solution-5.R

wickhamc

Fri Nov 13 12:31:47 2015

library(reshape2)
library(coin)

## Loading required package: survival

library(ggplot2)

# 1

tips$percent <- with(tips, tip/total\_bill \* 100)

t.test(percent ~ sex, data = tips, var.equal = TRUE)

##
## Two Sample t-test
##
## data: percent by sex
## t = 1.0834, df = 242, p-value = 0.2797
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.7232898 2.4913277
## sample estimates:
## mean in group Female mean in group Male
## 16.64907 15.76505

t.test(percent ~ sex, data = subset(tips, percent < 40), var.equal = TRUE)

##
## Two Sample t-test
##
## data: percent by sex
## t = 1.4994, df = 240, p-value = 0.1351
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.2973175 2.1921332
## sample estimates:
## mean in group Female mean in group Male
## 16.35817 15.41076

wilcox\_test(percent ~ sex, data = tips, conf.int = TRUE)

##
## Asymptotic Wilcoxon-Mann-Whitney Test
##
## data: percent by sex (Female, Male)
## Z = 1.495, p-value = 0.1349
## alternative hypothesis: true mu is not equal to 0
## 95 percent confidence interval:
## -0.2875831 2.2197801
## sample estimates:
## difference in location
## 0.9639608

wilcox\_test(percent ~ sex, data = subset(tips, percent < 40), conf.int = TRUE)

##
## Asymptotic Wilcoxon-Mann-Whitney Test
##
## data: percent by sex (Female, Male)
## Z = 1.4485, p-value = 0.1475
## alternative hypothesis: true mu is not equal to 0
## 95 percent confidence interval:
## -0.3046447 2.1642487
## sample estimates:
## difference in location
## 0.9159365

The Wilcoxon Rank Sum test does appear more resistent to outliers than the two sample t-test.

The changes in the p-value from with outliers to without are:
for t-test: .2797 to .1351
for wilcoxon rank sum test: .1349 to .1475

The changes in the confidence intervals from with outliers to without are:
for t-test: (-0.72, 2.49) to (-0.30 2.19)
for wilcoxon rank sum test: (-0.29, 2.22) to (-0.30, 2.16)

The t-test p-value has almost halved but the Wilcoxon Rank Sun p-value has only changed by about 10%. Since the effect on the p-value of the outliers for the Wilcoxon Rank Sum is small, we say it is resistant. Similarly when we look at the confidence intervals, the endpoints change a lot for the t-test but not much for the Wilcoxon Rank Sum.

# 2

## 1.

library(ggplot2)

qplot(group, weight, data = PlantGrowth) + theme\_grey(10)



# ## 2\.
(avgs <- with(PlantGrowth, tapply(weight, group, mean)))

## ctrl trt1 trt2
## 5.032 4.661 5.526

(sds <- with(PlantGrowth, tapply(weight, group, sd)))

## ctrl trt1 trt2
## 0.5830914 0.7936757 0.4425733

(ns <- with(PlantGrowth, tapply(weight, group, length)))

## ctrl trt1 trt2
## 10 10 10

## 3.

(sp <- sqrt(sum((ns-1)\*sds^2)/sum(ns - 1)))

## [1] 0.6233746

(df <- sum(ns) - length(ns))

## [1] 27

The pooled standard deviation is 0.62 with 27 degrees of freedom.

## 4.

PlantGrowth$overall\_avg <- with(PlantGrowth, mean(weight))
PlantGrowth$group\_avg <- with(PlantGrowth, ave(weight, group))
(ss\_total <- with(PlantGrowth, sum((weight - overall\_avg)^2)))

## [1] 14.25843

(ss\_within <- with(PlantGrowth, sum((weight - group\_avg)^2)))

## [1] 10.49209

The residual sum of squares for the reduced model (total sum of squares) is 14.2584 and the residuals sum of squares for the full model (within group sum of squares) is 10.4921.

## 5. (a)

sp^2 \* df

## [1] 10.49209

## 5. (b)

sd(PlantGrowth$weight)^2 \* (sum(ns) - 1)

## [1] 14.25843

## 6.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source of variation | Sum Sq | Df | Mean Sq | F statistic | p-value |
| **Between groups** | 3.7663 | 2 | 1.8831 | 4.8459 | 0.0159 |
| **Within groups** | 10.4921 | 27 | 0.3886 |  |
| **Total** | 14.2584 | 29 |  |  |

Analysis of Variance Table

## 7.

There is moderate evidence that the treatment effects of the three conditions are not all equal (one-way ANOVA F-test, p-value = 0.016).

## 8.

(ci\_1 <- (avgs["ctrl"] - avgs["trt1"]) + c(-1, 1)\*qt(0.975, df)\*sp\*sqrt(1/ns["ctrl"]+ 1/ns["trt1"]))

## [1] -0.2010126 0.9430126

With 95% confidence the treatment effect of Treatment 1 is between increasing weight by 0.20 units and decreasing weight by 0.94 units, compared to the control.