

ANOVA and multiple comparison

Minato Nakazawa, Ph.D. (Department of Public Health)

June 5, 2019

1 Testing the difference of locations among 3 or more groups

The Analysis of Variance (ANOVA) has already been done by the homework from 26 May to 9 June, for the parallel group comparison design. Today, including such simple ANOVA, testing the location parameters among 3 or more groups, with *post hoc* test of multiple pairwise comparisons of location parameters.

To compare the means among 3 or more groups, you must not simply repeat *t*-tests for all possible pairs. In the statistical hypothesis testing, setting significance level as 5% in each comparison provides much more type I error for overall comparisons.

To solve this problem, there are two different approaches. (1) Evaluate the effect of a group variable on a quantitative variable. (2) Adjust the type I errors for multiple comparisons. Traditionally these approaches are conducted as two-steps: Only when the effect of the group is statistically significant, the pairwise comparisons with adjustment for multiple comparisons will be done. In that meaning, the latter step is called as *post hoc* analysis of the former step. However, it is recommended now that you should select more appropriate approach according to the purpose of the analysis.

1.1 One-way Analysis of Variance (ANOVA)

The typical analysis of the former approach to test the means among 3 or more groups is one-way analysis of variance (ANOVA). The one-way ANOVA decomposes the variance of the data into the variance by the group and the variance of random errors. If the ratio of these variance is significantly different from 1, we can judge the effect of the group variable is statistically significant.

The movie in the YouTube, https://www.youtube.com/watch?v=-yQb_ZJnFXw will give a very detailed explanation of the calculation.

For example, the heights of 37 adult males in 3 villages in the South Pacific were as follows. You can read the data from the web site^{*1}.

*1 <http://minato.sip21c.org/grad/sample2.dat>

Unique ID (PID)	Village (VG)	Height (cm) (HEIGHT)	Weight (kg) (WEIGHT)
10101	X	161.5	49.2
10201	X	167.0	72.8
⋮			
30301	Z	166.0	58.0
⋮			
70312	Y	155.5	53.6

In Rgui console, to conduct one-way ANOVA of the null-hypothesis that VG has no significant effect on HEIGHT, type as follows.

```
sp <- read.delim("http://minato.sip21c.org/grad/sample2.dat")
summary(aov(HEIGHT ~ VG, data=sp))
```

Then you get the following result, so-called “ANOVA summary table”.

```

          Df Sum Sq Mean Sq F value    Pr(>F)
VG          2  422.72   211.36   5.7777 0.006918 **
Residuals  34 1243.80    36.58
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Here the number of * at right is the codes for significance shown in the bottom line, but the $\Pr(>F)$, the significant probability, is more important. The column of Sum Sq means the sum of squared difference of each value with the mean. The value of VG’s Sum Sq is 422.72, which is the sum of number of individuals times the squared difference between the mean in each village and the overall mean. This value is called as inter-group (inter-class) variation. The Residuals’s Sum Sq, 1243.80, is the sum of squared difference between each individual’s height and the mean height of the village where the individual belongs to, which is called as within-group (within-class) variation or error variation. The Mean Sq is mean squared difference, which is the value of Sum Sq divided by Df, where Df is degree of freedom. The Mean Sq is the variance, so that VG’s Mean Sq, 211.36 is the inter-group variance and Residuals’s Mean Sq, 36.58 is the error variance. Then, the F value is the ratio of variances, which is the value of inter-group variance divided by error variance. The ratio of variances obeys F -distribution, The significant probability $\Pr(>F)$ is obtained from F -distribution. Here the $\Pr(>F)$ is 0.00692, so that the effect of VG can be judged as statistically significant at 5% significance level. We can reject the null-hypothesis, and the heights of them significantly differ by village.

In EZR, to read `sample2.dat` into `sp` data set, select [File], [Import data], [from text file, clipboard, or URL...], then type `sp` in the box of “Enter name for data set:”, check the ratio button beside “Internet URL”, and check the radio button beside [Tabs] of “Field Separator” and click [OK], then type `http://minato.sip21c.org/grad/sample2.dat` as Internet URL and click [OK].

Next, select [Statistical analysis], [Continuous variables], [One-way ANOVA...], then select HEIGHT as [Response Variable (pick one)] and VG as [Grouping Variables (pick at least one)], and click [OK].

Traditionally speaking, one-way ANOVA should be done only after testing the null-hypothesis that the variances of groups are not different (by Bartlett’s test or other test). In Rgui console, you can do so by typing `bartlett.test(HEIGHT ~ VG, data=sp)`. The obtained p-value, 0.5785 means that we cannot reject null-hypothesis. After confirming it, we can safely apply one-way ANOVA.

In EZR, select [Statistical analysis], [Continuous variables], [Bartlett’s test], then select HEIGHT as [Response Variable] and VG as [Groups], then click [OK].

However, such kind of two-steps analysis may cause multiple comparisons problem, so that Welch’s extended one-way ANOVA without testing equal variance is the most appropriate.

In Rgui console, type `oneway.test(HEIGHT ~ VG, data=sp)`, then you can get the p-value by the Welch’s extended one-way ANOVA. The p-value, 0.004002 is less than 0.05, so that we can judge the effect of village on the height is statistically significant at 5% significance level. Unfortunately, EZR and Rcmdr do not support `oneway.test()` now, unless typing the code into script window and submitting.

1.2 Kruskal-Wallis test and Fligner-Killeen test

The Kruskal-Wallis test is the nonparametric test to compare the locations of distributions among 3 or more groups, so that it seems a nonparametric alternative of ANOVA.

In Rgui console, type `kruskal.test(HEIGHT ~ VG, data=sp)`.

In EZR, select [Statistical analysis], [Nonparametric tests], [Kruskal-Wallis test], then select VG as [Groups] and HEIGHT as [Response Variable], and click [OK].

The Fligner-Killeen test is the test the null hypothesis that the variances in each of the groups are the same in nonparametric manner. It seems a nonparametric alternative of Bartlett’s test.

In Rgui console, type `fligner.test(HEIGHT ~ VG, data=sp)`. It is not supported in EZR nor Rcmdr, unless typing the code in script window and submitting.

1.3 Pairwise comparisons with adjustment of multiple comparisons

There are many methods to adjust type I errors such as Bonferroni's method, Holm's method, Tukey Honest Significant Differences (Tukey's HSD), and so on. Except Tukey's HSD, many adjustment methods are applicable in `pairwise.t.test()`, `pairwise.wilcox.test()`, or `pairwise.prop.test()`, where the `pairwise.prop.test()` is to test the proportions of 3 or more groups.

To compare means, pairwise *t*-test with Holm's adjustment or Tukey's HSD can be used. To compare medians, pairwise Wilcoxon's rank sum test with Holm's adjustment is appropriate.

For example, compare the means of height between the all pairs among 3 villages in `sp` data set.

In Rgui console, type either of the following lines:

```
pairwise.t.test(sp$HEIGHT, sp$VG, p.adjust.method="holm")
TukeyHSD(aov(HEIGHT ~ VG, data=sp))
```

In EZR, Select [Statistical analysis], [Continuous variables], [One-way ANOVA], then select VG as [Groups] and HEIGHT as [Response Variable] and check the box beside appropriate pairwise comparison methods (Bonferroni, Holm or Tukey), then click [OK]. After the result of ANOVA, the results of multiple comparisons will appear in the Output Window.

For non-parametric multiple comparison, Bonferroni, Holm or Steel-Dwass (non-parametric analogue of Tukey) can be specified in [Kruskal-Wallis test] menu.

Since almost all multiple comparison tests assume the equal variances for all groups, if the option of [No (Welch test)] for "Assume equal variances?" is selected, no pairwise (multiple) comparison is available in EZR. However, the method of Games-Howell does not assume the equal variances, so that it can be applied for the data with non-equal variances. Unfortunately, the Games-Howell method is not included in EZR.

Recently developed method for multiple comparison, controlling the **false discovery rate (FDR)** among the rejected hypotheses (Benjamini Y and Hochberg Y, "Controlling the false discovery rate: A practical and powerful approach to multiple testing." *J. Royal Stat. Soc. Ser. B*, 57(1): 289-300, 1995). It can be done by typing "fdr" instead of "holm" or "bon" in the `p.adjust.method=` option of `pairwise.t.test` or `pairwise.wilcox.test` functions.

To apply Games-Howell's method, there are several possibility: <http://aoki2.si.gunma-u.ac.jp/R/tukey.html> (http://www.gcf.dkf.unibe.ch/BCB/files/BCB_10Jan12_Alexander.pdf provides a good explanation for the script), <http://www.hs.hirosaki-u.ac.jp/~pteiki/research/stat/S/> (using customized Rcmdr), <https://gist.github.com/aschleg/ea7942efc6108aedfa9ec98aeb6c2096> (original script) or `posthocTGH()` function

of `userfriendlyscience` package.

When we use the `userfriendlyscience` package for the multiple comparison of chicken's weights among the 6 feeding types of the built-in data `chickwts`, you can type as follows.

```
if (require(userfriendlyscience)==FALSE) {
  install.packages("userfriendlyscience", dep=TRUE)
  library(userfriendlyscience) }
posthocTGH(chickwts$weight, chickwts$feed, method="games-howell")
```

1.4 Dunnett's multiple comparisons

Dunnett's method to adjust multiple comparison of means is used in the comparison of multiple treatment groups with a common control group. For example, randomly assign the 15 hypertension patients to 3 groups (each 5), which are treated by placebo, by usual drug, and by new drug. After 1 month treatment, the decrease of systolic blood pressures (mmHg) were [5, 8, 3, 10, 15] for the placebo group, [20, 12, 30, 16, 24] for the usual drug group, and [31, 25, 17, 40, 23] for the new drug group.

Here we can apply the Dunnett's multiple comparisons to compare the mean reductions of systolic blood pressure of placebo group with both usual drug group and new drug group. In Rgui console, type below. Unfortunately, it is not supported in Rcmdr.

```
bpdown <- data.frame(
  medicine=factor(c(rep(1,5),rep(2,5),rep(3,5))),
  labels=c("placebo","usualdrug","newdrug")),
  sbpchange=c(5, 8, 3, 10, 15, 20, 12, 30, 16, 24, 31, 25, 17, 40, 23))
summary(res1 <- aov(sbpchange ~ medicine, data=bpdown))
library(multcomp)
res2 <- glht(res1, linfct = mcp(medicine = "Dunnett"))
confint(res2, level=0.95)
summary(res2)
```

Basically, after loading `multcomp` library, give the resulted object of ANOVA to `glht()` function and specify "Dunnett" for `linfct=` option.

In EZR, you can import the data from <http://minato.sip21c.org/bpdown.txt>. Then select [Statistical analysis], [Continuous variables], [One-way ANOVA], then select medicine as [Groups] and sbpchange as [Response Variable] and check the box beside the pairwise comparison (Dunnett), then click [OK]. After the result of ANOVA, the results of multiple comparisons will appear in the Output Window.

For non-parametric multiple comparison, Steel (non-parametric analogue of Dunnett) can be specified in [Kruskal-Wallis test] menu.