STAT 218 – Final Exam

Dr. Allison Theobold

December 3, 2020

Name: _____

Instructor (circle one): Dr. Robinson Dr. Theobold

Section (circle one): 05 (MW 2pm) 06 (MW 4pm) 07 (TR 9am)

Read and Sign the Following Statement:

I understand that give or receiving help on this exam is a violation of academic regulations and is punishable by a grade of **F** in this course. This includes looking at other students' exams and / or allowing other students, actively or passively, to see answers on my exam. This also includes revealing, actively or passively, any information about the exam to any member of Professor Robinson or Professor Theobold's STAT 218 class who has not yet taken the exam. The use of cell phones is strictly prohibited.

I pledge not to do any of these things.

Signed: _____

Instructions.

- Read and sign the honesty pledge at the top of this page. Your exam will not be graded unless the honesty pledge is signed!
- You may use a calculator. You **may not** use your phone or any device that connects to the internet as a calculator.
- Show all work as clearly as possible. Point totals are shown in brackets next to each part. Formulas without values entered do not count as work.
- All answers should be reported in decimal form, rounded to three decimal places.
- For multiple choice and multi-select problems, completely fill in the provided circle (multiple choice) or square (multi-select) for your desired answer choice(s). If you change an answer, be sure to completely erase your initial selection.
- You have 2 hours and 50 minutes to complete this exam, which is ample time! If you get stuck on a problem, take a deep breath, say something positive about yourself, and write down what you know.

Golden Ticket

Scenario	One Categorical Response	Two Categorical Variables	One Quantitative Response	Two Quantitative Variables	Quant. Response and Categ. Explanatory
Type of plot	Bar plot	Dodged Bar plot, Stacked Bar plot, Filled Bar plot	Dot plot, Histogram, Boxplot	Scatterplot	Faceted Histograms, Side-by-side Boxplots
Summary measure	Proportion	Deviation between Observed Counts and Expected Counts (X^2)	Mean or Mean of Differences	Slope or Correlation	Difference in Means
Parameter notation	π	π_1, π_2, …, π_k	μ or μ_{diff}	Slope: β_1; Correlation: ρ	$\mu_1 - \mu_2$
Statistic notation	$\widehat{p}$	$\widehat{p}_1$, $\widehat{p}_2$, …, $\widehat{p}_k$	$\bar{x}$ or $\bar{x}_{\text{diff}}$	Slope: b_1; Correlation: r	$\bar{x}_1 - \bar{x}_2$
Statistical Method(s)	χ^2 Goodness of fit Test	χ^2 Test of Independence, χ^2 Test of Homogeneity, Permutation Test for X^2	t-test for One Mean, t-test for Paired Differences, Bootstrap Confidence Interval for One Mean	t-test for β_1, Permutation Test for β_1, Bootstrap Confidence Interval for β_1	t-test for $\mu_1 - \mu_2$, Permutation Test for $\mu_1 - \mu_2$, Bootstrap Confidence Interval for $\mu_1 - \mu_2$

Provided Formulas

$$IQR = Q3 - Q1$$

1.5 IQR Rule: above $Q3 + (1.5 \times IQR)$ or below $Q1 - (1.5 \times IQR)$

$$\hat{y} = b_0 + b_1 \times$$

$$\text{Residual} = y - \hat{y}$$

$$R^2 = r^2$$

general formula for a confidence interval: point estimate $\pm$ multiplier $\times$ SE(point estimate)

t-based confidence interval: $\bar{x} \pm t^*_{df} \times SE(\bar{x})$

$$SE(\bar{x}) = \frac{s}{\sqrt{n}}$$

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$F = \frac{\text{MSG}}{\text{MSE}}$$

$$\alpha^* = \frac{\alpha}{\#\text{ of comparisons}}$$

Expected Counts for One Categorical Variable

$$\text{Expected Count} = \text{total sample size } \times \text{null proportion for group k}$$

Expected Counts for Two Categorical Variables

$$\text{Expected Count} = \frac{(\text{row i total}) \times (\text{column j total})}{\text{total sample size}}$$

Chi-Square Test Statistic

$$X^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

Midterm 1 Question Bank

Q8 [2 pts] When you change from a 90% to a 95% confidence interval, which part(s) of the confidence interval change? (Select all that apply)

(a) Point Estimate (midpoint)

(b) Multiplier

(c) Standard error

Q5 [4 pts] Indicate whether each statement about a bootstrap resample is **TRUE** or **FALSE**.

(a) The bootstrap resample and original sample **must** be the same size. _____

(b) The bootstrap resample and original sample are **both** taken directly from the population.

_____

(c) The bootstrap resample can **only** use values that were in the original sample. _____

(d) The bootstrap resample uses **all** of the values that were in the original sample. _____

Q1 [12 pts] Dr. John Arbuthnot, an 18th century physician, writer, and mathematician is famous for performing the first hypothesis test of significance. Dr. Arbuthnot was interested in the ratio of newborn males to newborn females, so he gathered the baptism records for children born in London for every year from 1629 to 1710. Artbuthnot found that in every year, the number of males born in London exceeded the number of females.

(a) [2 pts] Describe the sampling method used by Dr. Arbuthnot.

(b) [2 pts] Describe how this sampling method could be biased.

(c) [3 pts] A preview of the dataset is provided below. Use this preview to address the following questions.

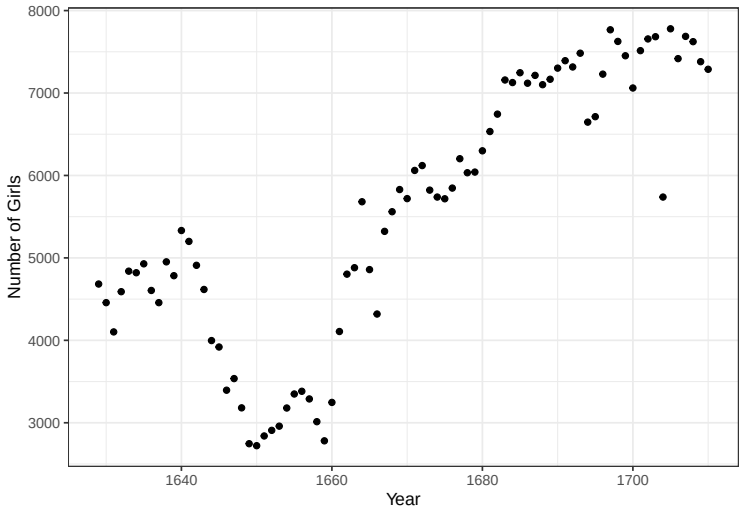
```
## # A tibble: 82 x 3
##     year  boys girls
##    <int> <int> <int>
##  1  1629  5218  4683
##  2  1630  4858  4457
##  3  1631  4422  4102
##  4  1632  4994  4590
##  5  1633  5158  4839
##  6  1634  5035  4820
##  7  1635  5106  4928
##  8  1636  4917  4605
##  9  1637  4703  4457
## 10  1638  5359  4952
## # ... with 72 more rows
```

- Identify the observational units in the data set.

- List the variables. Indicate whether each variable is categorical (c) or quantitative (q).

- What would the dimensions of the data set be? (number of rows by number of columns)

(d) [3 pts] A scatterplot displaying the number of girls born over time is displayed below. Describe the relationship you see in the scatterplot. Be sure to address the form, direction, strength, and outliers present.



(e) [2 pts] Would it be appropriate to model the relationship between the number of girls born and the year with a linear regression? Justify your belief!

Q4 [4 pts] Researchers are interested in the fish that reside in the Caspian Sea. They have plans to collect many fish and take multiple measurements on each. Match each statistical description on the right with each piece of information given. Put the letter of the statistical description in the blanks on the left.

_____ circumference of the fish	(a) quantitative variable
_____ species of the fish	(b) categorical variable
_____ average length of all fish in the area of consideration	(c) parameter: μ
_____ mean internal temperature of the fish collected in the sample	(d) statistic: $\bar{x}$
_____ one of the fish in the area of consideration	(e) observational unit
_____ method of only studying the fish caught in the net 3pm on Wednesday of the research time frame	(f) cluster sampling method
_____ method of selecting 5% of each species, known to be in the area of consideration, for the sample	(g) stratified sampling method
____ method of dividing up the whole location with netting and sampling 10 random netted areas	(h) convenience sampling method

Q6 [3 pts] The purpose of creating a null distribution is to: (Select all that apply)

(a) Discover what statistics might have occurred if the null hypothesis was true.

(b) To determine if the null hypothesis is true.

(c) To determine if the observed statistic is unlikely if the null was true.

Q2 [22 pts] I collected data on 512 different fast food items from Mcdonalds, Chick-Fil-A, Sonic, Arby's, Burger King, Dairy Queen, Subway, and Taco Bell. To obtain these data, for every restaurant I randomly sampled 64 items from their entire menu and recorded the nutritional content of each item selected (e.g., calories, saturated fat, calcium, protein, etc.).

(a) [2 pts] Describe the sampling method I used to obtain these 512 fast food items.

(b) [3 pts] I am interested in studying the linear relationship between the total calories of a food item and the amount of saturated fat that item contains.

Write the null hypothesis for my question of interest, using both words and notation.

(c) [2 pts] Is the alternative hypothesis one- or two-sided? Select one.

- One-sided
- Two-sided

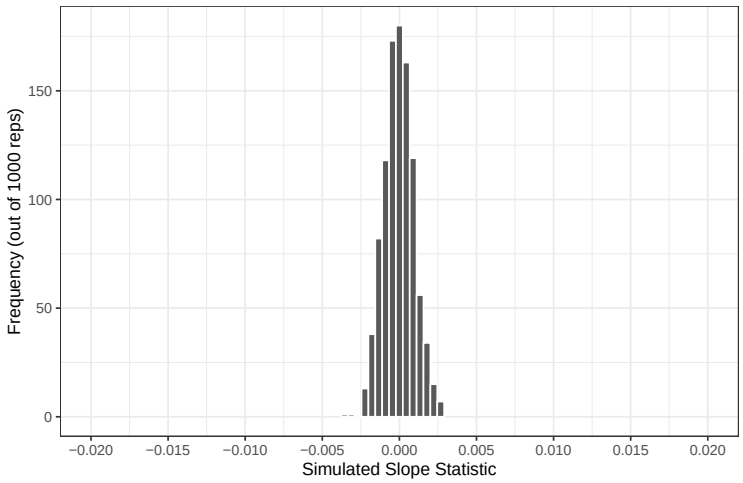
(d) [5 pts] On the following page is the plot of the simulated null distribution from `R`. Fill in the blanks below with one answer in each set of parentheses to correctly explain how one sample on the null distribution would be created. Blanks preceded by (#) should be filled in with a number, all other blanks should be filled in with the context of the study.

On (#) _____ cards, write _____ and _____ on the cards.

Assume the null hypothesis is true and _____.

Generate a new sample of 512 ordered pairs by _____.

Calculate and plot the _____ from each simulated sample.



(e) [2 pts] Using the regression output below, draw a vertical line where the observed statistic falls on the null distribution.

term	estimate	std_error
intercept	-0.771	0.406
calories	0.017	0.001

(f) [2 pts] Shade the location of the plot you would use to calculate the p-value.

(g) [1 pts] Estimate the p-value associated with this hypothesis test.

(h) [3 pts] Which of the following is a correct interpretation of the p-value obtained? (Circle one)

- If there is no linear relationship between the total calories and the saturated fat of a fast food item, we would observe a sample slope of 0.017 or more extreme with a probability of less than 1 in 1000.
- If there is a linear relationship between the total calories and the saturated fat of a fast food, we would observe a sample slope of 0.017 or more extreme with a probability of less than 1 in 1000.
- The probability of seeing a sample slope between the total calories and the saturated fat of a fast food item of 0.017 or more extreme is less than 0.1
- The probability that there is no linear relationship between the total calories and the saturated fat of a fast food item, is less than 0.1

(i) [2 pts] Given the p-value for the hypothesis test, would the 95% confidence interval for β_1 contain 0? Be sure to justify your choice!

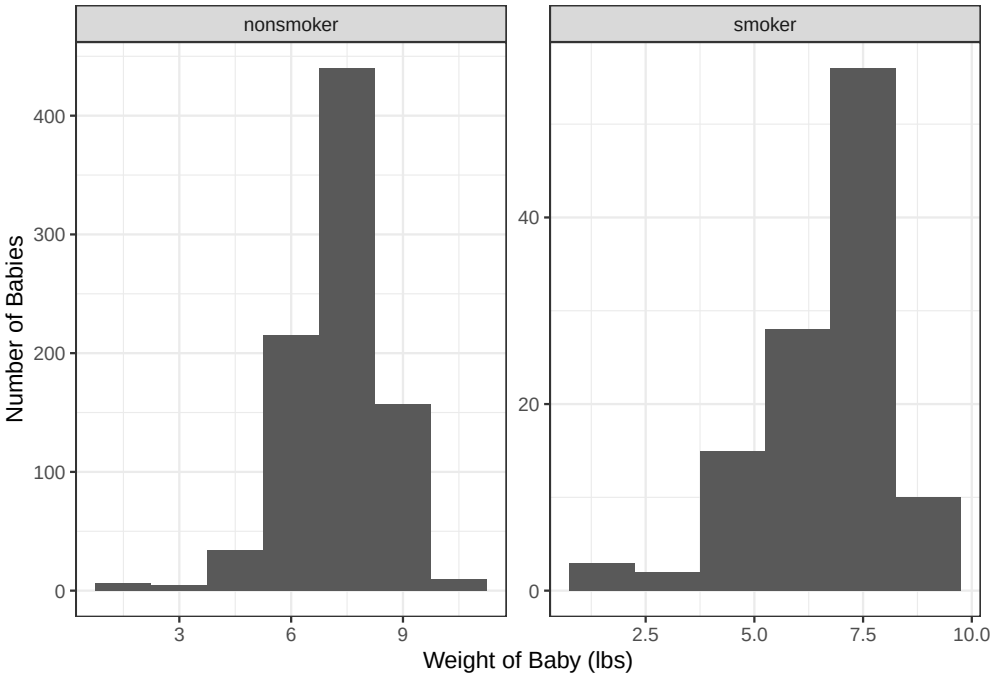
Midterm 2 Question Bank

Q3 [24 pts] In 2004, the state of North Carolina released to the public a large dataset containing information on births recorded in this state. This data set has been of interest to medical researchers who are studying the relation between habits and practices of expectant mothers and the birth of their children. This analysis will focus on a random sample of 981 observations from the published dataset.

(a) [3 pts] Hospital administration at Duke University Hospital are interested in the difference in the mean baby birth weight between mothers who do not smoke and mothers who do smoke. Using the table below, report the observed statistic for this comparison. Indicate in your answer what notation should be used for this statistic.

$$\underbrace{\qquad\qquad\qquad\qquad}_{\text{(notation)}} = \underbrace{\qquad\qquad}_{\text{(value)}}$$

habit	min	Q1	median	Q3	max	mean	sd	n	missing
nonsmoker	1.19	6.57	7.35	8.06	10.42	7.27	1.233	867	0
smoker	0.75	5.953	7.03	7.805	9.25	6.677	1.597	114	0



(b) [4 pts] These hospital administrators are interested in estimating the true difference in mean birth weight between mothers who do not smoke and mothers who do smoke. The administrators learned in their Statistics class how to obtain a confidence interval for a difference in means using a t-distribution. Using the plots above and your knowledge of how the data were collected evaluate whether it would be appropriate for the administrators to use a t-distribution to obtain a confidence interval for the true difference in means.

(c) [5 pts] The administrators contacted the Department of Statistics at Duke and requested a consultation. The Statistician they spoke with suggested they use bootstrapping instead of a t-distribution to obtain a confidence interval. Fill in the blanks below to explain to the administrators how one bootstrap (re)sample is found.

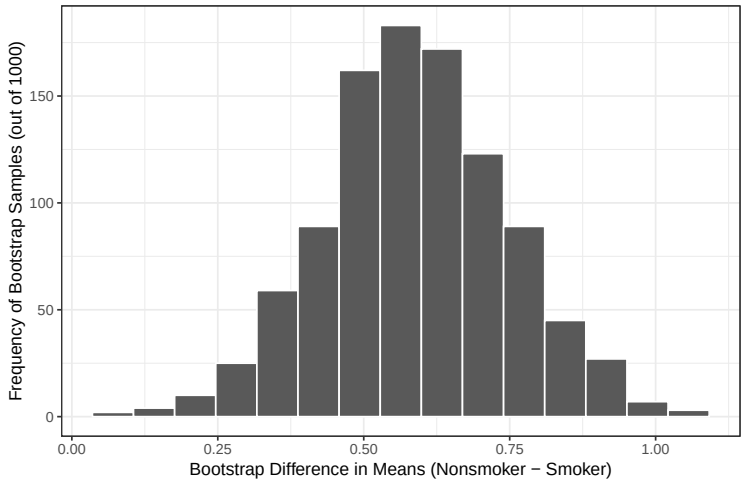
On (#) _____ cards, write _____ and _____ on the cards.

Generate a new sample by:

Calculate and plot the _____ from each bootstrap (re)sample.

(d) [3 pts] The Statistician carried out a simulation with 1000 bootstrap resamples using the order (`nonsmoker` - `smoker`) and obtained the bootstrap distribution below.

Where is the bootstrap distribution centered? Why is the distribution centered at that value?



(e) [3 pts] The table below presents percentiles for the bootstrap distribution shown above. Circle the two values which will construct a 99% confidence interval.

Quantile	Value
0.5%	0.1587
1%	0.2132
2.5%	0.2801
5%	0.3335
90%	0.7916
95%	0.8649
97.5%	0.9019
99%	0.9477
99.5%	1.004

(f) [4 pts] Interpret the 99% confidence interval found in part (f) **in the context of this study**.

(g) [2 pts] Based on your confidence interval in (f), which of the following is the most likely p-value for a two-sided hypothesis test? Circle one.

(i) 0.20 (ii) 0.10 (iii) 0.05 (iv) 0.01 (v) 0.005

Q4 [21 pts] As you may be aware, many individuals are concerned about the presence of BPA in plastics, especially plastics that make contact with food and drinks. Currently, these is an incomplete understanding of how exposure to BPA affects our ingestion. Last year Dr. Hagobian in the Kinesiology and Public Health carried out a study to investigate the role of Bisphenol A (BPA) in metabolism and endocrine disruption.

Dr. Hahobian recruited 11 subjects, each of whom ate two types of cookies on two separate visits, one visit in December and the second in February. On one visit they ate the BPA-laced cookie and, on a different day, a placebo cookie (with no BPA). Thirty minutes after eating the cookie on each occasion, they were given a glucose tolerance test to measure their glucose metabolism.

A summary of the glucose test results (mmol/L) after eating each type of cookie as well as the difference in glucose results for each subject is shown below.

Cookie	Mean	Standard Deviation	n
Placebo	5.259	0.762	11
BPA	5.355	1.462	11
Difference: Placebo - BPA	-0.095	1.153	11

(a) [4 pts] For simplicity, Dr. Hagobian could have given all subjects the BPA cookie on their first visit in December, and the Placebo cookie on the second visit in February, but he didn't. Instead, when a subject came for their first visit, he flipped a coin. If it was heads, they received BPA on that visit (and Placebo on their second visit). If it was tails they received the Placebo cookie first. Why did he add this extra coin flipping step instead of the simpler approach of just giving everyone one type of cookie in December and the other type in February?

(b) [2 pts] Dr. Hagobian is interested in testing whether BPA causes a shift in glucose levels. Which analysis would be more appropriate? Circle one.

Difference in Two Independent Means **Mean of the Paired Differences**

(c) [3 pts] Based on your answer to (b), write out the null and alternative hypotheses for Dr. Hagobian's test using **notation**. *Be sure to indicate the order of subtraction being used!*

H_0**:**

H_A**:**

(d) [3 pts] To perform the analyses you selected in (b), what conditions does Dr. Hagobian need to check before obtaining a p-value? Circle all that apply.

(i) Independence of the differences

(ii) Independence of the observations within each group

(iii) Independence of the observations between the groups

(iv) Independence of the variables

(v) Equal variance between the groups

(vi) Linear relationship between the variables

(vii) Normality of the differences

(viii) Normality of the observations within each group

(e) [3 pts] Using `R`, Dr. Hagobian obtained the following table.

statistic	p_value	estimate	lower_ci	upper_ci
1.096	0.299	0.417	-0.2729	1.107

Which of the following would be the best overall conclusion in the context of Dr. Hagobian's study? Your selection should reflect the hypotheses you wrote in part (c)!

(i) With such a large p-value, we have significant evidence to reject the null hypothesis. We conclude the true mean of the differences in glucose between eating a BPA cracker and a Placebo cracker is not 0.

(ii) With such a large p-value, we have insufficient evidence to reject the null hypothesis. We conclude the true mean of the differences in glucose between eating a BPA cracker and a Placebo cracker is 0.

(iii) With such a large p-value, we have insufficient evidence to reject the null hypothesis. We do not have evidence to suggest the mean of the differences in glucose between eating a BPA cracker and a Placebo cracker is different from 0.

(vi) With such a large p-value, we have significant evidence to reject the null hypothesis. We conclude the true mean glucose after eating a BPA cracker is different from the true mean glucose after eating a Placebo cracker.

(v) With such a large p-value, we have insufficient evidence to reject the null hypothesis. We conclude there is no difference in the true mean glucose after eating a BPA cracker and the true mean glucose after eating a Placebo cracker.

(vi) With such a large p-value, we have insufficient evidence to reject the null hypothesis. We do not have evidence to suggest the true mean glucose after eating a BPA cracker is different from the true mean glucose after eating a Placebo cracker.

(f) [2 pts] Based on the decision you reached in (e), what type of error could you have made? Circle one.

Type I Error **Type II Error** **No error was made**

(g) [2 pts] If instead Dr. Hagobian had 100 subjects, the chance of the error described in part (f) would

increase **decrease** **stay the same**

(h) [2 pts] In a different study, Dr. Hagobian obtained a p-value of 0.0425 and a 95% confidence interval of (-1.129, 0.0437). Which of the following statements about these findings is true? Circle one.

(i) The results of the hypothesis test and the confidence interval tend to agree with each other at the 5% significance level. Four percent of the time we would obtain a statistic like the one we saw somewhere in the interval of -1.129 mmol/L to 0.0437 mmol/L.

(ii) The results of the hypothesis test and the confidence interval are conflicting at the 5% significance level. With a p-value of 0.0425 we have evidence to reject the null hypothesis, which would mean that our confidence interval would not contain 0.

(iii) The results of the hypothesis test and the confidence interval are conflicting at the 5% significance level. There's a 95of 0.0425 would be in the interval (-1.129, 0.0437).

(vi) The results of the hypothesis test and the confidence interval seem to agree with one another at the 5% significance level. With a p-value of 0.0425 we do not have evidence to reject the null hypothesis, thus indicating that 0 should be in our interval.

Final Exam Question Bank

Q1 [17 pts] Wild mushrooms, such as chanterelles or morels, are delicious, but eating wild mushrooms carries the risk of accidental poisoning. Even a single bite of the wrong mushroom can be enough to cause fatal poisoning. An amateur mushroom hunter is interested in finding an easy rule to differentiate poisonous and edible mushrooms. They think that the mushroom's gills (the part which holds and releases spores) might be related to a mushroom's edibility. They used a data set of 8124 mushrooms and their descriptions. For each mushroom, the data set includes whether it is edible or poisonous and the spacing of the gills (Broad or Narrow).

Please Note: According to The Audubon Society Field Guide to North American Mushrooms, there is no simple rule for determining the edibility of a mushroom; no rule like "leaflets three, leave them be" for Poisonous Oak and Ivy.

Class	Broad	Narrow	Total
Edible	3920	288	4208
Poisonous	1692	2224	3916
Total	5612	2512	8124

(a) [4 pts] Fill in each blank with one of the options in parentheses to best describe the variables collected.

Whether the mushroom is edible or poisonous is the (explanatory / response) _____ and it is (categorical / quantitative) _____.

Gill size (Broad or Narrow) is the (explanatory / response) _____ and it is (categorical / quantitative) _____.

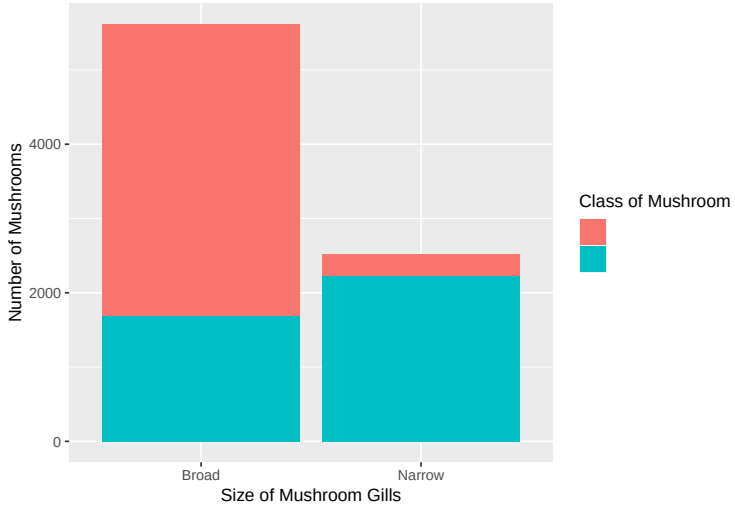
(b) [3 pts] Calculate the proportion of mushrooms with a broad gill size that are poisonous. *Leave your value in* ***unreduced*** *fraction form.*

$$\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad}_{\text{(notation)}} = \underbrace{\qquad\qquad\qquad}_{\text{(value)}}$$

(c) [3 pts] Calculate the proportion of mushrooms with a narrow gill size that are poisonous. *Leave your value in* ***unreduced*** *fraction form.*

$$\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad}_{\text{(notation)}} = \underbrace{\qquad\qquad\qquad}_{\text{(value)}}$$

(d) [2 pts] Using your answers to (b) and (c), fill in the correct names next to each color, to label the bar chart showing the relationship between gill size (broad or narrow) and whether the mushroom is edible.



(e) [3 pts] Based on the plot, describe the relationship between a mushrooms gill size and whether it is edible or not.

(f) [2 pts] Suppose the Chi-Squared test resulted in a "significant" p-value. Which of the following would be the correct scope of inference for this study?

(i) It can be inferred for all mushrooms that gill size causes a mushroom to be poisonous.

(ii) It can be inferred for all mushrooms that gill size is associated with whether a mushroom is poisonous.

(iii) It can be inferred for this sample of mushrooms that gill size causes a mushroom to be poisonous.

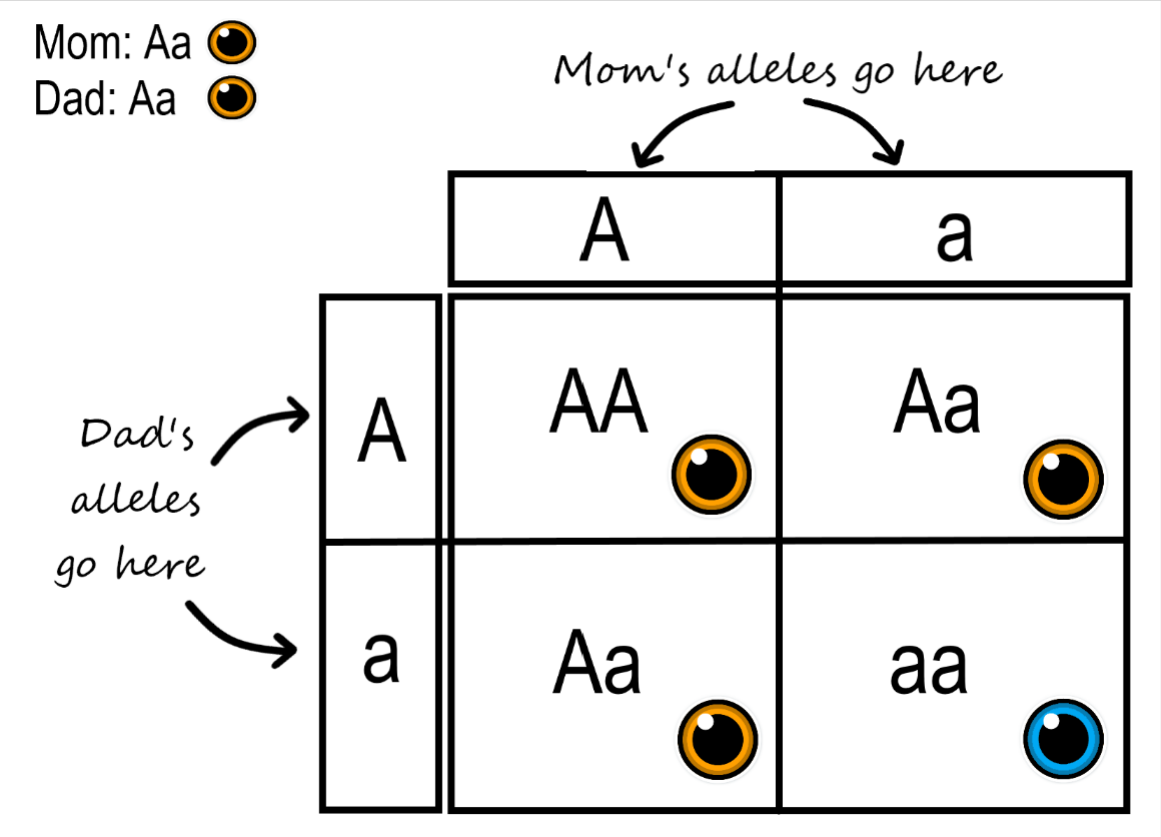
(iv) It can be inferred for this sample of mushrooms that gill size is associated with whether a mushroom is poisonous.

Q2 [21 pts]

Mendelian inheritance refers to certain patterns of how traits are passed from parents to offspring. These general patterns were established by the Austrian monk Gregor Mendel, who performed thousands of experiments with pea plants in the 19th century. Mendel's discoveries of how traits (such as color and shape) are passed down from one generation to the next introduced the concept of dominant and recessive modes of inheritance.

Mendelian inheritance refers to the inheritance of traits controlled by a single gene with two alleles, one of which may be completely dominant to the other. You can use a Punnett square to determine the expected ratios of possible genotypes in the offspring of two parents.

In the table below, we see an example of eye color inheritance. In this case, both parents are heterozygotes (Aa) for the gene. Half of the gametes produced by each parent will have the A allele, and half will have the a allele, shown on the side and the top of the Punnett square. Filling in the cells of the Punnett square gives the possible genotypes of their children. It also shows the most likely ratios of the genotypes, which in this case is 25% AA, 50% Aa, and 25% aa.



(a) [2 pts] When Mendel crossed his pea plants, he learned that tall (T) was dominant to short (t). Suppose in your Biology course you carried out an experiment to test if the plot offspring would follow Mendelian inheritance.

Fill in the cells of Punnett square to give the possible genotypes for plant tallness.

	T	t
T		
t		

(b) [3 pts] If the Mendelian inheritance is true, what proportions would you expect for each of the following genotypes? Insert the corresponding values in each cell.

TT	Tt	tt
$\pi_{TT} =$	$\pi_{Tt} =$	$\pi_{tt} =$

(c) [2 pts] Actually, our table could be a bit simpler. Both the TT and Tt genotypes will present as "tall" plants, whereas tt genotypes will present as "short" plants.

Compress your previous table into a new table with only two levels of tallness.

Tall	Short
$\pi_{Tall} =$	$\pi_{Short} =$

(d) [3 pts] If the table above represents what Mendelian inheritance assumes to be true about tallness under H_0, state the alternative hypothesis using words.

(e) [2 point] After you cross your plants, you measure the characteristics of the 400 offspring. You note that there are 305 tall pea plants and 95 short pea plants.

Fill in the table summarizing these observed counts.

_____	_____	Total
		400

(f) [2 pts] Fill in the table below, summarizing the expected counts for these 400 plants.

_____	_____	Total
		400

(g) [4 pts] Calculate how far "off" was your observed number of tall and short plants were from what you expected if H_0 was true. Use these values to report the X^2 statistic for your experiment.

Tall:

Short:

X^2 **statistic**:

(h) [3 pts] The p-value associated with your X^2 statistic is 0.5645424. Your Biology textbook suggests you interpret this value as:

> *The large p-value proves that Mendelian inheritance is true.*

What issue(s) to you have with this interpretation?

Q5 [11 pts] The effect of guilt on how a decision maker focuses on a problem was investigated in the *Journal of Behavioral Decision Making* (January 2007). A total of 155 volunteer students participated in the study, where each was randomly assigned to one of three emotional states (guilt, anger, or neutral) through a reading / writing task. Immediately after the task, the students were presented with a decision problem (e.g., whether or not to spend money on repairing a very old car). The researchers found that a higher proportion of students in the guilty-state group chose not to repair the car than those in the neutral-state and anger-state groups.

(a) [3 pts] What is the study design? Select one.

(i) Observational study. The researchers did not take a random sample of students.

(ii) Observational study. There is no random assignment of students to emotional state.

(iii) Experiment. The students are a representative sample of all students.

(iv) Experiment. The students were randomly assigned to emotional state.

(b) [3 pts] Which types of sampling bias may be present in this study? Select all that are present, or if you believe there is no bias present, select option (iv) No bias.

(i) Selection bias

(ii) Non-response bias

(iii) Response bias

(iv) No bias

(c) [3 pts] The researchers found a difference in the proportion of students in the guilty-state group chose not to repair the car than those in the neutral-state and anger-state groups. Can we conclude that the emotional state group caused a higher proportion of students to not repair the car? Select one.

(i) Yes, because these data are from a representative sample.

(ii) No, because the students are volunteers.

(iii) Yes, because the researchers evened out confounding variables across emotional state group by randomly assigning emotional state.

(iv) No, because the sample size is not large enough.

(d) [2 pts] Which type of plot would be the **most** appropriate to display the relationship between each level of emotional state and whether students choose to repair the car? Select one.

(i) Scatterplot

(ii) Filled bar plot

(iii) Pie chart

(iv) Side-by-side boxplot

Q7 [2 pts] Suppose you reject the null hypothesis at the 0.05 level of significance. A colleague had planned to use a 0.01 level of significance instead. Will your colleague also reject the null hypothesis?

(a) Yes

(b) No

(c) Maybe

(d) Changing a level of significance cannot affect decisions.

Q9 [12 pts] For each of the following, select the single most appropriate analysis for the situation described. You may use an analysis for more than one situation. (2 pts each)

Chi-Square Test of Independence	One-Way ANOVA
Simple Linear Regression	Chi-Square Test of Homogeneity
Chi-Squared Goodness-of-Fit test	Paired t-test
Confidence interval for μ	Hypothesis test for μ
Hypothesis test for $\mu_1 - \mu_2$	

(a) Researchers are interested in investigating how the number of visitors to Yellowstone National Park in a year impacts the local economy in Livingston. To do this they count the number of yearly visitors to Yellowstone and measure the dollars spent by tourists in Livingston for the year.

_____

(b) A study of honeybees looked at whether the species of honeybees varied by state. Ten states were used in the study, and 100 honeybees were randomly sampled in each state, and 7 different species were seen in the data set.

_____

(c) An attorney in Boston observes that some judges seem to select juries that contain few women. She collects data on 20 randomly selected juries from each of 10 judges, and the number of women on each jury for each judge.

_____

(d) Researchers are interested in determining if the yield of a tomato plant differs among three tomato varieties.

_____

(e) You are interested in deciding if you should rent a new apartment off campus. As this will be your first time living off campus, you are anxious to know the average amount of time it should take you to walk to campus. What is the best **method** to estimate the average time it will take you to walk to campus?

_____

(f) Matchmaking data scientists are always investigating what characteristics of a person can produce better matches. Data scientists at Tinder are interested in looking into the relationship between someone's sexual orientation and whether they would date someone who is taller than them.

_____

You are striking! Congratulations on completing STAT 218!

Have a great winter break!