# Final Practice Math 11, UCSD, Winter 2018 (Do not turn in!)

Note: The length of this practice does not reflect the actual length of the exam.

It is meant to cover the variety of topics seen in class in the entire quarter.

### Exercise 1

Census data shows that the average income level in the vicinity of a mall is \$33,950. The owners of the mall are interested in determining whether mall shoppers have a higher income level, and are not interested in any result that shoppers have a lower income. They perform a simple random sample of 50 shoppers, who had an average income of \$34,076 with a standard deviation of \$474 and the observations were not too skewed.

1. Can they conclude that their shoppers have a higher income level at a 5% significance level?

2. What about a 10% significance level, or a 1% significance level?

3. At the 5% significance level, what is the power (probability of rejecting the null hypothesis) of this test against the alternative hypothesis that the actual income of shoppers averages 34,000?

### Exercise 2

In a forest, the non-edible mushrooms are four times as many as edible mushrooms.

1. We randomly pick 20 mushrooms. What is the probability that, among these 20 mushrooms, at most 13 of them are non-edible?

2. We randomly pick 100 mushrooms. What is the probability that, among these 100 mushrooms, we got between 15 and 20 edible mushrooms?

3. How many mush rooms must we pick in order to get a probability of picking at least 40 edible mush rooms larger than 90%?

### Exercise 3

In a second-generation flower culture obtained by breeding white-flower plants (from white-flower lineage) with yellow-flower plants (from yellow-flower lineage), the color of the flowers is either white, yellow, or striped white and yellow.

1. If the coloration of these flower is governed by two alleles (AA white; aa yellow; Aa streaked white and yellow), explain why the expected proportions are  $p_{AA} = 1/4$ ,  $p_{aa} = 1/4$ , and  $p_{Aa} = 1/2$ . (*Hint: a tree may help*)

2. Build a test that would allow us to know whether the coloration of these flower is governed by two alleles (AA white; aa yellow; Aa streaked white and yellow).

3. On an experiment of 1000 flowers in this culture, we find 221 whites, 230 yellows, and 549 streaked white and yellow. What is your conclusion?

Many people believe that gender, weight, drinking habits, and many other factors are much more important in predicting blood alcohol content (BAC) than simply considering the number of drinks a person consumed. Here we examine data from sixteen student volunteers at Ohio State University who each drank a randomly assigned number of cans of beer. These students were evenly divided between men and women, and they differed in weight and drinking habits. Thirty minutes later, a police officer measured their blood alcohol content (BAC) in grams of alcohol per deciliter of blood.23 The scatterplot and regression table summarize the findings.



	Estimate	Std. Error	t value	$\Pr(> t )$
(Intercept)	-0.0127	0.0126	-1.00	0.3320
beers	0.0180	0.0024	7.48	0.0000

1. Describe the relationship between the number of cans of beer and BAC.

2. Write the equation of the regression line. Interpret the slope and intercept in context.

3. Do the data provide strong evidence that drinking more cans of beer is associated with an increase in blood alcohol? State the null and alternative hypotheses, report the *p*-value, and state your conclusion.

4. The correlation coefficient for number of cans of beer and BAC is 0.89. Calculate  $R^2$  and interpret it in context.

5. Suppose we visit a bar, ask people how many drinks they have had, and also take their BAC. Do you think the relationship between number of drinks and BAC would be as strong as the relationship found in the Ohio State study?

#### Exercise 5

Let X be a continuous random variable whose probability density function is

$$f(x) = \begin{cases} c|x|^3, & \text{if } -1 \le x \le 1\\ 0 & \text{otherwise,} \end{cases}$$

where c is a positive constant.

- 1. What value must c be for f(x) to be a valid density function?
- 2. Calculate the probability that X falls between 0 and 1/2.
- 3. Compute the expected value and standard deviation of X.

4. Let  $X_1, X_2, \ldots, X_{100}$  be independent random variables with the density f(x). Calculate (an approximation of)  $P(\bar{X}_{100} < 0.1)$ .

The hourly sales of fried chicken at Big Kahuna Fried Chicken are normally distributed with mean 2000 pieces and standard deviation 500 pieces. From one hour to another, the number of sales is independent.

1. What is the probability that in a 9-hour day more than 20,000 pieces will be sold?

2. What is the minimum number of pieces to sell in an hour to be among the top 5% fruit-ful hours?

#### Exercise 7

TwitterCounter is a data analytics tool that allows you to explore statistics for any public Twitter account. For example, in the last 40 days, President Trump has tweeted a total of 190 times. Assume that he continues at this rate and that he is equally likely to issue a tweet at any time of day (including, during the night).

1. What is the probability that he tweets exactly 6 times on a random day?

2. What is the probability he can go the next 12 hours without issuing a tweet?

3. Suppose you have the computer randomly choose five of his old tweets. What is the probability at least two of the five tweets will have a time stamp between 2:00 PM and 5:00 PM?

### Exercise 8

Suppose IQ scores for individuals are normally distributed with mean 100 and SD 15. You decide you are only interested in dating a "genius" (someone with an IQ of 135 or above). Your plan is to go on dates with random people until you finally meet a genius.

1. How many dates should you expect to go on in order to meet your genius?

2. What is the probability that it will take an even number of dates to finally meet your genius?

3. On one of your dates, your date (correctly) claims to be smarter than you. If your IQ is 120, what is the probability that your date is a "genius"?

"Fake news" is a term that describes any type of news media that is deliberately constructed to mislead people. It includes stories that are completely made up, as well as reporting that selectively promotes certain biased viewpoints. In the May 2017 Harvard-Harris national poll, respondents were asked if they agreed with the claim "There is a lot of fake news in the mainstream media". 498 of 624 Republicans agreed with the claim, while 370 of 696 Democrats agreed with the claim.

1. Find a 70% C.I. for the difference in Republican and Democratic support for this claim.

2. Another prompt in the above study was "Many of my friends intentionally post fake news or false information on social media". You are curious if a majority of the 624 Republicans who took part in the survey agreed with this claim. What is the smallest number of Republicans that would need to agree with the statement in order to get a statistically significant hypothesis test with  $\alpha = 0.10$ ?

### Exercise 10

In the early 1990's, researchers in the UK collected data on traffic accident related emergency room admissions on Friday the 13th and the previous Friday, Friday the 6th. The distributions of these counts from Friday the 6th and Friday the 13th are shown below for six such paired dates along with summary statistics. You may assume that conditions for inference are met.

	6th	13th	Difference
Mean	7.5	10.83	-3.33
SD	3.33	3.6	3.01
n	6	6	6

1. Conduct a hypothesis test to evaluate if there is a difference between the average numbers of traffic accident related emergency room admissions between Friday the 6th and Friday the 13th.

2. Calculate a 95% confidence interval for the difference between the average numbers of traffic accident related emergency room admissions between Friday the 6th and Friday the 13th.

3. The conclusion of the original study states, "Friday 13th is unlucky for some. The risk of hospital admission as a result of a transport accident may be increased by as much as 52%. Staying at home is recommended." Do you agree with this statement? Explain your reasoning.

Suppose a random number generator with a uniform distribution gives real numbers in the interval [0,3]. You have it a produce a random number x, and then you draw a square with vertices (0,0), (x,0), (0,x), and (x,x).

- 1. What is the expected value for the area of your square?
- 2. What is the probability that your square contains the point (2,1)?

### Exercise 12

In the real world, the lifetime of electronics has been shown to frequently follow an exponential distribution. Suppose a certain brand of toaster breaks down, on average, after 10 years.

1. Find  $P(X \ge 4 \text{ years})$ .

2. After how much time will your toaster be among the top 20% of longest-lasting toasters (without a breakdown)?

## Exercise 13

In the 2016 season, the Pittsburgh Penguins (an ice hockey team) scored roughly 3.37 goals per game.

1. Hockey fans tend to have a good time at a game if their team scores at least two goals during that game. What is the probability that a Penguins fan has a good time at a game?

2. favorite hockey team scores g goals per game, on average. If his team is just as likely to score 0 goals as they are to score 2 goals, what must g be?