# Statistics II
# Factorial ANOVA

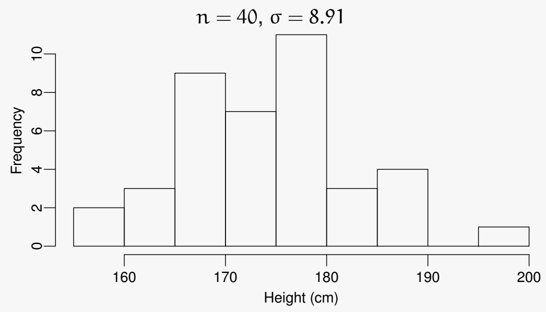
Çağrı Çöltekin

University of Groningen
Information Science

April 22, 2014



## From the survey: the histogram of the height








## Class activity

- Your height measurements are in the bag.
- Pick four height measurements randomly
- Write them down.
- Pass it to your neighbor.
- Calculate the 95% confidence interval for the mean of the numbers that you sampled.
- During the break, draw a line on the board representing the confidence interval you have calculated.





## Factorial ANOVA

- Factorial ANOVA is used when there are more than one categorical variables (multiple factors, or grouping dimensions).
  - treatment and type of illness
  - instruction method and gender
  - education and socio-economic status.
- Factorial (n-way) ANOVA follows essentially the same logic as single (one-way) ANOVA.





## Example problems for Factorial ANOVA

- Compare time needed for lexical recognition in
  1. healthy adults
  2. patients with Wernicke's aphasia
  3. patients with Broca's aphasia

  and gender of the subject.
- Usability of an application based on different user interfaces and input methods.
- Language development of children based on their parent's education and socio-economic status.
- Compare Dutch proficiency scores of second language learners based on their native language and profession.





## Why not multiple one-way ANOVAs?

- Efficiency: answer more questions with smaller sample size. For example, we want to choose between two web designs, and two background colors
  - Two one-way ANOVAs:

| design 1 | design 2 |
|---|---|
| 30 | 30 |

| dark bg | light bg |
|---|---|
| 30 | 30 |

Total participants needed: 120

  - One two-way ANOVA:

| | design 1 | design 2 |
|---|---|---|
| dark background | 15 | 15 |
| light background | 15 | 15 |

Total participants needed: 60

- Interactions: effects of different factors are not always additive.





## Interactions

Interactions occur when change in one of the variables depends on the change in another.

- A particular treatment may have different effects on different illnesses.
- Living in big cities may increase life expectancy for people with low socio-economic status (SES), but may have no or reverse effect for people with higher SES.
- A new teaching method may be more effective with respect to the old one for girls but less effective for boys.

When there is an interaction, interpretation of main effects alone is incomplete and can be misleading.





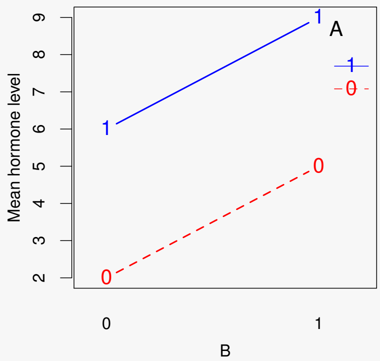
## An example for interaction

Two drugs, A and B, are tested with a factorial design. Each drug is administered in doses 0 and 1.
In other words, four groups receive none, A, B and A and B together respectively.

| | | drug A | |
|---|---|---|---|
| | | 0 | 1 |
| drug B | 0 | control | A only |
| | 1 | B only | A and B |

Response measure: blood level of some hormone.

## Types of interaction (1)

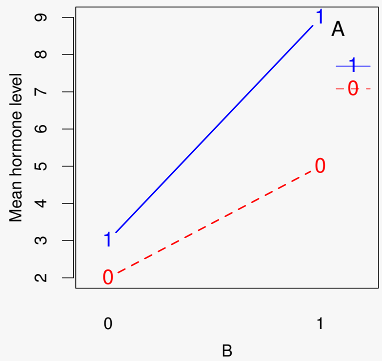



- both drugs have positive effects
- combined effect is additive
- no interaction





## Types of interaction (2)

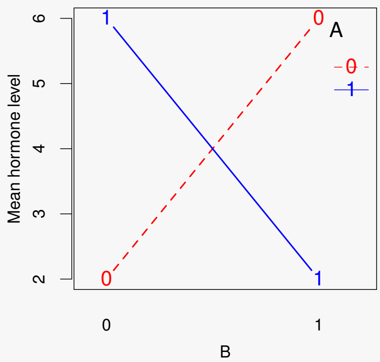



- both drugs have positive effects
- combined effect is stronger than sum of separate effects
- interaction





## Types of interaction (3)

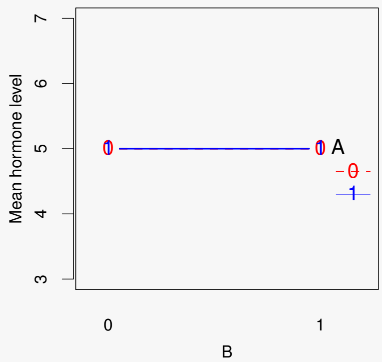



- both drugs have positive effects separately
- combination cancel out each other's effect
- interaction





## Types of interaction (4)




- drugs show no effect
- either separately or in combination
- null hypothesis is true
- no interaction





## ANOVA: partitioning the variance

In single ANOVA, we partition the total variance ($SS_T$) as variance due to group means (or, due to the groups, or the model $SS_M$) and the variance around the group means (or, residual variance, $SS_R$).

$$SS_T = SS_M + SS_R$$

The F-test used in single ANOVA is based on,

$$F = \frac{MS_M}{MS_R}$$

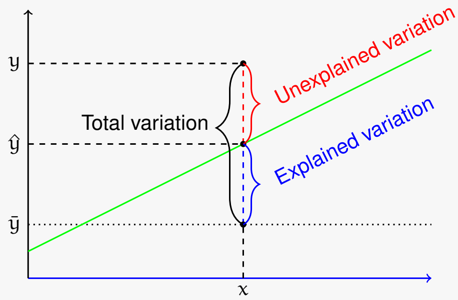
Associated degrees of freedom, for $n$ observations, and $k$ groups, are:

$$\begin{array}{ccccc} DF_T & = & DF_M & + & DF_R \\ n-1 & = & k-1 & + & n-k \end{array}$$





## Reminder: partitioning the variance in regression

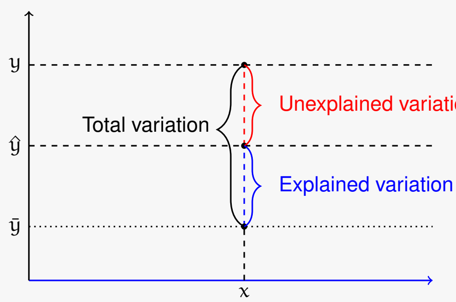



$$\begin{array}{ccccc} \text{Total variation} & = & \text{Unexplained variation} & + & \text{Explained variation} \\ y - \bar{y} & = & y - \hat{y} & + & \hat{y} - \bar{y} \end{array}$$





## Reminder: partitioning the variance in ANOVA

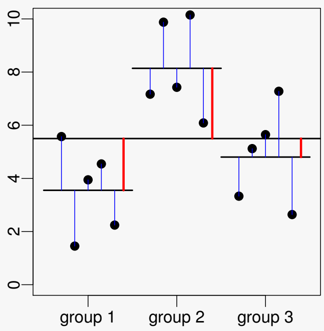



$$\begin{array}{ccccc} \text{Total variation} & = & \text{Unexplained variation} & + & \text{Explained variation} \\ y - \bar{y} & = & y - \hat{y} & + & \hat{y} - \bar{y} \end{array}$$





## Reminder: partitioning the variance in single ANOVA




$$F = \frac{\frac{SS_M}{n-k}}{\frac{SS_R}{k-1}}$$

where $k$ is the number of groups, and $n$ is the number of observations.

Degrees of freedom values in the formula are wrong. It should be: $\frac{\frac{SS_M}{k-1}}{\frac{SS_R}{n-k}}$.

## Factorial ANOVA: partitioning the variance

Factorial ANOVA partitions the $SS_M$ further.

- For two-way ANOVA, with factors $A$ and $B$, $SS_M$ is partitioned as:

$$SS_M = \underbrace{SS_A + SS_B}_{\text{main effects}} + \underbrace{SS_{A\times B}}_{\text{interaction}}$$

- For three-way ANOVA, with factors $A$, $B$ and $C$, $SS_M$ is partitioned as:

$$SS_M = \underbrace{SS_A + SS_B + SS_C}_{\text{main effects}} + \underbrace{SS_{A\times B} + SS_{A\times C} + SS_{B\times C}}_{\text{2-way intections}} + \underbrace{SS_{A\times B\times C}}_{\text{3-way inter.}}$$





## Factorial ANOVA: degrees of freedom

As in single ANOVA:

$$\begin{array}{rcccc} DF_T & = & DF_M & + & DF_R \\ n-1 & = & k-1 & + & n-k \end{array}$$

If we have $k_A$ levels due to factor $A$, and $k_B$ levels due to factor $B$, total number of groups is $k = k_A \times k_B$. We can now further partition the $DF_M$ as,

$$\begin{array}{rcccccc} DF_M & = & DF_A & + & DF_B & + & DF_{A\times B} \\ k-1 & = & k_A - 1 & + & k_B - 1 & + & (k_A - 1)\times(k_B - 1) \end{array}$$





## Factorial ANOVA: F-statistics

Once we have calculated sums of squares, and degrees of freedom values, we can calculate the estimated variance (mean squares) for each component as $MS = \frac{SS}{DF}$.
For two-way ANOVA we will get three F-tests:

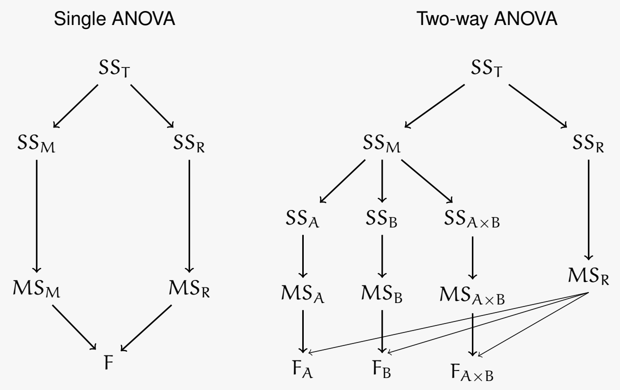
$$\begin{array}{rcl} F_A & = & \frac{MS_A}{MS_R} \\ F_B & = & \frac{MS_B}{MS_R} \\ F_{A\times B} & = & \frac{MS_{A\times B}}{MS_R} \end{array}$$

For three-way ANOVA there will be 7 F-tests (three main effects, three two-way interactions and one three-way interaction).





## Factorial ANOVA: the picture








## Factorial ANOVA: an example

We return to our 'web design' example.

- We have two new web page designs.
- We also want know the effect of dark or light background.
- This is a two-way ANOVA with two levels at each dimension: commonly called 2×2 (experiment) design.
- If we also wanted to know the effect of age (young, middle aged, old), we would do a three-way, 2×2×3, ANOVA.





## Example: participants

We gather a random sample of 60 people from our target audience, and randomly assign equal number of participants to one of the following groups (15 in each):

| | | Design | |
|---|---|---|---|
| | | 1 | 2 |
| BG color | light | design 1, light BG | design 2, light BG |
| | dark | design 1, dark BG | design 2, dark BG |

The response is the average opinion of each participant assessed through a 7-point questionnaire with multiple questions.





## Example: data

We have a numeric response variable (opinion) and two categorical variables (design and background color), both with two levels.

| participant | opinion | design | background |
|---|---|---|---|
| 1 | 6.2 | 1 | light |
| 2 | 5.8 | 1 | dark |
| ⋮ | ⋮ | ⋮ | ⋮ |
| 59 | 4.8 | 2 | light |
| 60 | 6.4 | 2 | dark |

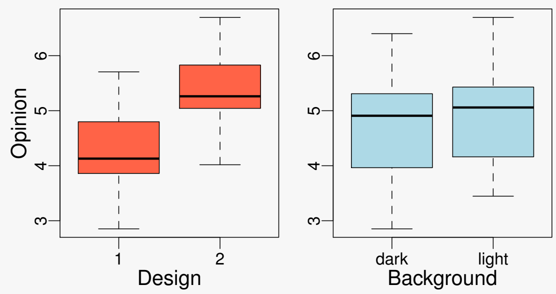
Important:

- participants are randomly selected and randomly assigned to a combination of design and background color
- each participant provides a single observation

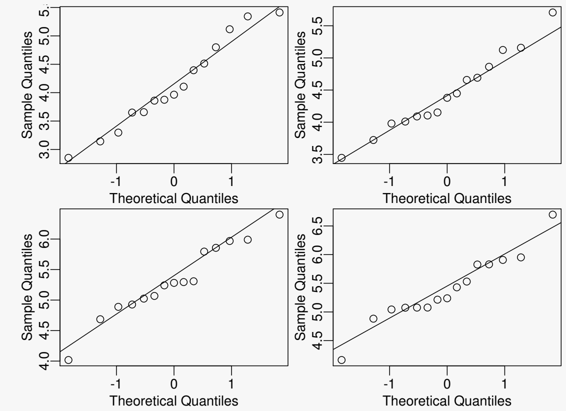




## Example: visualizing the data



## Example: checking for normality








## Example: checking homogeneity of variance

- Box plots already indicate that the variances are similar.
- Two possible tests for homogeneity of variance (output is from R, In SPSS enable the option in ANOVA dialog):

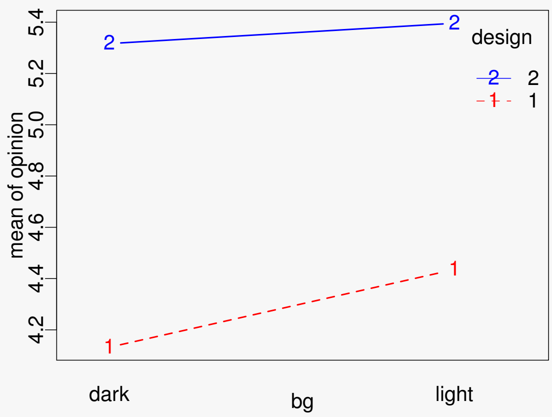
```
> leveneTest(opinion~design*bg)
Levene's Test for Homogeneity of Variance (center =
    median)
      Df F value Pr(>F)
group  3  0.6342 0.5961
      56
> bartlett.test(opinion~design*bg)
        Bartlett test of homogeneity of variances
data:  opinion by design by bg
Bartlett's K-squared = 0.9009, df = 1, p-value =
    0.3426
```

- Both tests support the visual inspection. No significant evidence of non-homogeneity.





## Example: visualizing the interaction








## Example: two-way ANOVA

```
Analysis of Variance Table
Response: opinion
          Df  Sum Sq Mean Sq F value    Pr(>F)
design     1 17.2513 17.2513 40.5053 3.859e-08 ***
bg         1  0.5493  0.5493  1.2898    0.2609
design:bg  1  0.1880  0.1880  0.4414    0.5092
Residuals 56 23.8506  0.4259
```

- ‘design’ has a significant effect.
- the background color does not have significant effect.
- there is no evidence for interaction.





## Example: post-hoc comparisons after factorial ANOVA

Multiple comparisons with ‘Tukey’s Honest Significant Differences’:

```
  Tukey multiple comparisons of means
    95% family-wise confidence level
Design
        diff       lwr    upr    p adj
2-1        1.072  0.735 1.410       0
Background color
              diff    lwr    upr     p adj
light-dark 0.191 -0.146 0.529    0.26092
Design x Background
                  diff    lwr     upr   p adj
2:dark-1:dark     1.184  0.553  1.815 0.00004
1:light-1:dark    0.303 -0.328  0.934 0.58382
2:light-1:dark    1.264  0.633  1.895 0.00001
1:light-2:dark   -0.881 -1.512 -0.250 0.00273
2:light-2:dark    0.079 -0.552  0.710 0.98710
2:light-1:light   0.961  0.330  1.591 0.00095
```





## (factorial) ANOVA and effect size

Simplest form of effect size for ANOVA is called $\eta^2$ (eta-squared). $\eta^2$ is equivalent to $r^2$ for regression.

$$\eta^2 = \frac{SS_M}{SS_T}$$

For factorial ANOVA, we can calculate partial-$\eta^2$ for each factor.

$$\eta^2_A = \frac{SS_A}{SS_A + SS_R}$$

Like $r^2$, $\eta^2$ increases as number of levels/factors increase. An adjusted effect size measure, called $\omega^2$ (omega-squared), corrects for chance increase caused by additional factor levels.
Statistical software (typically) will give you both numbers.





## (factorial) ANOVA and effect size (2)

As in t-test *Cohen's d* can be specified as the effect size for pairwise comparisons.

In general, for standardized effect size measures, the rule of thumb for interpretation is,

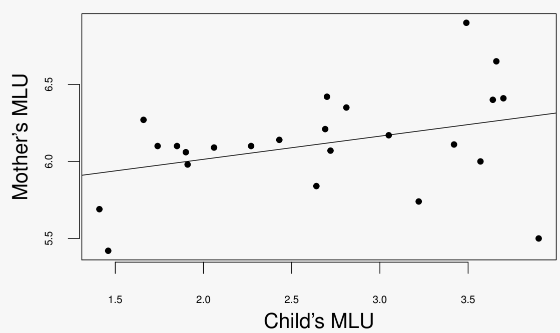
| | |
|---|---|
| less than 0.1 | weak effect |
| between 0.1 and 0.6 | medium-size effect |
| greater than 0.6 | large effect |

Effect sizes are best interpreted with considering the particular problem at hand. For example, obtaining small effect sizes may be important in some problems.





## Lab 1: scatter plot



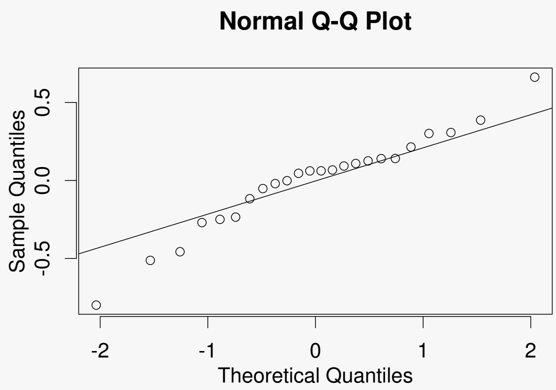
## Lab 1: the regression fit

```
lm(formula = mot.mlu ~ chi.mlu, data = d)
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 5.7133    0.2326 24.559  <2e-16 ***
chi.mlu     0.1503    0.0839  1.791  0.0871 .
---
Residual standard error: 0.3182 on 22 degrees of freedom
Multiple R-squared: 0.1272, Adjusted R-squared: 0.08757
F-statistic: 3.207 on 1 and 22 DF, p-value: 0.08708
```





## Lab 1: Q-Q plot of residuals








## Factorial ANOVA: summary

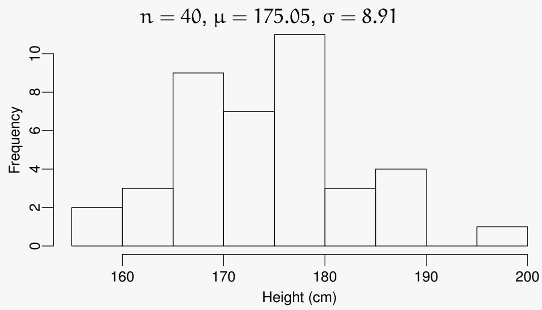
- Factorial ANOVA is a generalization of single ANOVA (or t-test).
- Compare groups along more than one dimension.
- Assumptions: the response variable in all groups
  - is (approximately) normally distributed
  - have (approximately) equal variances
- Efficient in use of subjects.
- Allows to investigate interaction.

Next week: Repeated-measures ANOVA. Reading: 19.1–19.5 (3e), 20.1–20.5 (4e).





## From the survey: the histogram of the height (including the mean)








## From the survey: confidence intervals of the mean of height