Two Group Comparisons and Beyond

Rick White

February 4, 2015
Outline

- Some Theory
- Testing the location parameter (t-tests)
- More than two groups (ANOVA)
- Questions
Our data

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 $\mu_x$ and $\mu_y$
- The variances are denoted by $\sigma_x^2$ and $\sigma_y^2$
- The covariance is denoted by $\sigma_{xy}$
- 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 (\( \sigma_{xy} \)) and correlation (\( \rho_{xy} \)) = 0
The 2 group experimental setup

We have a random sample of data from 2 populations (observational study) or we have a random sample 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: $X = x_1, \ldots, x_n$ are iid $F_1(\mu_x, \sigma_x^2)$

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

iid means **Independent** Identically Distributed

Note: without independence our estimate of the variance would be inaccurate. The estimate of the mean is unaffected.
Estimating the parameters from the sample

In order to conduct the test we need to know the group means, variances and sometimes the correlation between the data in the two groups.

We use the sample average to estimate the group mean.

$$\hat{\mu}_x = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

We use the sample variance to estimate the group variances.

$$\hat{\sigma}^2_x = s^2_x = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Usually the groups are independent samples which means the data are independent.

If the data are paired \((n = m)\), we can estimate the covariance between the variables which allows us to compute the correlation.

$$\hat{\sigma}_{xy} = s_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$
Significance testing

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

A hypothesis test consists of a Null hypothesis \( (H_0) \) and an alternative hypothesis \( (H_1) \).

Assuming \( H_0 \) is true, we compute a test statistic and a p-value based on our data. If our p-value is less than \( \alpha \) we reject the Null hypothesis in favour of the alternative \( (H_1) \).

\( \alpha \) specifies the chance of a Type 1 error or a false positive result. If we set \( \alpha = 0 \), we will never reject \( H_0 \).

Note: the test does not confirm the Null when it does not provide any evidence against \( H_0 \).
Sample Size and Power calculations

$\alpha$ (or significance) is the probability of rejecting $H_0$ when it is true. Most studies use 5% but there is no hard and fast rule. It does not depend on the sample size.

$1 - \beta$ (or power) is the probability of rejecting $H_0$ when it is false. As $N$ increases so does the power. Power calculations require you to fully specify the true parameters of the model. Information can come from previous studies, educated guesses or some other source. The more complicated the analysis the more parameters you need to specify.

Software to calculate sample size or power is available. For a list of software see Wikipedia (Statistical Power).
The distribution of the sample mean.

With a large sample size the sample average will converge to a normal distribution (bell curve) for almost any distribution of the original data.

$$\sqrt{n}(\bar{x} - \mu_x) \xrightarrow{d} N(0, \sigma^2_x)$$

How quickly it converges depends on the distribution of data.

**On line Example**

Since the t-test is based on sample averages, this property makes it very robust to the normality assumption if the sample size is reasonably large.
t-test for two uncorrelated samples

Stated assumptions:

- Data are normally distributed
- Equal variance in the two groups
- Data are independent

The Null Hypothesis is the group means are equal \( (H_0 : \mu_x = \mu_y) \). The alternative is usually two sided, the group means are not equal \( (H_1 : \mu_x \neq \mu_y) \) but sometimes a one-sided test is used.

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

$\mu = 10$

$\sigma^2 = 100$

Highly skewed
Non-normal
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 $Y \sim N(0, 100)$

Uniform distribution

p-value from t-test

$n = 10$

$m = 100$

$n = 100$

$m = 100$

$n = 100$

$m = 10$
Independence assumption violated

\( X \sim N(0, 1), Y \sim N(0, 1) \) correlation \( \rho_{(i,i+1)} = 0.5 \)
More on t-tests for 2 independent groups.

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. This is usually not easy to do.

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 used instead. The Welch’s t-test computes a separate variance for each group while the standard two sample t-test computes a single variance.

The normality assumption is not critical for the t-test and can essentially be ignored. Violation of this assumption does not affect the power of the test under alternative hypothesis. If the data is non-normal the t-test is still valid but there may be more powerful options for testing the data.
Paired t-test

Sometimes we do not have two independent samples. The most common situation occurs when data are observed in pairs. This means for each $x$ there is a specific $y$ that is related to it.

If the data are observed in pairs it usually means there is correlation ($\rho_{xy} \neq 0$). In order to compare the two groups in this case we need to adjust for the correlation. This is done by a paired t-test. This should be a standard test in most statistical software.

A paired t-test is a special case of a one sample t-test. If we take the difference in the observed values in each pair, we reduce our two sample problem to a one sample problem. We then test if the mean of our sample is equal to 0.

We compute $z_i = x_i - y_i$; then test if $\mu_z = 0$. 
Non-Parametric tests

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

Instead of a two-sample t-test a **Wilcoxon rank sum test** or equivalently a **Mann-Whitney test** can be used. The Mann-Whitney test assumes two independent samples are drawn, both have the same scale parameter and the observations within each group are independent. There is no assumption about the distribution of the data.

**Wilcoxon signed rank test** can be used instead of a paired t-test. It makes similar assumptions as the paired t-test except for normality. While the paired t-test looks at the mean of the distribution, the signed rank test looks at the median, specifically the median of the pairwise differences. Note while \( \mu_{x-y} = \mu_x - \mu_y \), the same is not true for the median.
The key factor in choosing between a t-test and a Mann-Whitney test is the statistical power under the alternative hypothesis.

A lack of symmetry causes the greatest difference in the power of a t-test and Mann-Whitney test. For symmetric data, including the normal distribution, the t-test is slightly more powerful than the Mann-Whitney test but if the data are skewed, the Mann-Whitney test can be substantially more powerful than the t-test.

The advantage of the t-test is that it looks at a very specific hypothesis. It will only reject the Null if the group means are sufficiently different. The Mann-Whitney test looks at a more general hypothesis and can reject the Null for reasons other than a difference in the group means.
# Power of the two sample t-test vs Mann-Whitney

We can compare the power of the two tests by simulation. We randomly generate samples from various distributions under the alternative hypothesis to estimate the probability we reject the Null. In all distributions $\mu_1 - \mu_2 = 2$, $\sigma^2 = 100$ and $n = 250$ in each group.

<table>
<thead>
<tr>
<th>Dist</th>
<th>Shape</th>
<th>T.test</th>
<th>MW.test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>symmetric</td>
<td>0.560</td>
<td>0.552</td>
</tr>
<tr>
<td>Uniform</td>
<td>symmetric</td>
<td>0.601</td>
<td>0.576</td>
</tr>
<tr>
<td>Gamma</td>
<td>low skew</td>
<td>0.638</td>
<td>0.706</td>
</tr>
<tr>
<td>Gamma</td>
<td>high skew</td>
<td>0.611</td>
<td>0.911</td>
</tr>
</tbody>
</table>

Regardless of the distribution, the power of the t-test is about 60%, however the power of the Mann-Whitney test increases as the data become more skewed.
More about the Mann-Whitney test

While the Mann-Whitney can detect a shift in the distributions, it may reject $H_0$ if the data come from different distributions even if the means of the distributions are the same.

Exponential

Normal

-5 0 5 10 15 20

## 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

While the paired t-test compares the mean of the difference in the paired data, the signed rank test compares the median of the difference in the paired data. If the data are symmetric the median equals the mean but if the data are skewed, the mean is different from the median. Also note \( \text{med}(x - y) \neq \text{med}(x) - \text{med}(y) \).

```r
## mean(x) = -0.10499, median(x) = 0
## mean(y) = 1.048733, median(y) = 0

## If z = x-y then:
## mean(z) = -1.153723, median(z) = -0.62956
## p-value from a paired t-test = 7.551932e-16
## signed rank test = 3.186196e-10
```
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 rely heavily on the normality assumption.

**Levene’s Test** is less susceptible to the normality assumption and can test many groups.

**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**.
One-Way ANOVA

One-way **ANalysis Of VAriance** (ANOVA) is a way to compare more than 2 groups defined by a single categorical predictor variable. An observation is associated with one and only one level of the factor.

$Y$ is a continuous response variable.

$A$ is a categorical variable that indicates several ($I$) distinct groups.

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

$H_0 : \text{all } \alpha_i = 0, \ H_1 : \text{at least one } \alpha_i \neq 0$

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

$N(0, \sigma^2)$
$N(0, \sigma_i^2)$
$\Gamma(1, 1/10)$

Based on 1000 simulated data sets
Normality and equal variance (unbalanced case)

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

- $N(0, \sigma^2)$
- $N(0, \sigma_i^2)$
- $\Gamma(1, 1/10)$

Based on 1000 simulated data sets

Uniform distribution

ANOVA p-value

Based on 1000 simulated data sets
When fitting an ANOVA we must select a parametrization for the model.

The usual method is to select one of the groups to be the reference (usually a control or existing treatment). All other levels of the factor are compared to this group.

With only 2 groups, the model computes the mean for the reference group and the difference between the means of the treatment and reference groups. We are usually only interested in the difference.

However if there are more than 2 groups, none of the non-reference groups are compared to each other. These comparisons can be made as posthoc tests.
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 \( I = 2 \) is equivalent to a two sample t-test.

```r
## Response = Plant Weight
##
##   Df  Sum Sq Mean Sq  F value Pr(>F)
## group     1 0.688   0.688 1.4191  0.249
## Residuals 18 8.729  0.485

## 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

The model can fit the mean of each group. Note: the standard error in each group is the same because the sample size in each group is the same and a common variance is assumed.

```
##   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
```

However the parameters the model uses are below.

```
##   Estimate Std.Error  tvalue  Pr(>|t|)
##  Ctl    5.032    0.2202  22.8501    0.0000
## Trt - Ctl -0.371    0.3114 -1.1913    0.2488
```

Usually we only care about the difference in the group means.
Selecting parameters for the model

With ANOVA, there are many possible parameter choices for the model, each will give the same ANOVA table but the interpretation of the parameter will change.

The most common method is to set a specific level of a factor as a reference. Another method is to set the average effect over all the levels of the factor to be zero. There are many other contrasts possible.

The model will have a single parameter to reflect a “typical” value and \( l - 1 \) parameters that model the deviations from this value.

**Important:** When interpreting the output from an ANOVA model make sure you know which parametrization is used by your software.
ANOVA Example: Sepal length in 3 species of Iris

There is a clear indication of a difference in the sepal length between the species.
ANOVA results for the iris data

## ANOVA Table

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species</td>
<td>2</td>
<td>13.789</td>
<td>6.8943</td>
<td>29.314</td>
</tr>
<tr>
<td>Residuals</td>
<td>27</td>
<td>6.350</td>
<td>0.2352</td>
<td></td>
</tr>
</tbody>
</table>

The ANOVA test for the significant differences between the 3 species has only 2 degrees of freedom. This is the number of variables you need to estimate to compare all the level of the factor. In general testing a factor with \( l \) levels will have \( l - 1 \) degrees of freedom.

The residual term represents the error in the model. The degrees of freedom here represents the sample size available to estimate the error in the model. The degrees of freedom for error mainly determines the power of the ANOVA test.
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.

<table>
<thead>
<tr>
<th></th>
<th>Diff</th>
<th>SE</th>
<th>Tstat</th>
</tr>
</thead>
<tbody>
<tr>
<td>versicolor - setosa</td>
<td>0.87</td>
<td>0.217</td>
<td>4.01</td>
</tr>
<tr>
<td>virginica - setosa</td>
<td>1.66</td>
<td>0.217</td>
<td>7.65</td>
</tr>
<tr>
<td>virginica - versicolor</td>
<td>0.79</td>
<td>0.217</td>
<td>3.64</td>
</tr>
</tbody>
</table>

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. The number of pairwise comparisons increases dramatically with the number of levels in the factor. For example with 4 levels there are 6 pairwise comparisons and with 5 levels there are 10 pairwise comparisons.
ANOVA with a blocking factor

ANOVAs may contain more than one factor. Factors that are not of interest are considered blocking factors. In general, blocking factors do not interact with factors of interest. They 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, $A$ is a factor with $I$ levels, $B$ is the blocking factor with $J$ levels. Usually every level of $A$ is observed at every level of $B$.

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

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

Our hypothesis of interest is

$H_0 : \text{all } \alpha_i = 0, \ H_1 : \text{at least one } \alpha_i \neq 0$
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

Both the block and the drug are significant. 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

Note that block ANOVA with \( a = 2 \) and only a single observation for each group in each block is equivalent to a paired t-test.

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

We can quantify the difference in the group means.

```r
## 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. In fact apply a Kruskal-Wallis test when there are only 2 groups is identical to a Mann-Whitney test.

Friedman rank-sum test can be used as a non-parametric way to analyse unreplicated \((n = 1)\) complete blocked data. However if there are only 2 groups, this 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 as we have seen an ANOVA model is still valid even if the normality assumption is violated. However the power of the non-parametric tests may be better. This is especially true if the data is highly skewed.
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, \( A \) is a factor with \( I \) levels, \( B \) is a factor with \( J \) levels. Usually every level of \( A \) is observed at every level of \( B \) (fully crossed).

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

\[ \mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} \]

We are interested in testing

\( H_0 : \) 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 \( \alpha_i = 0, \) \( H_1 : \) at least one \( \alpha_i \neq 0 \)

\( H_0 : \) all \( \beta_j = 0, \) \( H_1 : \) at least one \( \beta_j \neq 0 \)
Meaning of the interaction

We cannot interpret the main effects of a factor in a model where that factor interacts with another because that interaction means the effects of the two factors depend on each other. An interaction between factor $A$ and $B$ means the effect of $A$ depends on which level of $B$ is applied and the effect of $B$ depends on which level of $A$ is applied.

A main effect need not be significant in the presence of a significant interaction term. That does not mean the main effect can be removed from the model. In fact 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.

Treatments are Vitamin C dose and delivery method.
The Interaction plot

An interaction plot is a simple way to check if an interaction exists. We plot the mean of the response in each of the combined factor levels \((2\times3=6)\). We select one factor for the X axis and use different colour lines to indicate the other factor.
## ANOVA Table

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>supp</td>
<td>1</td>
<td>205.4</td>
<td>205.4</td>
<td>15.572</td>
<td>0.000231</td>
</tr>
<tr>
<td>dose</td>
<td>2</td>
<td>2426.4</td>
<td>1213.2</td>
<td>92.000</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>supp:dose</td>
<td>2</td>
<td>108.3</td>
<td>54.2</td>
<td>4.107</td>
<td>0.021860</td>
</tr>
<tr>
<td>Residuals</td>
<td>54</td>
<td>712.1</td>
<td>13.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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, the main effects cannot be interpreted directly and posthoc comparisons must be done within the levels of the other factor.

For our example, the effect of dose depends on the delivery method of Vitamin C. There are 6 groups in total 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 have both the delivery method and the dose change between the 2 groups being compared. These comparisons are usually not of interest.
## Within Dose Level

<table>
<thead>
<tr>
<th></th>
<th>Diff</th>
<th>SE</th>
<th>Tstat</th>
</tr>
</thead>
<tbody>
<tr>
<td>VC:0.5 - OJ:0.5</td>
<td>-5.25</td>
<td>1.62</td>
<td>-3.2327</td>
</tr>
<tr>
<td>VC:1.0 - OJ:1.0</td>
<td>-5.93</td>
<td>1.62</td>
<td>-3.6514</td>
</tr>
<tr>
<td>VC:2.0 - OJ:2.0</td>
<td>0.08</td>
<td>1.62</td>
<td>0.0493</td>
</tr>
</tbody>
</table>

## Within delivery method

<table>
<thead>
<tr>
<th></th>
<th>Diff</th>
<th>SE</th>
<th>Tstat</th>
</tr>
</thead>
<tbody>
<tr>
<td>OJ:1.0 - OJ:0.5</td>
<td>9.47</td>
<td>1.62</td>
<td>5.83</td>
</tr>
<tr>
<td>OJ:2.0 - OJ:0.5</td>
<td>12.83</td>
<td>1.62</td>
<td>7.90</td>
</tr>
<tr>
<td>OJ:2.0 - OJ:1.0</td>
<td>3.36</td>
<td>1.62</td>
<td>2.07</td>
</tr>
<tr>
<td>VC:1.0 - VC:0.5</td>
<td>8.79</td>
<td>1.62</td>
<td>5.41</td>
</tr>
<tr>
<td>VC:2.0 - VC:0.5</td>
<td>18.16</td>
<td>1.62</td>
<td>11.18</td>
</tr>
<tr>
<td>VC:2.0 - VC:1.0</td>
<td>9.37</td>
<td>1.62</td>
<td>5.77</td>
</tr>
</tbody>
</table>
More Complicated ANOVA

ANOVA models can have any number of factors. However 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.

For example if we have 3 factors of interest, one with 3 levels, one with 4 levels, and one with 5 levels; then there are 60 unique groups in our experiment \((3 \times 4 \times 5 = 60)\).

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.

This experiment looks at the effect of nitrogen (N), phosphate (P) and potassium (K) on the growth of peas. The experiment was conducted in 6 plots of land each subdivided into 4 sections.

The blocking factor (plot) does not interact with any of the 3 other factors. But the N, P and K factors may interact with each other. Since we have 3 factors, each with 2 levels (present/absent), there are 8 groups in total. However 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 that confounds the three way interaction between the 3 elements.

Generally, three way and higher order interactions are hard to interpret.
The model contains a main effect for the block and three elements factors, and 3 two way interactions between the 3 element factors. Block is significant but none of the two way interactions are significant. We cannot interpret the main effects although it looks like N is significant
Since we see no evidence of interactions, we should refit the model excluding interactions so we can interpret the main effects.

## Response: yield

<table>
<thead>
<tr>
<th></th>
<th>Sum Sq</th>
<th>Df</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>189.282</td>
<td>1</td>
<td>11.8210</td>
<td>0.00366</td>
</tr>
<tr>
<td>P</td>
<td>8.402</td>
<td>1</td>
<td>0.5247</td>
<td>0.47999</td>
</tr>
<tr>
<td>K</td>
<td>95.202</td>
<td>1</td>
<td>5.9455</td>
<td>0.02767</td>
</tr>
<tr>
<td>Residuals</td>
<td>240.185</td>
<td>15</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Now we have significant effects for both N and K. We can look at the estimated effects.

|       | 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|
Questions?

- www.stat.ubc.ca/SCARL
- STAT 551 - Stat grad students taking this course offer free statistical advice. Fall semester every academic year.
- SOS Program - An hour of free consulting to UBC graduate students. Funded by the Provost and VP Research.
- Short Term Consulting Service - Advice from Stat grad students. Fee-for-service on small projects (less than 15 hours).
- Hourly Projects - SCARL professional staff. Fee-for-service consulting.

The End