BIOS6611-Homework7-Randy-20191024

Randy

10/24/2019

#### Question1a

A study based on 1,500 observations using the SOC identified a median LOS of 9 days. For both hospitals, test if their median LOS is significantly different from the historic median of 9 days using a function in R (e.g., wilcox.exact, binom.test, SIGN.test, etc.) and interpret your results.

library(exactRankTests)

## Package 'exactRankTests' is no longer under development.  
## Please consider using package 'coin' instead.

library(BSDA)

## Loading required package: lattice

##   
## Attaching package: 'BSDA'

## The following object is masked from 'package:datasets':  
##   
## Orange

NewLOS<-c(3, 3, 4, 5, 5, 5, 6, 7, 7, 8, 9, 15)  
SOCLOS<-c(6, 7, 7, 7, 8, 8, 8, 9, 9, 10, 10, 11, 13, 13, 15)  
SIGN.test(NewLOS, m=9)

##   
## One-sample Sign-Test  
##   
## data: NewLOS  
## s = 1, p-value = 0.01172  
## alternative hypothesis: true median is not equal to 9  
## 95 percent confidence interval:  
## 4.106364 7.893636  
## sample estimates:  
## median of x   
## 5.5   
##   
## Achieved and Interpolated Confidence Intervals:   
##   
## Conf.Level L.E.pt U.E.pt  
## Lower Achieved CI 0.8540 5.0000 7.0000  
## Interpolated CI 0.9500 4.1064 7.8936  
## Upper Achieved CI 0.9614 4.0000 8.0000

SIGN.test(SOCLOS, m=9)

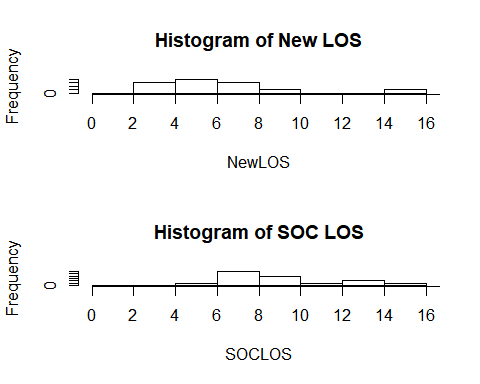
##   
## One-sample Sign-Test  
##   
## data: SOCLOS  
## s = 6, p-value = 1  
## alternative hypothesis: true median is not equal to 9  
## 95 percent confidence interval:  
## 7.178168 10.821832  
## sample estimates:  
## median of x   
## 9   
##   
## Achieved and Interpolated Confidence Intervals:   
##   
## Conf.Level L.E.pt U.E.pt  
## Lower Achieved CI 0.8815 8.0000 10.0000  
## Interpolated CI 0.9500 7.1782 10.8218  
## Upper Achieved CI 0.9648 7.0000 11.0000

For the New LOS median is significantly different from the historic median of 9 days with Sign test; however, the SOC LOS median is the same as the historic median of 9 days with Sign test.

#### Question1b

reate a 2x1 panel figure showing histograms of the LOS for Cauchy General and Skellam Memorial that have a range from 0 to 16 for their x-axis, have bins of width 2 (e.g., “breaks” at 0, 2, 4, 6, …, 14, and 16 in the base R hist function), and informative titles and labels. [Hint: One option to creating a panel figure is to use par( mfrow=c( number rows, number columns ) ).]

par(mfrow=c(2, 1))  
hist(NewLOS, main="Histogram of New LOS", xlim=c(0, 16), xaxt = 'n', breaks=c(seq(0, 16, by=2)))  
axis(1, at = 2\*(0:20))  
hist(SOCLOS, main="Histogram of SOC LOS", xlim=c(0, 16), xaxt = 'n', breaks=c(seq(0, 16, by=2)))  
axis(1, at = 2\*(0:20))



#### Question1c

Carry out a nonparametric procedure for testing the hypothesis that lengths of stay are comparable in the two hospitals by using a function in R (e.g., wilcox.exact, binom.test, SIGN.test, etc.). For your interpretation, if appropriate, reference your histogram from part (b) for discussion of the shape of the two groups.

wilcox.test(NewLOS, SOCLOS, alternative = "two.sided", paired = FALSE, exact = F, correct = F)

##   
## Wilcoxon rank sum test  
##   
## data: NewLOS and SOCLOS  
## W = 33.5, p-value = 0.005517  
## alternative hypothesis: true location shift is not equal to 0

wilcox.exact(x=NewLOS, y=SOCLOS, alternative = "two.sided", paired = FALSE, exact = F, correct = T)

##   
## Asymptotic Wilcoxon rank sum test  
##   
## data: NewLOS and SOCLOS  
## W = 33.5, p-value = 0.005517  
## alternative hypothesis: true mu is not equal to 0

According to the test, the p-value is less than 0.05 so we reject the null hyothesis that the two sample have the same distribution. According to the graphs, we can see the New’s distribution is left shifted than the SOC, which is consistant with the results.

#### Question2

After correction, which SNPs show statistically significant effects?

SNPpvalue <- c(0.04,0.100,0.400,0.550,0.340,0.620,0.001,0.010,0.800,0.005)  
Number <- 1:length(SNPpvalue)  
SNPpvalueAdj <- p.adjust(SNPpvalue, method="BH", n=length(SNPpvalue))  
SNPAdj<-as.data.frame(cbind(Number, SNPpvalue, SNPpvalueAdj))  
SNPAdj[SNPpvalue<=0.05,]

## Number SNPpvalue SNPpvalueAdj  
## 1 1 0.040 0.10000000  
## 7 7 0.001 0.01000000  
## 8 8 0.010 0.03333333  
## 10 10 0.005 0.02500000

SNPAdj[SNPpvalueAdj<=0.05,]

## Number SNPpvalue SNPpvalueAdj  
## 7 7 0.001 0.01000000  
## 8 8 0.010 0.03333333  
## 10 10 0.005 0.02500000

So the FDR adjustment gives a larger p-values. the SNPs still have a p-value less than 0.05 are

#### Question3a

Assume that the variances across the groups are equal and test the hypothesis that there is an overall mean difference in bronchial reactivity among the three lung-function groups.

library(tidyverse)

## -- Attaching packages ---------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.2.1 v purrr 0.3.2  
## v tibble 2.1.3 v dplyr 0.8.3  
## v tidyr 1.0.0 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

## -- Conflicts ------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

FEVFVC <- data.frame(Group = c( rep("A", 5), rep("B", 12), rep("C", 5)), Volume = c(20.8, 4.1, 30.0, 24.7, 13.8, 7.5, 7.5, 11.9, 4.5, 3.1, 8.0, 4.7, 28.1, 10.3, 10.0, 5.1, 2.2, 9.2, 2.0, 2.5, 6.1, 7.5))  
FEVFVCaov <- aov(Volume~Group, data = FEVFVC)  
summary(FEVFVCaov)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Group 2 503.5 251.77 4.989 0.0181 \*  
## Residuals 19 958.8 50.46   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#### Question3b

If justified, compare the means of each pair of groups using the Tukey HSD method and summarize the results. Otherwise note why it isn’t justified.

library(MASS)

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':  
##   
## select

library(DescTools)  
pairwise.t.test(x=FEVFVC$Volum, g=FEVFVC$Group, p.adjust="none", alternative = c("two.sided"))

##   
## Pairwise comparisons using t tests with pooled SD   
##   
## data: FEVFVC$Volum and FEVFVC$Group   
##   
## A B   
## B 0.0151 -   
## C 0.0084 0.4203  
##   
## P value adjustment method: none

TukeyHSD(FEVFVCaov)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = Volume ~ Group, data = FEVFVC)  
##   
## $Group  
## diff lwr upr p adj  
## B-A -10.105 -19.71110 -0.4988964 0.0382469  
## C-A -13.220 -24.63375 -1.8062481 0.0217454  
## C-B -3.115 -12.72110 6.4911036 0.6932026

PostHocTest(FEVFVCaov, method=c('hsd') )

##   
## Posthoc multiple comparisons of means : Tukey HSD   
## 95% family-wise confidence level  
##   
## $Group  
## diff lwr.ci upr.ci pval   
## B-A -10.105 -19.71110 -0.4988964 0.0382 \*   
## C-A -13.220 -24.63375 -1.8062481 0.0217 \*   
## C-B -3.115 -12.72110 6.4911036 0.6932   
##   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

TukeyHSD (Tukey Honest Significant Differences) gives us the adjusted p-values.

#### Question3c

Carry out part (a) assuming that the variances across the groups are not equal. If justified, describe a way to compare the means of each pair of groups, but do not carry out any further analysis.

bartlett.test(log(Volume) ~ Group, data=FEVFVC)

##   
## Bartlett test of homogeneity of variances  
##   
## data: log(Volume) by Group  
## Bartlett's K-squared = 0.16335, df = 2, p-value = 0.9216

FEVFVC\_logaov <- aov(log(Volume) ~ Group, data = FEVFVC)  
summary(FEVFVC\_logaov)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Group 2 3.858 1.9288 3.916 0.0377 \*  
## Residuals 19 9.357 0.4925   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

oneway.test(Volume ~ Group, data=FEVFVC, var.equal=FALSE)

##   
## One-way analysis of means (not assuming equal variances)  
##   
## data: Volume and Group  
## F = 3.9682, num df = 2.0000, denom df = 8.9319, p-value = 0.05845

pairwise.t.test(x=log(FEVFVC$Volum), g=FEVFVC$Group, p.adjust="BH", alternative = c("two.sided"), var.equal=FALSE)

##   
## Pairwise comparisons using t tests with pooled SD   
##   
## data: log(FEVFVC$Volum) and FEVFVC$Group   
##   
## A B   
## B 0.066 -   
## C 0.041 0.298  
##   
## P value adjustment method: BH

First I will run the BOx-Cox test for the transformation, and run the log transformation for the data. After the transformation, based on the Bartlett, the data is with equal variance. Then do the ANOVA test on the log-transformed data.

Otherwise, we can directly use the Welch’s ANOVA for unequal varianced data, and do the pairwised Welch’s t tests for the post hoc tests with approperate adjustment.