Statistics 2

Analysis of Variance (ANOVA)

Casper Albers & Jorge Tendeiro

Lecture 11, 2019 – 2020

university of groningen

Overview

Two-way ANOVA

- Types of effect
- Visualizing
- Hypotheses tested
- Partitioning the variance
- Effect size in ANOVA models
- Three, four, more way ANOVA

Literature for this lecture

Read:

Agresti, Section 12.4

One-way ANOVA

- **Principle:**
  Study differences in the means of g independent groups.

$$y_i \sim \mathcal{N}(\mu_i, \sigma)$$

- **Test:**

$$\mathcal{H}_0 : \mu_1 = \mu_2 = \cdots = \mu_g$$

  versus

$$\mathcal{H}_a : \text{ Not all } \mu\text{'s are equal.}$$

- **Procedure:**
  Compare the between and the within group variances using the F-test

Source	SS	df	MS	F
Group	$\sum_{ij}(\overline{y}_i - \overline{y})^2$	$g-1$	GSS/df_1	GMS/RSS
Residual	$\sum_{ij}(y_{ij} - \overline{y}_i)^2$	$n-g$	RSS/df_2	
Total	$\sum_{ij}(y_{ij} - \overline{y})^2$	$n-1$		

$F \sim F(g-1, n-g)$

Residual MS also denoted as 'Mean Square Error' (MSE).

- F large $\longrightarrow$ Reject $\mathcal{H}_0$ $\longrightarrow$ There are differences between groups.
- But where are these differences? $\longrightarrow$ Use contrasts (Lecture 9)

Two way ANOVA

Two way ANOVA $\longrightarrow$ Two factors (= categorical IVs)

- Group membership is defined by 2 factors, say A and B.
- A has I categories, B has J categories.
  - ✓ In total there are $I \times J$ groups.
- In each (i, j) group:

$$y_{ij} \sim N(\mu_{ij}, \sigma)$$

normality assumption

homoscedasticity assumption

- Test effects:
  - ✓ Main effects: Factor A, Factor B.
  - ✓ Interaction effect: Between factors A and B.

Two way ANOVA: Example

- Small-scale ($n = 77$ undergraduate students) follow-up study[1] on the 'anger expression' theory by Bushman et al. (JPSP; 2001).
- DV: Anger Expression
  (to what extend does one express it when one has an angry mood).
- IVs:
  - ✓ GENDER (male vs. female).
  - ✓ ATHLETE (student is vs. is not an athlete).

	Anger Expression	
	Female	Male
Valid	48	30
Missing	0	0
Mean	36.90	37.17
Std. Deviation	13.11	12.89
Minimum	7.000	14.00
Maximum	68.00	67.00

Main effect A:

Differences between A's marginal means

	Anger Expression	
	Athlete	Non-athlete
Valid	25	53
Missing	0	0
Mean	30.96	39.85
Std. Deviation	10.54	13.07
Minimum	14.00	7.000
Maximum	50.00	68.00

Main effect B:

Differences between B's marginal means

[1] http://onlinestatbook.com/2/case_studies/angry_moods.html

Two way ANOVA: Example

- DV: Anger Expression
  (to what extend does one express it when one has an angry mood).
- IVs:
  - ✓ GENDER (1 = Male, 2 = Female).
  - ✓ ATHLETE (is the student an athlete?; 1 = Yes; 2 = No).

Gender	Sports	Mean	SD	N
Female	Athlete	30.21	10.75	14
	Non-athlete	39.65	13.13	34
Male	Athlete	31.91	10.70	11
	Non-athlete	40.21	13.32	19

Interaction effect A $\times$ B:

Differences between the 2×2 group means.

Use a cross table to display the means.

		SPORTS		
		Athl.	Non	
GENDER	Male	31.91 ($n = 11$)	40.21 ($n = 19$)	37.17 ($n = 30$)
	Female	30.21 ($n = 14$)	39.65 ($n = 34$)	36.90 ($n = 48$)
		30.96 ($n = 25$)	39.85 ($n = 53$)	37.00 ($n = 78$)

- Main effect SPORTS:

  Are these two marginal means significantly different from each other?

  These are means weighted by sample size:

  - ✓ Athletes: $30.96 = (11 \times 31.91 + 14 \times 30.21)/25$.
  - ✓ Non-athletes: $39.85 = (19 \times 40.21 + 34 \times 39.65)/53$.

Two way ANOVA: Main effect II

Use a cross table to display the means.

		SPORTS		
		Athl.	Non	
GENDER	Male	31.91 ($n = 11$)	40.21 ($n = 19$)	37.17 ($n = 30$)
	Female	30.21 ($n = 14$)	39.65 ($n = 34$)	36.90 ($n = 48$)
		30.96 ($n = 25$)	39.85 ($n = 53$)	37.00 ($n = 78$)

- Main effect GENDER:

  Are these two marginal means significantly different from each other?

  These are means weighted by sample size:

  ✓ Men: $37.17 = (11 \times 31.91 + 19 \times 40.21)/30$.

  ✓ Women: $36.90 = (14 \times 30.21 + 34 \times 39.65)/48$.

Use a cross table to display the means.

	Ath.		Non-A.
M	31.91	$\xrightarrow{+8.30}$	40.21
F	30.21	$\xrightarrow{+9.44}$	39.65

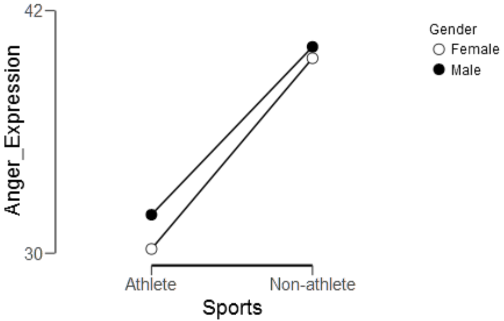
	Ath.	Non-A.
M	31.91	40.21
	$\downarrow -1.70$	$\downarrow -0.56$
F	30.21	39.65

- Interaction effect:
  Are there differences between differences in means?

  – The Athlete vs. Non-Athlete difference is not the same for men and women.

  – The gender difference is not the same for athletes and non-athletes.

Means plot: Visualize model effects.

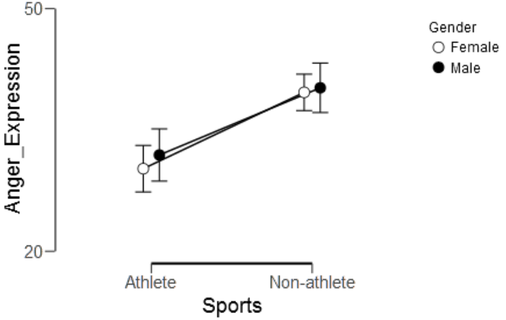


Interaction effect:

"Is the change of Anger_Expression scores across the levels of SPORTS the same for both GENDERs?"

Slope of ○-line and ●-line not parallel.

Means plot: Visualize model effects including error bars.



Slope of ○-line and ●-line not parallel.

Some differences in slope exist.

Two way ANOVA: Hypotheses testing

Two way ANOVA tests the significance of three effects.

1. Main effect A

   $\mathcal{H}_0$: There is no main effect for factor A.

   $\mathcal{H}_a$: There is a main effect for factor A.

2. Main effect B

   $\mathcal{H}_0$: There is no main effect for factor B.

   $\mathcal{H}_a$: There is a main effect for factor B.

3. Interaction effect

   $\mathcal{H}_0$: There is no interaction effect $A \times B$.

   $\mathcal{H}_a$: There is an interaction effect $A \times B$.

Two way ANOVA: Partitioning the variance

Recall One-way ANOVA (i for the factor, j for the person):

$$\underbrace{\sum_{ij}(y_{ij}-\overline{y})^2}_{\text{TSS}} = \underbrace{\sum_{ij}(\overline{y}_i-\overline{y})^2}_{\text{GSS}} + \underbrace{\sum_{ij}(y_{ij}-\overline{y}_i)^2}_{\text{RSS}}$$

Two way ANOVA is analogous:

$$\boxed{\text{TSS} = \underbrace{\text{A SS} + \text{B SS} + \text{A}\times\text{B SS}}_{\text{explained}} + \underbrace{\text{RSS}}_{\text{error}}}$$

Two way ANOVA: Partitioning the variance

Factor A with I levels; Factor B with J levels.

Source	SS	df	MS	F
Factor A	A SS	$I-1$	A SS/df$_A$	A MS/RMS
Factor B	B SS	$J-1$	B SS/df$_B$	B MS/RMS
Factor $A \times B$	A$\times$ B SS	$(I-1)(J-1)$	A $\times$ B SS/df$_{AB}$	A$\times$ B MS/RMS
Residual	RSS	$n-IJ$	RMS=RSS/df$_R$	
Total	TSS	$n-1$	Var(y)	

Two way ANOVA: Partitioning the variance

Q: How do the SS's of one- and two way ANOVA relate?

- One way ANOVA:
- Two way ANOVA:

$$
\begin{tikzcd}
\mathrm{TSS} \arrow[d, leftrightarrow, "?"] & = & \mathrm{A\ SS} \arrow[d, leftrightarrow, "?"] & + & \mathrm{RSS} \arrow[drrrr, leftrightarrow, "?"] & & & & \\
\mathrm{TSS} & = & \mathrm{A\ SS} & + & \mathrm{SSB} & + & \mathrm{SSAB} & + & \mathrm{RSS}
\end{tikzcd}
$$

A: Very tricky question!

- The SST's coincide, of course.
- If all n_{ij} are equal: $\mathrm{A\ SS}_{\text{one way}} = \mathrm{A\ SS}_{\text{two way}}$.
- Otherwise: A SS and RSS depend on the so-called types of SS.
  - ✓ In Stats II we will typically use Type III SS.
  - ✓ More on this topic: Statistics III.
- In general: The RSS is reduced when factor B is added.

Two way ANOVA: $\mathcal{H}_0$ rejected, now what?

Rejecting $\mathcal{H}_0$ implies:

- Significant difference between groups.
- At least one groups is different from the others.

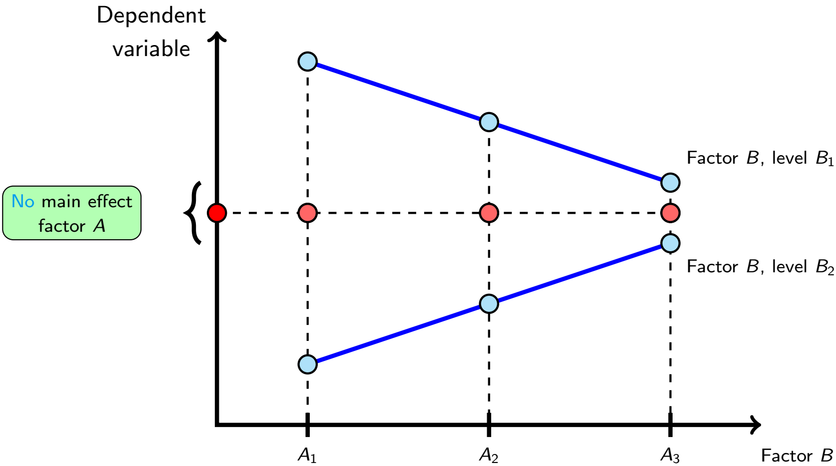
Q: But which group(s) is(are) different?
Which factor(s) is (are) responsible for these differences?

A: Further investigation is required:

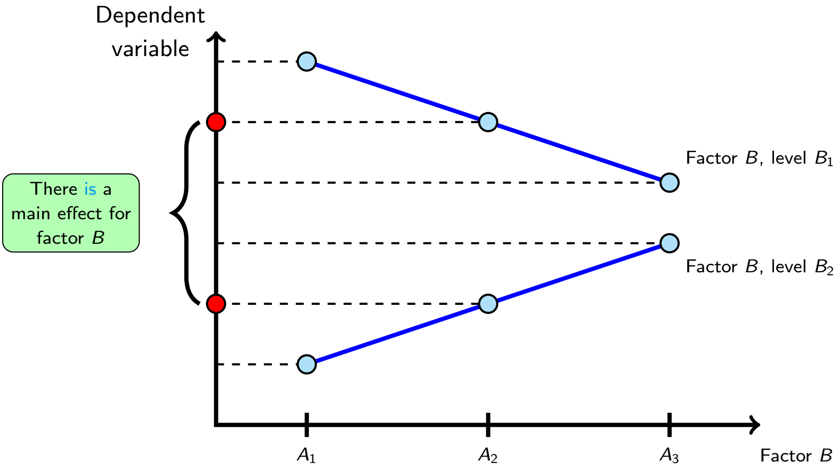
- Visually — use plots.
- Perform statistical inference.
  - Planned comparisons: Contrasts (more details in Stats III).
  - Post hoc comparisons: Multiple comparisons (only for main effects).

Two way ANOVA: Visual inspection (means plots)

Interpret main effect factor A

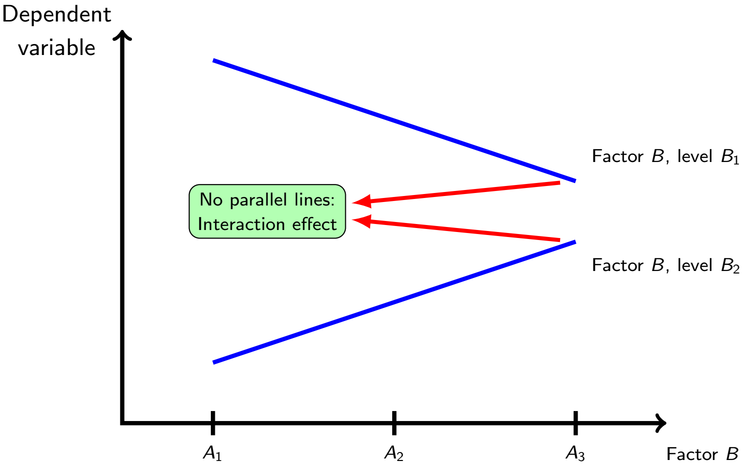


Interpret main effect factor B

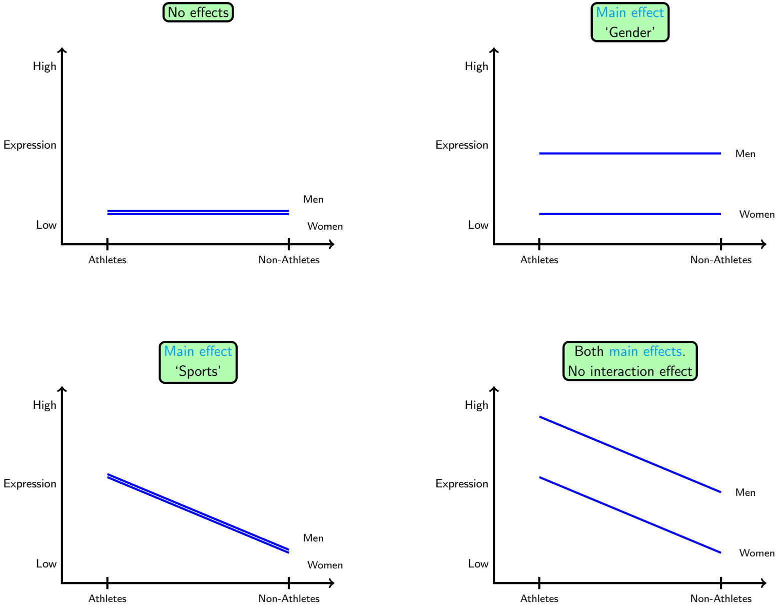


Two way ANOVA: Visual inspection (means plots)

Interpret interaction effect

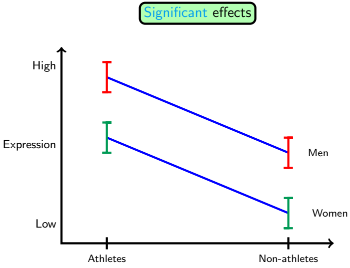


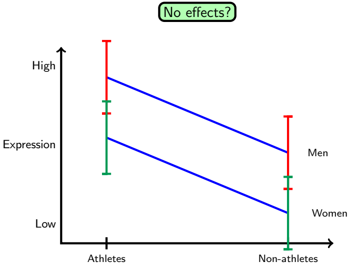
Two way ANOVA: Visual inspection (means plots)



Two way ANOVA: Visual inspection (means plots)

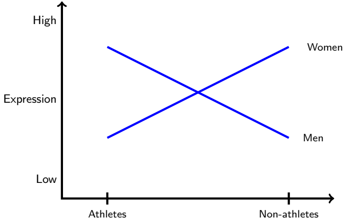
Remember: Take within-groups variance into account (sampling variation).



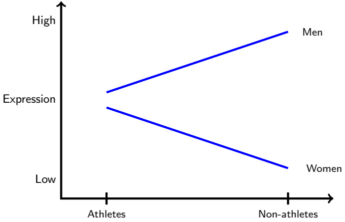


Two way ANOVA: Visual inspection (means plots)

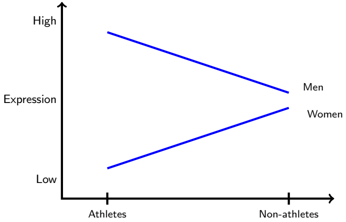
No main effects.
Interaction effect



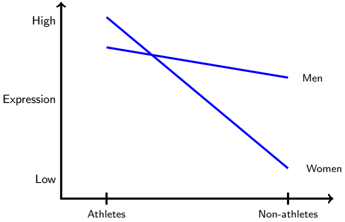
Main effect ‘Gender’.
Interaction effect



Main effect ‘Gender’.
Interaction effect



Main effects (both).
Interaction effect



ANOVA in general: Effect size

There are several effect size indices for ANOVA models.

- Eta squared (η^2)
  Proportion of the total sample variance explained by the effect.

$$\eta^2 = \frac{\text{SS}_{\text{effect}}}{\text{SST}}$$

✓ $\eta^2 = R^2$ in one way ANOVA.
✓ **Advantage:** Effects are additive for balanced designs (same n per group).

$$\sum_{\text{All effects}} SS_{\text{effect}} = \text{SSM}$$

✓ **Disadvantages:**
  - η^2 depends on the number and size of the remaining effects.
  - η^2 does not estimate the proportion of variance accounted for in the population: biased estimator

ANOVA in general: Effect size

There are several effect size indices for ANOVA models.

- Partial eta squared (η_p^2)
  Proportion of the (effect + error) sample variance explained by the effect.

$$\eta_p^2 = \frac{\text{Effect SS}}{\text{Effect SS} + \text{RSS}}$$

  - ✓ $\eta_p^2 = R^2$ in one way ANOVA.
  - ✓ **Advantage:** η_p^2 does not depend so much on the remaining effects.
  - ✓ **Disadvantage:**
    - Effects are not additive for balanced designs.
    - η_p^2 does not estimate the proportion of variance accounted for in the population.

ANOVA in general: Effect size

There are several effect size indices for ANOVA models.

- Omega squared (ω^2)
  Estimate of the proportion of the variance explained by the effect in the population (no repeated measures, balanced design).

$$\omega^2 = \frac{\text{Effect SS} - \text{Effect df} \times \text{RMS}}{\text{RMS} + \text{TSS}}$$

  - ✓ **Advantage:** It no longer overestimates the population effects, as η^2 and η_p^2 did.
  - ✓ **Disadvantages:** Effects are not additive. Estimate can be negative.

ANOVA in general: Effect size

In general, $\omega^2 < \eta^2 < \eta_p^2$.

How to interpret an effect size?

- As 'Percentage variance accounted for' (holds for η^2, η_p^2, and ω^2).
- Don't rely on rules of thumb! Context matters.

ANOVA in general: Effect size

Cases	SS	df	MS	F	p	η^2	η^2_p	ω^2
Gender	20.868	1	20.868	0.134	0.715	0.002	0.002	0.000
Sports	1286.934	1	1286.934	8.258	0.005	0.100	0.100	0.087
Gender × Sports	5.237	1	5.237	0.034	0.855	0.000	0.000	0.000
Residual	11532.189	74	155.840					
Total	12845.228	77						

E.g., for 'SPORTS':

- $\eta^2 = \frac{\text{Effect SS}}{\text{TSS}} = \frac{1286.934}{12845.228} = .100.$
- $\eta^2_p = \frac{\text{Effect SS}}{\text{Effect SS}+\text{RSS}} = \frac{1286.934}{1286.934+11532.189} = .100.$
- $\omega^2 = \frac{\text{Effect SS}-\text{Effect df}\times\text{RMS}}{\text{RMS}+\text{TSS}} = \frac{1286.934-1\times155.840}{155.840+11532.189} = .087.$

Report finding:

> *There is evidence for a difference in anger expression between athletes and non-athletes,* $F(1,74) = 8.258$, $p = .005$, $\omega^2 = .087$.

Three, four, more way ANOVA

- ANOVA also possible with more than two factors.
- The underlying principles are unchanged.
- The 'interaction' concept is extended.
  E.g., with three factors A, B, and C:
  - ✓ Second-order effects: $A \times B$, $A \times C$, $B \times C$
  - ✓ Third-order effects: $A \times B \times C$.
- Higher-order effects (say, above 3rd order) are hard to interpret. Avoid if possible.
- Analysis straightforward in JASP:
  Similar as two way, but with more factors.

For the next lecture

Contents:

- Bayesian Statistics

Reading material:

Kruschke, J. K. & Liddell, T. M. (2018). Bayesian data analysis for newcomers. *Psychonomic Bulletin & Review*, *25*, 155-177. doi:10.3758/s13423-017-1272-1