

Two Group Comparisons and ANOVA models

Biljana Jonoska Stojkova
Applied Statistics and Data Science Group (ASDa)
Department of Statistics, UBC

December 4, 2017

Resources for statistical assistance

Department of Statistics at UBC:

www.stat.ubc.ca/how-can-you-get-help-your-data

SOS Program - An hour of free consulting to UBC graduate students. Funded by the Provost and VP Research Office.

STAT 551 - Stat grad students taking this course offer free statistical advice. Fall semester every academic year.

Short Term Consulting Service - Advice from Stat grad students. Fee-for-service on small projects (less than 15 hours).

Hourly Projects - ASDa professional staff. Fee-for-service consulting.

Hands-on workshops in R: <http://ecoscope.ubc.ca/events/>

Outline

Some theory

Testing the location parameter (t-tests)

More than two groups (ANOVA)

Our data

<http://www.zoology.ubc.ca/~whitlock/Kingfisher/SamplingNormal.htm>

We assume our data is drawn at random from a probability distribution

The data have a mean and a variance and variables can be correlated

If X and Y are our samples from 2 groups:

- ▶ The means are denoted by μ_x and μ_y
- ▶ The variances are denoted by σ_x^2 and σ_y^2
- ▶ The covariance is denoted by σ_{xy}
- ▶ The correlation is $\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$ is a number between -1 and 1

Properties of the mean and variance

The mean of the sum is the sum of the mean $\mu_{x+y} = \mu_x + \mu_y$

The mean of the difference is the difference of the mean

$$\mu_{x-y} = \mu_x - \mu_y$$

The variance of the sum is usually **not** the sum of the variance

$$\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2 + 2\sigma_{xy}$$

The variance of the difference is **not** the difference of the variances.

$$\sigma_{x-y}^2 = \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}$$

If X and Y are independent then the covariance (σ_{xy}) and correlation (ρ_{xy}) = 0

The 2 group experimental setup

We have a random sample of data from 2 populations (observational study) or from a population randomized into 2 groups (controlled experiment)

We measure a variable of interest on each member of the sample and want to determine if the mean of that variable is different in the two groups

Group 1: $X = x_1, \dots, x_n$ are iid $F_1(\mu_x, \sigma_x^2)$

Group 2: $Y = y_1, \dots, y_m$ are iid $F_2(\mu_y, \sigma_y^2)$

iid means **Independent** Identically Distributed

Estimating the parameters from the sample

We use the sample average to estimate the group mean

$$\hat{\mu}_x = \bar{x} = \sum_1^n x_i/n$$

We use the sample variance to estimate the group variances

$$\hat{\sigma}_x^2 = s_x^2 = \sum_1^n (x_i - \bar{x})^2/(n - 1)$$

If the data are paired, we can estimate the covariance between the variables which allows us to compute the correlation

$$\hat{\sigma}_{xy} = s_{xy} = \sum_1^n (x_i - \bar{x})(y_i - \bar{y})/(n - 1)$$

Significance testing, Hypothesis

We compare the means of the two group by a hypothesis test

null hypothesis, H_0 :

- ▶ the opposite of what you want to prove
- ▶ rejected if the evidence proves otherwise

alternative hypothesis, H_1 :

- ▶ the claim you want to prove

Example:

- ▶ H_0 : $\mu_x = \mu_y$ or $\mu_x - \mu_y = 0$
- ▶ H_1 : $\mu_x \neq \mu_y$ or $\mu_x - \mu_y \neq 0$
 - ▶ $\mu_x > \mu_y$ or $\mu_x - \mu_y > 0$
 - ▶ $\mu_x < \mu_y$ or $\mu_x - \mu_y < 0$

Significance testing, Test Statistic

Use the observed sample estimates (\bar{x} , s_x^2 , \bar{y} , s_y^2) to calculate a test statistic

| Test For | Null Hypothesis (H_0) | Test Statistic | Distribution | Use When |
|-----------------------------------------------|-------------------------------------------|-----------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------|-------------------------------------------------------------------------------|
| Population mean (μ) | $\mu = \mu_0$ | $\frac{(\bar{x} - \mu_0)}{\sigma/\sqrt{n}}$ | Z | Normal distribution or $n > 30$; σ known |
| Population mean (μ) | $\mu = \mu_0$ | $\frac{(\bar{x} - \mu_0)}{s/\sqrt{n}}$ | t_{n-1} | $n < 30$, and/or σ unknown |
| Population proportion (p) | $p = p_0$ | $\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ | Z | $n\hat{p}$, $n(1-\hat{p}) \geq 10$ |
| Difference of two means ($\mu_1 - \mu_2$) | $\mu_1 - \mu_2 = 0$ | $\frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ | Z | Both normal distributions, or $n_1, n_2 \geq 30$; σ_1, σ_2 known |
| Difference of two means ($\mu_1 - \mu_2$) | $\mu_1 - \mu_2 = 0$ | $\frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ | t distribution with $df =$ the smaller of $n_1 - 1$ and $n_2 - 1$ | $n_1, n_2 < 30$; and/or σ_1, σ_2 unknown |
| Mean difference μ_d (paired data) | $\mu_d = 0$ | $\frac{(\bar{d} - \mu_d)}{s_d/\sqrt{n}}$ | t_{n-1} | $n < 30$ pairs of data and/or σ_d unknown |
| Difference of two proportions ($p_1 - p_2$) | $p_1 - p_2 = 0$ | $\frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ | Z | $n\hat{p}$, $n(1-\hat{p}) \geq 10$ for each group |

Statistical significance, P-value

Calculate the chance of getting the observed test statistic, assuming H_0 is true: p-value

How small should this chance be before you reject H_0 ? 1%? 5%?
Need to decide the chance of a false positive or Type I error, α .

If p-value is less than α :

- ▶ reject H_0 , accept H_1
- ▶ there is enough evidence statistically that H_1 is true (H_0 is not true)

If p-value is greater than α :

- ▶ do not reject H_0
- ▶ there is not enough evidence statistically that H_1 is true (H_1 is not true)
- ▶ NOTE that you can't say that there is enough evidence statistically to accept H_0 or to prove that H_0 is true

The distribution of the sample mean

http:

[//www.zoology.ubc.ca/~whitlock/Kingfisher/CLT.htm](http://www.zoology.ubc.ca/~whitlock/Kingfisher/CLT.htm)

With a large enough sample size the sample average will converge to a normal distribution for almost any distribution of the original data

$$\bar{x} \xrightarrow{d} N(\mu_x, \sigma_x/\sqrt{n})$$

$$(\bar{x} - \mu_x)/\sigma_x/\sqrt{n} \xrightarrow{d} N(0, 1)$$

How quickly it converges depends on the distribution of data

The distribution of the sample mean - cont

← → www.zoology.ubc.ca/~whitlock/Kingfisher/CLT.htm ☆

SAMPLE 10 INDIVIDUALS

CALCULATE MEAN

MEANS FOR MANY SAMPLES

n 10

Use the mouse to draw your own distribution of individuals at the right.

COFFEE

NORMAL

FLU

VOTE

CUSTOM

Frequency

Individual X value

True mean = 43.4

HIDE NORMAL APPROXIMATION

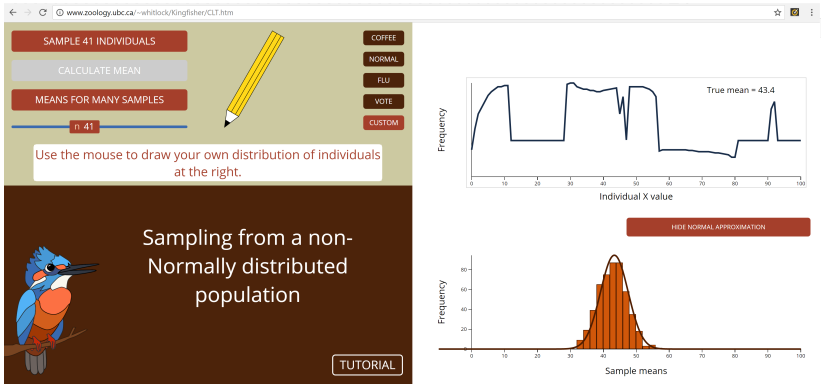
Frequency

Sample means

Sampling from a non-Normally distributed population

TUTORIAL

The distribution of the sample mean - cont



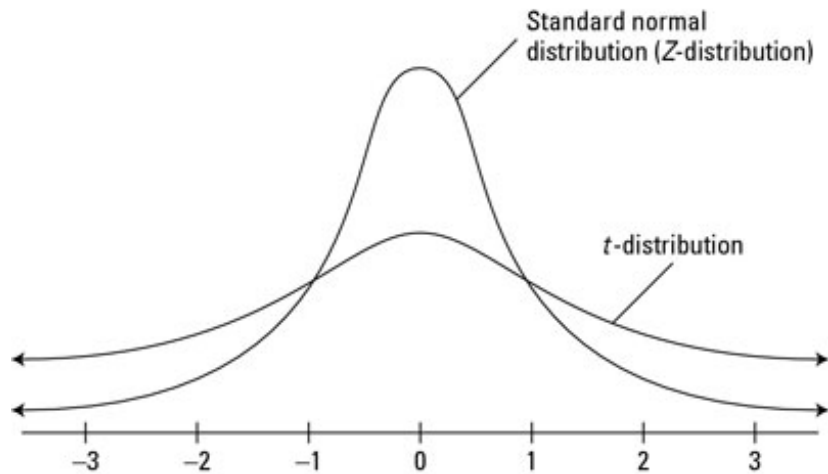
The screenshot shows a web browser window with the URL www.zoology.abc.ca/~whitlock/KIngfisher/CLT.htm. The interface includes several interactive elements:

- Buttons for "SAMPLE 41 INDIVIDUALS", "CALCULATE MEAN", and "MEANS FOR MANY SAMPLES".
- A yellow pencil icon for drawing a distribution.
- A dropdown menu showing "n: 41".
- A text box with the instruction: "Use the mouse to draw your own distribution of individuals at the right."
- A vertical menu of options: "COFFEE", "NORMAL", "FLU", "VOTE", and "CUSTOM".

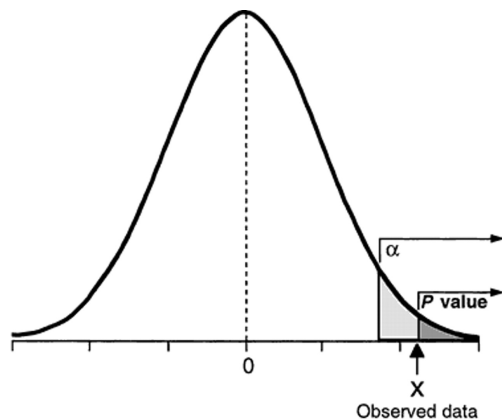
The main content area is split into two panels:

- Left Panel:** Features a blue kingfisher illustration and the text "Sampling from a non-Normally distributed population". A "TUTORIAL" button is located at the bottom right.
- Right Panel:** Contains two plots:
 - Top Plot:** A line graph of "Frequency" vs "Individual X value" (0-100). It shows a jagged, non-normal distribution with a peak around 10 and another around 90. A text label indicates "True mean = 43.4".
 - Bottom Plot:** A histogram of "Frequency" vs "Sample means" (0-100). The histogram bars are orange, and a smooth black normal distribution curve is overlaid, centered around 43.4. A red button labeled "HIDE NORMAL APPROXIMATION" is positioned above the plot.

Normal versus t-distribution



Statistical significance



http://www.nature.com/news/statisticians-issue-warning-over-misuse-of-p-values-1.19503?WT.mc_id=SFB_NNEWS_1508_RHBox#ref-link-1
<http://amstat.tandfonline.com/doi/pdf/10.1080/00031305.2016.1154108?needAccess=true>

t-test for two uncorrelated samples

Stated assumptions:

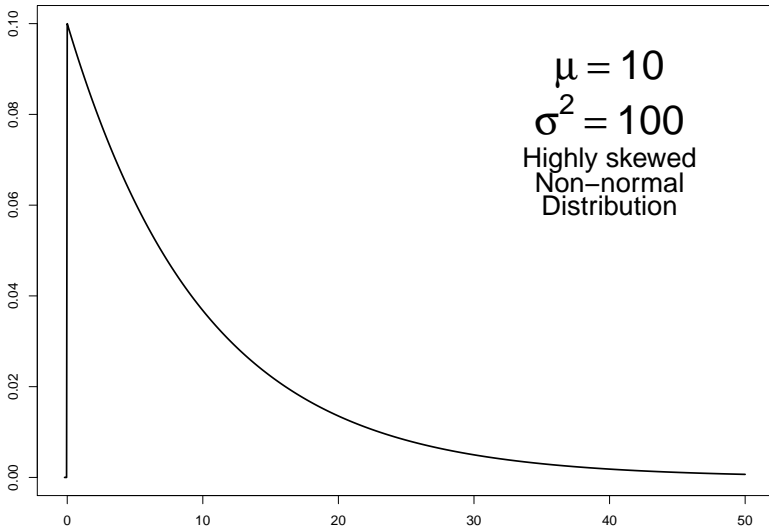
- ▶ data are normally distributed
- ▶ equal variance in the two groups
- ▶ Both samples are simple random samples from their respective populations and are independent of each other.

What happens if the assumptions are violated?

We can test this by simulating data and computing the p-value

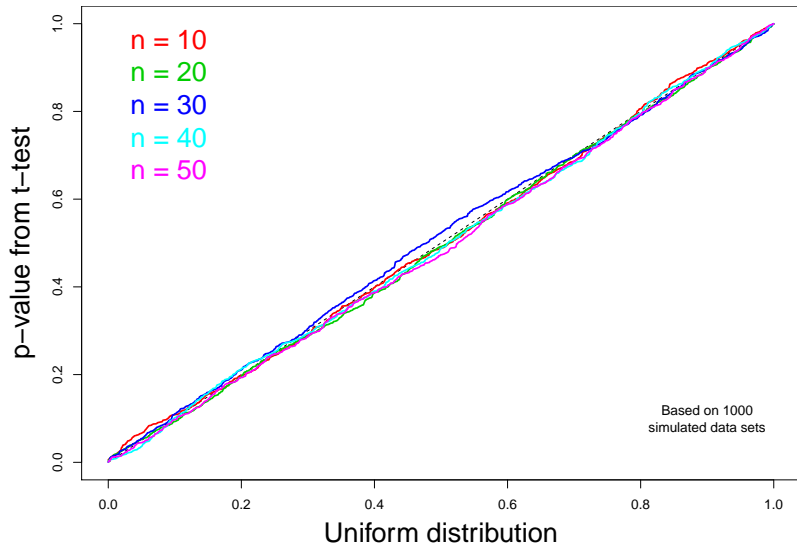
If we plot the p-values against the quantiles of a uniform distribution we should get a straight line

This is a $\Gamma(1, 1/10)$ distribution



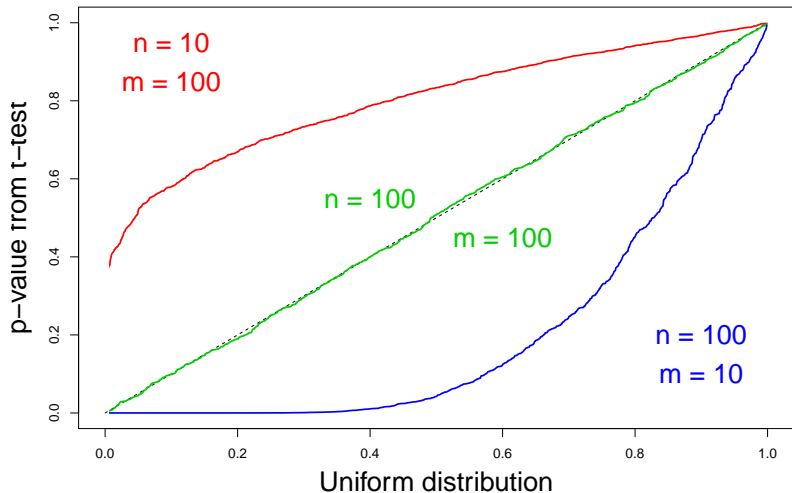
Normality assumption violated

$$X \sim \Gamma(1, 1/10), Y \sim \Gamma(1, 1/10)$$



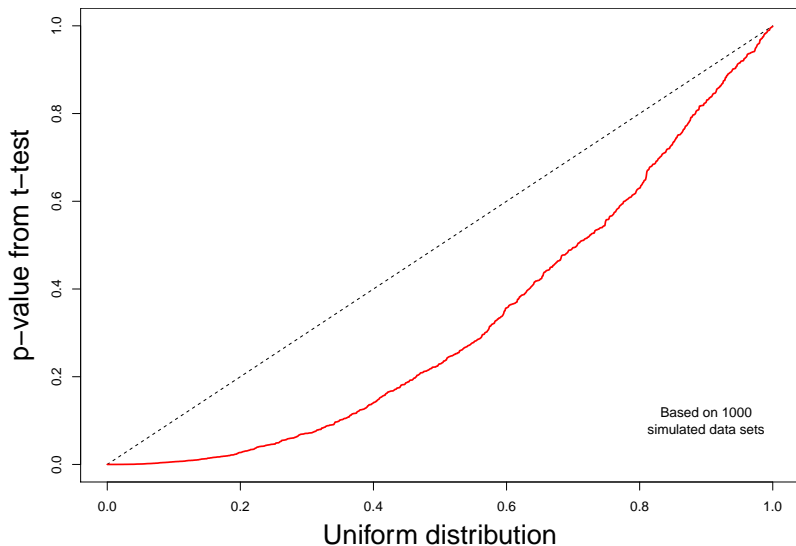
Equal variance assumption violated

n samples $X \sim N(0, 1)$ versus m samples $Y \sim N(0, 10)$



Independence assumption violated

$X \sim N(0, 1)$, $Y \sim N(0, 1)$ not simple random samples from their respective populations, correlation $\rho_{(i,i+1)} = 0.5$



Conclusions about 2 sample t-tests

The independence assumption is critical for the t-test to be valid

- ▶ If the data within a group are not independent then the dependence must be estimated and adjusted for

The equal variance assumption is not critical if the sample size in each group is similar

- ▶ If the variances and the sample sizes in the two groups are different, the **Welch's t-test** can be used instead

The normality assumption is not critical for the t-test and can essentially be ignored

- ▶ Violation of this assumption can affect the power of the test
- ▶ If data is skewed either transform or use an alternative test

Paired t-test

Paired data means for each x there is a specific y related to it

This usually means there is correlation ($\rho_{xy} \neq 0$).

We need to adjust for the correlation by using a paired t-test

This should be a standard test in most statistical software

If we take the difference in the observed values in each pair, a paired t-test becomes a one sample t-test

We compute $z_i = x_i - y_i$ and test:

- ▶ $H_0: \mu_z = 0$
- ▶ $H_1: \mu_z \neq 0$ or $\mu_z < 0$ or $\mu_z > 0$

Non-Parametric tests

Non-parametric tests can be used as an alternative to a two-sample or paired t-test.

- ▶ Two-sample t-test -> **Wilcoxon rank sum test**

Wilcoxon test is the same as **Mann-Whitney U test**

- ▶ Paired t-test -> **Wilcoxon signed rank test**

Wilcoxon tests makes similar assumptions as the t-test except for normality

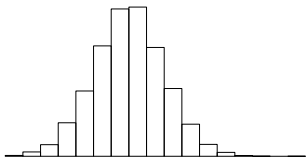
Two sample t-test versus Wilcoxon rank sum test

The key factor in choosing between a t-test and a Mann-Whitney test is the statistical power under the alternative hypothesis

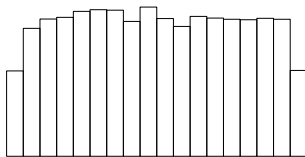
- ▶ For symmetric data, including the normal distribution, the t-test is slightly more powerful than the Mann-Whitney test
- ▶ If the data are skewed, the Mann-Whitney test can be substantially more powerful than the t-test
- ▶ t-test only reject H_0 if the group means are different
- ▶ The Mann-Whitney test can reject H_0 for reasons other than a difference in the group means

Sample distributions

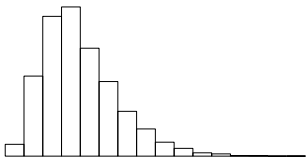
Normal



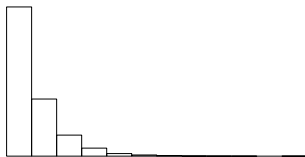
Uniform



Gamma (Low Skew)



Gamma (High Skew)



Power of the two sample t-test vs Mann-Whitney

We can compare the power of the two tests by simulation

In all distributions $\mu_1 - \mu_2 = 2$, $\sigma^2 = 100$ and $n = 250$ in each group

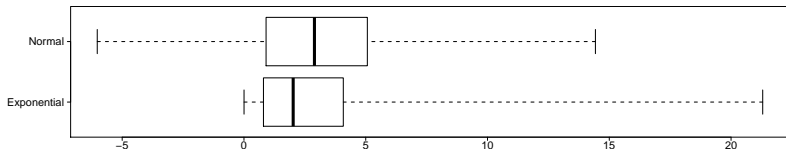
| ## | Dist | Shape | T.test | MW.test |
|------|---------|-----------|--------|---------|
| ## 1 | Normal | symmetric | 0.634 | 0.603 |
| ## 2 | Uniform | symmetric | 0.607 | 0.590 |
| ## 3 | Gamma | low skew | 0.608 | 0.693 |
| ## 4 | Gamma | high skew | 0.613 | 0.894 |

Regardless of the distribution, the power of the t-test is about 60%

The power of the Mann-Whitney test increases as the data become more skewed

More about the Mann-Whitney test

Mann-Whitney may reject H_0 if the data come from different distributions even if the means are the same



```
##           Dist      mean variance  median
## 1      Normal 2.965056 9.639457 2.894027
## 2 Exponential 2.982511 9.005277 2.020666
```

```
## t.test = 0.8982921 ,   wilcoxon = 0.004800793
```

Paired t-test versus Wilcoxon signed rank test

Paired t-test assesses the mean of the difference in the paired data

Signed rank test assesses the median of the difference in the paired data

If the data are skewed, the mean is different from the median

```
## mean(x) = -0.0744075, median(x) = 0
```

```
## mean(y) = 0.9284962, median(y) = 0
```

```
## With  $z_i = x_i - y_i$ :
```

```
## mean(z) = -1.002904, median(z) = -0.1748501
```

```
## p-value from a paired t-test = 0.01515364
```

```
## p-value from a signed rank test = 0.08973196
```

Testing the difference in the scale parameter

$$H_0 : \sigma_x^2 = \sigma_y^2 \text{ versus } H_1 : \sigma_x^2 \neq \sigma_y^2$$

The **F test** is the ratio of two variance estimates

Bartlett's test can test equality of variances in many groups

The two tests above assume normality

Levene's Test is a test for many groups that is less dependent on the assumption of normality

Brown–Forsythe test is similar to Levene's test but is even more robust to the distributional assumptions

Non-parametric tests are also available **Fligner-Killeen test**, **Ansari-Bradley test** and **Mood test**

A statistical model

Attempts to describe the data generating process



**"All models are wrong
but some are useful"**

George Box, 1978

The word "model" implies simplification or approximation, it will not reflect all of reality

One-Way ANOVA

https:

[//www.khanacademy.org/math/statistics-probability/analysis-of-variance-anova-library/v/anova-1-calculating-sst-total-sum-of-squares](https://www.khanacademy.org/math/statistics-probability/analysis-of-variance-anova-library/v/anova-1-calculating-sst-total-sum-of-squares)

Data: 1, 2, 3, 3, 4, 4, 5, 5, 6, 6, 7, 8

| | Group I | | |
|--------|---------|-----|-----|
| Unit J | 1 | 2 | 3 |
| 1 | 1 | 3 | 5 |
| 2 | 2 | 4 | 6 |
| 3 | 3 | 5 | 7 |
| 4 | 4 | 6 | 8 |
| Mean | 2.5 | 4.5 | 6.5 |

Where is most of the variability in the data? Within the groups or between the groups?

One-Way ANOVA

| | Sums of Squares (Sum Sq) | Degrees of Freedom (Df) |
|----------------|--------------------------|-------------------------|
| Within Groups | 15 | $(J - 1) \times I = 9$ |
| Between Groups | 32 | $I - 1 = 2$ |
| Total | 47 | $(I \times J) - 1 = 11$ |

| | Mean Square (Sum Sq / Df) |
|--------------------------|---------------------------|
| Within groups | 1.667 |
| Between groups | 16 |
| Between/Within (F value) | 9.6 |

One-Way ANOVA

One-way **AN**alysis **Of VA**riance (ANOVA) is a way to compare more than 2 groups

Y is a continuous response variable

T is a categorical variable with K distinct groups or levels

We assume Y_{ij} are independent $N(\mu_i, \sigma^2)$

$$\mu_i = \mu + \tau_i$$

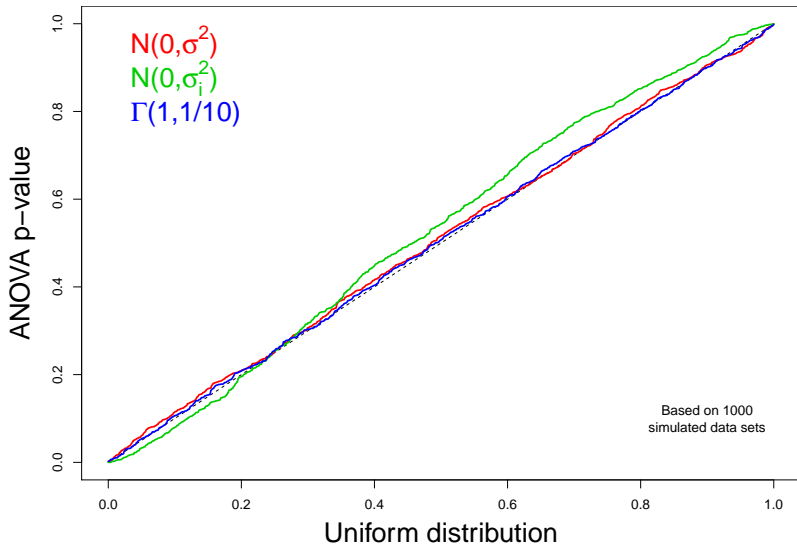
H_0 : all $\tau_i = 0$, H_1 : at least one $\tau_i \neq 0$

Like the two sample t-test:

- ▶ ANOVA is robust to the normality assumption
- ▶ Balanced ANOVA is robust to unequal variance
- ▶ Independence is a very important assumption

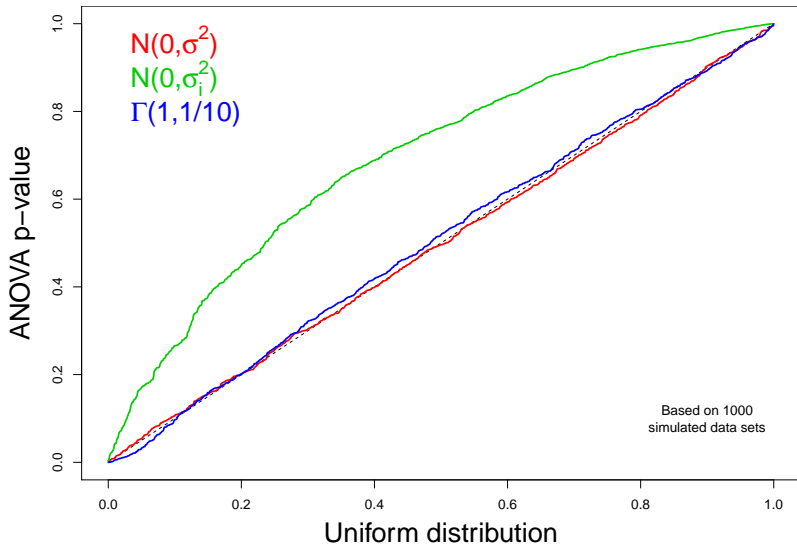
Normality and equal variance (balanced case)

Five Groups, $n = 10$ in each group



Normality and equal variance (unbalanced case)

Five Groups, $n = 2, 6, 10, 12, 20$ (unbalanced)



ANOVA models

How the parameters are specified can vary

The usual method is to select one of the groups to be the reference

- ▶ All other levels of the factor are compared to this group

With only 2 groups, the model computes

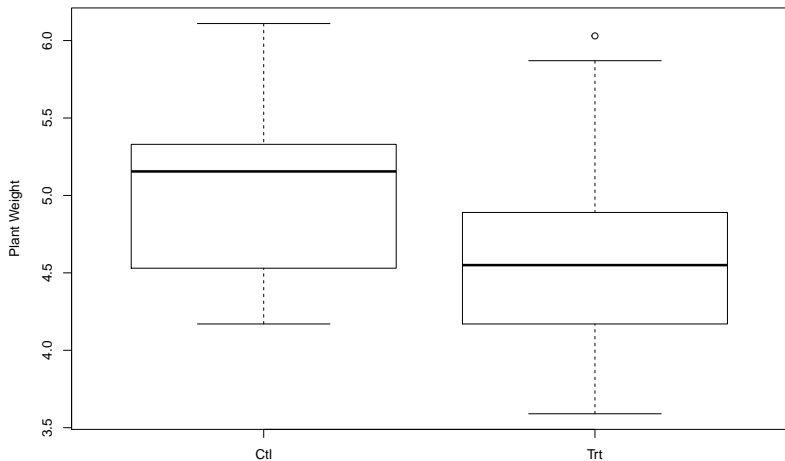
- ▶ the mean for the reference group
- ▶ the difference between the mean of the treatment and reference group

We are usually only interested in the difference

If there are more than 2 groups, none of the non-reference groups are compared to each other

Example Plant Weight Data

With any analysis you should try plotting the data first. Here are boxplots of the data in each group.



One-Way ANOVA

ANOVA with 2 groups is equivalent to a two sample t-test

```
## Response = Plant Weight
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
```

```
## group      1  0.688  0.6882   1.419  0.249
```

```
## Residuals 18  8.729  0.4850
```

```
## Two sample t-test
```

```
## t = 1.19126 , df = 18 , p value = 0.2490232
```

Note: $\sqrt{1.419} = \pm 1.191$

In fact a $t_{\nu}^2 \sim F_{1,\nu}$

Quantifying the difference

Here are the fitted parameters for the model used:

```
##           Estimate Std.Error  tvalue Pr(>|t|)
## Ctl           5.032    0.2202 22.8501   0.000
## Trt - Ctl     -0.371    0.3114 -1.1913   0.249
```

The fitted model parameters are used to calculate the mean of each group

```
##      Estimate Std.Error   2.5%  97.5%
## Ctl    5.032    0.2202 4.5693 5.4947
## Trt    4.661    0.2202 4.1983 5.1237
```

Usually we only care about the difference in the group means

Selecting parameters for the model

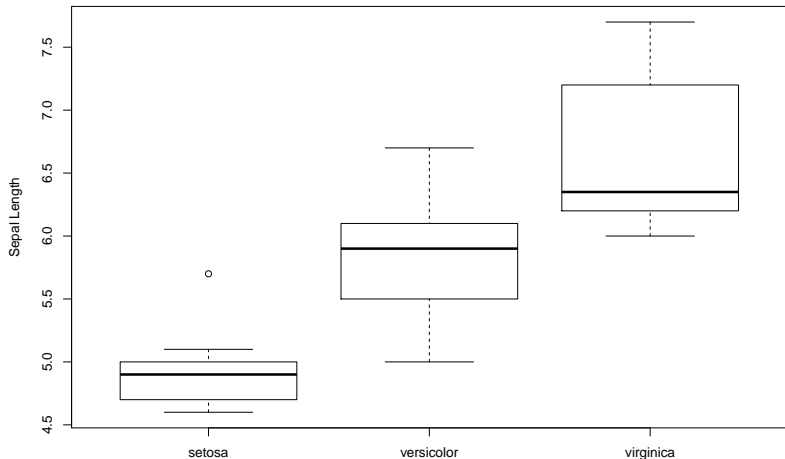
There are many possible parameter choices for an ANOVA model

Important: Make sure you know which parametrization is used by your software before you interpret the output from the model

- ▶ Most common method is to set a specific level of a factor as a reference
- ▶ Alternately set the average effect over all the levels of the factor to be zero

The model will have a single parameter to reflect a “typical” value and $K - 1$ parameters that model the deviations from this value

ANOVA example: Sepal length in 3 Iris species



There is a clear indication of a difference in the sepal length between the species

ANOVA results for Iris data

```
## Response = Sepal Length
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Species    2 13.789   6.8943   29.314 1.71e-07
## Residuals 27   6.350   0.2352
```

The Species term gives the Sums of Squares and Mean Square BETWEEN the 3 species

The Residuals term gives the Sums of Squares and Mean Square WITHIN each of the 3 species

- ▶ represent the error in the model

Estimated coefficients for Iris data

```
## Setting a reference group
```

| ## | Estimate | Std. Error | t value | Pr(> t) |
|---------|----------|------------|---------|----------|
| ## MU | 6.59 | 0.153 | 42.97 | 2.13e-26 |
| ## Var1 | -1.66 | 0.217 | -7.65 | 3.12e-08 |
| ## Var2 | -0.79 | 0.217 | -3.64 | 1.13e-03 |

```
## Making the sum of the group effects = 0
```

| ## | Estimate | Std. Error | t value | Pr(> t) |
|---------|----------|------------|---------|----------|
| ## MU | 5.7733 | 0.0885 | 65.205 | 3.05e-31 |
| ## Var1 | -0.8433 | 0.1252 | -6.735 | 3.13e-07 |
| ## Var2 | 0.0267 | 0.1252 | 0.213 | 8.33e-01 |

Without more information these parameters don't mean much

Posthoc tests

If we look at the pairwise comparisons of the groups in either model the results are the same

| ## | | Diff | SE | Tstat |
|----|------------------------|------|-------|-------|
| ## | versicolor - setosa | 0.87 | 0.217 | 4.01 |
| ## | virginica - setosa | 1.66 | 0.217 | 7.65 |
| ## | virginica - versicolor | 0.79 | 0.217 | 3.64 |

We can compute p-values for these comparisons but need to be adjusted for multiple comparisons

The amount of adjustment increases with the number of pairwise comparisons

ANOVA with a blocking factor

ANOVAs may contain more than one factor

Factors that are not of interest are considered blocking factors

Blocking factors are primarily used to control other sources of error that otherwise might hide significant effects in our factor of interest

Y is a continuous response, T is a factor with I levels, B is the blocking factor with J levels

We assume Y_{ijk} are independent $N(\mu_{ij}, \sigma^2)$

$$\mu_{ij} = \mu + \tau_i + \beta_j$$

H_0 : all $\tau_i = 0$, H_1 : at least one $\tau_i \neq 0$

Example Students's Sleep data

The data shows the effect of two drugs (group) on the amount of extra sleep hours. The blocks are subjects (ID) who act as their own control.

| ## | | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|----|-----------|----|--------|---------|---------|---------|
| ## | group | 1 | 12.48 | 12.482 | 16.501 | 0.00283 |
| ## | ID | 9 | 58.08 | 6.453 | 8.531 | 0.00190 |
| ## | Residuals | 9 | 6.81 | 0.756 | | |

If we ignore ID then we do not see a significant effect for group

| ## | | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|----|-----------|----|--------|---------|---------|--------|
| ## | group | 1 | 12.48 | 12.482 | 3.463 | 0.0792 |
| ## | Residuals | 18 | 64.89 | 3.605 | | |

More on the block design

ANOVA with a blocking factor and 2 groups and only 1 observation per group in each block is equivalent to a paired t-test

```
## Paired t-test  
## t = -4.062128 , df = 9 , p value = 0.00283289
```

The difference in the group means, with and without blocking

| ## | Estimate | Std. Error | t value | Pr(> t) |
|-----------------|----------|------------|---------|----------|
| ## Block Design | 1.58 | 0.3890 | 4.062 | 0.002833 |
| ## One-Way | 1.58 | 0.8491 | 1.861 | 0.079187 |

Note the estimated difference is the same for the block design and the One-Way ANOVA but the standard error is different

Non-parametric generalizations

Kruskal-Wallis is a rank based version of a One-Way ANOVA

A Kruskal-Wallis test with only 2 groups is identical to a Mann-Whitney test

Friedman rank-sum test is a non-parametric way to analyse unreplicated ($n = 1$) complete blocked data

- ▶ If there are only 2 groups, it is not the same as a signed rank test

Once you are outside these 2 special cases there are very few non-parametric methods available

However we have seen an ANOVA model is still valid even if the normality assumption is violated

Two-Way ANOVA

Two-Way ANOVA is similar to a block design except both factors are of interest and can interact with each other

Y is a continuous response, T is a factor with K levels, B is a factor with J levels

We assume Y_{ijk} are independent $N(\mu_{ij}, \sigma^2)$

$$\mu_{ij} = \mu + \tau_i + \beta_j + \gamma_{ij}$$

H_0 : all $\gamma_{ij} = 0$, H_1 : at least one $\gamma_{ij} \neq 0$

If there is no interaction then we are interested in

H_0 : all $\tau_i = 0$, H_1 : at least one $\tau_i \neq 0$

H_0 : all $\beta_j = 0$, H_1 : at least one $\beta_j \neq 0$

Meaning of the interaction

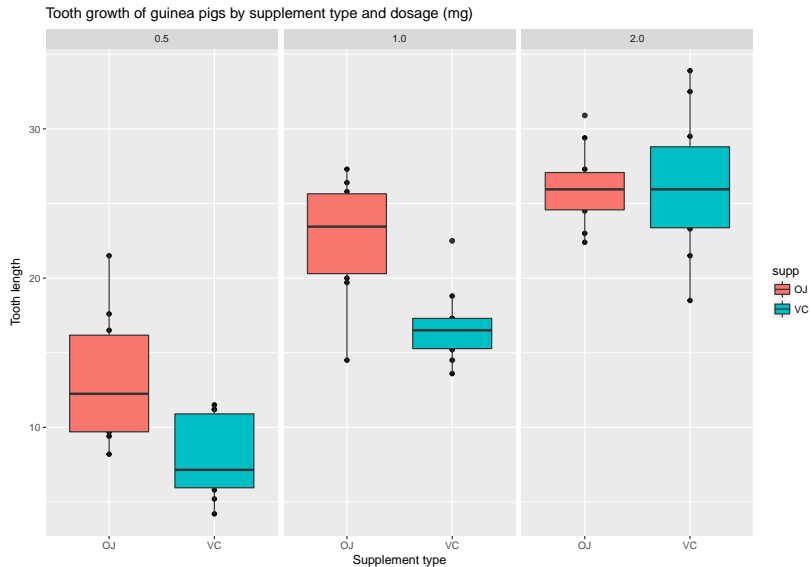
An interaction between factor T and B means the effect of T depends on the level of B and the effect of B depends on the level of T

If an interaction between T and B is in the model then we cannot interpret the main effects of either factor

A main effect need not be significant in the presence of an interaction term

But the main effects must remain in the model otherwise we cannot interpret any of the interaction terms that involve that factor

Example Tooth Growth in Guinea Pigs



Two-Way ANOVA

```
## ANOVA Table
##           Df Sum Sq Mean Sq F value    Pr(>F)
## supp      1  205.4   205.4  15.572 0.000231
## dose      2 2426.4  1213.2  92.000 < 2e-16
## supp:dose  2  108.3    54.2   4.107 0.021860
## Residuals 54  712.1    13.2
```

Setting a reference group, we can see the estimated effects

```
##           Estimate Std. Error t value Pr(>|t|)
## Ref(OJ:0.5)   13.230     1.148  11.521 3.60e-16
## VC            -5.250     1.624  -3.233 0.00209
## 1.0             9.470     1.624   5.831 3.18e-07
## 2.0            12.830     1.624   7.900 1.43e-10
## VC:1.0        -0.680     2.297  -0.296 0.76831
## VC:2.0         5.330     2.297   2.321 0.02411
```

Posthoc comparisons in a Two-Way ANOVA

If there is an interaction between the factors posthoc comparisons must be done within the levels of the other factor

In our example, there are 6 groups which means 15 pairwise comparisons are possible

- ▶ Each of the 3 levels of Dose contain a single comparison of delivery methods (VC-OJ)
- ▶ Each of the 2 delivery methods have 3 possible dose comparisons (1.0-0.5, 2.0-0.5, 2.0-1.0)
- ▶ The other 6 pairwise comparisons are usually not of interest because both the delivery method and the dose change between the 2 groups being compared

Within Dose Level

| ## | | Diff | SE | Tstat |
|----|-----------------|-------|------|---------|
| ## | VC:0.5 - OJ:0.5 | -5.25 | 1.62 | -3.2327 |
| ## | VC:1.0 - OJ:1.0 | -5.93 | 1.62 | -3.6514 |
| ## | VC:2.0 - OJ:2.0 | 0.08 | 1.62 | 0.0493 |

Within delivery method

| ## | | Diff | SE | Tstat |
|----|-----------------|-------|------|-------|
| ## | OJ:1.0 - OJ:0.5 | 9.47 | 1.62 | 5.83 |
| ## | OJ:2.0 - OJ:0.5 | 12.83 | 1.62 | 7.90 |
| ## | OJ:2.0 - OJ:1.0 | 3.36 | 1.62 | 2.07 |
| ## | VC:1.0 - VC:0.5 | 8.79 | 1.62 | 5.41 |
| ## | VC:2.0 - VC:0.5 | 18.16 | 1.62 | 11.18 |
| ## | VC:2.0 - VC:1.0 | 9.37 | 1.62 | 5.77 |

More Complicated ANOVA

ANOVA models can have any number of factors

- ▶ As the number of factors of interest increases the number of factor level combinations increases dramatically
- ▶ This is not true for blocking factors because they do not interact with the other factors

When the number of factors is large it may become impossible to observe every combination of factor levels possible

We can reduce the sample size by assuming certain higher order interactions are negligible and design an experiment that confounds these effects

This leads to incomplete block designs, fractional factorial designs, Latin square designs and others

Example: 3 binary factors in 6 blocks

Response is growth of peas

factors: nitrogen (N), phosphate (P), potassium (K)

blocks: 6 plots of land each subdivided into 4 sections

We have 3 factors each with 2 levels (present/absent) so there are 8 groups in total

We can only observe 4 groups in each plot

In order to maximize our statistical power when estimating the main effects and two-way interaction between the 3 elements, we use a fractional factorial design in each plot


```

## Response: yield
##           Sum Sq Df F value    Pr(>F)
## block    343.29  5  4.4467 0.015939
## N        194.04  1 12.5672 0.004033
## P         0.35   1  0.0225 0.883288
## K         7.35   1  0.4758 0.503435
## N:P       21.28   1  1.3783 0.263165
## N:K       33.13   1  2.1460 0.168648
## P:K        0.48   1  0.0312 0.862752
## N:P:K                0
## Residuals 185.29 12

```

The model contains a main effect for the block and three element factors, and 3 two way interactions between the 3 element factors

Block is significant but none of the two way interactions are significant

Since we see no evidence of interactions, we refit the model excluding interactions so we can interpret the main effects

```
## Response: yield
##           Sum Sq Df F value  Pr(>F)
## N           189.282  1 11.8210 0.00366
## P             8.402  1  0.5247 0.47999
## K           95.202  1  5.9455 0.02767
## Residuals 240.185 15
```

Now we have significant effects for both N and K

```
##           Estimate Std. Error t value Pr(>|t|)
## N Present    5.6167     1.6336  3.4382 0.00366
## P Present   -1.1833     1.6336 -0.7244 0.47999
## K Present   -3.9833     1.6336 -2.4383 0.02767
```

Summary

ANOVA are used to compare numeric responses by categorical predictors

Predictors can be of interest (factors) or not (blocks)

ANOVA is robust to the normality assumption

Balanced ANOVA is robust to the equal variance assumption

Independent observations is an important assumption

Questions?

Department of Statistics at UBC:

www.stat.ubc.ca/how-can-you-get-help-your-data

SOS Program - An hour of free consulting to UBC graduate students. Funded by the Provost and VP Research Office.

STAT 551 - Stat grad students taking this course offer free statistical advice. Fall semester every academic year.

Short Term Consulting Service - Advice from Stat grad students. Fee-for-service on small projects (less than 15 hours).

Hourly Projects - ASDa professional staff. Fee-for-service consulting.

Hands-on workshops in R: <http://ecoscope.ubc.ca/events/>