



The following resources are associated:
 Checking normality in R, ANOVA in R, Interactions and the Excel dataset 'Diet.csv'

Two-way (between-groups) ANOVA in R

Dependent variable: Continuous (scale/interval/ratio),

Independent variables: Two categorical (grouping factors)

Common Applications: Comparing means for combinations of two independent categorical variables (factors).

Data: The data set Diet.csv contains information on 78 people who undertook one of three diets. There is background information such as age, gender (Female=0, Male=1) and height. The aim of the study was to see which diet was best for losing weight but it was also thought that the best diets for males and females may be different so the independent variables are diet and gender. To open the file use the `read.csv()` command.

| | gender | Age | Height | pre.weight | Diet | weight6weeks |
|---|--------|-----|--------|------------|------|--------------|
| 1 | NA | 41 | 171 | 60 | 2 | 60.0 |
| 2 | NA | 32 | 174 | 103 | 2 | 103.0 |
| 3 | 0 | 44 | 174 | 58 | 2 | 60.1 |
| 4 | | 37 | 172 | 58 | | 56.0 |
| 5 | | | 159 | | | 54.2 |

Female = 0

Diet 1, 2 or 3

You will need to change the command depending on where you have saved the file.

```
dietR<-read.csv("D:\\diet.csv",header=T,sep=",")
```

Tell R to use the diet dataset until further notice using `attach(dataset)` so 'Height' can be used instead of `dietR$Height` for example.

```
attach(dietR)
```

Calculate the weight lost by person (difference in weight before and after the diet) and add the variable to the dataset. Then attach the data again.

```
dietR$weightlost<-pre.weight-weight6weeks
```

```
attach(dietR)
```

There are three hypotheses with a two-way ANOVA. There are the tests for the main effects (diet and gender) as well as a test for the interaction between diet and gender.



Checking the assumptions for two-way ANOVA

| Assumptions | How to check | What to do if the assumption is not met |
|--|---|---|
| Residuals should be normally distributed | Save the residuals from the <code>aov()</code> command output and produce a histogram or conduct a normality test (see checking normality in R resource) | If the residuals are very skewed, the results of the ANOVA are less reliable. Try to fix them by using a simple or Box-Cox transformation or try running separate ANOVAs or Kruskal-Wallis tests by one independent (e.g. gender). |
| Homogeneity of Variance (Levene's Test) | Use the Levene's test of equality of variances through the <code>leveneTest()</code> command (See the One way ANOVA in R resource) | If $p < 0.05$, the results of ANOVA are less reliable. There is no equivalent test but comparing the p-values from the ANOVA with 0.01 instead of 0.05 is acceptable. |

Steps in R

To carry out a two way ANOVA with an interaction, use

`aov(dependent~as.factor(independent1)*as.factor(independent2), data=filename)` and give the ANOVA model a name e.g. **anova2**. The `as.factor()` tells R that the two independents are categorical.

```
anova2<-aov(weightlost~
as.factor(gender)*as.factor(Diet), data=dietR)
```

Ask for the residuals (difference between each individual and their diet/gender combination mean) and give them a name (res).

```
res<-anova2$residuals
```

Produce a histogram of the residuals.

```
hist(res, main="Histogram of
residuals", xlab="Residuals")
```

The Levene's test for equality of variances is in the additional 'car' package.

```
library(car)
```

If this command does not work, go to Packages -->

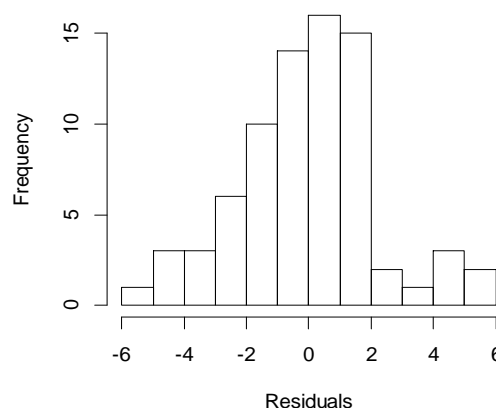
Install package(s) and select the UK (London)

CRAN mirror. Once loaded, carry out Levene's test.

```
leveneTest(weightlost~
as.factor(gender)*as.factor(Diet), data=dietR)
```

```
> leveneTest(weightlost~as.factor(gender)*as.factor(Diet), data=dietR)
Levene's Test for Homogeneity of Variance (center = median)
  Df F value Pr(>F)
group 5  0.3867 0.8563
    70
```

Histogram of residuals



The p value is 0.8563 which is greater than 0.05, so equal variances can be assumed.



The ANOVA output

To view the ANOVA table use the `summary()` command

```
summary(anova2)
```

```
> summary(anova2)
```

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------------------------------|----|--------|---------|---------|------------|
| as.factor(gender) | 1 | 0.3 | 0.278 | 0.052 | 0.82062 |
| as.factor(Diet) | 2 | 60.4 | 30.209 | 5.619 | 0.00546 ** |
| as.factor(gender):as.factor(Diet) | 2 | 33.9 | 16.952 | 3.153 | 0.04884 * |
| Residuals | 70 | 376.3 | 5.376 | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Main effect of gender
p = 0.82

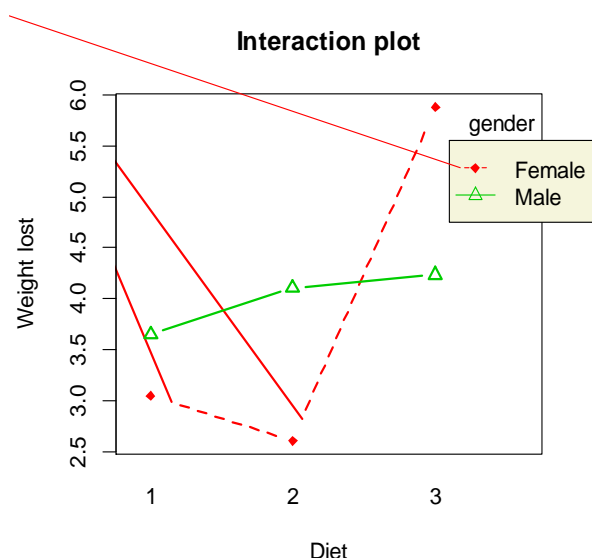
Main effect of diet
p = 0.005

Interaction p = 0.049

The results of the two-way ANOVA and post hoc tests are reported in the same way as one way ANOVA for the main effects and the interaction e.g. there was a statistically significant interaction between the effects of Diet and Gender on weight loss [$F(2, 70)=3.153$, $p = 0.049$]. Since the interaction effect is significant ($p = 0.049$), the 'Diet' effect cannot be generalised for both males and females together.

The easiest way to interpret the interaction is to use a means or interaction plot which shows the means for each combination of diet and gender (see the **Interactions** resource for more details). Before producing an interaction plot, tell R the labels for gender.

```
gender<-factor(gender,c(0,1), labels=c('Female','Male'))
```



There are lots of options for line colour, style and legend placing with an interaction plot.

Use `?interaction.plot` to find out more.

To produce this interaction plot use:

```
interaction.plot(Diet,gender,weight
lost,type="b",col=c(2:3),leg.bty="o",
leg.bg="beige",lwd=2,pch=c(18,24),
xlab="Diet",ylab="Weight
lost",main="Interaction plot")
```

The means (or interaction) plot clearly shows a difference between males and females in the way that diet affects weight lost, since the lines are not parallel. The differences between the mean weight lost on the diets is much bigger for females.

Post hoc tests

The `TukeyHSD(anova2)` command will produce post hoc tests for the main effects and interactions. Only interpret post hoc tests for the significant factors from the ANOVA. If the interaction is NOT significant, interpret the post hoc tests for significant main effects but if it is significant, only interpret the interactions post hoc tests.

Post hoc tests for main effects of diet and gender



```
> TukeyHSD(anova2)
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = weightlost ~ as.factor(gender) * as.factor(Diet), data = dietR)

$`as.factor(gender)`
      diff      lwr      upr    p adj
1-0 0.1221283 -0.9480861 1.192343 0.8206233

$`as.factor(Diet)`
      diff      lwr      upr    p adj
2-1 -0.03484966 -1.6215073 1.551808 0.9984761
3-1  1.84475570  0.2871469 3.402365 0.0162482
3-2  1.87960536  0.3385771 3.420634 0.0128844
```

The interaction was significant so the main effects are not interpreted here but if your data does not have a significant interaction,

interpret these in the same way as post hoc tests on the one-way ANOVA resource.

The following output for post hoc interactions tests has been adjusted in Excel to make it easier to read.

| | Groups being compared | Difference in means | Lower confidence interval | Upper confidence interval | Adjusted p-value |
|--|-----------------------|---------------------|---------------------------|---------------------------|------------------|
| | | diff | lwr | upr | p adj |
| comparisons with females on diet 1 (0:1) | 1:1-0:1 | 0.60 | -2.21 | 3.41 | 0.9888 |
| | 0:2-0:1 | -0.44 | -3.01 | 2.13 | 0.9958 |
| | 1:2-0:1 | 1.06 | -1.68 | 3.80 | 0.8657 |
| | 0:3-0:1 | 2.83 | 0.31 | 5.35 | 0.0191 |
| | 1:3-0:1 | 1.18 | -1.49 | 3.86 | 0.7855 |
| comparisons with males on diet 1 (1:1) | 0:2-1:1 | -1.04 | -3.86 | 1.77 | 0.8852 |
| | 1:2-1:1 | 0.46 | -2.51 | 3.43 | 0.9975 |
| | 0:3-1:1 | 2.23 | -0.54 | 5.00 | 0.1863 |
| | 1:3-1:1 | 0.58 | -2.33 | 3.49 | 0.9916 |
| Comparisons with females on diet 2 (0:2) | 1:2-0:2 | 1.50 | -1.24 | 4.24 | 0.5963 |
| | 0:3-0:2 | 3.27 | 0.75 | 5.80 | 0.0040 |
| | 1:3-0:2 | 1.63 | -1.05 | 4.30 | 0.4833 |
| Comparisons with males/ diet 2 | 0:3-1:2 | 1.77 | -0.93 | 4.47 | 0.3965 |
| | 1:3-1:2 | 0.12 | -2.71 | 2.96 | 1.0000 |
| Male 3: Female 3 | 1:3-0:3 | -1.65 | -4.28 | 0.98 | 0.4514 |

There are 6 combinations of diet and gender. The interactions post hoc tests compare each pair of combinations. This shows that the only significant differences are for females and are between diets 1 and 3 ($p=0.0191$) and diets 2 and 3 ($p=0.004$). Women on diet 3 lose on average 2.83kg more than those on diet 1 and 3.27kg more than those on diet 2.

Reporting ANOVA

A two way ANOVA was carried out on weight lost by diet type and gender. There was a statistically significant interaction between the effects of Diet and Gender on weight loss [$F(2, 70)=3.153$, $p = 0.049$]. Tukey's HSD post hoc tests were carried out. For females, diet 3 was significantly different to diet 1 ($p = 0.0191$) and diet 2 ($p = 0.004$) but there is no evidence to suggest that any diets differed for males. Women on diet 3 lost on average 2.83kg more than those on diet 1 and 3.27kg more than those on diet 2.

Normality checks and Levene's test were carried out and the assumptions were met.

