HW4

Math 189

Friday of Week 4, 04/22/2022

0. Run the following `R` commands again:

```
# install the packages if needed by using
# install.packages("...")
library(tidyr)
library(readr)
library(tidytuesdayR)
urlRemote <- 'https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/'
pathGithub <- 'data/2020/2020-07-28/'
fileName <- 'penguins.csv'
penguins <- paste0(urlRemote, pathGithub, fileName) %>% read.csv(header = TRUE)
dfr <- drop_na(as.data.frame(penguins))
head(dfr)
```

1. Use `R` to find the following values:

   a) The 0.05 upper quantile of an F distribution with 3 and 6 degrees of freedom.

   b) Assume that in a Hotelling's T^2 test, the value of the Hotelling's T^2 statistic is 10, and $n = 15$, $p = 4$. Find the p-value.

2. Let `dfr` be defined as in Question 0. Let $\boldsymbol{X} = \mathrm{dfr}[1:50, 3:6]$ be the first 50 observations of the dataset for the 3rd to 6th variables. Let $\boldsymbol{\mu}$ be the population mean of these four variables, and let $\boldsymbol{\mu}_0 = (44, 17, 200, 4207)$. Assume that normality assumption is satisfied.

   a) We want to test whether $\boldsymbol{\mu}_0$ is a plausible value for the population mean $\boldsymbol{\mu}$. State the null hypothesis and the two-sided alternative hypothesis.

   b) Calculate the Hotelling's T^2 test statistic.

   c) Give the specific distribution of T^2 under H_0 is true.

   d) Find the p-value this hypothesis test. What conclusion you can make based on this p-value given the significance level is 0.05?

   e) Construct a rejection region given the significance level is 0.1.

   f) Determine the 90% confidence ellipsoid for $\boldsymbol{\mu}$. (Note: Please provide only the mathematical expression of the confidence ellipsoid, and no need to draw the graph.)

   g) Obtain the 90% simultaneous confidence intervals for the four variables.

   h) Obtain the 90% Bonferrini confidence intervals for the four variables.