1. (5 points) Confidence intervals. A student is suspicious of the formula of confidence intervals her sloppy professor wrote down on a napkin over lunch. She decided to conduct an experiment to check on the correctness of the formula. She generated random samples from computer and used the professor's formula to calculate a 95% confidence interval from each sample. Then she checks for each sample if the true parameter falls within the confidence interval. After checking 98 samples, the confidence interval contains the true parameter 93 times. "Ha," she said, "now the confidence interval has to contain the true parameters in all next two times to be correct." When it turns out that in the next two samples, none of the confidence intervals contain the true parameter, she gleefully told her classmates that the professor made another mistake. Do you agree with her? Why and why not? What is the probability that a 95% confidence interval contains the true parameter in exactly 95 out of 100 random samples?

Solution: The 95% confidence interval interprets that 95% of the intervals obtained from random samples generated will contain the true population parameter. Therefore, out of 98 samples, if 93 times the interval contains the true population parameter, then the percentage is: $\frac{93}{98} \times 100 = 94.89 \sim 95$. Therefore, the student is not right.

There are $n = 100$ independent, random trials, with probability of success, $p = 0.9489$. The probability of success is constant throughout the trials. The specific number of success in n trials is denoted by r.

$$P(X, r) = \binom{n}{r} (p)^r (1-p)^n - r$$

$$\text{where,} \qquad \binom{n}{r} = \frac{n!}{r!\,(n-r)!}$$

$$P(X = 95) = \binom{100}{95} (0.9489)^{95} (1 - 0.9489)^5 = 0.1798$$

2. (5 points each)

Exercises 10.8.4. Briefly explain the relationship between confidence intervals and hypothesis testing.

Solution: A confidence interval is a range of values that is likely to contain the true value of a population parameter, such as a population mean or proportion. The confidence interval is computed from a sample statistic, such as a sample mean or proportion, and is accompanied by a confidence level, which is a measure of the level of certainty or confidence that the interval actually contains the true population parameter. For example, a 95% confidence interval for a population mean indicates that we are 95% confident that the true population mean falls within the interval.

Hypothesis testing, on the other hand, is a method for testing a statistical hypothesis about a population parameter, such as whether the population mean is equal to a certain value or whether the population proportions are equal between two groups. In hypothesis testing, we typically start with a null hypothesis, which is a statement that there is no significant difference between the parameter and a specific value, and we test this hypothesis against an alternative hypothesis, which is a statement that there is a significant difference.

Confidence intervals and hypothesis testing are related in that confidence intervals can be used to perform hypothesis tests, and hypothesis tests can be used to compute confidence intervals. For example, a two-sided hypothesis test at the 5% significance level is equivalent to constructing a 95% confidence interval for the population parameter of interest. Similarly, if a confidence interval does not contain a certain value, such as zero, then we can reject a null hypothesis that the population parameter is equal to that value at a certain level of significance.

Exercise 10.8.5. Under what circumstances might you use a one-sided test of hypothesis rather than a two-sided test?

Solution: A one-sided test of hypothesis, also known as a one-tailed test, is used when the researcher is interested in testing whether the sample statistic is significantly larger or smaller than a hypothesized value in only one direction. In contrast, a two-sided test of hypothesis, also known as a two-tailed test, is used when the researcher is interested in testing whether the sample statistic is significantly different from a hypothesized value in either direction.

The choice between a one-sided test and a two-sided test depends on the research question and the hypothesis being tested. A one-sided test is appropriate when the research question specifically relates to a directional hypothesis or when there is prior knowledge or theory that supports a directional hypothesis. For example, a researcher might hypothesize that a new treatment is more effective than a standard treatment, and use a one-sided test to determine whether there is sufficient evidence to support this hypothesis. Similarly, a researcher might hypothesize that a new advertising campaign will result in a significant increase in sales, and use a one-sided test to determine whether the increase is statistically significant.

In contrast, a two-sided test is appropriate when the research question does not specify a direction for the hypothesis, or when the hypothesis is based on limited prior knowledge or theory. For example, a researcher might hypothesize that the mean weight of apples produced by a certain farm is different from a certain value, and use a two-sided test to determine whether there is sufficient evidence to support this hypothesis.

It is important to note that the choice between a one-sided test and a two-sided test can have implications for the type I error rate (the probability of rejecting a true null hypothesis) and the statistical power (the probability of correctly rejecting a false null hypothesis) of the test. As a result, the choice of test should be carefully considered based on the research question, the hypothesis being tested, and the available data.

3. **(10 points) Exercises 10.8.11. Body mass index is calculated by dividing a person's weight by the square of his or her height; it is a measure of the extent to which the individual is overweight. For the population of middle-aged men who later develop diabetes mellitus, the distribution of baseline body mass indices is approximately normal with an unknown mean μ and standard deviation σ. A sample of 58 men selected from this group has mean overline $\overline{x} = 25.0\ kg/m^2$ and standard deviation $s = 2.7\ kg/m^2$ [197].**
   (a) Construct a 95% confidence interval for the population mean μ.
   (b) At the 0.05 level of significance, test whether the mean baseline body mass index for the population of middle-aged men who do develop diabetes is equal to 24 kg/m^2 the mean for the population of men who do not. State the test statistic, probability distribution of the test statistic, and the *p*-value.
   (c) What do you conclude?
   (d) Based on the 95% confidence interval, would you have expected to reject or not reject the null hypothesis? Why?

Solution: Given, $\bar{x} = 25.0, s = 2.7, n = 58$

Level of significance, $\alpha = 0.05$

(a) Using the following formula to compute the 95% confidence interval for μ.

$$95\%CI = \bar{x} \pm t_{\alpha/2,\text{n-1}} \left(\frac{s}{\sqrt{n}}\right)$$
$$= 25 \pm t_{0.05/2,58-1} \left(\frac{2.7}{\sqrt{58}}\right)$$
$$= 25 \pm (2.002)(0.3545)$$
$$= 25 \pm 0.7097$$
$$= (24.2903, 25.7097)$$

[Calculated using Excel (=TINV(0.05,57)]

(b) The null and alternative hypothesis are,

$$H_0: \mu = 24$$
$$H_\text{a}: \mu \neq 24$$

The degree of freedom is: $df = n - 1 = 58 - 1 = 57$

Using the following formula to compute the test statistic:

$$t_{\text{cal}} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{25 - 24}{2.7/\sqrt{58}} = \frac{1}{0.3545} = 2.82$$

Computing the *p*-value.

$$p - value = 2P(t > t_{cal})$$
$$= 2P(t > 2.82)$$
$$= 0.007$$

[Calculated using Excel (=TDIST(2.82,57,2)]

(c) Decision rule: Reject the null hypothesis if the *p*-value is less than level of significance $\alpha = 0.05$

Since $0.007 < 0.05$, that means the *p*-value is less than level of significance $\alpha = 0.05$, reject the null hypothesis. Therefore, there is no sufficient evidence to conclude that the mean baseline body mass index for the population of middle-aged man who do develop diabetes is equal to 24.0 kg/m^2.

(d) From the above results, the 95% confidence interval of μ is (24.2903, 25.7097). Here the population mean (24) does not fall in the above interval, thus, reject the null hypothesis. Therefore, there is no sufficient evidence to conclude that the mean baseline body mass index for the population of middle-aged men who do develop diabetes is equal to 24.0 kg/m^2.

4. (10 points) Exercises 10.8.15. In Norway, the distribution of birth weights for full-term infants whose gestational age is 40 weeks is approximately normal with mean μ = 3500 grams and standard deviation σ = 430 grams [177]. An investigator plans to conduct a study to determine whether the birth weights of full-term babies whose mothers smoked throughout pregnancy have the same mean. A two-sided test conducted at the 0.05 level of significance will be used.

(a) If the true mean birth weight for the infants whose mothers smoked is as low as 3200 grams, then the investigator wants to risk only a 10% chance of failing to detect this difference. What sample size is needed for this study?

(b) What sample size would be required if the investigator wants to risk only a 5% chance of failing to detect a difference of 300 grams?

(c) If the true mean birth weight for the infants whose mothers smoked is 3300 grams and the investigator wants to risk only a 10% chance of failing to detect this difference, what sample size is required?

(d) Suppose that the investigator only has the resources to study 30 infants whose mothers smoked throughout pregnancy. If the true population mean birth weight for these infants is 3300 grams, what would be the power of this study?

Solution: (a) We will use the formula for the sample size required for a two-sided t-test to solve for the required sample size n. The formula is:

$$n = \left[\frac{(Z_{\alpha/2} + Z_{\beta})\sigma}{\delta}\right]^2$$

where $Z_{\alpha/2}$ is the critical value of the standard normal distribution at the 0.025 level of significance (since it is a two-sided test), Z_β is the critical value of the standard normal distribution that corresponds to a power of 0.9, σ is the standard deviation, and δ is the difference in means we want to detect.

Substituting the given values, we get:

$$n = \left[\frac{(1.645 + 1.28) \times 430}{3500 - 3200}\right]^2 = 36.54$$

Therefore, the required sample size is n = 37.

(b) Using the same formula, we get:

$$n = \left[\frac{(1.96 + 1.645) \times 430}{300}\right]^2 = 68.62$$

Therefore, the required sample size is n = 69.

(c) Again, using the same formula, we get:

$$n = \left[\frac{(1.28 + 1.645) \times 430}{3500 - 3300}\right]^2 = 26.36$$

Therefore, the required sample size is n = 27.

(d) To calculate the power of the study, we first need to calculate the effect size, which is the difference in means divided by the standard deviation. The effect size is:

$$d = (3500 - 3300)/430 = 0.47$$

Using a t-distribution with 29 degrees of freedom (since we have a sample size of 30), we can find the critical value of t for a two-sided test at the 0.05 level of significance to be 2.045. The non-centrality parameter, λ, can be calculated as:

$$\lambda = \sqrt{n} \times d\,/\,\sigma = \sqrt{30} \times 0.47\,/\,430 = 0.33$$

Using a t-distribution calculator, we can find that the power of the study is approximately 0.49 or 49%. Therefore, if the true population mean birth weight for these infants is 3300 grams, there is a 49% chance that we will correctly reject the null hypothesis and detect a difference in means of 200 grams.

5. **(10 points) Exercises 10.8.17. The Bayley Scales of Infant Development yield scores on two indices - the Psychomotor Development Index (PDI) and the Mental Development Index (MDI) - which can be used to assess a child's level of functioning in each of these areas at approximately one year of age. Among normal healthy infants, both indices have a mean value of 100. As part of a study assessing the development and neurologic status of children who underwent reparative heart surgery during the first three months of life, the Bayley Scales were administered to a sample of one-year-old infants born with congenital heart disease. The data are contained in the data set bayley [189]; PDI scores are saved under the variable name pdi, while MDI Scores are saved under mdi.**
(a) At the 0.05 level of significance, test the null hypothesis that the mean PDI score for children born with congenital heart disease who undergo reparative heart surgery during the first three months of life is equal to 100, the mean score for healthy children. Use a two-sided test. What is the p-value? What do you conclude?
(b) Conduct the analogous test of hypothesis for the mean MDI score. What do you conclude?
(c) In Chapter 9 you constructed 95% confidence intervals for the true mean PDI score and the true mean MDI Score for this population of children with congenital heart disease. What complementary information is provided by the hypothesis tests and confidence intervals?

Solution: Given, $n = 143$

(a) The relevant null and alternative hypothesis are,

$$H_0: \mu_{pdi} = 100$$
$$H_a: \mu_{pdi} \neq 100$$

$$\text{mean}, \bar{x}_{pdi} = 94.78322 \text{ and standard deviation}, s_{pdi} = 15.85104$$

$$t = \frac{\bar{x}_{pdi} - \mu_{pdi}}{s_{pdi}/\sqrt{n}} = \frac{94.78322 - 100}{15.85104/\sqrt{143}} \sim -3.9356$$

Degree of freedom $df = n - 1 = 143 - 1 = 142$
Calculating the p-value
p – value = 0.000127

[Calculated using Excel (=TDIST(3.94, 142, 2)]

For a two-sided test, this corresponds to a *p*-value of 0.000127. We can reject H_0, as the *p*-value is less than 0.05. Hence the mean PDI score for children born with congenital heart disease who undergo reparative heart surgery during the first 3 months of life is different from 100.

(b) The relevant null and alternative hypothesis are,

$$H_0: \mu_{mdi} = 100$$
$$H_a: \mu_{mdi} \neq 100$$
$$\text{mean}, \bar{x}_{mdi} = 104.7063 \text{ and standard deviation}, s_{mdi} = 15.6551$$
$$t = \frac{\bar{x}_{mdi} - \mu_{mdi}}{s_{mdi}/\sqrt{n}} = \frac{104.7063 - 100}{15.6551/\sqrt{143}} \sim 3.5949$$

Degree of freedom $df = n - 1 = 143 - 1 = 142$
Calculating the *p*-value
p – value = 0.000455

[Calculated using Excel (=TDIST(3.59, 142, 2)]

For a two-sided test, this corresponds to a *p*-value of 0.000455. We can reject H_0, as the *p*-value is less than 0.05. Hence the mean MDI score for children born with congenital heart disease who undergo reparative heart surgery during the first 3 months of life is different from 100.

(c) 95% confidence intervals for the true mean PDI score for this population of children with congenital heart disease is (92.18522, 97.38121)
95% confidence intervals for the true mean MDI Score for this population of children with congenital heart disease is (102.1404, 107.2722)
Therefore, neither of these intervals contains the value 100. We would not expect the intervals to contain 100 since both null hypothesis were rejected.

R Code:
```
library(haven)
bayley     <-     read_dta("C:/Users/abhil/OneDrive/Desktop/MSAM-Northeastern/MATH7343/Homeworks/4/bayley.dta")
bayley

pdi_mean <- mean(bayley$pdi)
pdi_std <- sd(bayley$pdi)

# 95% Confidence Interval for pdi
```

```
alpha <- 0.05
z_critical <- qnorm(1 - alpha/2)
CI_95_pdi <- pdi_mean + c(-z_critical, z_critical) * (pdi_std/
sqrt(nrow(bayley)))
CI_95_pdi

mdi_mean <- mean(bayley$mdi)
mdi_std <- sd(bayley$mdi)

# 95% Confidence Interval for mdi
alpha <- 0.05
z_critical <- qnorm(1 - alpha/2)
CI_95_mdi <- mdi_mean + c(-z_critical, z_critical) * (mdi_std /
sqrt(nrow(bayley)))
CI_95_mdi
```

6. **(15 points) Exercise 11.5.13. A study was conducted to compare physical and psychological characteristics of monozygotic (identical) twins [213]. One question of interest was whether brain volume differed by birth order. The dataset `twins` contains total brain volumes measured in cm^3 for 10 pairs of monozygotic twins. The brain volumes of the first-born twins are saved under the variable name `volume_1`, and the brain volumes of the second-born twins under the name `volume_2`.**
   (a) Are the two samples of total brain volume measurements paired or independent?
   (b) What are the appropriate null and alternative hypotheses for a two-sided test?
   (c) Conduct the test at the 0.05 level of significance. What is the p-value? What is the probability distribution of the test statistic, assuming that the null hypothesis is true?
   (d) What do you conclude about the relationship between birth order and brain volume in monozygotic twins?

Solution: (a) The two samples of total brain volume measurements are paired because each pair of measurements represents two measurements from the same set of twins (one measurement for the first-born twin and one for the second-born twin).

(b) The appropriate null hypothesis for a two-sided test is that the mean difference in brain volume between first-born twins and second-born twins is zero.
The alternative hypothesis is that the mean difference in brain volume between first-born twins and second-born twins is not zero.

$$H_0: \mu_1 - \mu_2 = 0$$
$$H_a: \mu_1 - \mu_2 \neq 0$$

where μ_1 is the mean brain volume of the first-born twins, and μ_2 is the mean brain volume of the second-born twins.

(c) To conduct the test, we can use a paired t-test with a significance level of 0.05. The test statistic is calculated as:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s/\sqrt{n}}$$

where $\bar{x}_1$ and $\bar{x}_2$ are the sample means of the brain volumes for the first-born and second-born twins, s is the sample standard deviation of the differences in brain volumes, and n is the number of pairs of twins.

Using R, we can calculate the test statistic and p-value as follows:

R Code:

```
library(haven)
twins <- read_dta("C:/Users/abhil/OneDrive/Desktop/MSAM-Northeastern/MATH7343/Homeworks/4/twins.dta")

t.test(twins$volume_1,twins$volume_2, alternative="two.sided", conf.level=0.95, paired=TRUE)
```

Output:

```
Paired t-test

data:  twins$volume_1 and twins$volume_2
t = -0.37206, df = 9, p-value = 0.7185
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -47.43705  34.03705
sample estimates:
mean difference
           -6.7
```

The test statistic is t = -0.37206 with 9 degrees of freedom, and the p-value is 0.7185. The probability distribution of the test statistic, assuming that the null hypothesis is true, is a t-distribution with 9 degrees of freedom.

(d) Since the p-value is greater than the significance level of 0.05, we fail to reject the null hypothesis. There is insufficient evidence to conclude that there is a significant difference in brain volume between first-born and second-born monozygotic twins.

7. (10 points) Exercise 11.5.16. The Bayley Scales of Infant Development provide scores on two indices - the Psychomotor Development Index (PDI) and the Mental Development Index (MDI) - which can be used to assess a child's level of functioning at approximately one year of age. As part of a study investigating the development and neurologic status of children who had undergone reparative heart surgery during the first three months of life, the Bayley Scales were administered to a sample of one-year-old infants born with congenital heart disease. The children had been randomized to one of two different treatment groups, known as "circulatory arrest" and "low-flow bypass." The groups differed in the specific way in which the reparative surgery was performed. Unlike circulatory arrest, which stops blood flow through the brain for a short period of time, low-flow bypass maintains continuous circulation. Although it is felt to be preferable by some physicians, it also has its own associated risk of brain injury. The data for this study are saved in the data set `bayley` [189]. PDI scores are saved under the variable name `pdi`, MDI scores under mdi, and indicators of treatment group under `trtment`. For this variable, O represents circulatory arrest and 1 is low-flow bypass.

(a) At the 0.05 level of significance, test the null hypothesis that the mean PDI score at one year of age for the circulatory arrest treatment group is equal to the mean PDI score for the low-flow group. What is the *p*-value for this test? What do you conclude?

(b) Test the null hypothesis that the mean MDI scores are identical for the two treatment groups. What is the *p*-value? What do you conclude?

(c) What do these tests suggest about the relationship between a child's surgical treatment group during the first three months of life and his or her subsequent developmental status at one year of age?

Solution:

(a) Test the null hypothesis that the mean PDI score at one year of age for the circulatory arrest treatment group is equal to the mean PDI score for the low-flow group.

The null and alternative hypothesis are:

$$H_0: \mu_C = \mu_L$$
$$H_a: \mu_C \neq \mu_L$$

Where μ_C and μ_L are the population means for PDI scores in the circulatory arrest and low flow groups respectively.

Using R, performing t test.

Output:

```
Two Sample t-test

data:  pdi by treatment
t = -2.2385, df = 141, p-value = 0.02676
alternative hypothesis: true difference in means between group 0
and group 1 is not equal to 0
```

```
95 percent confidence interval:
 -11.0232390  -0.6840017
sample estimates:
mean in group 0 mean in group 1
       91.91781        97.77143
```

The p-value for this test is 0.02676, which is smaller than the significance level of 0.05. Therefore, we reject the null hypothesis and conclude that there is difference in the mean PDI scores between the circulatory arrest and low-flow groups at one year of age.

(b) Test the null hypothesis that the mean MDI score at one year of age for the circulatory arrest treatment group is equal to the mean MDI score for the low-flow group.
The null and alternative hypothesis are:

$$H_0: \mu_{\mathrm{C}} = \mu_{\mathrm{L}}$$
$$H_{\mathrm{a}}: \mu_{\mathrm{C}} \neq \mu_{\mathrm{L}}$$

Where μ_C and μ_L are the population means for MDI scores in the circulatory arrest and low flow groups respectively.

Using R, performing t test.
Output:

```
Two Sample t-test

data:  mdi by treatment
t = -1.2696, df = 141, p-value = 0.2063
alternative hypothesis: true difference in means between group 0
and group 1 is not equal to 0
95 percent confidence interval:
 -8.484009  1.848393
sample estimates:
mean in group 0 mean in group 1
       103.0822        106.4000
```

The p-value for this test is 0.2063, which is larger than the significance level of 0.05. Therefore, we fail to reject the null hypothesis and conclude that there is no significant difference in the mean MDI scores between the circulatory arrest and low-flow groups at one year of age.

(c) Based on the results of the two tests in parts (a) and (b), we can conclude that there is no significant relationship between a child's surgical treatment group during the first three months of life and his or her subsequent developmental status at one year of age, as measured by the PDI and MDI scores on the Bayley Scales. However, it is important to note that this study only investigated one-year-old infants born with congenital heart disease who had undergone reparative heart surgery, and the results may not be generalizable to other populations or age groups.

R Code:

```
library(haven)

bayley <- read_dta("C:/Users/abhil/OneDrive/Desktop/MSAM-Northeastern/MATH7343/Homeworks/4/bayley.dta")


# Conduct two-sample t-test with equal variances on pdi

t.test(pdi ~ treatment, data = bayley, var.equal = TRUE)


# Conduct two-sample t-test with equal variances on mdi

t.test(mdi ~ treatment, data = bayley, var.equal = TRUE)
```