One way ANOVAs in R

Meredith Dodd <meredith.dodd@woz-u.com>

Tue 8/4/2020 12:53 PM

To: Meredith Dodd <meredith.dodd@woz-u.com>

library("dplyr")

library("rcompanion")

library("car")

library("ggplot2")

library("IDPmisc")

See if the number of people who cross differs by the way they cross

Testing Assumptions

plotNormalHistogram(BorderCrossing\$Value)

That is not normally distributed at all - positively skewed

Square root

BorderCrossing\$ValueSQRT <- sqrt(BorderCrossing\$Value)

plotNormalHistogram(BorderCrossing\$ValueSQRT)

Log

BorderCrossing\$ValueLOG <- log(BorderCrossing\$Value)

BorderCrossing2 <- NaRV.omit(BorderCrossing)

plotNormalHistogram(BorderCrossing2\$ValueLOG)

Log looks great, go with that

Bartlett's Test for homogeneity of variance

bartlett.test(ValueLOG ~ Measure, data=BorderCrossing2)

If it's significant, that's bad, you have violated homogeneity (like we have here)

Sample size - we have 355,000 something rows, so good there

Run the test, use the Welch's test because we violated homogeneity of variance

ANOVA <- Im(ValueLOG ~ Measure, data=BorderCrossing2)

Anova(ANOVA, Type="II", white.adjust=TRUE)

It is significant, meaning that the number of people does vary based on their method of travel across the border

Post hocs

pairwise.t.test(BorderCrossing\$Value, BorderCrossing\$Measure, p.adjust="bonferroni", pool.sd = FALSE)

Every single group differs from every single other group. That is so not helpful!

Look at the means

crossingMeans <- BorderCrossing %>% group_by(Measure) %>% summarize(Mean = mean(Value)) %>% arrange(desc(Mean))

The fewest people come by trains, the most come by car. This needs more wrangling to get the best picture of all, though!

Meredith Dodd, Ph.D. | Data Science Program Chair and Instructor

meredith.dodd@woz-u.com

o: 480-291-8068



https://woz-u.com