Parametric and Non-parametric tests

Yebelay Berehan

Biostatistician

yebelay.ma@gmail.com

2022-06-18

For statistical tests, we use two types of variables:

Independent Variable

- variation does not depend on another variable

  - Usually denoted as X.

  - Typically represents what the researcher set up (treatment, group, etc.)

Dependent Variable

- value depends on another variable (the independent one).

  - Represents the variable that the researcher is interested in

  - Output or outcome, Usually denoted as Y

Almost all statistical tests give three important pieces of information

Test statistic

- Variable calculated from sample data and used in hypothesis test
- Used to determine whether a test was significant or not

Degrees of Freedom

- Number of values of quantities that can be assigned to a statistical distribution
- Should be reported with test results

P-value

- Measure of significance for the test statistic
- Typically, 0.05 is the cutoff value

Hypothesis testing

- What is hypothesis testing in statistics?

- Hypothesis Testing is a type of statistical analysis in which you put your assumptions about a population parameter to the test.

- There are two type of hypothesis testing

  - Null hypothesis
  - Alternative hypothesis

- The probability of reject the null hypothesis when it is true is called type-I error.

- Failing to reject the null hypothesis when it is false is type-II error

- The p-value is the smallest level of significance that would lead to rejection of the null hypothesis with the given data

Inferential Statistics:

- Draw conclusions about an entire population from information obtain from sample.

- To make inferences about a population based on data that we gather from a sample.

- Provide measures of how well your data support your hypothesis and if your data are generalizable beyond what was tested (significance tests).

- Inferential statistics allow us to determine how likely it is to obtain a set of results from a single sample is known as statistical significance.

T-test

- There are three type of t-test

- **The one sample t-test** determines whether the sample mean is statistical different from a known or hypothesized population mean

- **The Independent Samples t-Test** compares the means of two independent groups in order to determine whether there is statistical evidence that the associated population means are significantly different.

- **The Paired-Samples t-Test** compares two means that are from the same individual, object, or related units.

- The purpose of the test is to determine whether there is statistical evidence that the mean difference between paired observations on a particular outcome is significantly different from zero

Inferential Statistics (Analyzing our Data)

- When we move from categorical outcomes to variables measured on an interval or ratio scale, we become interested in means rather than frequencies.

- Comparing means is usually done with the *t*-test, of which there are several forms.

- The one-sample *t*-test is appropriate when a single sample mean is compared to a population mean but the population standard deviation is unknown.

- A sample estimate of the population standard deviation is used instead.

- The appropriate sampling distribution is the t-distribution, with N-1 degrees of freedom.

$$t_{df=N-1} = \frac{\bar{X} - \mu}{\frac{\hat{\sigma}}{\sqrt{N}}}$$

One-sample t-tests

- Some of the steps in t test and how this differs from a z-test.

	Z-test	**t-test**
μ	known	known
σ	known	unknown
sem or σ_M	$\dfrac{\sigma}{\sqrt{N}}$	$\dfrac{\hat{\sigma}}{\sqrt{N}}$ or $\dfrac{s}{\sqrt{N}}$
Probability distribution	standard normal	t
DF	none	$N-1$
Tails	One or two	One or two
Critical value ($\alpha = .05$ two-tailed)	1.96	Depends on DF

When you assume...

...you can run a parametric statistical test!

Assumptions of the one-sample *t*-test

Normality: We assume the sampling distribution of the mean is normally distributed.

Independence. Observations in the dataset are not associated with one another.

- i.e. collecting a score from Participant A doesn't tell me anything about what Participant B will say.

A brief example

- A random sample data of 22 students' weights was taken from student population.
- Then use the following test whether the average weight of student population is different from 140 lb.

Weight 135,119,106,135,180,108,128,160,143,175,170,205,195,185,
182,150,175,190,180,195,220,235

Hypotheses

$H_0 : \mu = 140\, lb$

$H_1 : \mu \neq 140\, lb$

```
weight <- c(135,119,106,135,180,108,128,160,143,175,170,
            205,195,185,182,150,175,190,180,195,220,235)
```

Hypothesis test for a test of normality

- Null hypothesis: The data is normally distributed.

- If p > 0.05, normality can be assumed.

- There are two common ways to check this assumption in R:

1. Visual Method(histogram or Q-Q plot)

- If the histogram is roughly **bell-shaped**, or if the points in the plot roughly fall along a straight diagonal line, then the data is assumed to be normally distributed.

1. Formal Statistical Test(Shapiro-Wilk Test & Kolmogorov-Smirnov Test)

- If the p-value of the test is greater than $\alpha = .05$, then the data is assumed to be normally distributed.

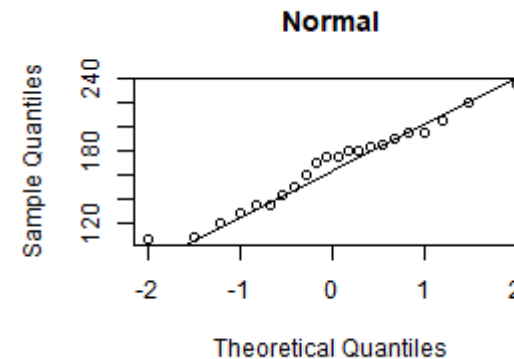The following codes show how to use each of these methods in practice.

- Create a Histogram

```
hist(weight, col='steelblue',
     main='Normal')
```
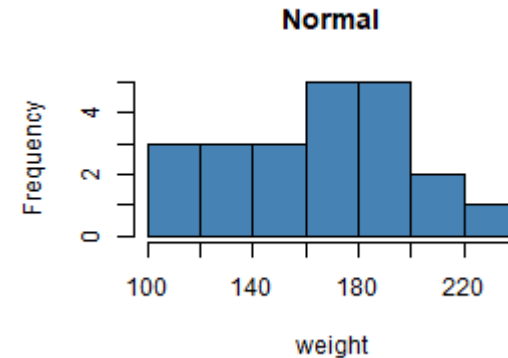
Normal

- Create a Q-Q plot

```
qqnorm(weight,
       main = 'Normal')
qqline(weight)
```

Normal

- Perform a Shapiro-Wilk Test

```
shapiro.test(weight)
```

```
##
##      Shapiro-Wilk normality test
##
## data:  weight
## W = 0.96918, p-value = 0.6922
```

- The p-value 0.6922 greater than 0.05 which imply that it is acceptable to assume that the weight distribution is normal .

- Therefore, normality can be assumed for this data set and, provided any other test assumptions are satisfied, an appropriate parametric test can be used.

```
t.test(weight, mu = 140, alternative = "two.sided")
```

```
##
##      One Sample t-test
##
## data:  weight
## t = 3.5819, df = 21, p-value = 0.001757
## alternative hypothesis: true mean is not equal to 140
## 95 percent confidence interval:
##   151.2668 182.4605
## sample estimates:
## mean of x
##   166.8636
```

Conclusion: based on one sample test table, we have enough evidence that the weight of student is different from the 140lb at 5/% level of significance (P-value = 0.002<0.05).

Two – sample test.

- It is used to compare the means between two different sample data.
- Example: Suppose that the Medical Doctor wanted to determine whether the average Cholesterol levels (mg/dl) of men with Type A differed significantly from Type B behaviors. Data are given below.

Type A		Type B	
233	276	344	226
291	234	185	175
312	181	263	242
250	248	246	252
246	252	224	183
197	202	212	137
268	218	188	202
224	212	250	194
239	325	148	213
254	239	169	

- Check the assumption of normality and equality of variance.
- State the hypothesis?
- What is the test statistic?
- What is the mean and standard error for the Cholesterol levels?

```
twotest <-
  data.frame(Type=rep(c('Type A','Type B'),20),
             Cholesterol= c(233, 344, 291, 185, 312, 263, 250, 246,
                            246, 224, 197, 212, 268, 188, 224, 250,
                            239, 148, 254, 169, 276, 226, 234, 175,
                            181, 242, 248, 252, 252, 183, 202, 137,
                            218, 202, 212,194,325,213,239,NA))
```

```
shapiro.test(twotest$Cholesterol)
```

```
##
##      Shapiro-Wilk normality test
##
## data:  twotest$Cholesterol
## W = 0.97852, p-value = 0.6497
```

- Based on the Shapiro-Wilk test the Cholesterol levels of the group is normal distribution at 5% level of significance.

test the hypothesis

- State the Null and Alternate Hypotheses
  - Ho=There is no difference between Type A and Type B on Cholesterol levels.
  - Ha =There is a difference between Type A and Type B on Cholesterol levels.

```
t.test(twotest$Cholesterol~twotest$Type)
```

```
##
##      Welch Two Sample t-test
##
## data:  twotest$Cholesterol by twotest$Type
## t = 2.3217, df = 33.782, p-value = 0.02642
## alternative hypothesis: true difference in means between group Type A and gro
## 95 percent confidence interval:
##    3.949556 59.518865
## sample estimates:
## mean in group Type A mean in group Type B
##              245.0500              213.3158
```

- From the output, the mean cholesterol level for Type-A is 245.05 and for Type-B is 213.32

- The P-value =0.026<0.05, indicates that there is statistically significant difference between the mean of cholesterol level of the type A and Type B at 5% level of significance.

Paired-Samples T Test

- The procedure computes the differences between values of the two variables for each case and tests whether the average differs from 0.

- The second observation is dependent upon the first since they come from the same population.

- Each subject has two measures, often called before and after measures or Pre-test/post-test .

Assumption

- **Independent Observations:** The scores from before and after the treatment must not be related
- **Normal Distribution:** The population of difference scores must be normally distributed.

Example

- Serum Cholesterol Levels for 12 subject before and after Diet-Exercise program

```
Serum Cholesterol level
Before      After
201         200
231         236
221         216
260         233
228         224
237         216
326         296
235         195
240         207
267         247
284         210
201         209
```

State the Null and Alternate Hypotheses

- Ho: The diet exercise program is not effective in reducing serum cholesterol levels.
- H1: the diet exercise program is effective in reducing serum cholesterol levels.

```
paired <-
  data.frame(
    Before=c(201,231,221,260,228,237,326,235,
240,267,284,201),
After=c(200,236,216,233,224,216,296,195,207,247,210,209))
```

```
t.test(paired$Before,paired$After, paired = TRUE)
```

```
##
##     Paired t-test
##
## data:  paired$Before and paired$After
## t = 3.0201, df = 11, p-value = 0.01165
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##   5.469691 34.863642
## sample estimates:
## mean difference
##        20.16667
```

- The Paired Samples Test result provides mean difference 20.2 and p-value 0.012.

- **Conclusion:** There is enough evidence that the diet exercise program is significantly effective in reducing serum cholesterol levels ($P<0.05$).

ANOVA Test

- Analysis of Variance (ANOVA) is a hypothesis-testing technique used to test the equality of two or more population (or treatment) means by examining the variances of samples that are taken.

- ANOVA is based on comparing the variance (or variation) between the data samples to variation within each particular sample.

- If the **between variation** is *much larger* than the **within variation**, the means of different samples will not be equal.

- If the between and within variations are approximately the same size, then there will be no significant difference between sample means.

- There are two types of ANOVA
  - One-way ANOVA
  - Two-way ANOVA

One-way ANOVA

- **Dependent variable:** Continuous (scale/interval/ratio),
- **Independent variable:** Categorical (at least 3 unrelated/ independent groups)

- **Common Applications:** Used to detect a difference in means of 3 or more independent groups. It can be thought of as an extension of the independent t-test for and can be referred to as 'between-subjects' ANOVA.

- The one-way ANOVA cannot determine which groups are statistically different, and Tukey's, LSD test is required to do so.

Assumptions of ANOVA:

- All populations involved follow a normal distribution.
- All populations have the same variance (or standard deviation).
- The samples are randomly selected and independent of one another.

Test of hypothesis

$$H_o : \mu_1 = \mu_2 = \ldots = \mu_k$$

$$H_1 : \mu_j \neq, at least one j \neq 0$$

- Test statistics: $F = MSTR/MSE$
- Decision rule: Reject H_o : if $p - value$ is less than α-Value

Example: use diet dataset

- The data set contains information on 76 people who undertook one of three diets (referred to as diet *A*, *B* and *C*).
- There is background information such as age, gender, and height. The aim of the study was to see which diet was best for losing weight.

- defining a new column *weight.loss*, the difference between columns `initial.weight` and `final.weight` of the dataset.

- displaying *weight loss* per *diet type* (column `diet.type`) by means of a boxplot.

```
diet = read.csv("diet.csv"); attach(diet)
diet$weight.loss = initial.weight - final.weight
diet$diet.type = factor(diet.type,levels=c("A","B","C"))
diet$gender = factor(gender,levels=c("Female","Male"))
boxplot(weight.loss~diet.type,data=diet,col="light gray",
        ylab = "Weight loss (kg)", xlab = "Diet type")
abline(h=0,col="blue")
```
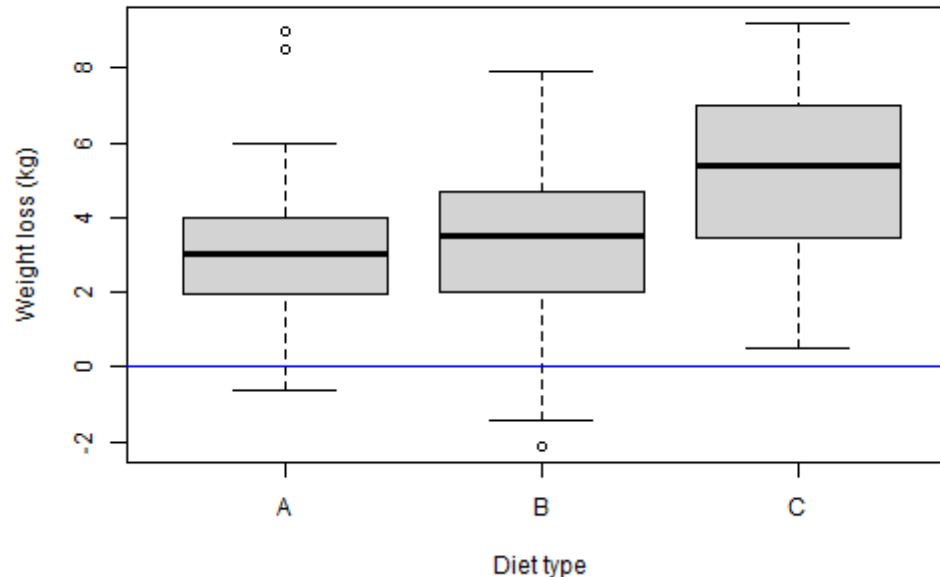
ANOVA Test

- perform a Fisher's, one-way ANOVA, by using the functions `aov()`.
- display and analyse the results: Use the function `summary()` to display the results of an R object of class `aov`.

```
anova1  = aov(weight.loss~diet.type,data=diet)
summary(anova1)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## diet.type    2   60.5  30.264   5.383 0.0066 **
## Residuals   73  410.4   5.622
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Note that, when the interest lies in the difference between two means, the Fisher's ANOVA (fonction `aov()`) and the Student's t-test (function `t.test()` with argument `var.equal` set to `TRUE`) leads to the same results.

- Let check this by comparing the mean weight losses of *Diet A* and *Diet C*.

```
summary(aov(weight.loss~diet.type,data=diet[diet.type!="B",]))
```

```
##               Df Sum Sq Mean Sq F value  Pr(>F)
## diet.type     1   43.4    43.4   8.036 0.00664 **
## Residuals    49  264.6     5.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
t.test(weight.loss~diet.type,data=diet[diet.type!="B",],var.equal=T)
```

```
##
##      Two Sample t-test
##
## data:  weight.loss by diet.type
## t = -2.8348, df = 49, p-value = 0.006644
## alternative hypothesis: true difference in means between group A and group C
## 95 percent confidence interval:
##  -3.1582988 -0.5379975
## sample estimates:
## mean in group A mean in group C
##        3.300000        5.148148
```

Mutiple comparisons

Post Hoc Tests

- ANOVA tests the null hypothesis 'all group means are the same' so the resulting p-value only concludes whether or not there is a difference between one or more pairs of groups.

- If the ANOVA is significant, further 'post hoc' tests have to be carried out to confirm where those differences are.

- The post hoc tests are mostly t-tests with an adjustment to account for the multiple testing.

- Tukey's is the most commonly used post hoc test but check if your discipline uses something else.

Use the command.

```
TukeyHSD(anova1)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = weight.loss ~ diet.type, data = diet)
##
## $diet.type
##           diff        lwr       upr      p adj
## B-A -0.032000 -1.6530850 1.589085 0.9987711
## C-A  1.848148  0.2567422 3.439554 0.0188047
## C-B  1.880148  0.3056826 3.454614 0.0152020
```

- Report each of the three pairwise comparisons e.g. there was a significant difference between diet C and diet A ($p = 0.02$).

Use the mean difference between each pair e.g. people on diet C lost on average 1.85 kg more than those on diet A or use individual group means to conclude which diet is best.

Checking the assumptions for one-way ANOVA

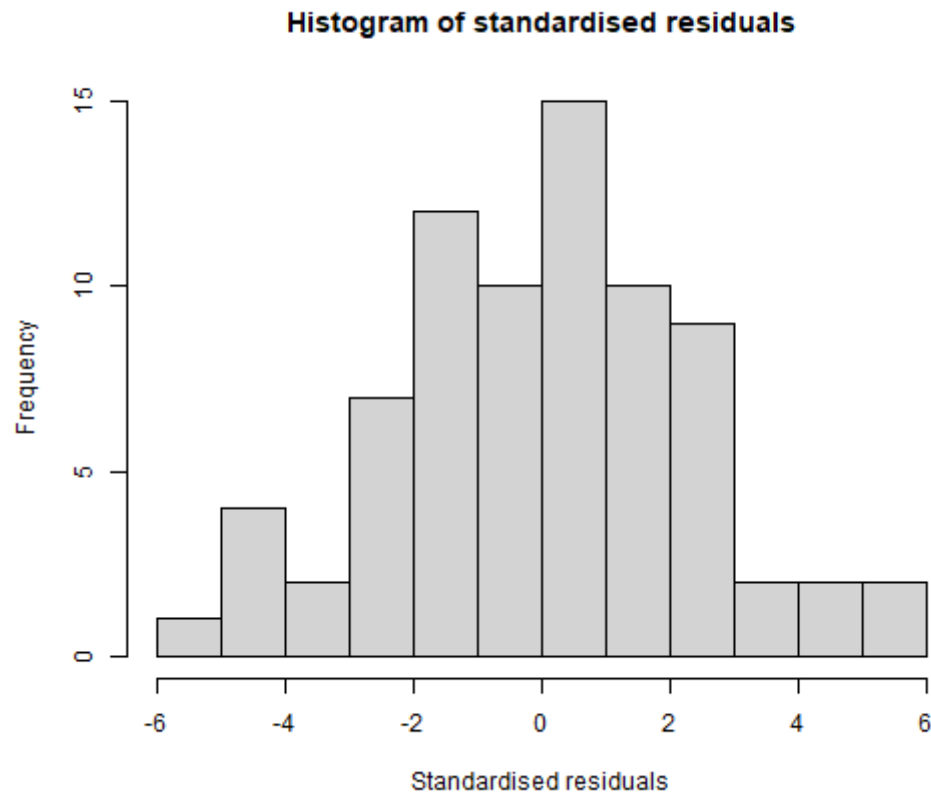A) Residuals should be normally distributed

- Use histogram, QQ plots and normality tests as diagnostic tools
  - If the residuals are very skewed, the results of the ANOVA are less reliable so the Kruskall- Wallis test should be used instead.

B) Homogeneity (equality) of variance: The variances (SD squared) should be similar for all the groups

- Use the Levene's test of equality of variances through the package car `library(car)` and function `leveneTest(weight.loss~diet.type)`
- If p - value > 0.05, equal variances can be assumed and the ANOVA results are valid
  - If p - value < 0.05, the results of the ANOVA are less reliable.
  - The Welch test is more appropriate and can be accessed via `oneway.test(weight.loss~diet.type)` in car package.

Produce a histogram of the residuals.

```
res<-anova1$residuals
hist(res, main="Histogram of standardised residuals",
     xlab="Standardised residuals")
```

Histogram of standardised residuals

The residuals are normally distributed

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

```
library(car)
leveneTest(weight.loss~diet.type, data = diet)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  2  0.4629 0.6313
##       73
```

- As p - value (0.6313) > 0.05, equal variances can be assumed

Reporting ANOVA

- A one-way ANOVA was conducted to compare the effectiveness of three diets.

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

- There was a significant difference in mean weight lost [$F_{(2,73)}=5.38$, $p = 0.0066$] between the diets.

- Post hoc comparisons using the Tukey test were carried out.

- There was a significant difference between diets A and C ($p = 0.019$) with people on diet C lost on average 1.85 kg more than those on diet A. - There was also a significant difference between diets B and C difference ($p = 0.015$) with people on diet C lost on average 1.88 kg more than those on diet B.

Two-way ANOVA

- Two-way analysis of variance (two-way ANOVA) is an extension of the one-way ANOVA to examine the influence of two different categorical independent variables on one continuous dependent variable.

- The purpose of a two-way ANOVA is

    - to determine how to factors impact a response variable
    - to determine whether or not there is an interaction between the two factors on the response variable.

- Examples:

    - The effect of sex and race on wages
    - The effects of religion and region on income

Type of effects and hypotheses

- In two-way ANOVA with two factors (independent variables) A and B, we can test two main effects and one interaction effect:

- **Main effect of A:** whether the population means are the same for different levels (groups) of A?

    - **The null hypothesis:** the population means are the same for different levels of A.

- **Main effect of B:** whether the population means are the same for different levels (groups) of B?

    - **The null hypothesis:** the population means are the same for different levels of A.

- **The interaction effect of A and B:** whether the difference in the population means for difference level of A depends on the level of B or vice visa?

    - **The null hypothesis:** the effect of A or B on the outcome does not depend on B or A.

- Always we have to start to test interaction effects.

- Hypothesis test for Interaction: The interaction effect hypotheses are as follows:

  - H_o: There is no interaction effect.
  - H_1: There is an interaction effect.

- Hypothesis test for Main Effects

- The **main effects** hypotheses are: For factor A:

  - H_o : Population means are equal across levels of Factor A.
  - H_1 : Population means are not equal across levels of Factor A.

- The main effects hypotheses are: For Factor B:

  - H_o : Population means are equal across levels of Factor B.
  - H_1 : Population means are not equal across levels of Factor B.

Two-way ANOVA F tests

- Like in one-way ANOVA, F test is used in hypothesis testing.
- The test is constructed based on the decomposition of the sum of squares of the outcome variables. Specifically, we have

$$SS_T = SS_A + SS_B + SS_{A \times B} + SS_W$$

- SS_T or SS_{total} is the sum of squares for the outcome variable.
- SS_A is the sum of squares attributes to the factor A.
- SS_B is the sum of squares attributes to the factor B.
- $SS_{A \times B}$ is the sum of squares attributes to the joint effect of A and B.
- SS_W or SS_{within} is the residual sum of squares that cannot be explain by A, B or their interaction.

- perform a two-way ANOVA to assess if the weight loss means are different per levels of the factors *diet* and/or *gender*.

```
anova2 = aov(weight.loss~diet.type*gender,data=diet)
summary(anova2)
```

```
##                   Df Sum Sq Mean Sq F value  Pr(>F)
## diet.type          2   60.5  30.264   5.629 0.00541 **
## gender             1    0.2   0.169   0.031 0.85991
## diet.type:gender   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
```

- Or using `lm()` function

```
anova(lm(weight.loss~diet.type*gender,data=diet))
```

Main effect of Diet types

- From the output, the F-statistic for testing the main effect of diet type is 5.6292.

- Based on the F distribution with degrees of freedom 2 and 70, the corresponding p-value is 0.005, which indicates a significant main effect for diet type.

- Therefore, one may conclude there is significant difference in weight loose among the three diet types.

Main effect of gender

- From the output, the F-statistic for testing the main effect of gender is 0.0314.

- Based on the F distribution with degrees of freedom 1 and 70, the corresponding p-value is 0.8599, which indicates an insignificant main effect sex.

- Therefore, one may conclude there is insignificant difference in weight loose between the male and female participants.

Interaction effect between diet.type and gender

- From the output, the F-statistic for testing the interaction effect diet.type:gender is 3.15.

- Based on the F distribution with degrees of freedom 2 and 70, the corresponding p-value is 0.0488, which indicates an significant interaction effect.

- Therefore, one may conclude the difference in weight loose among the three diet types does depend on gender.

Assumptions

- The populations from which the samples were obtained must be normally or approximately normally distributed.
- The samples must be independent.
- The variances of the populations must be equal.
- The groups must have the same sample size.

```
#- Produce a histogram of res.
res<-anova2$residuals
hist(res,
main="Histogram of residuals",
    xlab="Residuals")
```
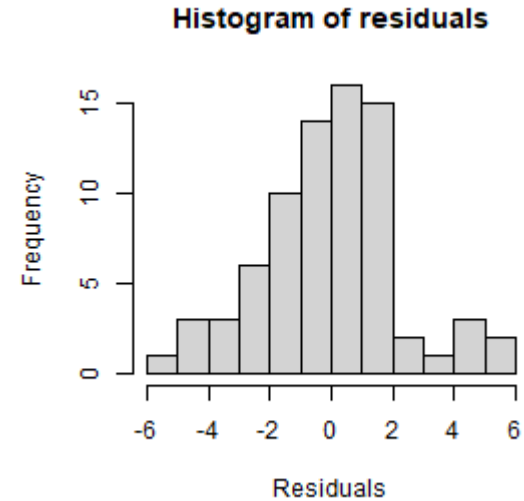
Histogram of residuals

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

```
library(car)
leveneTest(
  weight.loss~gender*diet.type,
          data=diet)
```

```
## Levene's Test for Homogeneity of Varia
##       Df F value Pr(>F)
## group  5  0.3867 0.8563
##       70
```

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

Treatment Groups

- Treatement Groups are formed by making all possible combinations of the two factors.
- For example, if the first factor has 3 levels and the second factor has 2 levels, then there will be 3x2=6 different treatment groups.

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.

```
TukeyHSD(anova2,ordered = T, which = 'diet.type:gender')
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##     factor levels have been ordered
##
## Fit: aov(formula = weight.loss ~ diet.type * gender, data = diet)
##
## $diet.type:gender
##                       diff        lwr      upr      p adj
## A:Female-B:Female 0.4428571 -2.1250148 3.010729 0.9958151
## A:Male-B:Female   1.0428571 -1.7701056 3.855820 0.8852416
## B:Male-B:Female   1.5019481 -1.2354126 4.239309 0.5963201
## C:Male-B:Female   1.6261905 -1.0465354 4.298916 0.4833188
## C:Female-B:Female 3.2728571  0.7481458 5.797569 0.0040103
## A:Male-A:Female   0.6000000 -2.2129628 3.412963 0.9887997
## B:Male-A:Female   1.0590909 -1.6782698 3.796452 0.8656520
## C:Male-A:Female   1.1833333 -1.4893925 3.856059 0.7855223
## C:Female-A:Female 2.8300000  0.3052886 5.354711 0.0191170
## B:Male-A:Male     0.4590909 -2.5093998 3.427582 0.9975014
## C:Male-A:Male     0.5833333 -2.3256625 3.492329 0.9915569
## C:Female-A:Male   2.2300000 -0.5436187 5.003619 0.1863470
## C:Male-B:Male     0.1242424 -2.7117126 2.960197 0.9999949
```

- 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.

Reporting ANOVA

- A two way ANOVA was carried out on weight loss 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 C was significantly different to diet A ($p = 0.0191$) and diet B ($p = 0.004$) but there is no evidence to suggest that any diets differed for males.

- Women on diet C lost on average 2.83kg more than those on diet A and 3.27 kg more than those on diet B.