

506 Group Project Proposal

Thomas Fiore, Matthew Farrow, Xinje Jiang, Hu Sun

November 13, 2018

Title: Anova for Means with Normality and Medians without Normality

Topics

The topics to cover are:

- Find p -value for significantly different means (One-Way Anova),
- Confidence intervals for differences between means of all groups,
- Visualization of confidence intervals for differences between means of all groups, and
- Kruskal-Wallis Test for significant difference between medians without normality assumption.

Commands

The relevant commands in R for this outline are as follows (using the R Cookbook as a source).

```
df = stack(list(factor1name=v1, factor2name=v2, factor3name=v3))
# creates a two-column data frame df with columns "values" and "ind"
# by concatenating the vectors v1, v2, v3 in the "values" column
# and their factor names in the "ind" column

oneway.test(x~f)
# assuming normality in each group, this will tell if the means of the groups are
# significantly different, where x is a vector of numeric values, and
# f is a factor vector, but this does NOT tell us which groups are
# different, nor the differences

model_object=aov(x~f)
TukeyHSD(model_object)
# computes the difference of means between each 2 groups, and the lower
# and upper endpoints of confidence intervals for those means,
# again assuming normality

model_object=aov(values~ind,data=df)
# this is another way to make the model object if our data is obtained
# by stacking vectors v1,v2,v3 to make df as in the outset above,
# instead of x,f

plot(TukeyHSD(model_object))
# plots the confidence intervals of the differences of means

kruskal.test(x~f)
# if we do NOT have normality in each group, but the distributions of
# the groups have similar shapes, the output p-value will tell us if
# there is a significant difference between the medians of two or more
# groups

In the unlikely event we have time, we might also look at an additional command.

glht(model_object)
# here model_object is the aov output
```

and glht is from the multcomp package

Data for Example: ratSurvival.csv

We would like to use the data from the website

<http://sia.webpopix.org/statisticalTests2.html>

from Marc Lavielle's 2017 course *Statistics in Action with R* at the Ecole Polytechnique (France). He describes the data set as follows.

The health effects of a Roundup-tolerant genetically modified maize, cultivated with or without Roundup, and Roundup alone, were studied during a 2 years study in rats.

For each sex, one control group had access to plain water and standard diet from the closest isogenic non-transgenic maize control; six groups were fed with 11, 22 and 33% of GM NK603 maize either treated or not with Roundup. The final three groups were fed with the control diet and had access to water supplemented with different concentrations of Roundup.

A sample of 200 rats including 100 males and 100 females was randomized into 20 groups of 10 rats of the same sex. Within each group, rats received the same diet. For each sex, there are therefore nine experimental groups and one control group.

The file ratSurvival.csv reports the lifespan (in days) for each animal. Here, the experiment stopped after 720 days. Then, the reported survival time is 720 for those animals who were still alive at the end of the experiment.

The groups means we are interested in are the mean lifespans.

Languages

R (base R commands are listed above)

Stata

MATLAB