Analysis of variance (ANOVA)

September 6, 2017

we consider comparisons among more than two groups parametrically, using analysis of variance, as well as nonparametrically, using the KruskalWallis test. Furthermore, we look at two-way analysis of variance in the case of one observation per cell.

One-way analysis of variance

Let x_{ij} denote observation no. j in group i, so that x_{35} is the fifth observation in group 3; $\bar{x}_i$ is the mean for group i, and $\bar{x}_\cdot$ is the grand mean (average of all observations). We can decompose the observations as

$$x_{ij} = \bar{x}. + (\bar{x}_i - \bar{x}_\cdot) + (x_{ij} - \bar{x}_i)$$

informally corresponding to the model $X_{ij} = \mu + \alpha_i + \epsilon_{ij}, \qquad \epsilon_{ij} \sim N(0, \sigma^2)$
in which the hypothesis that all the groups are the same implies that all α_i are zero.
Now consider the sums of squares of the underbraced terms, known as *variation within groups*

$$SSDW = \sum\sum (x_{ij} - \bar{x}_i)^2 \tag{0.1}$$

and *variation between groups*

$$SSDB = \sum\sum (\bar{x}_i - \bar{x}_\cdot)^2 \tag{0.2}$$

It is possible to prove that

$$SSDB + SSDW = SSD_{total}$$

That is, the total variation is split into a term describing differences between group means and a term describing differences between individual measurements within the groups.
Accordingly, you can normalize the sums of squares by calculating mean squares:

$$MSW = SSDW/(N-k)$$

$$MSB = SSDB/(k-1)$$

MSW is the pooled variance obtained by combining the individual group variances and thus an estimate of σ^2. In the absence of a true group effect, MSB will also be an estimate of σ^2, but if there is a group effect, then the differences between group means and hence MSB will tend to be larger. Thus, a test for significant differences between the group means can be performed by

comparing two variance estimates. This is why the procedure is called *analysis of variance* even though the objective is to compare the group means. You calculate

$$F = MSB/MSW$$

Simple analyses of variance can be performed in R using the function *lm*, which is also used for regression analysis. For more elaborate analyses, there are also the functions *aov* and *lme* (linear mixed effects models, from the *nlme* package).

Example

Let consider the red cell folate data (IsWR package) from Altman (1991, p. 208). To use lm, it is necessary to have the data values in one vector and a factor variable describing the division into groups. The red.cell.folate data set contains a data frame in the proper format.

```
> ?red.cell.folate
> attach(red.cell.folate)
> summary(red.cell.folate)
folate          ventilation
Min.   :206.0   N2O+O2,24h:8
1st Qu.:249.5   N2O+O2,op :9
Median :274.0   O2,24h    :5
Mean   :283.2
3rd Qu.:305.5
Max.   :392.0
```

Recall that *summary* applied to a data frame gives a short summary of the distribution of each of the variables contained in it. The format of the summary is different for numeric vectors and factors, so that provides a check that the variables are defined correctly.

```
> anova(lm(folate~ventilation))
Analysis of Variance Table

Response: folate
Df Sum Sq Mean Sq F value  Pr(>F)
ventilation  2  15516  7757.9  3.7113 0.04359 *
Residuals   19  39716  2090.3
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Here we have *SSDB* and *MSB* in the top line and SSDW and MSW in the second line. In statistics textbooks, the sums of squares are most often labelled *between groups* and *within groups*. Like most other statistical software, R uses slightly different labelling. Variation between groups is labelled by the name of the grouping factor (*ventilation*), and variation within groups is labelled *Residual*.

Example

consider the data set juul (IsWR package). Notice that the *tanner* variable in this data set is a numeric vector and not a factor.

```
> attach(juul)
> anova(lm(igf1~tanner))                                    ## WRONG!
Analysis of Variance Table

Response: igf1
Df   Sum Sq  Mean Sq F value    Pr(>F)
tanner      1 10985605 10985605  686.07 < 2.2e-16 ***
Residuals 790 12649728    16012
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

The DF related to *tanner* does not reflect the number of levels for this covariate. Things can be fixed as follows:

```
> juul$tanner <- factor(juul$tanner,labels=c("I","II","III","IV","V")) ##Let fix this issue
> detach(juul)    ##### drop the previous data
> attach(juul)      #### consider the new one
> summary(tanner)
I   II  III   IV    V NA's
515  103   72   81  328  240

> anova(lm(igf1~tanner))    # Use anova on this fitted model
Analysis of Variance Table

Response: igf1
Df   Sum Sq Mean Sq F value    Pr(>F)
tanner      4 12696217 3174054  228.35 < 2.2e-16 ***
Residuals 787 10939116   13900
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Pairwise comparisons and multiple testing

If the F test shows that there is a difference between groups, the question quickly arises of where the difference lies. It becomes necessary to compare the individual groups.
Part of this information can be found in the *regression coefficients*. You can use *summary* to extract regression coefficients with standard errors and t tests. These coefficients *do not have their usual meaning* as the slope of a regression line but have a special interpretation, which is described below.

```
> summary(lm(folate~ventilation))

Call:
lm(formula = folate ~ ventilation)

Residuals:
Min      1Q  Median      3Q     Max
-73.625 -35.361  -4.444  35.625  75.375

Coefficients:
```

```
Estimate Std. Error t value Pr(>|t|)
(Intercept)            316.63      16.16  19.588 4.65e-14 ***
ventilationN2O+O2,op   -60.18      22.22  -2.709   0.0139 *
ventilationO2,24h      -38.63      26.06  -1.482   0.1548
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 45.72 on 19 degrees of freedom
Multiple R-squared:  0.2809,Adjusted R-squared:  0.2052
F-statistic: 3.711 on 2 and 19 DF,  p-value: 0.04359
```

The interpretation of the estimates is that the intercept is the mean in the first group *(N2O+O2,24h)*, whereas the two others describe the difference between the relevant group and the first one. Among the t tests in the table, you can immediately find a test for the hypothesis that the first two groups have the same true mean *($p = 0.0139$)* and also whether the first and the third might be identical ($p = 0.1548$). However, a comparison of the last two groups cannot be found.

Compare all groups

If we want to compare all groups, we ought to correct for multiple testing. Performing many tests will increase the probability of finding one of them to be significant; that is, the p-values tend to be exaggerated. A common adjustment method is the *Bonferroni correction*, which is based on the fact that the probability of observing at least one of n events is less than the sum of the probabilities for each event. Thus, by dividing the significance level by the number of tests or, equivalently, multiplying the p-values, we obtain a conservative test where the probability of a significant result is less than or equal to the formal significance level.
A function called *pairwise.t.test* computes all possible two-group comparisons. It is also capable of making adjustments for multiple comparisons and works like this:

```
> pairwise.t.test(folate, ventilation, p.adj="bonferroni")

Pairwise comparisons using t tests with pooled SD

data:  folate and ventilation

N2O+O2,24h N2O+O2,op
N2O+O2,op 0.042      -
O2,24h    0.464      1.000

P value adjustment method: bonferroni
```

The default method for *pairwise.t.test* is actually not the *Bonferroni correction* but a variant due to Holm.

R-Code

```
?red.cell.folate
attach(red.cell.folate)
summary(red.cell.folate)
anova(lm(folate~ventilation))
```

```
attach(juul)
anova(lm(igf1~tanner)) ## WRONG!
juul$tanner <- factor(juul$tanner,labels=c("I","II","III","IV","V")) ##Let fix this issue
detach(juul)    ##### drop the previous data
attach(juul)      #### consider the new one
summary(tanner)
anova(lm(igf1~tanner))    # Use anova on this fitted model
################################################ comparison when we reject Ho
summary(lm(folate~ventilation))
pairwise.t.test(folate, ventilation, p.adj="bonferroni")
?pairwise.t.test
pairwise.t.test(folate,ventilation)   ### default method is Holm
```

Exercise

1. In the *lung*data (located in the IsWR package), do the three measurement methods give systematically different results? If so, which ones appear to be different?