Homework 3

P6103 Fall 2022

**Due: Friday, December 02 (11:59 p.m.)**

Please do all problems below showing all work unless you are explicitly told to use R. You can use R to check your solutions for those problems that you do by hand if you’d like but we would like to see the steps of each solution clearly explained.

**Module 07 Questions**

1. Cortisol level determinations were made on two samples of women at childbirth. Group 1 subjects underwent emergency cesarean section following induced labor. Group 2 subjects delivered by either cesarean section or the vaginal route following spontaneous labor. The samples sizes, mean cortisol levels, and standard deviations were as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| Sample | N | Mean | St. Dev |
| 1 | 10 | 435 | 61 |
| 2 | 12 | 645 | 88 |

Assume that the cortisol levels are approximately normally distributed in each group.

1. Do these data provide sufficient evidence to indicate a difference in the mean cortisol levels in the populations represented? Perform a hypothesis test at the 5% significance level.

Be sure to provide the following:

* Define the parameter(s) of interest.
* State the null and alternative hypotheses.
* State the significance level.
* State what type of test you will perform, why you chose to use that test, and provide a check for any necessary assumptions.
* State your test statistic and its null distribution.
* Compute the test statistic BY HAND. Make a decision to reject or fail to reject based on either a p-value or by comparing the test statistic to a critical value (you can choose whichever you prefer).
* State your conclusion in the context of the problem.

1. Construct and interpret a 95% confidence interval for the difference in the population means.

2. The Bayley Scales of Infant Development yield scores on two indices – the Psychomotor Development Index (PDI) and the Mental Development Index (MDI) – that can be used to assess a child’s level of functioning at approximately one year of age. Among normal healthy infants, both indices have a mean value of 100. 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 failure. 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, low-flow bypass maintains continuous circulation through the brain; 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 in the dataset “Heart” in Canvas. PDI scores are saved under the variable name *pdi*, MDI scores under *mdi*, and indicators of treatment group under *trtment*. For this variable, 0 represents circulatory arrest and 1 is for low-flow bypass.

At the 5% significance level, 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. Also provide and interpret the corresponding 95% confidence interval.

* Define the parameter of interest.
* State the null and alternative hypotheses.
* State the significance level.
* State what type of test you will perform and why you chose to use that test.
* State and check any necessary assumptions. (Even if the sample sizes are sufficient, check normality of the values in the two groups.)
* State your test statistic and its null distribution.
* Give the output of the hypothesis test from R. Verify the calculation of the test statistic by hand. You do not have to verify the P-value by hand. You also do not have to state the decision rule.
* State your conclusion in the context of the problem. Be sure to include the P-value and confidence interval.

3. A crossover study was conducted to investigate whether oat bran helps to lower serum cholesterol levels in hypercholesterolemic males. Fourteen such individuals were randomly placed on a diet that included either oat bran or corn flakes; after two weeks, their low-density lipoprotein (LDL) (in mmol/l) cholesterol levels were recorded. Each man was then switched to the alternative diet. After a second two-week period, the LDL cholesterol level of each individual was again recorded. We want to see if oat bran lowers cholesterol more than cornflakes. The data is in “Oatbran” in Canvas. Perform the appropriate test using R.

* Define the parameter of interest.
* State the null and alternative hypotheses.
* State the significance level.
* State what type of test you will perform and why you chose to use that test.
* State and check any necessary assumptions.
* State your test statistic and its null distribution.
* State your decision rule.
* Give the output of the hypothesis test from R. Verify the calculation of the test statistic by hand. You do not have to verify the P-value by hand.
* State your conclusion in the context of the problem. Be sure to include the P-value.

**Module 08 Questions**

4. At what age do babies learn to crawl? Does it take longer to learn in the winter when babies are often bundled in clothes that restrict their movement? Data were collected from parents who brought their babies into the University of Denver Infant Study Center to participate in one of a number of experiments between 1988 and 1991. Parents reported the birth month and the age at which their child was first able to creep or crawl a distance of four feet within one minute. The resulting data were grouped by month of birth. The data are for January, May, and September

Average

Birth month crawling age SD n

January 29.84 7.08 32

May 28.58 8.07 27

September 33.83 6.93 38

Crawling age is given in weeks. Assume the data come from three independent simple random samples, one from each of the three populations (babies born in a particular month) and that the populations of crawling ages have normal distributions. A partial ANOVA table is given below:

Analysis of Variance for crawling age

Source df Sums of squares Mean square F-ratio

Birth month XX 505.26 XXXXXX XXXX

Error XX XXXXXX XXXXXX

Total XX 5529.56

Note the following: F2,94,0.05 = 3.09.

1. Complete the partial ANOVA table. Show your calculations.
2. Is the corresponding p-value greater or less than 5%? Justify your response. (Note that this question is not asking you to write out all steps of the test…just answer the question and justify your response.)

5. A study was conducted to investigate the risk factors for peripheral arterial disease among persons 55-74 years of age. The file “LDL” in Canvas contains data pertaining to LDL cholesterol levels (mmol/liter) from four different subgroups of subjects:

(intermittent) Patients with intermittent claudication (cramping in leg muscles)

(major) Major asymptomatic disease cases

(minor) Minor asymptomatic disease cases

(no) Those with no disease

Carry out the appropriate test to compare the four groups with respect to LDL cholesterol levels simultaneously. If you obtain significant differences, perform pair-wise tests. Be sure to provide the following:

* Define the parameters of interest.
* State the null and alternative hypotheses.
* State the significance level.
* State what type of test you will perform and why you chose to use that test.
* State and check any necessary assumptions.
* State your test statistic and its null distribution.
* Give the output of the hypothesis test from R. Verify the calculation of the test statistic by hand. You do not have to verify the P-value by hand. You also do not have to state the decision rule.
* State your conclusion in the context of the problem. Be sure to include the P-value.

**Module 09 Questions**

6. In December 2012 Gallup Poll conducted a survey of 1,015 Americans to determine if they had delayed seeking healthcare treatment due to the associated costs. Of the participants, 325 reported delaying seeking treatment due to costs. Create a 95% confidence interval for the proportion of all Americans who have delayed seeking healthcare treatment due to costs. Interpret your interval. (Do not use R to construct the interval.)

7. Although arsenic is known to be a poison, it also has some beneficial medicinal uses. In one study of the use of arsenic to treat acute promyelocytic leukemia (APL), a rare type if blood cancer, APL patients were given an arsenic compound as part of their treatment. Of those receiving arsenic, 32% were in remission and showed no signs of leukemia in subsequent examination. It is known that 18% of APL patients go into remission after conventional treatment. Suppose that the study had included 100 randomly selected patients. Is there sufficient evidence to conclude that the proportion in remission for the arsenic treatment is greater than 0.18, the remission proportion for conventional treatment?

1. Test the relevant hypothesis at the 0.01 significance level. Perform computations both by hand and in R. Be sure to show all steps for the hypothesis test.
2. Would your decision **rule** change if the sample size were only 35 patients? Why or why not? (You do not have to redo all the steps from part (a).)
3. Construct and interpret a 99% confidence interval for the true proportion of APL patients who go into remission after using arsenic. Does the inference drawn from this interval match the conclusion from the test that you did in part (a)?
4. Suppose that we are interested in hypothesis testing and confidence intervals for a single proportion. **In your own words**, explain why it possible for the conclusion of a hypothesis test done at significance level 0.05 to be different from the inference drawn from a 95% confidence interval that is constructed from the same data used to conduct the test.

8. The numbers of medical coding errors on bills from 200 randomly selected hospitals were recorded and the results are given in the table below.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Number of Errors | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | > 8 |
| Frequency | 5 | 12 | 31 | 40 | 38 | 29 | 22 | 14 | 5 | 4 |

Test the claim at the 5% significance level that a suitable model for these data is a Poisson distribution with a mean of 4. You may use R for any/all calculations except for calculating the corresponding test statistic (calculate the test statistic by hand). This data is stored in the “miscode.data” text file. (You are encouraged to discuss approaches to this problem with other students.)