STA 100 Homework 3

Due 11:59 pm Friday, July 21 onto Gradescope

1. A fair coin is to be tossed 20 times. Find the probability that 10 of the tosses will fall heads and 10 will fall tails,

   (a) using the binomial distribution formula.

   (b) using the normal approximation.

   (c) using the normal approximation with continuity correction.

2. As part of a study of natural variation in blood chemistry, serum potassium concentrations were measured in 81 randomly selected healthy women. The sample mean concentration was 4.36 mEq/l, and the sample standard deviation was 0.42 mEq/l.

   (a) Find the 95% confidence interval for the true serum potassium concentrate in healthy women.

   (b) Interpret your confidence interval found in (a) in terms of the problem.

   (c) Does your interval support the claim that normal serum potassium concentrations are above 2.3 mEq/l?

   (d) If we were to build a 99% confidence interval instead, would it widen or narrow?

3. Continue with the data from Problem 2. Assume $t_{n-1}(\alpha/2) = 2$.

   (a) If we wanted our confidence interval to have a margin of error of 0.1 mEq/l at 95% confidence, how many women should we sample (at least)?

   (b) If we wanted our confidence interval to have a margin of error of 0.05 mEq/l at 95% confidence, how many women should we sample (at least)?

   (c) What tends to happen to the sample size we need as the margin of error decreases? You may assume everything else remains constant.

   (d) What tends to happen to the sample size we need as the standard deviation increases? You may assume everything else remains constant.

4. Two treatment costs for a particular medical issue are being compared. The average cost of a sample of 50 randomly selected subjects for treatment A was \$490, with a standard deviation of \$32. The average cost for a sample of 70 randomly selected subjects for treatment B was \$500, with standard deviation \$48. Assume the degrees of freedom $\nu = 100$.

   (a) Find the 99% confidence interval for the difference in costs.

   (b) Interpret your interval from (a) in terms of the problem.

   (c) Does your interval suggest that the costs are different?

   (d) If we increased the confidence level (i.e, $1 - \alpha$), would the new interval be wider or narrower than that found in (a)?

5. Two baby-food manufacturers are competing, and manufacturer of brand A believes that their brand causes more weight gain in infants than manufacturer of brand B. A sample of size 15 fed their babies brand A for two months, and saw an average weight gain of 36.93 ounces, with a standard deviation of 4.23 ounces. A sample of size 25 fed their babies brand B and saw an average weight gain of 31.36 ounces with a standard deviation of 3.35 ounces. Assume the degrees of freedom $\nu = 24$.

(a) Find the 95% confidence interval for the difference in average weight gain between the two baby food brands.

(b) Interpret your interval from (a) in terms of the problem.

(c) Does your interval suggest one brand of baby food causes more weight gain than the other on average?

(d) If the sample sizes increased to 100, but everything else remained constant, would our interval widen or narrow?

6. Consider the following research topic: The goal is to determine if there is a statistically significant increase in the average weight gain of anorexic patients for a new treatment (μ_1) when compared to a standard treatment (μ_2).

(a) State the null hypothesis.

(b) State the alternative hypothesis.

(c) Interpret a Type I error in terms of the problem.

(d) Interpret a Type II error in terms of the problem.

(e) If we wanted to minimize the probability of a Type I error, what action should we take?

(f) If we wanted to minimize the probability of a Type II error, what action should we take?

7. A medical researcher wishes to see whether the pulse rates of smokers are higher than the pulse rate for non-smokers. Random samples of 70 smokers and 75 non-smokers are selected, and the pulse rate results are shown below. Test the researchers claim at $\alpha = 0.01$. Assume the degrees of freedom $\nu = 140$.

	Smokers	Non-smokers
Sample mean	90	88
Sample standard deviation	5.2	6.3
Sample size	70	75

(a) State the null hypothesis.

(b) State the alternative hypothesis.

(c) Calculate the test statistic for this problem.

(d) Calculate the range of the p-value for your test statistic.

(e) Interpret the p-value in terms of the problem.

(f) Do you reject the null, or fail to reject?

(g) State the conclusion in terms of the problem.

8. A scientist recorded the tail lengths (in cm) of two varieties of woodland salamanders: red-backed and lead-backed. The table below shows the summary:

	Red-backed	lead-backed
Sample mean	3.2	4.0
Sample standard deviation	0.9	0.7
Sample size	60	27

Assume the degrees of freedom $\nu = 80$. The researchers are interested in if the average tail lengths of the two salamanders are equal.

(a) State the null hypothesis.

(b) State the alternative hypothesis.

(c) Calculate the test statistic for this problem.

(d) Calculate the range of the p-value for your test statistic.

(e) Interpret the p-value in terms of the problem.

(f) Do you reject the null, or fail to reject?

(g) State the conclusion in terms of the problem.

R is necessary for the remaining questions. Attach source codes and any plots you produce to your homework submission. You may write down your numerical results.

9. (a) Consider a binomial random variable $X \sim B(1000, 0.5)$, Find the first and third quartiles of X. Hint: `qbinom()`.

   (b) The binomial distribution $B(n, p)$ can be well approximated by a normal distribution with mean np and variance $np(1-p)$ if both np and $n(1-p)$ are at least equal to 5. In the case of $B(1000, 0.5)$, obtain the approximated first and third quartiles using normal approximation and compare with the results in (a). Hint: `qnorm()`.

10. On Canvas (Files $\rightarrow$ Data) you will find the data `student.csv`. It contains information on 1548 Introduction to Statistics students, and has the following columns:

    Column 1: `height`: The height of the student.

    Column 2: `hsGPA`: The high school GPA of the student.

    Column 3: `pulse`: The pulse rate of the student when measured in class.

    (a) Using R, find the 95% confidence interval for the average students height.

    (b) What is the highest average height you expect a student to have, based on the confidence interval from (a)?

    (c) Using R, find the 99% confidence interval for the average students pulse.

    (d) What is the lowest average pulse you expect a student to have, based on the confidence interval from (c)?