lock^5 2e
Solutions are included in this document.

1. **Parameter/Statistic** In each of the following, state whether the quantity described is a parameter or a statistic and give the correct notation.

   A. Average household income for all houses in the US, using data from the US Census.
   B. Proportion of people who use an electric toothbrush, using data from a sample of 300 adults.
   C. Proportion of registered voters in a county who voted in the last election, using data from the county voting records.
   D. Average number of television sets per household in North Carolina, using data from a sample of 1000 households.

2. **Fish Consumption and Intelligence**: In 2000, a study was conducted on 4000 Swedish 15-year-old males. The boys were surveyed and asked, among other things, how often they consume fish each week. Three years later, these answers were linked to the boys' scores as 18-year-olds on an intelligence test. The study found that boys who consume fish at least once a week scored higher on the intelligence test.

   A. Is this an experiment or an observational study? Explain.
   B. What are the explanatory and response variables?
   C. Give an example of a potential confounding factor.
   D. Does this study provide evidence that eating fish once a week improves cognitive ability?

3. **Identifying the Method of Analysis**: In each of the following identify the method of analysis needed to answer the question. Indicate whether we should conduct a hypothesis test or find a confidence interval and also indicate whether the analysis will be done on a proportion, a mean, a difference in proportions, a difference in means, or a matched pairs difference in means.

   A. Use data collected at a retail store to estimate the average amount of money people spend in the store.
   B. Use results collected at a supermarket to see whether there is a difference in the average amount of time customers have to wait in line between two different check-out cashiers.
   C. Use data from an experiment on mice to see if there is evidence that mice fed a high-sugar diet are more likely to be classified as insulin-resistant than mice fed a normal diet.
   D. Use data collected at an online shopping site to estimate the proportion of people visiting the site who make a purchase.
   E. Use data collected from a sample of applicants at a college admissions office to measure how large the difference is in the average size of the financial aid package between early decision applicants and regular decision applicants.

F. Use data from a study done at a college fitness center in which muscle mass of participants was measured before and after a 6-week program working with resistance bands to estimate the mean increase in muscle mass.
G. Use a sample of students at a large university to determine whether the proportion of students at the university who are left-handed is different from the national US proportion of 12%.
H. Use results from a survey to estimate the difference in the proportion of males and females who say they are trying to lose weight.

4. **Be Sure to Get Your Beauty Sleep!** New research supports the idea that people who get a good night's sleep look more attractive. In the study, 23 subjects ages 18 to 31 were photographed twice, once after a good night's sleep and once after being kept awake for 31 hours. Hair, make-up, clothing, and lighting were the same for both photographs. Observers then rated the photographs for attractiveness, and the average rating under the two conditions was compared. The researchers report in the British Medical Journal that "Our findings show that sleep-deprived people appear less attractive compared with when they are well rested."

   A. What is the explanatory variable? What is the response variable?
   B. Is this an experiment or an observational study? If it is an experiment, is it a randomized comparative design or a matched pairs design?
   C. Can we conclude that sleep deprivation causes people to look less attractive? Why or why not?

5. **Employer-Based Health Insurance:** A report from a Gallup poll 2011 started by saying, "Forty-five percent of American adults reported getting their health insurance from an employer…" Later in the article we find information on the sampling method, "a random sample of 147,291 adults, aged 18 and over, living in the US," and a sentence about the accuracy of the results, "the maximum margin of sampling error is ±1 percentage point."

   A. What is the population of interest?
   B. What is the sample? Is the sample a representative of the population?
   C. What is the population parameter of interest? What is the relevant statistic?
   D. Use the margin of error to give an interval estimate for the parameter of interest. Interpret it in terms of getting health insurance from an employer.

6. **Interpreting a P-value:** In each case, indicate whether the statement is a proper interpretation of what a p-value measures.

   A. The probability the null hypothesis H_0 is true.
   B. The probability that the alternative hypothesis H_a is true.
   C. The probability of seeing data as extreme as the sample, when the null hypothesis H_0 is true.
   D. The probability of making a Type I error if the null hypothesis $H_0$is true.
   E. The probability of making a Type II error if the alternative hypothesis H_a is true.

7. **Carbo Loading:** It is commonly accepted that athletes should "carbo load," that is, eat lots of carbohydrates, the day before an event requiring physical endurance. Is there any truth to this? Suppose you want to design an experiment to find out for yourself: "Does carbo loading actually improve athletic performance the following day?" You recruit 50 athletes to participate in your study.

A. How would you design a randomized comparative experiment?
B. How would you design a matched pairs experiment?

8. Arsenic in Chicken: A test to determine if the mean level of arsenic in chicken meat is above 80 ppb. If a restaurant chain finds significant evidence that the mean arsenic level is above 80, the chain will stop using that supplier of chicken meat. The hypotheses are

$$H_0: \mu = 80 \text{ and } H_a: \mu > 80$$

where μ represents the mean arsenic level in all chicken meat from that supplier. Samples from two different suppliers are analyzed, and the resulting p-values are given:

Sample from Supplier A; p-value is 0.0003
Sample from Supplier B; p-value is 0.3500

A. Interpret each p-value in terms of the probability of the results happening by random chance.
B. Which p-value shows stronger evidence for the alternative hypothesis? What does this mean in terms of arsenic and chickens?
C. Which supplier, A or B, should the chain get chickens from in order to avoid too high a level of arsenic?

9. Which design is better? Which design do you think is better for this situation? Why? Some methods may be used to make a confidence interval wider or narrower. Check the following methods that would decrease the width of a confidence interval for a mean, if all else stays the same. For each choice you select, explain why that would decrease the confidence interval.

A. Increase the sample size.
B. Decrease the sample size.
C. Increase the level of confidence.
D. Decrease the level of confidence.

10. *t* -test assumptions. A student research project on adolescent girls' self esteem included a one-sample t-test with a sample size of 15 girls. What assumption should the students have checked before using the t-test?

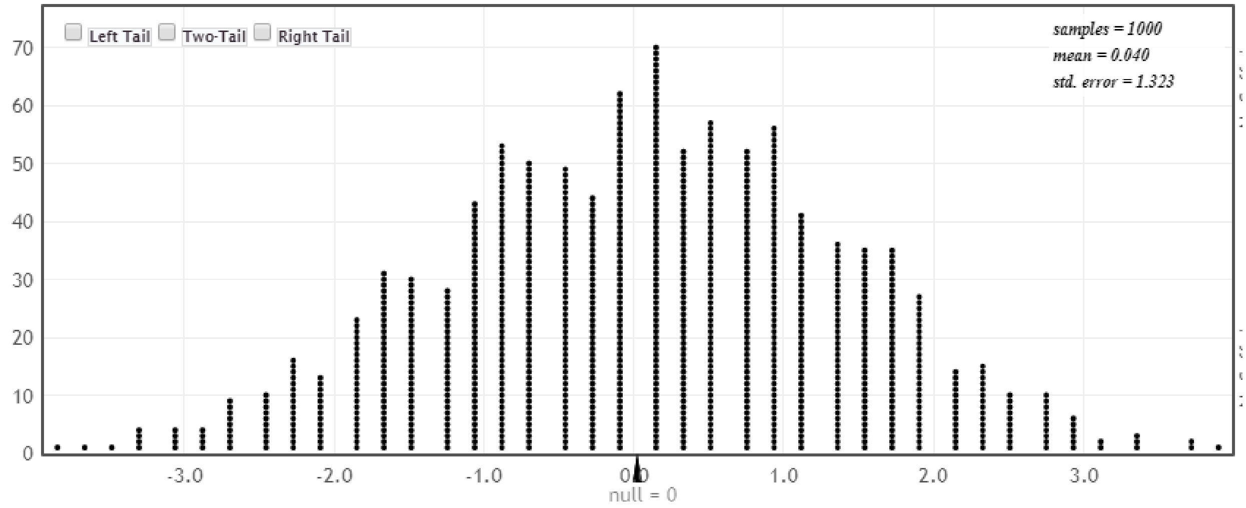
11. Finger Tapping and Caffeine: Many people feel they need a cup of coffee or other source of caffeine to “get going” in the morning. The effects of caffeine on the body have been extensively studied. In one experiment, researchers trained a sample of male college students to tap their fingers at a rapid rate. The sample was then divided at random into two groups of 10 students each. Each student drank the equivalent of about two cups of coffee, which included about 200 mg of caffeine for the students in one group but was decaffeinated coffee for the second group. After a 2-hour period, each student was tested to measure finger tapping rate (taps per minute). The students did not know whether or not their drinks included caffeine and the person measuring the tap rates was also unaware of the groups. This was a double-blind experiment with only the statistician analyzing the data having information linking the group membership to the observed tap rates. The goal of the experiment was to determine whether caffeine produces an increase in the average tap rate. The finger-tapping rates measured in this experiment are summarized in the table below and stored in CaffeineTaps.

Caffeine: 246 248 250 252 248 250 246 248 245 250
No caffeine: 242 245 244 248 247 248 242 244 246 242

Let μ_c and μ_n represent the average tap rate of the people who have had coffee with caffeine and without caffeine, respectively, the null and alternative hypothesis are:
$H_0: \mu_c = \mu_n$ and $H_a: \mu_c > \mu_n$

A. Find the sample mean of each group and calculate the difference, $\bar{x}_c - \bar{x}_n$, in the simulated sample means.
B. The difference in sample means found in part A is one data point in a randomization distribution. A sketch of the randomization distribution is shown below. Locate your randomization statistic on the sketch.

Randomization Dotplot of $\bar{x}_1 - \bar{x}_2$, Null hypothesis: $\mu_1 = \mu_2$

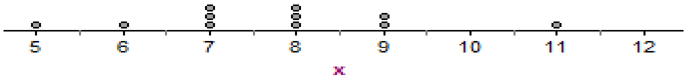


C. Use the randomization distribution to find a p-value for the test, and write its meaning in the context of this study.
D. Does the p-value give strong evidence that caffeine increases average finger-tapping? Explain.
E. If the test were a two-tailed test, what would be the p-value?
F. Use $t^* = 2.102$ and SE=1.323, which is an estimate of the standard error of the statistic from the bootstrap distribution, to construct a 95% confidence interval for difference in average tap rate between the caffeine and no-caffeine groups and interpret in the context of this study.
** Note that this answer will be the same as if you entered the summary statistics in your calculator under the menu option: 0:2-SampleT-Interval.*

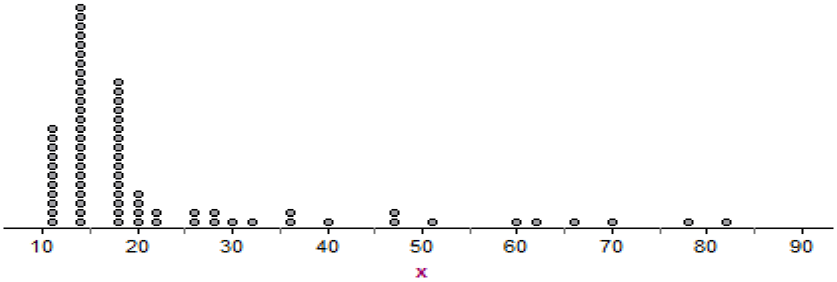
12. Voter polling. The company Mason-Dixon Polling and Research, Inc. conducted an opinion poll for the St. Paul Pioneer Press and Minnesota Public Radio from Oct. 30 through Nov. 1, 2002, just prior to the election on Nov. 5, 2002. The poll surveyed 625 potential voters. One of the questions asked if the person thought Walter Mondale was the best choice to replace Paul Wellstone as the Democratic candidate for Minnesota senator. Of the 625 people, 344 answered YES to this question. Does the evidence from this sample support the hypothesis that more than half of all potential voters thought Mondale was the best choice? Justify your answer with an appropriate statistical procedure.

13. Is a t-Distribution Appropriate? We give summary statistics and a dotplot for a sample. In each case (A-C), indicate whether or not it is appropriate to use the t-distribution. If it is appropriate, give the degrees of freedom for the t-distribution and give the estimated standard error.

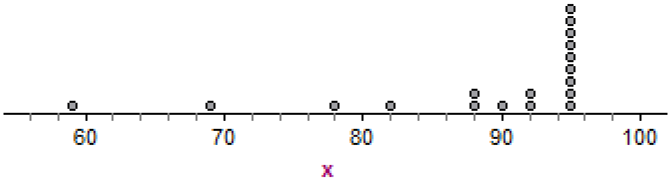
A. A sample with $n = 12$, $\bar{x} = 7.72$, and $s = 1.6$



B. A sample with $n = 75$, $\bar{x} = 22.85$, and $s = 16.51$



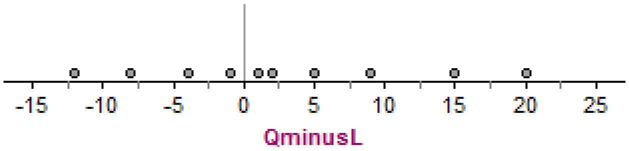
C. A sample with $n = 18$, $\bar{x} = 88.5$, and $s = 10.3$



14. Quiz vs Lecture Pulse Rates: Do you think your pulse rate is greater when you are taking a quiz than when you are sitting in a lecture? The data in the table below show pulse rates collected from 10 students in a class lecture and then from the same students during a quiz. The data are stored in QuizPulse10. The dotplot for the difference of the Quiz and the lecture pulse rates is shown below.

Student	1	2	3	4	5	6	7	8	9	10	Mean	Std. Dev
Quize(Q)	75	52	52	80	56	90	76	71	70	66	68.8	**12.5**

Lecture (L)	73	53	47	88	55	70	61	75	61	78	66.1	12.8
d=QminusL	2	−1	5	−8	1	20	15	−4	9	−12	2.7	9.93



A. Use a t-distribution (or use ISIapplets-Theory Based Inference) if it is appropriate to construct a 95% confidence interval for the difference in mean pulse rate between students in a class lecture and taking a quiz.

B. Write the interpretation of the confidence interval that you constructed above in the context of this study.

15. **P-values and Confidence Intervals**: We want to test $H_0: \mu = 0$ against $H_a: \mu \neq 0$. Assume the p-value is 0.07, calculated based on a sample data and the given hypotheses. If you were to construct a 95% confidence interval for μ, why would you expect zero to be in the interval? Explain.

16. **Approval Rating for Congress** In a Gallup poll conducted in August 2010, a random sample of $n = 1013$ American adults were asked “Do you approve or disapprove of the way Congress is handling its job?” The proportion who said they approve is $\hat{p} = 0.19$. If we use a 5% significance level, what is the conclusion if we are:

A. Testing to see if there is evidence that the job approval rating is different than 20%.

B. Testing to see if there is evidence that the job approval rating is different than 14%.

C. If we were to calculate the p-values associated with questions A and B, which one would you expect to be smaller than 0.05? Explain.

17. **Properties of Confidence Intervals**: In estimating the mean score on a fitness exam, we use an original sample of size n=30 and a bootstrap distribution containing 5000 bootstrap samples to obtain a 95% confidence interval of 67 to 73. In each of A to D given below, a change in this process is described. If all else stays the same, which of the following confidence intervals (a, b, or c) is the most likely result after the change:

a. 66 to 74 b. 67 to 73 c. 67.5 to 72.5

A. Using the data to find a 99% confidence interval.

B. Using the data to find a 90% confidence interval.

C. Using an original sample of size n=45.

D. Using an original sample of size n=16.

18. **Are Grades Significantly Higher on the Second Quiz?** The table below gives a sample of grades on the first two quizzes in an introductory statistics course. We are interested in testing whether the mean grade on the second quiz is significantly higher than the mean grade on the first quiz. Use ISIapplets-Theory Based Inference or other technology for data analysis.

Are grades higher on the second quiz?

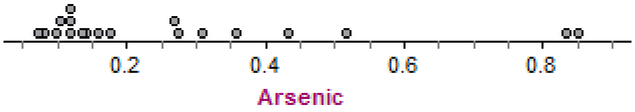
First Quiz 72 95 56 87 80 98 74 85 77 62
Second Quiz 78 96 72 89 80 95 86 87 82 75

A. Complete the test if we assume that the grades from the first quiz come from a random sample of 10 students in the course and the grades on the second quiz come from a different separate random sample of 10 students in the course. Clearly state the conclusion.
B. Now conduct the test if we assume that the grades recorded for the first quiz and the second quiz are from the same 10 students in the same order. (So the first student got a 72 on the first quiz and a 78 on the second quiz.)
C. Why are the results so different? Which is a better way to collect the data to answer the question of whether grades are higher on the second quiz?

19. Arsenic in Toenails: Arsenic is toxic to humans, and people can be exposed to it through contaminated drinking water, food, dust, and soil. Scientists have devised an interesting new way to measure a person’s level of arsenic poisoning: by examining toenail clippings. In a recent study in US, scientists measured the level of arsenic (in ppm) in toenail clippings of 19 people with private wells in New Hampshire. The following levels were recorded (the data are also available in ToenailArsenic):

0.119 0.118 0.099 0.118 0.275 0.358 0.080 0.158 0.310 0.105
0.73 0.832 0.517 0.851 0.269 0.433 0.141 0.135 0.175

A. A typical value in a sample or population data is defined to be a value that lies within one standard deviation of the mean in the data. Use Microsoft Excel to find all the typical values in the data given in this problem.
B. Find the z-score of the largest concentration and interpret it.
C. Use technology to find the five-number summary and sketch a boxplot.
D. What is the range? What is the IQR?
E. The figure below shows a dotplot of the arsenic concentrations. Which measures of center and spread are most appropriate for this distribution: the mean and standard deviation or the five-number summary? Explain.



F. Is it appropriate to use the general rule about having the data within two standard deviation of this distribution? Why or why not?

20. Comparing Global Internet Connections: The Nielsen Company measured connection speeds on home computers in nine different countries in order to determine whether connection speed affects the amount of time consumers spend online. The table below shows the percent of Internet users with a ‘‘fast” connection (defined as 2Mb or faster)and the average amount of time spent online, defined as total hours connected to the web from a home computer during the month of February 2011. The data are also available in the dataset GlobalInternet.

A. What would a positive association mean between these two variables? Explain why a positive relationship might make sense in this context.

B. What would a negative association mean between these two variables? Explain why a negative relationship might make sense in this context.

Internet connection speed and hours online

Country	Percent fast connection	Hours Online
Switzerland	88	20.18
United States	70	26.26
Germany	72	28.04
Australia	64	23.02
United Kingdom	75	28.48
France	70	27.49
Spain	69	26.97
Italy	64	23.59
Brazil	21	31.58

C. Make a scatterplot of the data, using connection speed as the explanatory variable and time online as the response variable. Is there a positive or negative relationship? Are there any outliers? If so, indicate the country associated with each outlier and describe the characteristics that make it an outlier for the scatterplot.

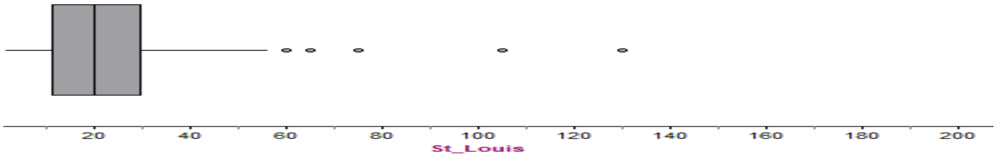
D. If we eliminate any outliers from the scatterplot, does it appear that the remaining countries have a positive or negative relationship between these two variables?

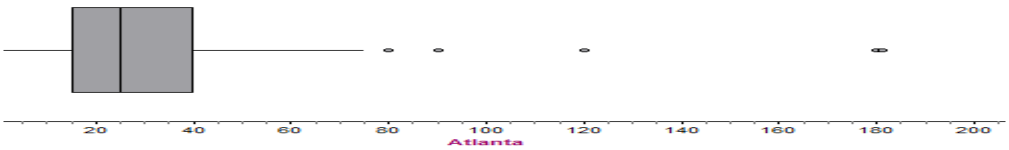
E. Use technology to compute the correlation. Is the correlation affected by the outliers?

F. Can we conclude that a faster connection speed causes people to spend more time online?

21. St. Louis vs Atlanta Commute Times: The datafile CommuteAtlanta contains a sample of commute times for 500 workers in the Atlanta area and the data in CommuteStLouis has similar information on the commuting habits of a random sample of 500 residents from metropolitan St. Louis. The summary and boxplots for the two samples are given below. We wish to estimate the difference in mean commute time between Atlanta and St. Louis.

Group	n	Mean	Std. Dev
Atlanta	500	29.11	20.72
St. Louis	500	21.97	14.23





Commute times in Atlanta and St. Louis

A. Discuss and compare the boxplots. Which city appears to have the longer average commute time?
B. Give notation for the parameter we are estimating and give the best point estimate from the data.
C. Describe how to compute one bootstrap statistic from this data.
D. Use StatKey or other technology to create a bootstrap distribution for the difference in mean commute times between the two cities and use the standard error to find and interpret a 95% confidence interval.
E. Is it appropriate to use a t-distribution(or ISIapplets-Theory Based Inference) for this situation?If yes, use it to compute a 95% confidence interval for $\mu_{atl} - \mu_{stl}$. Did you get a similar answer to that of question D?
F. Interprete the confidence interval in the context of this study.

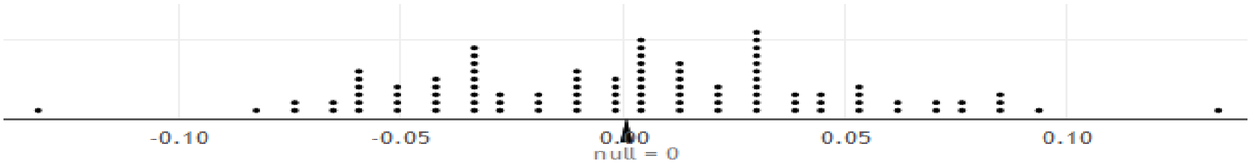
22. **Influencing Voters**: When getting voters to support a candidate in an election, is there a difference between a recorded phone call from the candidate or a flyer about the candidate sent through the mail? A sample of 500 voters is randomly divided into two groups of 250 each, with one group getting the phone call and one group getting the flyer. The voters are then contacted to see if they plan to vote for the candidate in question. A possible sample results are shown in the table below:

Is a phone call or a flyer more effective?

	Will Vote for Candidate	Will Not Vote for Candidate	Total
Phone call	152	98	250
Flyer	145	105	250

We wish to see if there is evidence that a recorded phone call is more effective than a flyer in persuading voters to vote for a particular candidate.

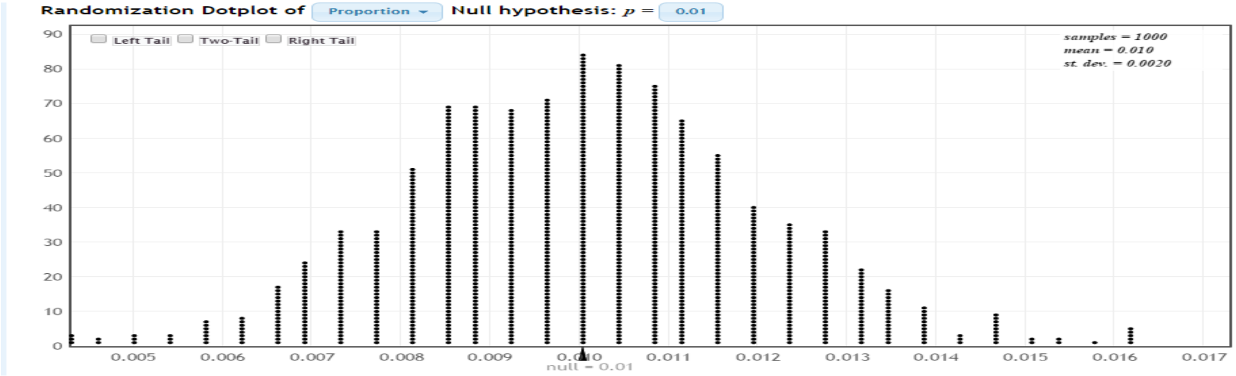
A. Define the relevant parameter(s) and state the null and alternative hypotheses.
B. Compute the two sample proportions:$\hat{p}_c$, the proportion of voters getting the phone call who say they will vote for the candidate, and $\hat{p}_f$, the proportion of voters getting the flyer who say they will vote for the candidate. Is there a difference in the sample proportions?
C. The sample statistic of interest is $D = \hat{p}_c - \hat{p}_f$. A null randomization distribution using 100 simulated samples for this test is shown below, and use it to find an approximate p-value.



D. If a z- distribution (or ISIapplets-Theory Based Inference) is appropriate, use it to find the p-value.
E. What is the conclusion in the context of this study?
F. In the conclusion, which type of error are we possibly making: Type I or Type II? Describe what that type of error means in this situation.

23. Colonoscopy, Anyone? A colonoscopy is a screening test for colon cancer, recommended as a routine test for adults over age 50. A new study provides the best evidence yet that this test saves lives. The proportion of people with colon polyps expected to die from colon cancer is 0.01. A sample of 2602 people who had polyps removed during a colonoscopy were followed for 20 years, and 12 of them died from colon cancer. Does this provide evidence that the proportion of people who die from colon cancer after having polyps removed in a colonoscopy is significantly less than the expected proportion (without a colonoscopy) of 0.01?

A. What are the null and alternative hypotheses?
B. What is the sample proportion?
C. The figure below shows a randomization distribution for this test. Use the fact that there are 1000 dots in the distribution to find the p-value. Explain your reasoning.


Randomization Dotplot of Proportion Null hypothesis: p = 0.01

samples = 1000
mean = 0.010
st. dev. = 0.0020

D. Does the p-value appear to show significant evidence that colonoscopies save lives?

24. Drink Tea for a Stronger Immune System: Drinking tea appears to offer a strong boost to the immune system. In a study, eleven healthy non-tea-drinking individuals were asked to drink five or six cups of tea a day, while ten healthy non-tea- and non-coffee-drinkers were asked to drink the same amount of coffee, which has caffeine but not the L-theanine that is in tea. The groups were randomly assigned. After two weeks, blood samples were exposed to an antigen and production of interferon gamma was measured. The results are shown in the table below and are available in ImmuneTea. The question of interest is whether the data provide evidence that production is enhanced in tea drinkers.

Immune system response in tea and coffee drinkers

Tea:	5	11	13	18	20	47	48	52	55	56	58
Coffee	0	0	3	11	15	16	21	21	38	52	

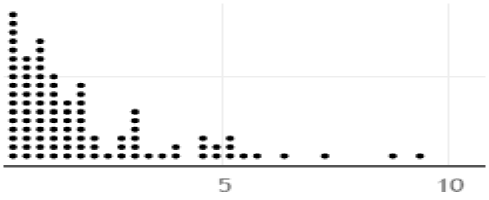
A. Is this an experiment or an observational study?
B. Is this a difference in means test with separate samples or a paired difference in means test?
C. What are the null and alternative hypotheses?
D. Find a standardized test statistic and use the t-distribution to find the p-value and make a conclusion.
E. Always plot your data! Look at a graph of the data. Does it appear to satisfy a normality condition?
F. Randomization test might be a more appropriate test to use in this case. Construct a randomization distribution for this test and use it to find a p-value and make a conclusion.
G. What conclusion can we draw?

25. Gender and Award Preference: A two-way table showing preferences for an award (Academy Award, Nobel Prize, Olympic gold medal) by gender for the students sampled in shown below. Test whether the data indicate there is some association between gender and preferred award. Use $\alpha = 0.05$.

Two-way table of gender and preferred award

	Academy	Nobel	Olympic	Total
Female	20	76	73	169
Male	11	73	109	193
Total	31	149	182	362

A. Set up the hypotheses and calculate the expected count.
B. Are the expected counts large enough for χ^2-test?
C. How many degrees of freedom do we have for this test?
D. Use StatKey to show that the observed χ^2 is 8.24, and the p-value is 0.016.
E. Write the meaning the p-value in the context of this study.
F. Do you reject the null hypothesis?
G. A randomization distribution for this test using 100 simulated samples is shown below. Circle the dots with values as high (or higher) than was observed in the sample.



H. State the conclusion in the context of the problem.
I. Explain what Type I and Type II errors mean in the context of this study.
J. If you made an error with the decisions in part F, is it a Type I or Type II error?

Solutions

1. **Parameter/Statistic**
   A. This mean is a population parameter; notation is μ.
   B. This proportion is a sample statistic; notation is $\hat{p}$.
   C. This proportion is a population parameter; notation is p.
   D. This mean is a sample statistic; notation is $\bar{x}$.
2. **Fish Consumption and Intelligence**

A. It is an observational study. The researcher asked the boys how often they ate fish and collected data on their intelligence test scores, but did nothing to change or determine their levels of fish consumption or intelligence.

B. The explanatory variable is whether or not fish is consumed at least once a week, and the response variable is the score on the intelligence test.

C. One possible confounding variable is the intelligence level of the parents. Families in which the parents are more intelligent may tend to eat more fish and also to have sons who score higher on an intelligence test. Other possible confounding variables might be whether boys live near the coast or inland, or how often the boys' parents provide home cooked meals. You can probably think of other possibilities. Remember that a confounding variable is a variable that might influence both the explanatory and the response variables.

D. No. Observational studies cannot yield causal conclusions.

3. **Identifying the Method of Analysis**

A. We are estimating mean amount spent in the store, so we use a confidence interval for a mean.

B. We are testing for a difference in mean time spent waiting in line, so we use a hypothesis test for a difference in means.

C. We are testing for a difference in the proportion classified as insulin-resistant between high sugar and normal diets, so we use a hypothesis test for a difference in proportions.

D. We are estimating the proportion who make a purchase, so we use a confidence interval for a proportion.

E. We are estimating the difference in the mean financial aid package between two groups of students, so we use a confidence interval for a difference in means.

F. We are estimating a difference in mean muscle mass, using before and after paired data. We use a confidence interval for matched pairs difference in means.

G. We are testing whether the proportion of left-handers is different from 12%, so we use a hypothesis test for a proportion.

H. We are estimating the difference in proportion trying to lose weight between males and females, so we use a confidence interval for a difference in proportions.

4. **Be Sure to Get Your Beauty Sleep!**

A. The explanatory variable is whether or not the person had a good night's sleep or is sleep-deprived. The response variable is attractiveness rating.

B. Since the explanatory variable was actively manipulated, this is an experiment. The two treatments are well-rested and sleep-deprived. Since all 23 subjects were photographed with both treatments, this is a matched pairs experiment.

C. Yes, we can conclude that sleep-deprivation causes people to look less attractive, because this is an experiment.

5. **Employer-Based Health Insurance**

A. The population is all people ages 18 and older living in the US.

B. The sample is the 147,291 people who were actually contacted and asked whether or not they got health insurance from an employer. Yes it is a representative of the population. Why?

C. The parameter of interest is p, the proportion of the entire population of US adults who get health insurance from an employer. The relevant statistic is $\hat{p}$= 0.45, the proportion of people in the sample who get health insurance from an employer.

D. An interval estimate is found by taking the best estimate ($\hat{p}$ = 0.45) and adding and subtracting the margin of error (±0.01). We are relatively confident that the population proportion is between 0.44 and 0.46, or that the percent of the entire population that receive health insurance from an employer is between 44% and 46%.

6. **Interpreting a P-value**

Only choice C "The probability of seeing data as extreme as the sample, when the null hypothesis,H_0, is true." matches the definition of the p-value. The other choices are common misinterpretations of a p-value. The p-value does not measure the probability of any hypothesis being true (of false) or the chance of making either type of error. It only measures how unusual the original data would be if the null hypothesis were true.

7. **Carbo Loading:**

    A. Randomly assign 25 people to carbo-load and 25 people to not carbo-load and then measure each person's athletic performance the following day.

    B. We would have each person carbo-load and not carbo-load, on different days (preferably different weeks). The order would be randomly determined, so some people would carbo-load first and other people would carbo-load second. In both cases athletic performance would be measured the following day and we would look at the difference in performance for each person between the two treatments.

    C. The matched pairs experiment is probably better because we are able to compare the different effects for the same person. It is more precise comparing one person's athletic performance under two different treatments, rather than different people's athletic performance under two different treatments.

8. **Arsenic in Chicken**

A. If the mean arsenic level is really 80 ppb, the chance of seeing a sample mean as high (or higher) than was observed in the sample from supplier A by random chance is only 0.0003. For supplier B, the corresponding probability (seeing a sample mean as high as B's when $\mu = 80$) is 0.35.

B. The smaller p-value for Supplier A provides stronger evidence against the null hypothesis and in favor of the alternative that the mean arsenic level is higher than 80 ppb. Since it is very rare for the mean to be that large when $\mu = 80$, we have stronger evidence that there is too much arsenic in Supplier A's chickens.

C. The chain should get chickens from Supplier B, since there is strong evidence that Supplier A's chicken have a mean arsenic level above 80 ppb which is unacceptable.

9. **Which design is better?** Which design do you think is better for this situation? Why? Some methods may be used to make a confidence interval wider or narrower. Check the following methods that would decrease the width of a confidence interval for a mean, if all else stays the same. For each choice you select, explain why that would decrease the confidence interval.
   E. Increase the sample size: creates a smaller standard error, which **decreases** the interval
   F. Decrease the sample size: increases the SE, which **widens** the interval
   G. Increase the level of confidence: **widens** the interval
   H. Decrease the level of confidence: **decreases** the interval

10. ***t*** **-test assumptions.** Because the sample size is n=15, which is small, distribution should be approximately normal, with no large outliers or skewing, in order to use the *t*-test.

11. **Finger Tapping and Caffeine**

A. Answers vary. For example, one possible randomization sample is shown below

caffeine 244 250 248 246 248 245 246 247 248 246 mean = 248.3
no caffeine 250 244 252 248 242 250 242 245 242 248 mean = 244.8

B. Answers vary. For the randomization sample above, $\bar{x}_c - \bar{x}_n$ = 248.3 – 244.8 = 3.5.

C. The sample difference of 3.5 for the randomization above would fall a bit to the right of the center of the randomization distribution.

D. p-value=$\frac{3}{1000} = 0.003$; 0.3% is the chance that the test statistic $\bar{x}_c - \bar{x}_n$ will take more extreme values than 3.5(3.5 or more) assuming that there is no difference in average tap rate between the caffeine and no-caffeine groups.

E. Yes since the p-value is less than 0.05.

F. p-value= $2 \cdot 0.003 = 0.006$.

G. $\bar{x}_c - \bar{x}_n \pm t * SE = 3.5 \pm 2.102(1.323) = (1.3326, 5.6674)$. We are 95% sure that average tap rate for caffeine group is higher from 1.3326 to 5.6674 than that of the no-caffeine group.

12. **Voter polling.** The company Mason-Dixon Polling and Research, Inc. conducted an opinion poll for the St. Paul Pioneer Press and Minnesota Public Radio from Oct. 30 through Nov. 1, 2002, just prior to the election on Nov. 5, 2002. The poll surveyed 625 potential voters. One of the questions asked if the person thought Walter Mondale was the best choice to replace Paul Wellstone as the Democratic candidate for Minnesota senator. Of the 625 people, 344 answered YES to this question. Does the evidence from this sample support the hypothesis that more than half of all potential voters thought Mondale was the best choice? Justify your answer with an appropriate statistical procedure.
Ho: p=0.5
Ha: p > 0.5
$\hat{p} = \frac{344}{625} = 0.5504$

Z = 2.52, p-value = 0.0058
Reject the null. There is strong evidence that more than half of all potential voters thought Mondale was the best choice.

13. Is a t-Distribution Appropriate?

A. The t-distribution is appropriate if the sample size is large $(n \geq 30)$ or if the underlying distribution appears to be relatively normal. We have concerns about the t-distribution only for small sample sizes and heavy skewness or outliers. In this case, the sample size is small $(n = 12)$ but the distribution is not heavily skewed and it does not have extreme outliers. A condition of normality is reasonable, so the t-distribution is appropriate. For the degrees of freedom df and estimated standard error SE, we have: $df = n - 1 = 12 - 1 = 11,$ and $SE = \frac{s}{\sqrt{n}} = \frac{1.6}{\sqrt{12}} = 0.46$

B. The t-distribution is appropriate if the sample size is large $(n \geq 30)$ or if the underlying distribution appears to be relatively normal. We have concerns about the t-distribution only for small sample sizes and heavy skewness or outliers. In this case, the sample size is large enough $(n = 75)$ that we can feel comfortable using the t-distribution, despite the clear skewness in the data. The t-distribution is appropriate. For the degrees of freedom df and estimated standard error SE, we have:

$$df = n - 1 = 75 - 1 = 74, \text{ and SE} = \frac{s}{\sqrt{n}} = \frac{16.51}{\sqrt{75}} = 1.91$$

C. The t-distribution is appropriate if the sample size is large$(n \geq 30)$ or if the underlying distribution appears to be relatively normal. We have concerns about the t-distribution only for small sample sizes and heavy skewness or outliers. In this case, the sample size is small $(n \geq 30)$ and the data is heavily skewed with some apparent outliers. It would not be appropriate to use the t-distribution in this case. We might try analyzing the data using simulation methods such as a bootstrap or randomization distribution.

14. **P-values and Confidence Intervals**

Here $\alpha = 1 - 0.95 = 0.05$. Since the p-value is greater than 0.05, we do not reject the null hypothesis, i.e., 0 is a plausible value for μ and therefore it should be one of the values in a 95% confidence interval for μ.

15. Quiz vs Lecture Pulse Rates

A. The distribution of differences appears to be relatively symmetric with no clear outliers, so we can use the t-distribution. From the Q − L differences in the table above we find that the mean difference is $\bar{x}_d = 2.7$ and the standard deviation of the differences is $s_d = 9.93$. Since there are 10 students in the sample, we find $t^* = 2.262$ using an area of 0.025 in the tail of a t-distribution with 10 − 1 = 9 degrees of freedom. To find the confidence interval for the mean difference in pulse rates we use $2.7 \pm 2.262 \frac{9.93}{\sqrt{10}}$ = 2.7 ± 7.1 = (−4.4, 9.8).

B. Based on these results, we are 95% sure that the mean pulse rate for students during a quiz is between 4.4 beats less and 9.8 beats more than the mean pulse rate during lecture.

16. **Approval Rating for Congress**

A. The hypotheses are $H_0: p = 0.20$ and $H_0: p \neq 0.20$. The hypothesized proportion of 0.20 lies inside the confidence interval 0.167 to 0.216, so 0.20 is a reasonable possibility for the population proportion given the sample results. Using a 5% significance level, we do not reject H0 and do not find evidence that Congressional approval is different than 20%.

B. The hypotheses are $H_0: p = 0.14$ and $H_0: p \neq 0.14$. The hypothesized proportion of 0.14 lies outside the confidence interval 0.167 to 0.216, so 0.14 is not a reasonable possibility for the population proportion given the sample results. Using a 5% significance level, we reject H0 and conclude that Congressional approval in August 2010 is different than the record low of 14%.

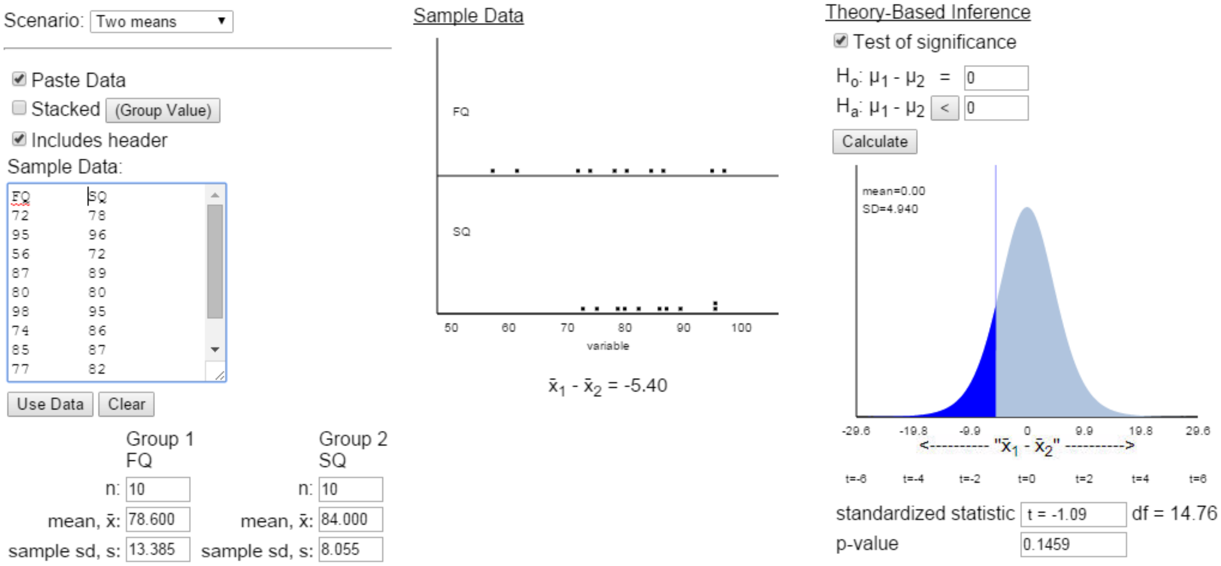
C. That of B since we rejected the null hypothesis at 5% significance level.

17. Are Grades Significantly Higher on the Second Quiz?

A. This is a difference in means test with separate samples. To conduct the test for the difference in means, the hypotheses are:

$$H_0: \mu_1 = \mu_2 \text{ and } H_a: \mu_1 < \mu_2$$

where μ_1 is the average grade for all students on the first quiz and μ_2 is the average grade for all students on the second quiz.



Scenario: Two means ▾

☑ Paste Data

☐ Stacked (Group Value)

☑ Includes header

Sample Data:

FQ	SQ
72	78
95	96
56	72
87	89
80	80
98	95
74	86
85	87
77	82

	Group 1 FQ	Group 2 SQ
n:	10	10
mean, $\bar{x}$:	78.600	84.000
sample sd, s:	13.385	8.055

Sample Data

$\bar{x}_1 - \bar{x}_2 = -5.40$

Theory-Based Inference

☑ Test of significance

$H_o: \mu_1 - \mu_2$ = 0

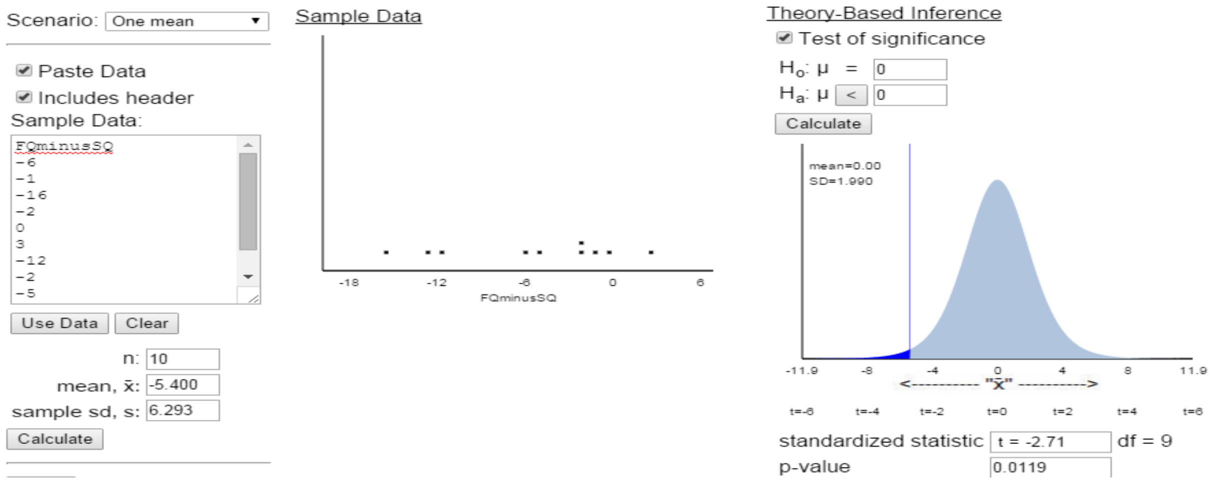
$H_a: \mu_1 - \mu_2$ < 0

standardized statistic t = -1.09 df = 14.76

p-value 0.1459

We see that the p-value for this lower-tail test is 0.146. We do not reject H_0 and do not find convincing evidence that the grades on the second quiz are higher.

B. This is a paired difference in means test. We begin by finding the differences: first quiz−second quiz: $-6 \quad -1 \ -16 \quad -2 \quad 0 \ 3 \quad -12 \quad -2 - 5 \quad -13$. The hypotheses are the same as for part A.



Scenario: One mean ▾

☑ Paste Data

☑ Includes header

Sample Data:

```
FQminusSQ
-6
-1
-16
-2
0
3
-12
-2
-5
```

Use Data | Clear

n: 10

mean, x̄: -5.400

sample sd, s: 6.293

Calculate

Theory-Based Inference

☑ Test of significance

H_o: μ = 0

H_a: μ < 0

Calculate

standardized statistic t = -2.71 df = 9

p-value 0.0119

We see that the p-value for this lower-tail test is 0.012. This is significant at a 5% level and almost at even a 1% level. We reject H_0 and find evidence that mean grades on the second quiz are higher.

C. The spread of the grades is very large on both quizzes, so the high variability makes it hard to find a difference in means with the separate samples. Once we know that the data are paired, it eliminates the variability between people. In this case, it is much better to collect the data using a matched pairs design.

18. **Properties of Confidence Intervals**

A. To find a 99% confidence interval, we go far out on either side than for a 95% confidence interval, so a is the most likely result.

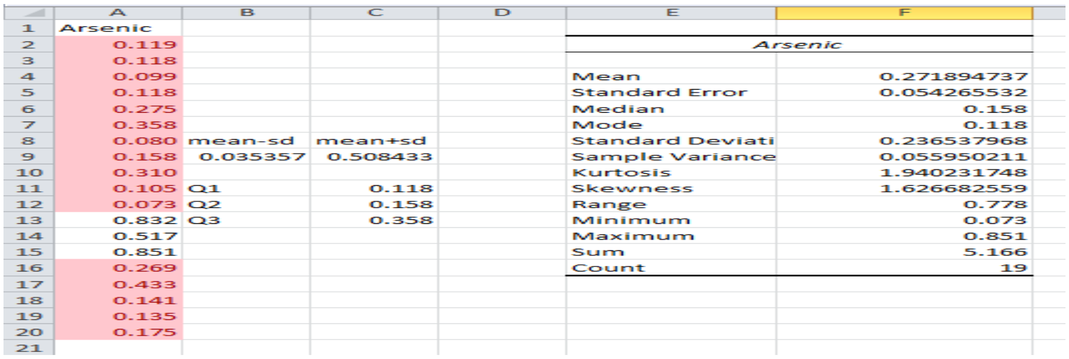
B. To find a 90% confidence interval, we go less far out on either side than for a 95% confidence interval, so c is the most likely result.

C. If the sample size is larger, we have more accuracy and the spread of the bootstrap distribution decreases, so the confidence interval will be narrower. Thus, c is the most likely result.

D. If the sample size is smaller, we have less accuracy and the spread of the bootstrap distribution increases, so the confidence interval will be wider. Thus, a is the most likely result.

19. **Arsenic in Toenails**

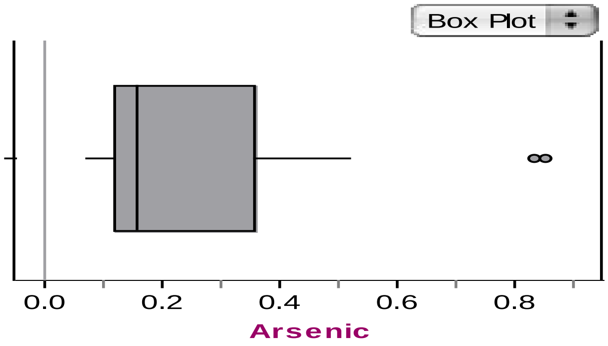
A. Those highlighted below will be typical values.



	A	B	C	D	E	F
1	Arsenic					
2	0.119				*Arsenic*	
3	0.118					
4	0.099				Mean	0.271894737
5	0.118				Standard Error	0.054265532
6	0.275				Median	0.158
7	0.358				Mode	0.118
8	0.080	mean-sd	mean+sd		Standard Deviati	0.236537968
9	0.158	0.035357	0.508433		Sample Variance	0.055950211
10	0.310				Kurtosis	1.940231748
11	0.105	Q1	0.118		Skewness	1.626682559
12	0.073	Q2	0.158		Range	0.778
13	0.832	Q3	0.358		Minimum	0.073
14	0.517				Maximum	0.851
15	0.851				Sum	5.166
16	0.269				Count	19
17	0.433					
18	0.141					
19	0.135					
20	0.175					
21						

B. $z = \frac{0.851-0.272}{0.237} = 2.44$. The largest value is almost two and a half standard deviations above the mean, and appears to be an outlier.

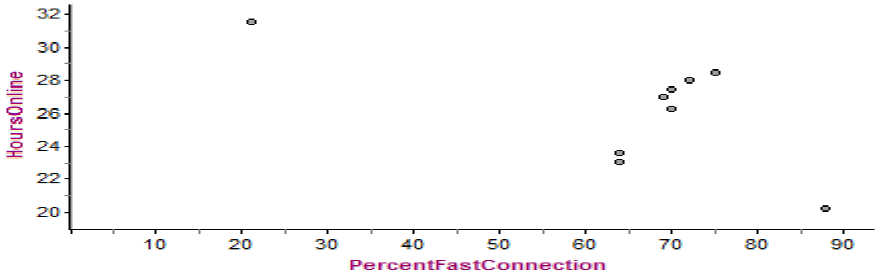
C. Using technology, we see that: Five number summary = (0.073, 0.118,0.158, 0.358, 0.851) and its boxplot:



D. The range is 0.851−0.073 = 0.778 and the interquartile range is IQR= 0.358−0.118 = 0.240.
E. The data is heavily skewed and there appear to be some large outliers. It is most appropriate to use the five number summary.
F. No, it is not appropriate to use that rule with this distribution. That rule is useful when data is symmetric and bell-shaped.

20. Comparing Global Internet Connections

A. A positive relationship implies that as connection speed goes up, time online goes up. This might make sense because being online is more enjoyable with a fast connection speed, so people may spend more time online.
B. A negative relationship implies that as connection speed goes up, time online goes down. This might make sense because if connection speed is fast, people can accomplish what they need to accomplish online in a shorter amount of time so they spend less time online waiting.
C. See the scatterplot below. These two variables have a negative association. There are two outliers. One, in the top left corner, corresponds to Brazil, which has a low percent with a fast connection and a high number of hours online. A second, in the bottom right, corresponds to Switzerland, which has a high percent fast connection and low hours online.
D. If we ignore the two outliers, the variables appear to have a positive relationship.
E. The correlation for this sample of countries is $r = -0.649$. The correlation is pretty strong and negative, so it is being heavily influenced by the two outliers.
F. No! This data comes from an observational study, so we cannot conclude that there is a causal association.



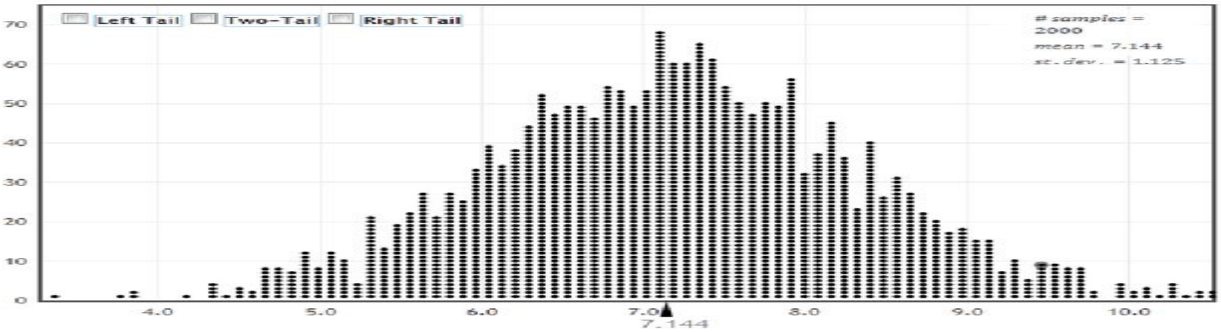
21. **St. Louis vs Atlanta Commute Times**

A. We see that both cities have a significant number of outliers, with very long commute times. The quartiles and median are all bigger for Atlanta than for St. Louis, so we expect that the mean commute time is larger for Atlanta.

B. We are estimating the difference between the cities in mean commute time for all commuters, $\mu_{atl} - \mu_{stl}$. We get a point estimate for the difference in mean commute times between the two cities with the difference in the sample means, $\overline{x}_{atl} - \overline{x}_{stl} = 29.11 - 21.97 = 7.14$ minutes.

C. Since the two samples were taken independently in different cities, for each bootstrap statistic we take 500 Atlanta times with replacement from the original Atlanta data and 500 St. Louis times with replacement from the original St. Louis sample, compute the mean within each sample, and take the difference. This constitutes one bootstrap statistic.

D. A bootstrap distribution for the difference in means with 2000 bootstrap samples is shown in the figure



The standard error for$\overline{x}_{atl} - \overline{x}_{stl}$, found in the upper corner of the figure, is SE=1.125. We find an interval estimate for the difference in the population means with

$$7.14 \pm 2 \cdot 1.125 = 7.14 \pm 2.25 = (4.89, 9.39)$$

E. From the boxplots given we see that both samples are right skewed and have numerous outliers. If these were smaller samples, we would be hesitant to model the difference in means with a t-distribution. However, with these large samples ($n_1 = n_2$=500) we can go ahead and use the t-distribution to find the interval.

Each sample has 500−1=499 degrees of freedom, so we find the t^* value with an area of 0.025 in the tail beyond it in a t-distribution with 499 degrees of freedom. This value is t^*=1.965 and is very close to the standard normal percentile of z^*=1.96.

Substituting into the formula for a confidence interval for a difference in means we get

$$(\overline{x}_{atl} - \overline{x}_{stl}) \pm t^* \left(\sqrt{\frac{s_{atl}^2}{n_1} + \frac{s_{stl}^2}{n_2}} \right)$$

$$29.11 - 21.97 \pm 1.965 * \left(\sqrt{\frac{20.72^2}{500} + \frac{14.23^2}{500}} \right)$$

$$7.14 \pm 2.21 \;\text{ or }\; 4.93 \;\text{ to } 9.35.$$

So, this is almost similar to that of the bootsrap confidence interval.

F. We are 95% confident that the average commuting time for commuters in Atlanta is somewhere between 4.93 and 9.35 minutes more than the average commuting time for commuters in St. Louis.

22. **Influencing Voters**

A. We define p_c to be the proportion supporting Candidate A after a phone call and p_f to be the proportion supporting Candidate A after a flyer. The hypotheses are:

$$H_0\colon p_c = p_f \quad \text{and} \quad H_a\colon p_c > p_f$$

B. We see that $\hat{p}_c$= 152/250 = 0.608 and $\hat{p}_f$= 145/250 = 0.580. The sample proportions are not equal.

C. The observed statistic D is 0.608−0.580, which is 0.028. An approximate p-value is the number of values of D that are greater or equal to 0.028 and then divided it by a 100. There are 32 dots to the right of 0.028 in the plot, so the p-value is 32/100 = 0.32.

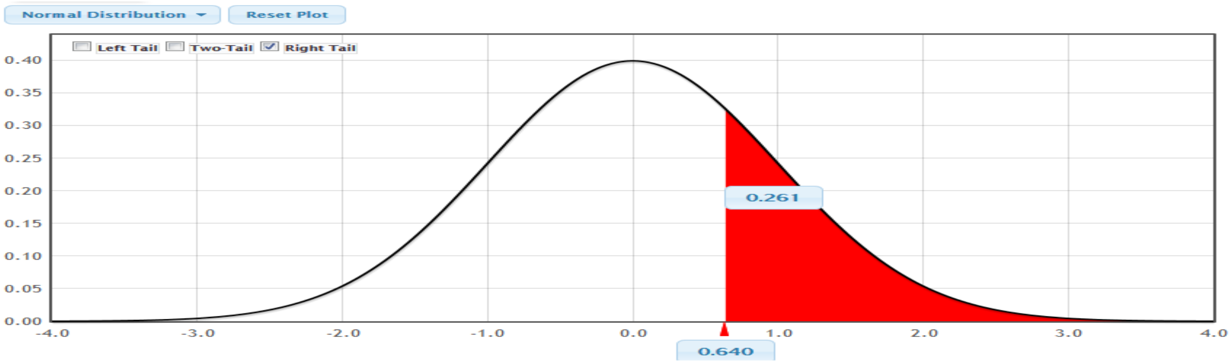
D. Since both samples are sufficiently large(10 success and 10 failures in each group), it is appropriate to use z-distribution to compute the p-value:

$\hat{p} = \frac{152+145}{250+250} = 0.594$ and SE= $\sqrt{0.594(1-0.594)\left(\frac{1}{250}+\frac{1}{250}\right)} = 0.044$ and therefore: $z = \frac{(0.608-0.580)-0}{0.044} = 0.64.$

For a one tail-alternative, the p-value is the area in the standard normal tail the right of 0.64. So we have p-value is 0.261(see the figure below).

E. Since the p-value is greater than 0.05, we do not reject H_0 and conclude that phone calls are not more effective at generating support for the candidate than flyers.

F. Since the decision in in question E is not to reject H_0, the error we might be making is a Type II error, which means we conclude that phone calls are not more effective when actually they are.



☐ Left Tail ☐ Two-Tail ☑ Right Tail

23. **Colonoscopy, Anyone?**

A. $H_0\colon p_c = 0.01$ and $H_a\colon p_c < 0.01.$

B. $\hat{p}_c = \frac{12}{2602} = 0.0046$

C. p-value=$\frac{5}{1000} = 0.005$

D. Yes, the p-value is smaller than 0.05.

24. **Drink Tea for a Stronger Immune System**

A. This is an experiment since the subjects are randomly assigned to drink tea or coffee.
B. This is a difference in means test with separate samples.
C. We are testing Hypotheses: $H_0: \mu_T = \mu_C$ and $H_a: \mu_T > \mu_C$
D. For the tea drinkers, we find that $\bar{x}_T$= 34.82 with s_T= 21.1 and n_T = 11. For the coffee drinkers, we have $\bar{x}_C$ = 17.70 with s_C= 16.7 and n_C= 10. The appropriate statistic is $\bar{x}_T - \bar{x}_C$ and the null parameter is zero, so we have:

$$t = \frac{\text{sample statistic} - \text{null parametr}}{\text{SE}}$$

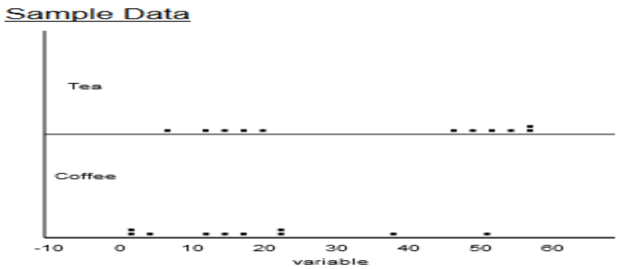
$$t = \frac{(\bar{x}_T - \bar{x}_C) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

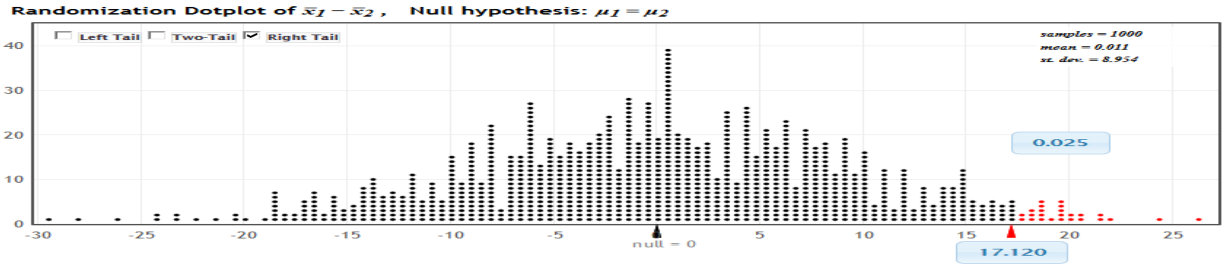
$$\frac{(34.82 - 17.70) - 0}{\sqrt{\frac{21.1^2}{11} + \frac{16.7^2}{10}}} = 2.07$$

The smaller sample size is 10, so we find the p-value using a t-distribution with df = 9. This is an upper tail test, so we find that the p-value is 0.0342. At a 5% level, we conclude that there is evidence that mean production of this disease fighting molecule is enhanced by drinking tea.

E. The sample sizes are small so it is hard to tell whether the data are normal or not. Dotplots for each group are shown below. Notice that most of the values are in the "tails" with few dots in the middle. This may cast some doubt on the normality condition. To be on the safe side, a randomization test might be more appropriate.

To create a randomization statistic we scramble the tea/coffee assignments (so they aren't related to the interferon gamma production values) and find the difference in means between the two groups. Repeating this 1000 times produces a randomization distribution of means such as the one below.





In this distribution, 25 of the 1000 randomizations produced a difference in means that was larger than the difference of 17.12 that was observed in the original sample. This gives a p-value of 0.025 which is fairly small, giving evidence that mean interferon gamma production is higher when drinking tea than when drinking coffee. Note that the p-value and strength of evidence are similar to what we found in part C with the t-distribution.

25. Gender and Award Preference

A. The hypotheses for testing an association between these two categorical variables are

H_0 : Award preference is not related to Gender
H_a : Award preference is related to Gender

Expected count:

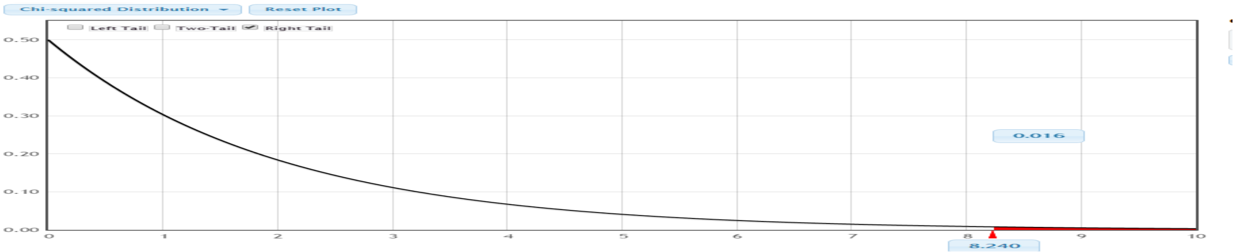
The table below shows the observed and expected counts for each cell. For example, the expected count for the (Female, Academy) cell is 31 · 169/362 = 14.5.

Academy	Academy	Nobel	Olympic	Total
Female	20(14.5)	76(69.6)	73(85.0)	169
Male	11(16.5)	73(79.4)	109(97.0)	193
Total	31	149	182	362

B. Yes.

C. df=2

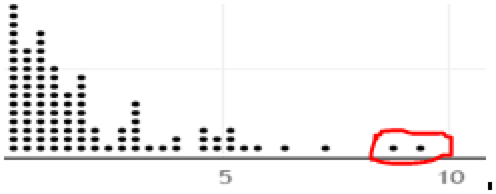
D. 0.016 (See the figure below)



E. If award preference is not related to Gender, the chance of seeing a sample chi-squared ($\chi^2(2)$) as high (or higher) than was observed in the sample by random chance is 0.016.

F. Yes!

G.



H. This is a fairly small p-value, less than 5%, so we have fairly strong evidence that the award preferences tend to differ between male and female students.

I. A Type I error would have occurred if we had rejected the null hypothesis, concluding that award preference were related to Gender, when they actually were not related. A Type II error would have occurred if we had failed to reject the null hypothesis, concluding that award preference were not related to Gender, when they actually were related.

J. If we made an error, it was a Type I error since we rejected the null hypothesis.