01NEX - Lecture 02
Comparing several treatment means, linear regression

Jiri Franc

Czech Technical University
Faculty of Nuclear Sciences and Physical Engineering
Department of Mathematics

Single factor experiment

Suppose we have a **treatment** - different **levels** of a single **factor** that we wish to compare.

There will be, in general, n observations under the ith treatment.

Typical Data for a Single Factor Experiment

Treatment (level)	Observations				Totals	Averages
1	y_{11}	y_{12}	$\cdots$	y_{1n}	$y_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	$\cdots$	y_{2n}	$y_{2.}$	$\bar{y}_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\vdots$	$\vdots$
a	y_{a1}	y_{a2}	$\cdots$	y_{an}	$y_{a.}$	$\bar{y}_{a.}$
TOTAL					$y_{..}$	$\bar{y}_{..}$

$$y_{i.} = \sum_{j=1}^{n} y_{ij} \quad y_{..} = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij} \quad \bar{y}_{i.} = \frac{y_{i.}}{n} \quad \bar{y}_{..} = \frac{y_{..}}{N} \quad i = 1, 2, \dots, a$$

and $N = (a \cdot n)$ is the total number of observations.

In this lesson we will mostly work with the **balance models**, all factor levels are replicated the same number of times.

Single factor experiment - models

Means model:

$$y_{ij} = \mu_i + \epsilon_{ij} \quad i = 1, 2, \ldots, a \quad j = 1, 2, \ldots, n_i,$$

where y_{ij} is the ijth observation, μ_i is the mean of the ith factor level and ϵ_{ij} is random error.

Effects model:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad i = 1, 2, \ldots, a \quad j = 1, 2, \ldots, n_i,$$

where μ is **overall mean** of the measurements and α_i is the ith effect of factor A.

Standard definition of the overall mean is:

$$\mu = \sum_{i=1}^{a} w_i \mu_i, \quad \text{where} \sum_{i=1}^{a} w_i = 1,$$

with the most frequent setting: $w_i = \frac{1}{a}$ for $i = 1, 2, \ldots, a$.

The weighted average is used when there are an unequal number of observations in each treatment. The weights w_i could be taken as $\frac{n_i}{N}$, for balanced model $n_i = n, \forall i$.

Single factor experiment - models

Regression model without intercept for a single factor experiment:

$$y_{ij} = \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_n x_{nj} + \epsilon_{ij} \quad i = 1, 2, \ldots, a \quad j = 1, 2, \ldots, n_i,$$

where the regression variables are indicators (i.e. take on the values 1, if observation j is from treatment i, and 0 otherwise)

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}_{n_1} \\ \mathbf{y}_{n_2} \\ \vdots \\ \mathbf{y}_{n_a} \end{pmatrix} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{1}_{n_a} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_a \end{pmatrix} + \begin{pmatrix} \epsilon_{n_1} \\ \epsilon_{n_2} \\ \vdots \\ \epsilon_{n_a} \end{pmatrix} = X\beta + \epsilon$$

The relationship between the parameters in the Means and Regression model:

$$\beta_i = \mu_i \quad i = 1, 2, \ldots, a.$$

Single factor experiment - models

Regression model with intercept for a single factor experiment:

$$y_{ij} = \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_n x_{nj} + \epsilon_{ij} \quad i = 1, 2, \ldots, a \quad j = 1, 2, \ldots, n_i,$$

where $x_{1j} = 1, \forall j$, and the others regression variables are indicators (i.e. take on the values 1, if observation j is from treatment i, and 0 otherwise)

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}_{n_1} \\ \mathbf{y}_{n_2} \\ \vdots \\ \mathbf{y}_{n_a} \end{pmatrix} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{1}_{n_2} & \mathbf{1}_{n_2} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{1}_{n_a} & \mathbf{0} & \ldots & \mathbf{1}_{n_a} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_a \end{pmatrix} + \begin{pmatrix} \epsilon_{n_1} \\ \epsilon_{n_2} \\ \vdots \\ \epsilon_{n_a} \end{pmatrix} = X\beta + \epsilon$$

The relationship between the parameters in the Effects and Regression model:

$$\beta_1 = \mu_1 \quad \text{and} \quad \beta_i = \mu_i - \mu_1 \quad i = 2, \ldots, a.$$

Analysis of Fixed effects model

We are interested in testing the equality of the a factor levels means:

$$E(y_{ij}) = \mu + \alpha_i = \mu_i \quad i = 1, 2, \ldots, a.$$

The appropriate hypothesis is:

$$\begin{aligned} H_0 &: \quad \mu_1 = \mu_2 = \cdots = \mu_a. \\ H_1 &: \quad \mu_i \neq \mu_j \quad \text{for at least one pair } (i, j). \end{aligned}$$

If we assume the homoscedasticity (the finite variance σ^2 is constant for all levels of factor and observations are mutually independent) and

$$y_{ij} \sim N(\mu_i, \sigma^2)$$

then the appropriate procedure for testing the equality of several means is: **analaysis of variance (ANOVA)**.

Decomposition of Total Sum of Squares

Name **ANOVA** comes from a partitioning of total variability into its components parts.

Overall total variability = between factor levels variability + within variability

$$SS_T = SS_A + SS_E$$

The total corrected sum of squares: $SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{..})^2$

The sum of squares due to treatment A: $SS_A = n \sum_{i=1}^{a} (\bar{y}_{i.} - \bar{y}_{..})^2$

The error sum of squares: $SS_E = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2 = SS_T - SS_A$

Mean squares:

$$MS_A = \frac{SS_A}{a-1}, \quad MS_E = \frac{SS_E}{N-a}.$$

Expected values of the mean squares:

$$E(MS_E) = \sigma^2, \quad E(MS_A) = \sigma^2 + \frac{n \sum_{i=1}^{a} \alpha_i^2}{a-1}.$$

MS_E estimates σ^2, and, if there are no differences in treatment means, MS_A is also an unbiased estimator of σ^2.

ANOVA table for Single - Factor, Fixed Effect Balanced Model

ANOVA table for Single - Factor, Fixed Effect Balanced Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Between treatments	$SS_A = n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2$	$a-1$	MS_A	$F_0 = \frac{MS_A}{MS_E}$
Within treatments	$SS_E = SS_T - SS_A$	$N-a$	MS_E	
Total	$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$	$N-1$		

If data are unbalanced, the manual computational formulas for SS_T and SS_A:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{..}^2}{N} \quad \text{and} \quad SS_A = \sum_{i=1}^{a}\frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N},$$

where n_i is number of observations taken under treatment i and $N = \sum_{i=1}^{a} n_i$.

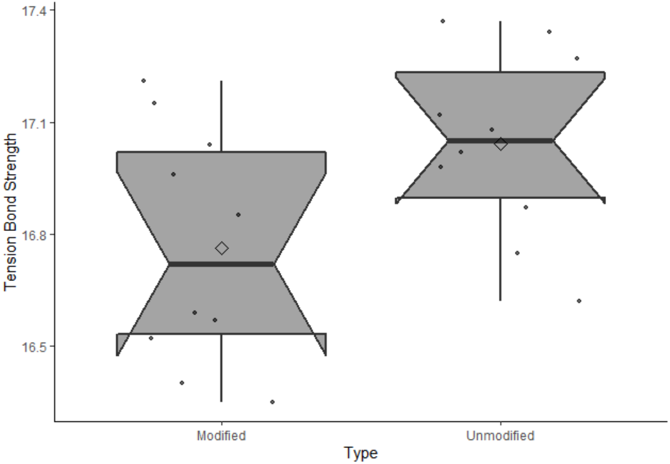
Cochran's theorem implies that $\frac{SS_A}{\sigma^2}$ and $\frac{SS_E}{\sigma^2}$ are independently distributed chi-square random variables and the **test statistic** is given by:

$$F_0 = \frac{\frac{SS_A}{a-1}}{\frac{SS_E}{N-a}} = \frac{MS_A}{MS_E}.$$

ANOVA for simple comparative experiment

In simple comparative experiment the ANOVA model changes into simple t-test.

Example from lecture 01:



ANOVA for simple comparative experiment

In simple comparative experiment (see lecture 01) the ANOVA model changes into simple t-test.

Example from lecture 01 and results in R:

```
> t.test(Factor1, Factor2 , alternative = "two.sided", mu = 0,
         paired = FALSE, var.equal = TRUE, conf.level = 0.95)

Two Sample t-test
data:  Modified and Unmodified
t = -2.1869, df = 18, p-value = 0.0422
alternative hypothesis: true difference in means is not equal
95 percent confidence interval:       -0.54507339 -0.01092661
sample estimates:  mean of x mean of y 16.764      17.042


> summary(aov(Response~Factor,data=data.cement))

            Df  Sum Sq   Mean Sq   F value  Pr(>F)
Factor       1  0.3864    0.3864    4.782   0.0422 *
Residuals   18  1.4544    0.0808
```

Example - Plasma Etching Experiment data

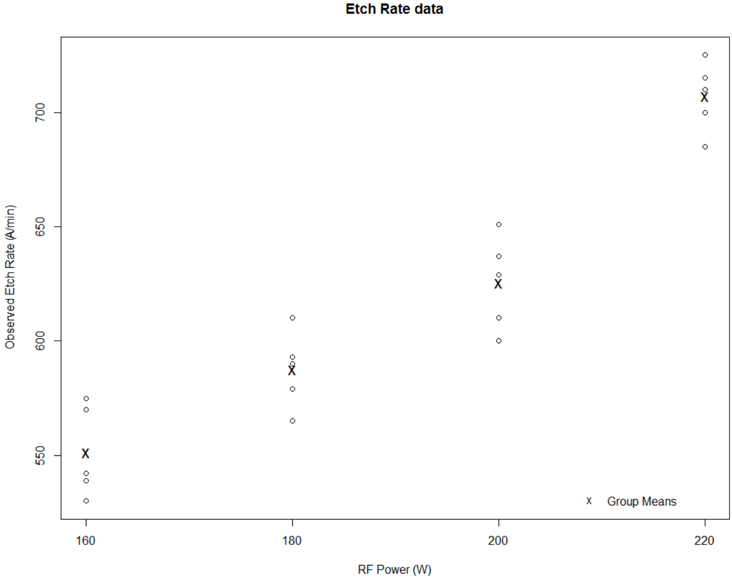
Etch Rate Data (in A/min) from Plasma Etching Experiment

Power	Observations (5 runs)						
(W)	1	2	3	4	5	Totals	Averages
160	575	542	530	539	570	2 756	551.2
180	565	593	590	579	610	2 937	587.4
200	600	651	610	637	629	3 127	625.4
220	725	700	715	685	710	3 535	707.0
						$y_{\cdot\cdot} = 12355$	$\bar{y}_{\cdot\cdot} = 617.75$

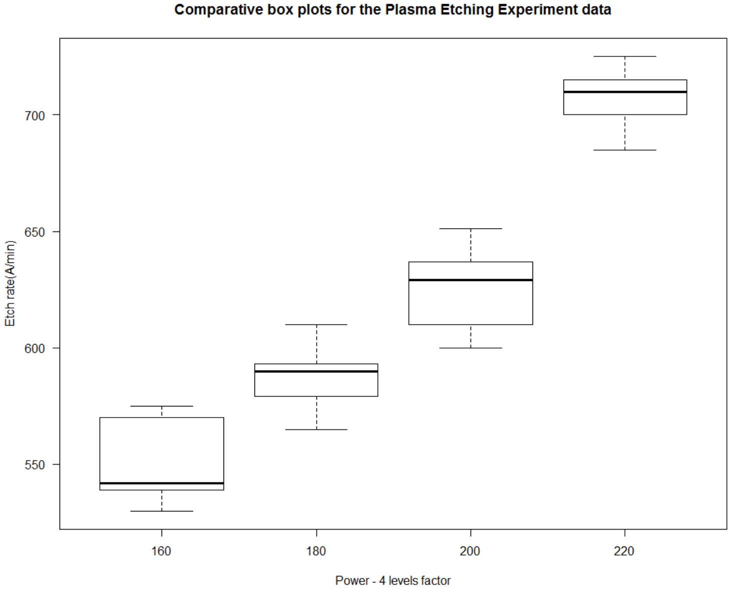
20 runs were made in random order.

Performing all 6 pairwise t-tests is inefficient. It takes a lot of effort and it inflates the type I error.

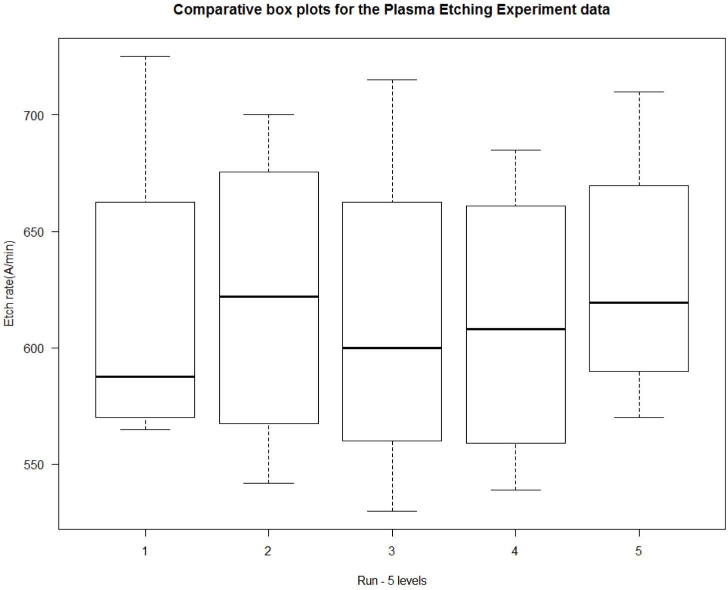
Example - Plasma Etching Experiment data



Example - Plasma Etching Experiment data



Example - Plasma Etching Experiment data



ANOVA for the Plasma Etching Experiment

```
> etch.rate.aov1 <- aov(rate~Power+Run,etch.rate)
> summary(etch.rate.aov1)
            Df Sum Sq Mean Sq F value   Pr(>F)
Power        3  66871   22290  62.369 1.37e-07 ***
Run          4   1051     263   0.735    0.586
Residuals   12   4289     357

> etch.rate.aov2 <- aov(rate~Power,etch.rate)
> summary(etch.rate.aov2)
            Df Sum Sq Mean Sq F value   Pr(>F)
Power        3  66871   22290    66.8 2.88e-09 ***
Residuals   16   5339     334

> anova(etch.rate.aov1,etch.rate.aov2)
Analysis of Variance Table
Model 1: rate ~ Power + Run
Model 2: rate ~ Power
     Res.Df   RSS       Df   Sum of Sq  F        Pr(>F)
1    12       4288.7
2    16       5339.2   -4   -1050.5    0.7348   0.5857
```

ANOVA for the Plasma Etching Experiment

```
>anova(lm(rate~Power,etch.rate))
# same as
> summary(aov(rate~Power,etch.rate))
            Df Sum Sq Mean Sq F value    Pr(>F)
Power        3  66871   22290    66.8 2.88e-09 ***
Residuals   16   5339     334
```

$$SS_E = SS_{Residuals} = SS_{Total} - SS_{Power} = 72209.75 - 66870.55 = 5339.20$$

$$F_0 = \frac{22290.18}{333.7} = 66.8$$

Because $F_0 = 66.8 > 5.29 = F_{0.01,3,16}$ we reject H_0 at the significance level $\alpha = 0.01$ and conclude that the treatment means differ.

Effect model for the Plasma Etching Experiment

```
> model1  = lm(rate~Power, data=etch.rate)
> summary(model1)
Call:
lm(formula = rate ~ Power, data = etch.rate)
Residuals:
   Min     1Q Median     3Q    Max
 -25.4  -13.0    2.8   13.2   25.6
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  551.200      8.169  67.471  < 2e-16 ***
Power180      36.200     11.553   3.133  0.00642 **
Power200      74.200     11.553   6.422 8.44e-06 ***
Power220     155.800     11.553  13.485 3.73e-10 ***

> confint(model1, level = 0.95))
                2.5 %    97.5 %
(Intercept) 533.88153 568.51847
Power180     11.70798  60.69202
Power200     49.70798  98.69202
Power220    131.30798 180.29202
```

Means model for the Plasma Etching Experiment

```
> model2 = lm(rate~Power - 1, data=etch.rate)
> summary(model2)
Call:
lm(formula = rate ~ Power - 1, data = etch.rate)
Residuals:
   Min     1Q Median     3Q    Max
 -25.4  -13.0    2.8   13.2   25.6
Coefficients:
         Estimate Std. Error t value Pr(>|t|)
Power160  551.200      8.169   67.47   <2e-16 ***
Power180  587.400      8.169   71.90   <2e-16 ***
Power200  625.400      8.169   76.55   <2e-16 ***
Power220  707.000      8.169   86.54   <2e-16 ***

> confint(model2, level = 0.95))
            2.5 %   97.5 %
Power160 533.8815 568.5185
Power180 570.0815 604.7185
Power200 608.0815 642.7185
Power220 689.6815 724.3185
```

Means model for the Plasma Etching Experiment

$100(1-\alpha)$ percent confidence interval on the ith treatment mean μ_i is:

$$\bar{y}_{i.} - t_{\frac{\alpha}{2},N-a}\sqrt{\frac{MS_E}{n}} \leq \mu_i \leq \bar{y}_{i.} + t_{\frac{\alpha}{2},N-a}\sqrt{\frac{MS_E}{n}}$$

```
> model2 = lm(rate~Power - 1, data=etch.rate)
              2.5 %  Estimate     97.5 %
Power160  533.8815   551.200   568.5185
Power180  570.0815   587.400   604.7185
Power200  608.0815   625.400   642.7185
Power220  689.6815   707.000   724.3185
```

Example of 95% confidence interval of treatment 4:

$$689.6815 = 707 - 2.120\sqrt{\frac{333.7}{5}} \leq \mu_4 \leq 707 + 2.120\sqrt{\frac{333.7}{5}} = 724.3185$$

Treatment effects in the Plasma Etching Experiment

```
> overall mean of Plasma Etching Experiment data
> (erate.mean <- mean(etch.rate$rate))
   617.75

> etch.rate.aov <- aov(rate~Power,etch.rate)
> model.tables(etch.rate.aov)
Tables of effects
Power
   160    180    200    220
-66.55 -30.35   7.65  89.25

> (MSe <- summary(etch.rate.aov)[[1]][2,3])
   333.7
> (SD <- sqrt(MSe/16))
   4.566864
```

ANOVA - Model Adequacy Checking

To check:

- Normality
- Independence
- Constant Variance

Useful instruments are **residual plots**:

- **Residuals**: $e_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \bar{y}_{i.}$
- **Standartized Residuals**: $d_{ij} = \frac{e_{ij}}{\sqrt{MS_E}}$
- An adequate model produces residual plots that are structureless.
- Plot of Residuals vs. Fitted Values, Factor Levels, or Time Order should be a band of “random noise”.
- Constant Variance

Residuals

Residuals vs. Fitted Values:

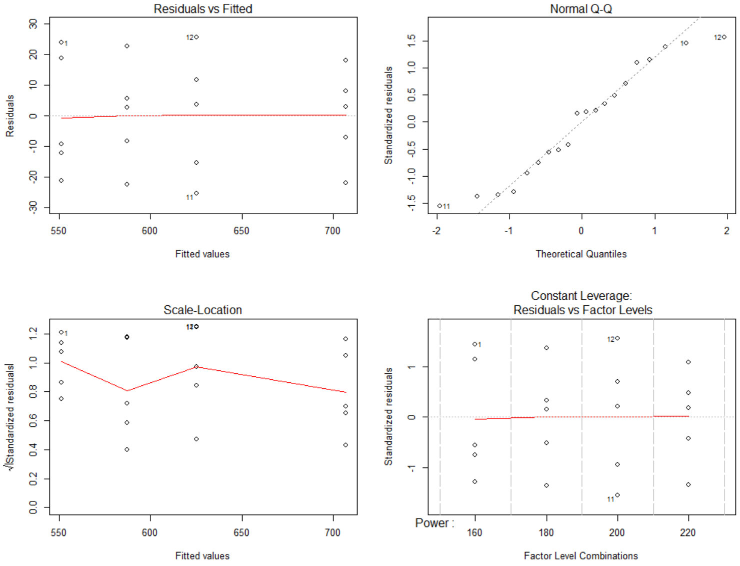
- Trumpet: Error variance depends on the mean (Transformation)
- Arch: Possible need for second order term or transformation
- Constant increase, decrease: Error in analysis (Possible omission of intercept)

Residuals vs. Factor Levels:

- Trumpet: Error variance depends on the factor (Variance Effect)
- Arch: Possible need for that factor's quadratic term or transformation
- Constant increase, decrease: Error in analysis (Possible need to include the main effect of the factor)

For more detailed discussion about the variance stabilizing transformations see Box-Cox transformation (Box and Cox (1964)).

Model Adequacy Checking for the Plasma Etching Experiment



Model Adequacy Checking for the Plasma Etching Experiment - Equality of Variance

Bartlett's test:

```
> bartlett.test(rate~RF,data=etch.rate)
Bartlett test of homogeneity of variances
data:  rate by RF
Bartlett's K-squared = 0.4335, df = 3, p-value = 0.9332
```

Bartlett's test is very sensitive to the normality assumption. When the validity of this assumption is doubtful, this test should not be used.

Levene test:

```
> leveneTest(etch.rate.aov)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  0.1959 0.8977
```

Levane's test statistic is simply the usual ANOVA F statistic for testing equality of means applied to the absolute deviations.

Model Adequacy Checking for the Plasma Etching Experiment - Normality

```
> y1 = etch.rate$rate[etch.rate$RF==160]
```

Shapiro-Wilk normality test:

```
       Shapiro-Wilk normality test
data:  y1
W = 0.8723, p-value = 0.2758
```

Kolmogorov-Smirnov test:

```
One-sample Kolmogorov-Smirnov test
data:  y1
D = 0.2771, p-value = 0.7519
alternative hypothesis: two-sided
```

The Plasma Etching Experiment - Regression Model

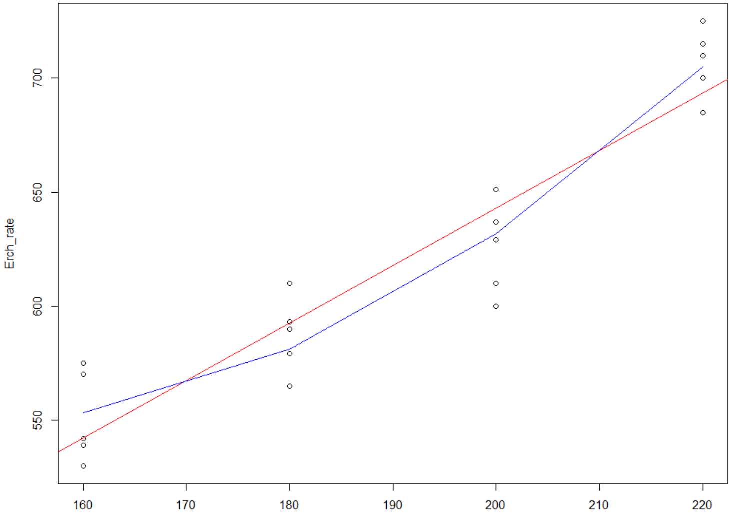
The factors involved in an experiment can be either quantitative or qualitative.

```
lm(formula = Erch_rate ~ Power1)
Residuals:    Min     1Q      Median  3Q      Max
             -43.02 -12.32   -1.21   16.71   33.06
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 137.6200    41.2108   3.339  0.00365 **
Power1        2.5270     0.2154  11.731 7.26e-10 ***
Residual standard error: 21.54 on 18 degrees of freedom
Multiple R-squared: 0.8843,     Adjusted R-squared: 0.8779

lm(formula = Erch_rate ~ Power1 + Power2)
Residuals:   Min      1Q       Median  3Q      Max
            -31.67 - 14.75   1.48    13.08   28.87
Coefficients:
              Estimate Std. Error t value  Pr(>|t|)
(Intercept) 1147.77000  368.52081   3.115  0.00631 **
Power1        -8.25550    3.91993  -2.106  0.05037 .
Power2         0.02838    0.01030   2.754  0.01356 *
Residual standard error: 18.43 on 17 degrees of freedom
Multiple R-squared:  0.92,      Adjusted R-squared: 0.9106
```

The Plasma Etching Experiment - Regression Model

Comparison of two regressions models, with and without curvature (square of explanatory variable).



Multiple comparisons problem (multiple testing problem)

History (from Wikipedia):
The interest in the problem of multiple comparisons began in the 1950s with the work of Tukey and Scheffé. New methods and procedures came out: Closed testing procedure (Marcus et al., 1976), Holm–Bonferroni method (1979). Later, in the 1980s, the issue of multiple comparisons came back (Hochberg and Tamhane (1987), Westfall and Young (1993), and Hsu (1996)).

Example:
For example, if one test is performed at the 5% level, there is only a 5% chance of incorrectly rejecting the null hypothesis if the null hypothesis is true. However, for 100 tests where all null hypotheses are true, the expected number of incorrect rejections is 5. If the tests are independent, the probability of at least one incorrect rejection is 99.4%.

In R:
You can use the function *pairwise.t.test* with different with different corrections: ("hochberg", "bonferroni", "holm", "hommel","BH", "BY", "fdr", "none"). For details see *p.adjust.methods*.

Post-ANOVA Comparison of Means for The Plasma Etching Experiment

```
> pairwise.t.test(Erch_rate,Power1,p.adjust.method="bonferroni
        Pairwise comparisons using t tests with pooled SD
data:  Erch_rate and Power1
    160     180     200
180 0.038   -       -
200 5.1e-05 0.028   -
220 2.2e-09 1.0e-07 1.6e-05
P value adjustment method: bonferroni

> pairwise.t.test(Erch_rate,Power1,p.adjust.method="hochberg")
        Pairwise comparisons using t tests with pooled SD
data:  Erch_rate and Power1
    160     180     200
180 0.0064  -       -
200 2.5e-05 0.0064  -
220 2.2e-09 8.5e-08 1.1e-05
P value adjustment method: hochberg
```

Tukey HSD for the The Plasma Etching Experiment

Tukey (1953) proposed a procedure for testing hypothesis for which the overall significance level is exactly α when sample size are n_i and n_j.

Studentized range statistic:

$$q = \frac{\bar{y}_{max} - \bar{y}_{min}}{\sqrt{\frac{MS_E}{2}(\frac{1}{n_i} + \frac{1}{n_j})}}$$

where $\bar{y}_{max}$ and $\bar{y}_{min}$ are the largest and smallest sample means.

Tukey's test declares two means significantly different if the absolute value of their sample differences exceeds

$$T_\alpha = q_\alpha(a, f)\sqrt{\frac{MS_E}{2}(\frac{1}{n_i} + \frac{1}{n_j})},$$

where a is number of levels and f is number of degrees of freedom for error. $q_\alpha(a, f)$'s are tabularized.

Plasma etching experiment example:

$$T_{0.05} = q_{0.05}(4, 16)\sqrt{\frac{MS_E}{n}} = 4.05\sqrt{\frac{333.7}{5}} = 33.09$$

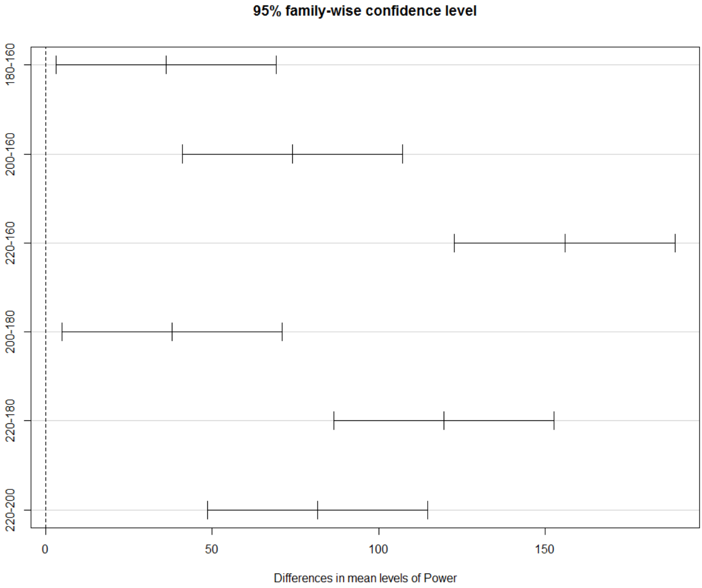
Thus, any pair of treatment averages that differ in absolute value by more than 33.09 would imply that the corresponding pair of population means are significantly different.

Tukey HSD for the The Plasma Etching Experiment

Create a set of confidence intervals on the differences between the means of the levels of a factor. The intervals are based on the Tukeys Honest Significant Difference method.

```
> TukeyHSD(etch.rate.aov, ordered = FALSE,conf.level = 0.95)
  Tukey multiple comparisons of means 95 confidence level
Fit: aov(formula = rate ~ Power, data = etch.rate)
Power   diff   lwr        upr       p adj
180-160  36.2   3.145624  69.25438 0.0294279
200-160  74.2  41.145624 107.25438 0.0000455
220-160 155.8 122.745624 188.85438 0.0000000
200-180  38.0   4.945624  71.05438 0.0215995
220-180 119.6  86.545624 152.65438 0.0000001
220-200  81.6  48.545624 114.65438 0.0000146
> plot(TukeyHSD(etch.rate.aov, ordered = FALSE,...
        conf.level = 0.95,las=1) )
```

Tukey HSD for the Plasma Etching Experiment



Fisher's LSD for the The Plasma Etching Experiment Method

Fisher's Least Significant Difference Method uses the t statistic for testing H_0 : $\mu_i = \mu_j$:

$$t_0 = \frac{\bar{y}_{i.} - \bar{y}_{j.}}{\sqrt{MS_E(\frac{1}{n_i} + \frac{1}{n_j})}}$$

Assuming a two-sided alternative. The quantity

$$LSD = t_{\frac{\alpha}{2},N-a}\sqrt{MS_E(\frac{1}{n_i} + \frac{1}{n_j})}$$

is called the least significance difference. If $|\bar{y}_{i.} - \bar{y}_{j.}| > LSD$, we conclude that the population means μ_i and μ_j differ.

Fisher's LSD for the The Plasma Etching Experiment Method

Fisher's Least Significant Difference Method at $\alpha = 0.05$ for he Plasma Etching Experiment Method is:

$$LSD = t_{0.025,16}\sqrt{\frac{2MS_E}{n}} = 2.120\sqrt{\frac{2(333.7)}{5}} = 24.49$$

Thus, any pair of treatment averages that differ in absolute value by more than 24.49 would imply that the corresponding pair of population means are significantly different.

Fisher's LSD

```
> LSD.test(etch.rate$rate, etch.rate$Power, 16,334)
Study: LSD t Test for etch.rate$rate
Mean Square Error:  334
etch.rate$Power,  means and individual ( 95 %) CI
    etch.rate.rate  std.err r      LCL      UCL Min. Max.
160          551.2 8.952095 5 532.2224 570.1776  530  575
180          587.4 7.487323 5 571.5276 603.2724  565  610
200          625.4 9.179325 5 605.9407 644.8593  600  651
220          707.0 6.819091 5 692.5442 721.4558  685  725
alpha: 0.05 ; Df Error: 16
Critical Value of t: 2.119905
Least Significant Difference 24.50302
Means with the same letter are not significantly different.
Groups, Treatments and means
a   220   707
b   200   625.4
c   180   587.4
d   160   551.2
```

Sample Size Determination

- Sample size depends on type of experiment, how it will be conducted, resources, and desired sensitivity
- Sensitivity refers to the difference in means that the experimenter wishes to detect
- Generally, increasing the number of replications increases the sensitivity or it makes it easier to detect small differences in means

Sample Size Determination

Operating characteristic (OC) curve is a plot of type II error probability of statistical test for a particular sample size versus a parameter that reflects the space to witch the null hypothesis is false. For single factor fixed effects model:

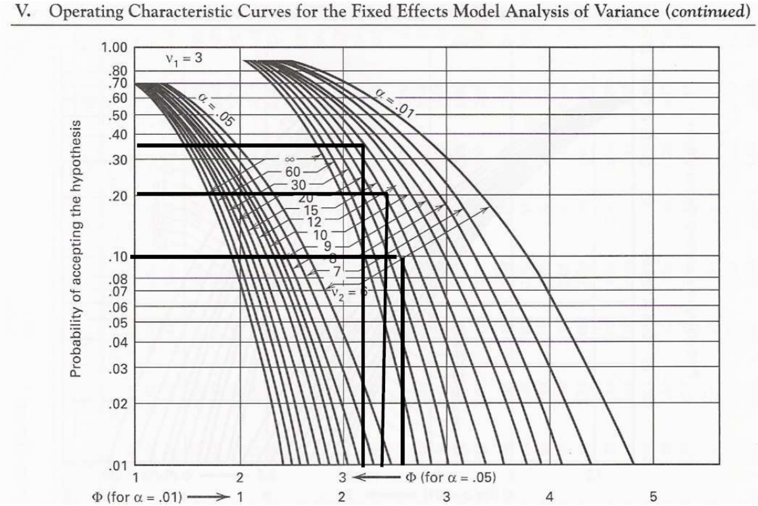
$$\beta = 1 - P\left[\text{Reject } H_0 \mid H_0 \text{ is false}\right] = 1 - P\left[F_0 > F_{\alpha,a-1,N-a} \mid H_0 \text{ is false}\right]$$

It can be shown that, if H_0 is false, the statistic $F_0 = \frac{MS_A}{MS_E}$ is distributed as a noncentral F random variable with $a-1$ and N-a degrees of freedom and the noncentrality parameter δ.

Operating characteristic curves plot β against a parameter Φ^2

$$\Phi^2 = \frac{n\sum_{i=1}^{a}(\mu_i - \bar{\mu})^2}{a\sigma^2}.$$

Sample Size Determination



Sample Size Determination for the Plasma Etching Experiment

One way how to use OC curves is to define a difference in two means D of interest, then the minimum value of Φ^2 is

$$\Phi^2 = \frac{nD^2}{2a\sigma^2}$$

In Plasma Etching Experiment, suppose we would like to reject the null hypothesis with a probability of at least 0.90 if any two treatment means differed by as much as 75 A/minute and $\alpha = 0.01$.

Sample Size Determination for the Plasma Etching Experiment

Sample size	Φ^2	Φ	DF for errors	Power
4	4.50	2.12	12	0.61
5	5.62	2.37	16	0.80
6	6.75	2.60	20	0.92

Sample Size Determination for the Plasma Etching Experiment

In the Plasma Etching Experiment, suppose we would like to reject the null hypothesis with a probability of at least 0.90 if any two treatment means differed by as much as 75 A/minute and $\alpha = 0.01$.

```
> nn             = seq(4,10,by=1)
> sd             = 25
> max_difference = 75
> DF             = 3
> beta <- c(NA,nr=length(sd),nc=length(nn))
> for (i in 1:length(sd))
+   beta[i,] <- power.anova.test(groups=4,n= nn,
               between.var = (max_difference^2)/(2*DF),
               within.var=(sd^2), sig.level=.01)$power
       4         5         6        7         8         9
Power  0.6064585 0.8048383 0.915384 0.9669989 0.9881851 0.9960
```

Sample Size Determination for the Plasma Etching Experiment

Power computation for given ANOVA table from the Plasma Etching Experiment:

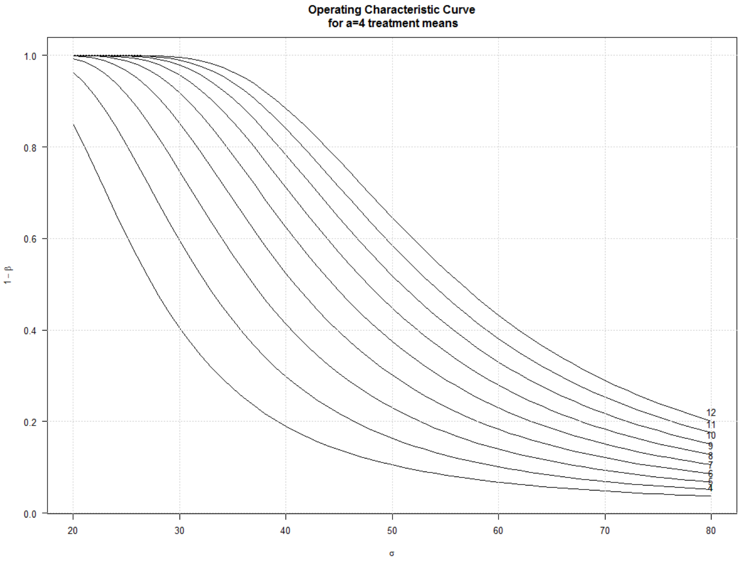
```
power.anova.test(groups=4, n=5, between.var = MS_A ,
                within.var= MS_E , sig.level=.01)$power
Balanced one-way analysis of variance power calculation
         groups = 4
              n = 5
    between.var = 22290
     within.var = 334
      sig.level = 0.01
          power = 1
 NOTE: n is number in each group
```

Sample Size Determination for the Plasma Etching Experiment

Example of sample size computation:

```
> power.anova.test(groups=4, power = 0.9,
                   between.var = 1000,
                    within.var=500 , sig.level=.01)
     Balanced one-way analysis of variance power calculation
         groups = 4
              n = 4.764321
    between.var = 1000
     within.var = 500
      sig.level = 0.01
          power = 0.9
 NOTE: n is number in each group
```

Sample Size Determination



Today Exercise & Next lecture

Today Exercise:

- Do exercise 3.7, 3.8, 3.9, and 3.10.
- Use the R to create and analyze given designs.

Data and exercises come from D.C. Montgomery: Design and Analysis of Experiment.

Next Lectures: Randomized blocks, Latin Squares.

- Randomized blocks,
- Latin Square design,
- Graeco-Latin Squaree design,
- Balanced Incomplete Block design,
- First Homework - real measurement during the lesson.

Exercises 3.07

The tensile strength of Portland cement is being studied. Four different mixing techniques can be used economically. A completely randomized experiment was conducted and the following data were collected:

Mixing	Technique Tensile Strength (lb/in2)			
1	3129	3000	2865	2890
2	3200	3300	2975	3150
3	2800	2900	2985	3050
4	2600	2700	2600	2765

1. Test the hypothesis that mixing techniques affect the strength of the cement. Use $\alpha = 0.05$.
2. Construct a graphical display as described in Section 3.5.3 to compare the mean tensile strengths for the four mixing techniques. What are your conclusions?
3. Use the Fisher LSD method with $\alpha = 0.05$ to make comparisons between pairs of means.
4. Construct a normal probability plot of the residuals. What conclusion would you draw about the validity of the normality assumption?
5. Plot the residuals versus the predicted tensile strength. Comment on the plot.
6. Prepare a scatter plot of the results to aid the interpretation of the results of this experiment.

Exercises 3.08 and 3.09

Reconsider the experiment in Problem 3.07.

1. Rework part (3) of Problem 3.07 using Tukey's test with $\alpha = 0.05$. Do you get the same conclusions from Tukey's test that you did from the graphical procedure and/or the Fisher LSD method?
2. Explain the difference between the Tukey and Fisher procedures.
3. Find a 95 percent confidence interval on the mean tensile strength of the Portland cement produced by each of the four mixing techniques. Also find a 95 percent confidence interval on the difference in means for techniques 1 and 3. Does this aid you in interpreting the results of the experiment?

Exercises 3.10

A product developer is investigating the tensile strength of a new synthetic fiber that will be used to make cloth for men's shirts. Strength is usually affected by the percentage of cotton used in the blend of materials for the fiber. The engineer conducts a completely randomized experiment with five levels of cotton content and replicates the experiment five times.

Cotton Weight Percent			Observations		
15	7	7	15	11	9
20	12	17	12	18	18
25	14	19	19	18	18
30	19	25	22	19	23
35	7	10	11	15	11

1. Is there evidence to support the claim that cotton content affects the mean tensile strength? Use $\alpha = 0.05$.
2. Use the Fisher LSD method to make comparisons between the pairs of means. What conclusions can you draw?
3. Analyze the residuals from this experiment and comment on model adequacy.