Biometry. Lecture 22

Alexey Shipunov

Minot State University

April 27, 2015

> setwd("<working folder>")
>   or
>   "Change dir"
>   in menu!

On Mac, be sure that startup option is working: getwd()
(getwd() checks if R is in working folder, dir() checks the folder content)

Two-dimensional statistics
ANalysis Of VAriation (ANOVA)

Categorical influence and numerical response

- What if there are three species in horsetail data? How to compare the diameter of stem of them all?
- Paired comparisons are "temptation" for p-value, and Bonferroni correction is sometimes of no help.
- There is a solution: analysis of variation (ANOVA) and its non-parametric twin, Kruskal-Wallis test.

ANOVA null and alternative hypotheses and assumptions

- Null is that all groups are not different, alternative is that **at least one group is different from all others**.
- All variables **should be normally distributed**. Small deviations from normality are typically accepted but Kruskal-Wallis test is preferable for all "non-normal" data.
- Variances should be at least **similar**. This is possible to overcome with `oneway.test()`
- Actually, t-test is just an ANOVA for only two groups. Wilcoxon test, in turn, is a special case of Kruskal-Wallis test for two groups only.

Introductory example: eight horsetails

```
> eq8 <- read.table("http://ashipunov.info/data/eq8.txt",
+ h=T)
> str(eq8); head(eq8)
> plot(DIA.ST ~ SPECIES, data=eq8)
> eq8.anova <- lm(DIA.ST ~ SPECIES, data=eq8)
> anova(eq8.anova)
# If variables are not normal:
> kruskal.test(DIA.ST ~ SPECIES, data=eq8)
```

Weight, height and hair color (artificial data)

```
> hwc <- read.table("http://ashipunov.info/data/hwc.txt", h=T)
> str(hwc)
> boxplot(WEIGHT ~ COLOR, data=hwc)
> anova(lm(WEIGHT ~ COLOR, data=hwc))
> kruskal.test(WEIGHT ~ COLOR, data=hwc)
```

kruskal.test() in a non-parametric alternative for anova()
Other related tests are oneway.test(), friedman.test()

Post-hoc tests and plots

```
> pairwise.t.test(hwc$HEIGHT, hwc$COLOR)
> pairwise.wilcox.test(hwc$HEIGHT, hwc$COLOR)
> w.c <- aov(lm(WEIGHT ~ COLOR, data=hwc))
> (w.c.hsd <- TukeyHSD(w.c))
> plot(w.c.hsd) # note confidence intervals, dashed line is 0
```

pairwise.wilcox.test() is a non-parametric post-hoc test
TukeyHSD() runs "Tukey Honest Significant Differences test", one of
the best post-hoc tests
aov() is another method of ANOVA which use in R is less frequent

Are people with different types of hair color differ also by their height?

```
> hwc <- read.table("http://ashipunov.info/data/hwc.txt",
+ h=T)
> str(hwc)
> boxplot(HEIGHT ~ COLOR, data=hwc)
> anova(lm(HEIGHT ~ COLOR, data=hwc))
> h.c <- aov(lm(HEIGHT ~ COLOR, data=hwc))
> (h.c.hsd <- TukeyHSD(h.c))
> plot(h.c.hsd)
```

Finishing...

Save your commands!

(savehistory(<todaysdate>.r) or File -> Save as... on Mac)

Summary: most important commands

- Post-hoc `paiwise.t.test()`, `paiwise.wilcox.test()`, `TukeyHSD()` help to understand **which** groups has a biggest difference
- For ANOVA, check normality AND similarity of variances, or go to `oneway.test()` (still needs normality) or even `kruskal.test()`

For Further Reading

A. Shipunov.
Biometry [Electronic resource].
2012—onwards.
Mode of access:
http://ashipunov.info/shipunov/school/biol_240

A. Shipunov, and others.
Visual statistics. Use R!
Ongoing translation from Russian.