Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Statistical Programming Using the R Language

Lecture 4

Experimental Design & ANOVA

Darren J. Fitzpatrick, Ph.D

April 2016

Solutions I
2.2 - 2.3

```
holder <- c()

for (i in 1:ncol(affected_genes)){

  t_p <- t.test(unaffected_genes[,i], affected_genes[,i], paired=T)$p.value
  w_p <- wilcox.test(unaffected[,i], affected[,i], paired=T)$p.value
  p_r <- cor(unaffected_genes[,i], affected_genes[,i], method='pearson')
  s_r <- cor(unaffected_genes[,i], affected_genes[,i], method='spearman')
  all_vals <- c(t_p, w_p, p_r, s_r)
  holder <- rbind(holder, all_vals)

}
```


Solutions II
2.4

```
> holder
                [,1]         [,2]       [,3]       [,4]
all_vals 1.681791e-07 0.0001449585  0.2235804  0.3250774
all_vals 1.481568e-07 0.0002137219 -0.2310295 -0.1927462
all_vals 3.698624e-01 0.0221995636  0.3679746  0.1628749
```

```
class(holder)

holder_df <- as.data.frame(holder)
names(holder_df) <- c('t_test', 'w_test', 'pear_r', 'spear_r')
row.names(holder_df) <- c('guanylin', 'pyrroline',
                          'apolipoprotein')
```


Solutions III
2.4

```
> holder_df
                    t_test       w_test     pear_r    spear_r
guanylin       1.681791e-07 0.0001449585  0.2235804  0.3250774
pyrroline      1.481568e-07 0.0002137219 -0.2310295 -0.1927462
apolipoprotein 3.698624e-01 0.0221995636  0.3679746  0.1628749
```


Overview

- Type I & Type II Errors
- Statistical Power
- Effect Sizes
- Power Calculations
- Multiple Hypothesis Testing
- ANOVA

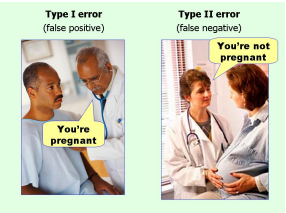
Type I & Type II Errors

H_0 : Null Hypothesis

H_1: Alternative Hypothesis

Type I error is a false positive, i.e., reject the null hypothesis when it is **true**. Type I errors are controlled by the p-value.

Type II error is a false negative, i.e., accept the null hypothesis when it is **false**. Type II errors are controlled by statistical power.



Type I & Type II Errors

H_0 : Null Hypothesis

H_1: Alternative Hypothesis

	H0 is True	H0 is False
Accept H0	True Positive p = 1-α	Type II Error p = β
Reject H0	Type I Error p = α	Correct Decision p = 1 – β (power)

- α is the Type I Error Rate. A p-value threshold of < 0.05 assumes a Type I Error Rate of 5%.
- β is the probability of accepting the null hypothesis if it is false.
- 1 – α is the probability of accepting the null hypothesis if it is true.
- 1 – β is the probability of rejecting the null hypothesis when it is false => **POWER**.

Statistical Power I

Statistical power, denoted *1-β*, is the probability of rejecting the null hypothesis when it is false.

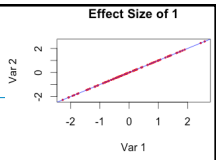
- Statistical power is an important part of experimental design.
- A study may not show a difference between groups because:
  1. There is no difference (true negative)
  2. The study failed to detect the difference (false negative)
- **A common reason for a false negative**: inadequate sample size to detect a significant difference.
- Power calculations enable us to estimate the sample size required to detect a difference of a given **effect size** at a given p-value threshold.

Statistical Power II

- There is a relationship between the four quantities:
  - sample size
  - **effect size**
  - significance level
  - power
- When you know three values, you can determine the fourth.
- By convention, power is usually expected to be at least 0.8.

But first, a brief mention of effect sizes!

Effect Sizes I



- Effect sizes attempt to give a sense of the difference between groups irrespective of whether or not it is statistically significant.
  - **correlation coefficients**
  - **regression coefficients**
- Effect sizes are daunting (at least to me) and outside regression modelling and clinical trials research, I hardly hear mention of them in the biological literature.
- They appear to be a big thing in the psychological/social science literature.
- Nonetheless, they are required for power calculations.

Effect Sizes II

- There are approaches to calculating effect sizes for hypothesis tests but they are outside the scope of this course.
  - For a t-test, you could compute Cohen's d value
  - For an ANOVA, you can compute the f-value
  - For a correlation, it's just the correlation coefficient.
- R has a package to compute effect sizes (`compute.es`)
- You can compute an estimated effect size from sample data in order to determine the sample size required for your study.
- For our purposes, the effect sizes for a t-test and an ANOVA range from 0 – 1 (this is in fact the case using Cohen's d and the f-value).

Power Calculations I

```
install.packages('pwr')
library(pwr)
```

R has a nice package for power calculations called pwr.

Function	Purpose
`pwr.anova.test()`	one-way ANOVA
`pwr.r.test()`	correlation
`pwr.t.test()`	t-tests

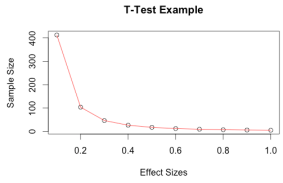
```
pwr.t.test(n = , d = , sig.level = 0.05, power = 0.8,
           type = c("two.sample", "one.sample", "paired"),
           alternative = c("two.sided", "less", "greater"))
```

Power Calculations II

```
e.size <-seq(0, 1, by=0.1)

holder <- c()

for (i in 1:length(e.size)){

      sample.size <- pwr.t.test(d=e.size[i],
                                sig.level=0.05,
                                power=0.8)$n

holder <- c(holder, sample.size)

}
```


Power Calculations III

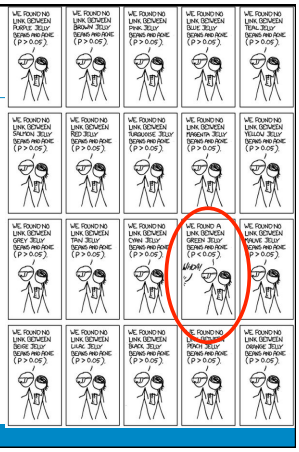
```
plot(e.size, holder, xlab='Effect Sizes',
     ylab='Sample Size', main='T-Test')
lines(e.size, holder, col='red')
```



The larger the effect size, the smaller the sample required to detect it.

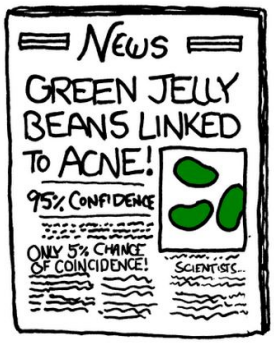
Multiple Hypothesis Testing I

A Fishing Expedition – if you wait long enough, say between now and infinity, you will probably catch a fish.



Multiple Hypothesis Testing II

Could this be **The Daily Mail**, I wonder?



Multiple Hypothesis Testing III

	H0 is True	**H0 is False**
Accept H0	True Positive $p = 1-\alpha$	Type II Error $p = \beta$
Reject H0	Type I Error $p = \alpha$	Correct Decision $p = 1-\beta$ (power)

- The Type I Error Rate ($\alpha = 0.05$) is applicable only to a single statistical test.
- If you perform multiple statistical tests on a data set, you have to adjust the Type I Error Rate in order to correct for the multiple tests.

Multiple Hypothesis Testing IV

The Family Wise Error Rate

- The FWER adjusts p-values so that it reflects the chance of at least 1 false positive. (**methods**: Bonferroni, Holm)
- A 5% FWER means there is a 5% chance that you have at least one false positive.

The False Discovery Rate

- The FDR adjusts p-values so as to control for the frequency of false positives permitted (**methods**: Benjamini- Hochberg).
- A 5% FDR means that you would expect 5% of your findings to be false positives.

FWER is more conservative than FDR.

Multiple Hypothesis Testing V

- R has a single function to adjust p-values for multiple testing.
  - `p.adjust()` takes two arguments
    - a numeric vector of p-values
    - a method

```
p.adjust(x, method='bonferroni')
p.adjust(x, method='BH')# FDR
```

- It returns a vector of corrected p-values.
- For Bonferroni, corrected p-values <= 0.05 → **FWER** of 5%
- For FDR, corrected p-values <= 0.05 → **FDR** of 5%

Analysis of Variance I

- Yesterday we looked at how to compare two samples (t-test, wilcoxon test).
- Sometimes we wish to compare the more than two groups.

```
    Protein Method Correct
1   Ubiquitin CF_AVG   0.467
2   Ubiquitin    GOR   0.645
3   Ubiquitin    PHD   0.868
4     DeoxyHb CF_AVG   0.472
5     DeoxyHb    GOR   0.844
6     DeoxyHb    PHD   0.879
7       Rab5c CF_AVG   0.405
8       Rab5c    GOR   0.604
9       Rab5c    PHD   0.787
10  Prealbumin CF_AVG   0.449
11 Prealbumin    GOR   0.772
12 Prealbumin    PHD   0.780
```

- 4 proteins
- 3 methods to evaluate protein secondary structure
- Proportion of times a method predicted the correct secondary structure

Analysis of Variance II

```
    Protein Method Correct
1   Ubiquitin CF_AVG   0.467
2   Ubiquitin    GOR   0.645
3   Ubiquitin    PHD   0.868
4     DeoxyHb CF_AVG   0.472
5     DeoxyHb    GOR   0.844
6     DeoxyHb    PHD   0.879
7       Rab5c CF_AVG   0.405
8       Rab5c    GOR   0.604
9       Rab5c    PHD   0.787
10  Prealbumin CF_AVG   0.449
11 Prealbumin    GOR   0.772
12 Prealbumin    PHD   0.780
```

We want to test whether the percent correct is different based on method.

We have three groups – we could do multiple T-Tests. ANOVA allows us to first determine if there is any difference.

ANOVA which is capable of comparing more than two groups.

Analysis of Variance III

```
    Protein Method Correct
1   Ubiquitin CF_AVG   0.467
2   Ubiquitin    GOR   0.645
3   Ubiquitin    PHD   0.868
4     DeoxyHb CF_AVG   0.472
5     DeoxyHb    GOR   0.844
6     DeoxyHb    PHD   0.879
7       Rab5c CF_AVG   0.405
8       Rab5c    GOR   0.604
9       Rab5c    PHD   0.787
10  Prealbumin CF_AVG   0.449
11 Prealbumin    GOR   0.772
12 Prealbumin    PHD   0.780
```

ANOVA looks at the response variable (**Correct**) and analyses the within group and between group variability.

Within group variability can be considered noise.

Accounting for within group variability, the between group variability is the signal or the variability due to method in this case.

Do we have a real difference between the methods (CF_AVG, GOR, PHD) relative to the noise?

Analysis of Variance IV

```
    Protein Method Correct
1   Ubiquitin CF_AVG   0.467
2   Ubiquitin    GOR   0.645
3   Ubiquitin    PHD   0.868
4     DeoxyHb CF_AVG   0.472
5     DeoxyHb    GOR   0.844
6     DeoxyHb    PHD   0.879
7       Rab5c CF_AVG   0.405
8       Rab5c    GOR   0.604
9       Rab5c    PHD   0.787
10  Prealbumin CF_AVG   0.449
11 Prealbumin    GOR   0.772
12 Prealbumin    PHD   0.780
```

In brief, an ANOVA looks at the ratio of between group variability and within group variability or the 'signal-to-noise'.

It uses this ratio to compute an **F-test statistic** to decide if there is a statistical difference between groups.

Analysis of Variance V

```
anova(lm(Correct~Method, data=df))

Analysis of Variance Table

Response: Correct
          Df   Sum Sq  Mean Sq F value    Pr(>F)
Method     2 0.305352 0.152676  28.581 0.0001263 ***
Residuals  9 0.048077 0.005342
```

0.152676 is our between group variability

0.005342 is our within group variability

The F-value is the ratio of these two

ANOVA tells us there is a difference somewhere but does not tell us which factor is causing that difference.

Analysis of Variance VI

To find where the difference lies, we have to do some post-hoc t-tests.

```
pairwise.t.test(df$Correct, df$Method, p.adjust.method='BH')

Pairwise comparisons using t tests with pooled SD

data:  df$Correct and df$Method

    CF_AVG  GOR
GOR 0.00086 -
PHD 0.00013 0.05793
```

Notice the adjustment for multiple testing!

Analysis of Variance VII

The Two-Way ANOVA

```
anova(lm(Correct~Method + Protein, data=df))
Analysis of Variance Table

Response: Correct
          Df   Sum Sq  Mean Sq F value    Pr(>F)
Method     2 0.305352 0.152676  44.587 0.0006495 ***
Protein    4 0.030955 0.007739   2.260 0.1975103
Residuals  5 0.017121 0.003424
```


Analysis of Variance VIII

The Kruskal-Wallis Non-Parametric One Way ANOVA

```
kruskal.test(Correct~Method, data=df)

        Kruskal-Wallis rank sum test

data:  Correct by Method
Kruskal-Wallis chi-squared = 8.7692, df = 2, p-value = 0.01247
```


Analysis of Variance IX

Post-Hoc Pairwise Wilcoxon Tests

```
pairwise.wilcox.test(df$Correct, df$Method, p.adjust.method='BH')

        Pairwise comparisons using Wilcoxon rank sum test

data:  df$Correct and df$Method

    CF_AVG GOR
GOR 0.043  -
PHD 0.043  0.114

P value adjustment method: BH
```


Lecture 4 Problem Sheet

- **A problem sheet entitled lecture_4_problems.pdf is located on the course website.**
- **Some of the code required for the problem sheet has been covered in this lecture. Consult the help pages if unsure how to use a function.**
- **Please attempt the problems for the next 30-45 mins.**
- **We will be on hand to help out.**
- **Solutions will be posted this afternoon.**

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Thank You