P8130 Biostatistical Methods Homework 5

Emil Hafeez (eh2928)

11/13/2020

Problem 1

```
#Read the CSV data into a dataframe
antibodies_df <- read.csv("./data/Antibodies.csv")

# Make AgeCategory, Smell, and Gender appopriate datatypes. Helps to ensure we know all unique values,
unique(antibodies_df$AgeCategory)
    antibodies_df <- antibodies_df %>%
        mutate(AgeCategory = factor(AgeCategory, labels = c("18-30", "31-50", "51+") ))
unique(antibodies_df$Smell)
    antibodies_df <- antibodies_df %>%
        mutate(Smell = factor(Smell, levels = c("Normal","Altered", "Unanswered/Others")))
    antibodies_df = antibodies_df %>% filter(Smell != "Unanswered/Others")
unique(antibodies_df$Gender)
    antibodies_df <- antibodies_df %>%
        mutate(Gender = factor(Gender, levels = c("Male","Female")))
```

In order to assess the difference in IgM levels between the two smell factor groups (Normal vs Altered) given non-normal distributions, we opt for a non-parametric test called the Wilcoxon Rank-Sum test. It's the nonparametric equivalent of the two-sample independent t-test, and examines if the medians of the two populations are equal versus not equal:

H_0 = the medians of the two groups' IgM levels are equal, and H_A = the medians of the two groups' IgM levels are not equal. The decision rule is that we reject H_0 = if $T > z_{1-(\alpha/2)}$.

The test statistic is computed, with a continuity correction, one of two ways; with no ties, the test statistic is

and with ties, the test statistic is

The p-value: The p-value is described by $2 * [1 - \Phi(T)]$.

interpret in context

Comment on the differences between the groups (ignoring unanswered and missing values)

Given the non-normal distributions, now you are asked to use an alternative, non-parametric test to assess and comment on the difference in Ig-M levels between the two groups (please ignore unanswered and missing values). You can use R to perform the calculations, but make sure you state test that you used, the hypotheses, test statistic, p-value and provide interpretation in the context of the problem.