Homework #2_FA2024_BMEN7340

First_Last_7340_HW2_FA24

2024-09-11

clean R environment

```
rm(list = ls(all=TRUE))
graphics.off()
shell("cls")
```

library(readxl)
library(ggplot2)
library(here)
library(epiR)
library(MASS)
library(nnet)
library(car)
library(PairedData)
library(pwr)
library(dplyr)
library(FSA) #posthoc

Question #1:

A nutrition researcher is interested in evaluating the effect of a new dietary supplement on blood pressure.
To do this, a study was conducted with randomly selected 14 participants who took the supplement for 8 weeks.
Blood pressure (mmHg) measurements were taken before the supplementation period and again at the end of the 8 weeks.
Use a 0.05 significance level to test the claim that there is a reduction in blood pressure after taking the dietary supplement.

1. State the Ho, Ha hypothesis.
2. Run the t.test()
3. Report the t-statistic, degrees of freedom, p-value, and 95% CI.
4. Based on the CI output ________, explain the t.test result, with the following language
   # If the Ho is true, the probability of having a ____________________
   # or more extreme, is ____________%
   # Since the value of Ho _________is_____________________ # we _______ null hypothesis.
   # We conclude that there is ______________________________

week0 <- c(130, 125, 140, 135, 138, 132, 137, 142, 129, 136, 133, 139, 140, 150)
week8 <- c(122, 118, 133, 127, 129, 124, 130, 135, 121, 128, 126, 132, 120, 123)

Question #2:

A company implemented a new training program to improve employee productivity. Before and after the training, each employee was measured for productivity (in terms of units produced per day).
The company wants to assess whether the new training program led to a significant change in productivity.
Use a 0.05 significance level to test the claim that there is a difference in productivity due to the training program.
before <- c(20, 22, 19, 5, 23, 21, 25, 22, 24, 10)
after <- c(15, 26, 21, 10, 27, 23, 60, 26, 27, 50)

1. State the Ho, Ha hypothesis.
2. Check assumptions and perform a proper statistical test.
3. Report the p value.
4. Based on the p-value _______, explain the result using the following language # If the Ho is true, the probability of having a ____________________
   # or more extreme, is ____________%
   # Since the value of p value_________is____________________ # we _______ null hypothesis. # We conclude that there is ______________________________

Question #3:

1. **Births.xlsx**
2. Test the claim that girls and boys have the same median birth weight, using a 0.05 significance level.
3. Include Ho, Ha, Assumption checks, plots, R output interpretations, and Conclusion.

Question #4:

Determine whether there is sufficient evidence to support the claim of a linear correlation between the two variables ("IQ" and Brain Volume "VOL").
Include Ho, Ha, Assumption checks, plots, R output interpretations, and Conclusion.

Question #5: You have a dataset that contains the following values representing the number of daily steps taken by 30 individuals over a month:

1. Visual Inspection: Plot a histogram of the step data and add a normal density curve to assess the shape of the distribution.

2. Transformation: Apply a log transformation to the step data. Plot a histogram of the transformed data and add a normal density curve.

3. Check Normality: Use the Shapiro test to check if the transformed data follows a normal distribution.

4. Comparison: Compare the histograms and normality tests of the original and transformed data. Discuss whether the transformation improved the normality of the data.

```r
# Original data
set.seed(123)
steps <- c(3200, 2900, 3100, 3000, 2800, 3500, 3100, 3000, 3300, 3400,
           2800, 2900, 3200, 3000, 3100, 3300, 3400, 3600, 3500, 3400,
           3000, 2900, 3100, 3200, 3100, 3000, 2800, 2900, 3200, 3100,
           3000, 3200, 3300, 3400, 3500, 3600, 3400, 3300, 3200, 3100)
```

Question #6

1. **Arsenic.xlsx**
2. Test the claim that the brown rice from Arkansas, California, and Texas have the same amount of arsenic.

Report the following:

1. Ho, Ha
2. assumption check
3. Mean squared between (MS between)
4. Mean squared within (Ms within)
5. F value
6. Interpretation the p value.
7. Make a conclusion (with the problem context!!!!), reject Ho, or FTR Ha, why?
8. If Ho is rejected, perform a post-hoc test.

Question #7:

Listed below are measured loads (in lb) on the left femur of crash test dummies.
Use a 0.05 significance level to test the null hypothesis that the different car categories have the same median.

1. **List Ho and Ha**
2. **Create a descriptive statistic plot**
3. **Perform a proper statistical test**
4. **Do these data suggest that larger cars are safer?**
5. **Why or why no?**

```
smallcar <- c(548,782,1188,707,324,320,634,501,274,437)
mcar <- c(194,280,1076,411,617,133,719,656,874,445)
largecar <- c(215,937,953,1636,937,472,882,562,656,433)
```

Question #8:

Suppose an equal number of girls at two poverty levels (below & above) will be recruited to study differences in calcium intake.
How many girls should be recruited to have an 80% chance of detecting a significant difference using a two-sided test with alpha=0.05?
In order to estimate the effect size, a pilot study is performed, the the result shown the mean calcium intake among 25 girls (age 12-14 years, below poverty level) is 6.56mg, sd=0.64mg the mean calcium intake among 40 girls (age 12-14 years, above poverty level) is 6.80mg, sd=0.76mg.

Question #9:

Suppose 50 girls above the poverty level and 50 girls below the poverty level are recruited for the study.
How much power will the study have of finding a significant difference using a two-sided test with alpha=0.05.
Assuming that the population parameters are the same as the sample estimates.