Abdulaziz Alqumayzi G200007615 Module-12 CT-5

Rudra S Bandhu

12/01/2021

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(ggplot2)
library(readxl)
```

Import datasets

```
problem_2 <- read_excel("CT5_problem.xlsx", sheet = "problem_2")
problem_3 <- read_excel("CT5_problem.xlsx", sheet = "problem_3", col_types =
c("text","numeric", "numeric"))
problem_4 <- read_excel("CT5_problem.xlsx", sheet = "problem_4")
```

Problem 1

```
# calculate p-value for 60 men
2*pbinom(40,100,0.5)# 40 coming from random sample of 100 and 60 men have
readings above 128. 100-60=40

## [1] 0.05688793
```

Point a:

p-value is 0.05688793

```
# calculate p-value for 70 men
2*pbinom(30,100,0.5)

## [1] 7.85014e-05
```

Point b:

p-value is 7.85014e-05

```
# calculate p-value for 80 men
2*pbinom(20,100,0.5)
```

```
## [1] 1.115909e-09
```

Point c:

p_value is 1.115909e-09

Problem 2

Point a:

The null hypothesis: there is no systematic difference in book prices between the two online booksellers.

The alternative hypothesis: there is a systematic difference in book prices between the two online booksellers.

Point b:

A hypothesis testing will be performed to compare the price mean of the two booksellers.

An independent test statistic will be performed to compare the two means at the level of significance of 0.05.

Point c:

```
# independent t-test
t.test(problem_2$`Seller A` , problem_2$`Seller B`,conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  problem_2$`Seller A` and problem_2$`Seller B`
## t = 0.064976, df = 18, p-value = 0.9489
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.261374  4.533374
## sample estimates:
## mean of x mean of y
##    14.624    14.488
```

Point d:

p-value= 0.9489

Since p-value > 0.05 at the level of significance(0.05), we fail to reject the null hypothesis.

We conclude that there is no systematic difference in book prices between the two online booksellers.

Problem 3

Point a:

The null hypothesis: the estimated prices they are receiving from Bob's garage are no different from other garages.

The alternative hypothesis: the estimated prices they are receiving from Bob's garage are higher than another garage.

Point b:

A hypothesis testing will be performed to compare if the estimated prices mean of Bob's garage is higher than another garage.

A dependent test statistic will be performed to compare the two means at the level of significance of 0.05.

Point c:

```
# paired t-test

t.test(problem_3$'Bob's Garage', problem_3$'Other Garage', alternative =
'greater',paired = T)

##
##  Paired t-test
##
## data:  problem_3$"Bob's Garage" and problem_3$"Other Garage"
## t = 3.1512, df = 9, p-value = 0.005858
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  47.68341      Inf
## sample estimates:
## mean of the differences
##                     114
```

Point d:

p-value= 0.005858

Since p-value < 0.05 at the level of significance(0.05), we reject the null hypothesis.

We conclude that the insurance adjusters have significant evidence of the estimated prices of Bob's garage are higher than another garage.

Problem 4

Point a:

The problem that I want to study is the effect of physical exercises to alleviate depression. I'm not an exercising person and I feel that is reflected in my depression. I found a published paper from Dr. Graham Hole that tested some depressed people. Each person randomly was allocated to one of three groups: no exercise; 20 minutes of jogging per day; or 60 minutes of jogging per day. At the end of a month, they asked each participant to rate

how depressed they now feel. I used that paper to make a non-parametric test (Kruskal-Wallis)

Point b:

The null hypothesis: doing exercises have no effect differences between the three groups on depression. The alternative hypothesis: doing exercises have effect differences between the three groups on depression.

Point c:

The dataset was collected manually from the paper. The table in the paper needs transformation so we can apply it to R kruskal.test() function. The table in the paper has three columns of No exercise, jogging(20min) and jogging(60min) and their values. The dataset was transformed into two columns (exercise and depression_scale). exercise column has the type of the group and depression_scale has the depression scale value.

Point d:

As mentioned in point c above. The dataset was transformed into two columns (exercise and depression_scale) exercise column is a character data type and has the type of the group (no, which means no exercises, jogging(20min) and jogging(60min). depression_scale column is a numeric data type and has the depression scale values.

The reason for putting the data in two columns is because of the kruskal.test() works with two column only.

Point e:

The appropriate test here is the Kruskal-Wallis test. We have three separate groups of participants, each of whom gives us a single score on a rating scale. Kruskal-Wallis test is used to compare three or more groups on a dependent variable that is measured on at least an ordinal level and so the data are not suitable for a parametric test.

Point f:

```
#  Kruskal-Wallis test
kruskal.test(problem_4$depression_scale ~ problem_4$exercise)

##
##  Kruskal-Wallis rank sum test
##
## data:  problem_4$depression_scale by problem_4$exercise
## Kruskal-Wallis chi-squared = 7.2903, df = 2, p-value = 0.02612
```

We compare our obtained value of H of 7.27 to the table of critical Chi-Square 5.99. With 2 degrees of freedom, a value of Chi-Square as large as 5.99 is likely to occur by chance only 5 times in a hundred. Our obtained value of 7.27 is even larger than this, and so this tells us that our value of H is even less likely to occur by chance. Our H will occur by chance with a probability of less than 0.05.

Point g:

Since p-value < 0.05 at the level of significance(0.05), we reject the null hypothesis.

We would conclude that there is a difference of some kind between our three groups.

References:

Field, A. P., Miles, J., & Field Zoë. (2017). Discovering statistics using R. W. Ross MacDonald School Resource Services Library.

Hole, G. (2021). The kruskal-wallis test - users.sussex.ac.uk. University of Sussex. Retrieved November 30, 2021, from https://users.sussex.ac.uk/~grahamh/RM1web/Kruskal-Wallis%20Handoout2011.pdf.

Peck, R., Short, T., & Olsen, C. (2020). Introduction to statistics and data analysis. Cengage.