1. A customer service representative seeking to improve waiting times at a large medical office collected a random sample of 500 patients and determined the time (in minutes) between when they arrived at the office until they were seen by a doctor. The average waiting time from that sample was 26.2 minutes, with a standard deviation of 7.6 minutes. Assume a normal distribution applies.

   a. What is the standard error for this sample (including measurement units)? _____

   b. What is the 90% confidence interval for the average waiting time in the emergency room (including measurement units)? _____

   c. True or false: We cannot be 99% confident that the average waiting time for the 500 sampled patients lies in the confidence interval that you calculated. _____

   d. True or false: The 95% confidence interval for the average waiting time is larger than the 90% confidence interval that you calculated. _____

   e. True or false: We are 90% confident that the average waiting time of all patients at this medical office is within the confidence interval that you calculated. _____

   f. If the sample had used only 250 patients (instead of 500), the margin of error in the sample would have been:

      1) Two times larger, or

      2) Two times smaller, or

      3) Neither of these.

2. In our case study on "Communication of Intent to Do Harm Preceding Mass Public Shootings" we saw the following data as part of Table 2: *Bivariate Results of Factors Associated with Leakage Among Perpetrators of Mass Shootings.*

Characteristic	Number (N = 170)	Perpetrator, number (%)		χ^2
		Did not leak (n = 91)	Leaked (n = 79)	
Location				
K-12 school	13	1 (7.7)	12 (92.3)	
College	9	5 (55.6)	4 (44.4)	
Government	6	4 (66.7)	2 (33.3)	
Place of worship	11	5 (45.5)	6 (54.6)	
Retail	29	14 (48.3)	15 (51.7)	19.9
Restaurant/club	22	15 (68.2)	7 (31.8)	
Workplace	52	27 (51.9)	25 (48.1)	
Other	14	8 (57.1)	6 (42.9)	
Outdoors	14	12 (85.7)	2 (14.3)	

We calculated the p-value for this chi-square value ($\chi^2 = 19.9$) as p = 0.01081.

a. At the 95% confidence level, would we reject, or fail to reject, the null hypothesis? _____

b. What is the appropriate null hypothesis. (The case study didn't specifically state it.) _____

c. There are 9 locations listed in the table. How many degrees of freedom should be used for the chi-square test? _____

d. If you sought to use leakage as an intervention method to prevent a mass shooting, which location do you think would have the highest likelihood of success for this intervention method, and why? _____
_____

e. When your classmate, Marina, described her chi-square calculation using R/RStudio, she got the following "warning" message: Chi-squared approximation may be incorrect. Why would she get such a warning?
_____

3. I once gave two variations of the same exam. My belief was that the two exams were equally difficult. Prior to giving them, the exams were shuffled to ensure each student received a random version (either Version A or Version B). Summary statistics for how students performed on these two exams are shown in the table below. The average grade for Version B (82.1%) was lower than the average grade for Version A (86.4%), but maybe not "statistically different." I will continue to believe that the exams were equivalent in difficulty unless there is a 90% chance that they were not. In other words:

H_0: $\mu_{ExamA} = \mu_{ExamB}$

H_A: $\mu_{ExamA} \neq \mu_{ExamB}$

Version	n	$\bar{x}$	s
A	26	86.4	9.2
B	24	82.1	8.5

a. Using a t-test, how many degrees of freedom should I use? _____

b. What standard error should I use? _____

c. What is the p-value for this hypothesis test? _____

d. Can I reject H_0 with 90% confidence? (That is, can I be 90% confident that the exams were not equally difficult?) _____

4. The *OpenIntro* website occasionally experiments with design of the position of hyperlink placement. They conducted an experiment testing three different position placements of a download link for the textbook on the book's main page to see which location, if any, led to the most downloads. The number of site visitors included in the experiment was 701 and is captured in one of the response combinations in the following table:

Observed	Download	No download	Total
Position 1	97	128	**225**
Position 2	102	130	**232**
Position 3	85	159	**244**
Total	**284**	**417**	**701**

With a null hypothesis: **H_0: all hyperlink position placements are equally likely to lead to a download**, complete the following matrix of expected outcomes.

Expected	Download	No download	Total
Position 1			
Position 2			
Position 3			
Total			

Perform a chi-square test and determine whether to reject the null hypothesis at the 95% confidence level. What are:

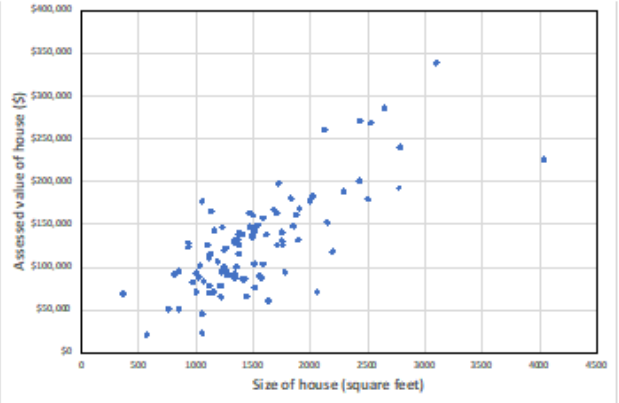
a. χ^2 = _____

b. Degrees of freedom = _____

c. p-value = _____

d. Reject H_0 or not? _____

5. It is useful to estimate the value of a house, based on its size. The following regression output was used to predict the assessed value (units = dollars) of a house, given the floor size of its living space area (units = square feet). The dataset is based on 50 houses in a particular geographic location.



	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9161.159	10759.786	0.851	0.397
Size_sq_ft	77.008	6.626	11.622	<2e-16

The regression equation can be expressed as: $Assessed\_value = \beta_0 + \beta_1 \cdot Size\_sq\_ft$

a. What is the point estimate for β_1? _____

b. What is the 95% confidence interval for β_1? _____

c. If a house on the market had 2000 square feet of living space , what would you estimate as its assessed value? (Include the measurement units with your answer.)
_____

d. If a house had 100 square feet more than the house next door to it, what is the estimate for the additional assessed value of the larger house? (Include the measurement units with your answer.) _____

e. The correlation coefficient for Assessed_value and Size_sq_ft is 0.7612. Calculate R^2 and interpret it in context.

R^2 = _____

Interpretation: _____

6. Public health researchers were interested in social factors that influence heart disease. They surveyed 15 towns and gathered data on the percentage of people in each town who smoke, the percentage of people in each town who bike to work, and the percentage of people in each town who have heart disease.

   A multiple linear regression was used to analyze the relationship between heart disease and these health-related factors. Both biking and smoking were coded as 0 = "no" and 1 = "yes". The regression summary is given below.

	Estimate	Std. Error	t value	Pr(>|t|)
(intercept)	14.984658	0.080137	186.99	<2e-16
biking	-0.200133	0.001366	-146.53	<2e-16
smoking	0.178334	0.003539	50.39	2.448e-15

   a. What percentage of people in these towns who smoke and bike to work would you estimate to have heart disease? _____

   b. How many degrees of freedom should be used for the t-distribution here? _____

   c. Calculate the 95% margin of error for the coefficient of **smoking**. (Use 4 decimal place accuracy.) _____

   d. Suppose that in a particular town, the percentage of people who do not smoke yet do bike to work is 13.2 percent. Determine if the model overestimates or underestimates the rate of heart disease in this town, and by how much.

      Over- or underestimate? _____

      By how much? _____