Repeated Measures Analysis of Variance (ANOVA)

Note on required packages: The following code required the package `fastR` to reorganize a data frame. If you have not already done so, download, install, and load the library with the following code:

```
install.packages("fastR")
```

```
library("fastR")
```

Repeated Measures Analysis of Variance (ANOVA) or **Within-Variation ANOVA** is a statistical method to estimate how the means differ for two or more outcome variable measured on the same scale.

It can be viewed as an extension the paired-samples t-test for difference in mean which estimates how the means for **two** variables differ from another.

Alternatively, it can be compared to a one-way ANOVA, where instead of comparing means across two or more independent groups, you compare means across two or more matched groups.

1. Assumptions

The following assumptions are necessary for the one-way ANOVA test:

- **Interval / Ratio scale**
- **Matched groups:** Each dependent variable is measured on the same scale and includes observations from exactly the same sampling units.
- **Sphericity / Homogeneity of variance:** Identical population variances for the differences in the outcome variables between each pair of categories
- **Central Limit Theorem assumptions:** There is a sufficiently large sample size and/or the population distribution for the outcome variable is normal

With **matched samples**, we have a single set of observations, but two or more measures for each observation. Obtaining two or more measures from each observation can result from the following situations:

1. **Across-time measures:** The same dependent variable is measured for each individual but at multiple time periods. For example, a sample of individuals may have their income measures once in 2013, again in 2014, and finally in 2015, and a researcher could ask whether there are differences in average income from one year to the next.

2. **Different conditions measures:** The same dependent variable is measured for each individual, but under multiple conditions. For example, a manager may measure worker productivity following various strategies to improve teamwork and morale.

3. **Related topics measures:** Two or more slightly different variables are measured for each individual. For example, a sample of students may have standardized test scores measured on the same scale for multiple academic subjects. A researcher could ask whether the test scores are different between different subjects.

2. Hypotheses

The null and alternative hypotheses for the RANOVA test are the following,

Null hypothesis The population mean for outcome variable is equal for each repeated-measures category

Alternative hypothesis The population mean for the outcome variable for at least one of the repeated-measures categories is different from the others

3. Example: Job Satisfaction

The data set below includes data on job satisfaction from 203 public sector employees in Israel. There are many variables in the data set. We will focus on the satisfaction variables that follow. Each of these variables are measured on an interval scale, where higher scores indicate higher levels of satisfaction regarding that characteristic.

- `Colleague`: Satisfaction employees have with their colleagues
- `Supervision`: Satisfaction employees have with their supervisors
- `Promotion`: Satisfaction employees have with the opportunities for promotion
- `Salary`: Satisfaction employees have with their salary

The following code downloads the data, opens it, and saves it in a data frame that we call `dat`.

```
dat <- read.csv("http://murraylax.org/datasets/jobsat.csv")
```

We can use this data to determine if average employee satisfaction is different for these aspects of the job.

4. Transforming the data set

The functions we use below to estimate a repeated measures ANOVA requires the data in a different format. Instead of having a single row for each sampling unit and multiple columns for job satisfaction characteristic, we need to transform the data so that there is a single outcome variable (single column) for job satisfaction. This means there will be four rows for each sampling unit. We also need a new categorical variable that identifies what type of job satisfaction that each row of our new single outcome variable refers to.

First we will create a subject ID variable that simply gives a unique identifier to each unique sampling unit in our original data frame. When we reformat the data frame, this variable can be used to determine with rows belong to the original sampling units. The code below creates a new variable for subject ID called `subid` and simply numbers them from 1 to the number of observations in the data frame (given by `nrow(dat)`.

```
dat$subid <- 1:nrow(dat)
```

The following call to `make.rm()` re-formats our data frame and saves the result in a new data frame object we call `redat`.

```
redat <- make.rm(constant=c("subid"),
          repeated = c("Colleague", "Supervision", "Promotion", "Salary"), data = dat)
```

We passed three parameters to the function `make.rm()`. The first parameter, `constant`, can refer to a single variable or a list of variables *besides the repeated-measures variables* to retain in the new data frame. The only variable we decided to copy to the new data frame was the subject id, `subid`. The second parameter, `repeated`, should be a list of variables that correspond to the repeated measures outcome. The variables `Colleague`, `Supervision`, `Promotion`, and `Salary` all refer to different types of employee satisfaction, measured on the same scale, and measured for each sampling unit. The final parameter, `data`, tells `make.rm()` in what data frame it can find all of these variables.

A call to the function `head()` will print to the screen the first few rows of a data frame. Let us look at what our new data frame looks like.

```
head(redat)
```

```
##   cons.col repdat contrasts
## 1        1     45        T1
```

```
## 2           2     61       T1
## 3           3     61       T1
## 4           4     45       T1
## 5           5     54       T1
## 6           6     50       T1
```

Unfortunately, our variable names were not preserved. The first column, named `cons.col` is the column of subject IDs. The second column, named `repdat` is the data for our repeated measures variable. This is the job satisfaction data. The final column, named `contrasts`, is a text description of which repeated measures category the observation belongs to. The labels "T1", "T2", "T3", and "T4" correspond to "Colleague", "Supervision", "Promotion", and "Salary", respectively.

We can repair our variable names and our category levels as follows:

```r
names(redat) <- c("subid", "satisfaction", "category")
levels(redat$category) <- c("Colleague", "Supervision", "Promotion", "Salary")
```

Now our data frame looks pretty and is in a format to estimate a RANOVA.

```r
head(redat)
```

```
##   subid satisfaction  category
## 1     1           45 Colleague
## 2     2           61 Colleague
## 3     3           61 Colleague
## 4     4           45 Colleague
## 5     5           54 Colleague
## 6     6           50 Colleague
```

5. Estimating the RANOVA Model

The following call to `aov()` can be used to estimate a repeated measures ANOVA.

```r
aovsat <- aov(satisfaction ~ category + Error(subid), data=redat)
```

The first parameter is a *formula* that says `satisfaction` in an outcome variable that depends on the categorical variable `category`. This formula is finished with `+ Error(subid)` to tell `aov()` we do not have independent groups, but rather the observations corresponding to the original subject ID (`subid`) are dependent on one another.

We can report a summary of the repeated measures analysis of variance:

```r
summary(aovsat)
```

```
##
## Error: subid
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals  1   1889    1889
##
## Error: Within
##            Df Sum Sq Mean Sq F value Pr(>F)
## category    3 170351   56784   617.3 <2e-16 ***
## Residuals 807  74237      92
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value is less than 2×10^{-16}. We reject the null hypothesis and conclude that we have sufficient statistical evidence that the mean job satisfaction is different for at least one aspect of employees' jobs.

6. Post-Hoc Tests

Having established that at least one aspect of employees' jobs has a different level of job satisfaction than the others, we can now determine which ones are different from the others. We can do this by running a series of paired-samples t-tests for each pair of satisfaction categories. Since there are 4 racial categories, there are 6 different pairs, and therefore 6 different paired samples t-tests to run.

Remember that any hypothesis test can lead to incorrectly rejecting the null hypothesis by statistical chance. This is called a Type 1 error. The probability of making a type 1 error is equal to the significance level you use for the hypothesis test, usually 5%. When we run 6 paired-samples t-tests, the probability of making a type 1 error in *at least one of the 6 tests* is larger. For example, if the chances of making a type 1 error are independent across the 6 statistical tests, the probability of making a type 1 error in at least one of the tests is more than 25%!

The **Bonferonni** p-value correction lets you run a series t-tests for differences in means so that the probability of making one or more type 1 errors over the 6 tests is not more than 5%.

The function call `pairwise.t.test()` that follows estimates paired-samples tests difference for means for each pair of job satisfaction categories, making a Bonferonni corrections to the p-values.

```
pairwise.t.test(redat$satisfaction, redat$category, p.adjust.method="bonf", paired = TRUE)
```

```
##
##  Pairwise comparisons using paired t tests
##
## data:  redat$satisfaction and redat$category
##
##            Colleague Supervision Promotion
## Supervision 1.8e-07   -           -
## Promotion   < 2e-16   < 2e-16     -
## Salary      < 2e-16   < 2e-16     < 2e-16
##
## P value adjustment method: bonferroni
```

The result is a matrix of p-values for each pairwise test. All the p-values are less than 5%, so we can conclude that we have statistical evidence that every type of job satisfaction has a different mean than every other other type.

Finally, we may wish to review the sample means for job satisfaction for each of our categories. We can do this with a call to `mean()`.

```
mean(redat$satisfaction ~ redat$category)
```

```
##   Colleague Supervision   Promotion      Salary
##    48.81281    54.36453    30.24138    18.14286
```