Factorial Analysis of Variance

"ANOVA is more important than ever because it represents a key idea in statistical modeling of complex data structures—the grouping of predictor variables and their coefficients into batches."
— Andrew Gelman

👨‍🏫 Statistical Reasoning Lecture #4
Alexander Savi, 2024
📚 Whitlock/Schluter, Ch. 15 & 18.3; Field, Ch. 13

Split Grid by Nicola Rennie (aRt package) 🎨

📢 Announcements

🤸 Warming Up

Split Grid by Nicola Rennie (aRt package) 🎨

🔁 Recap

Predictor / Independent

Response / Dependent		Categorical	Continuous
	Categorical	Chi-squared test …	Logistic regression …
	Continuous	Student’s *t*-test ANOVA … Pairwise *t*-test Repeated measures	Correlation Linear regression …

Analysis of Variance

For Starters

by Koen Derks (aRtsy package) 🎨

Height

> “ Men are taller on average than women, but that may not be a trait that evolved through selection. It might be a purely incidental result of estrogen’s effects on bone growth. — Quanta

Q. Can this effect be replicated in psychobiology students?

H. Male psychobiology students are taller than female psychobiology students.

E. In a sample of psychobiology students attending second-year lectures on statistical reasoning, I expect males to be taller than females.

Photo by fauxels

Decomposition of variability

Total variability

Explained variability
Model variability

Unexplained variability
Error variability

explained variance / unexplained variance

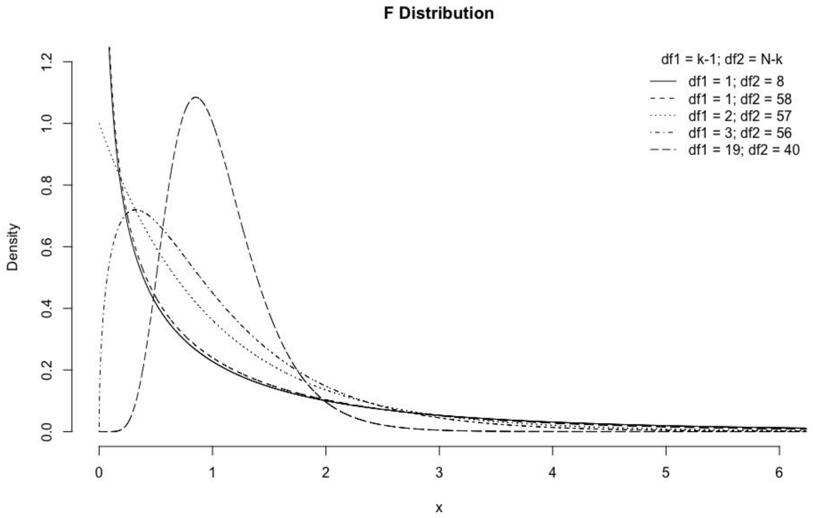
=

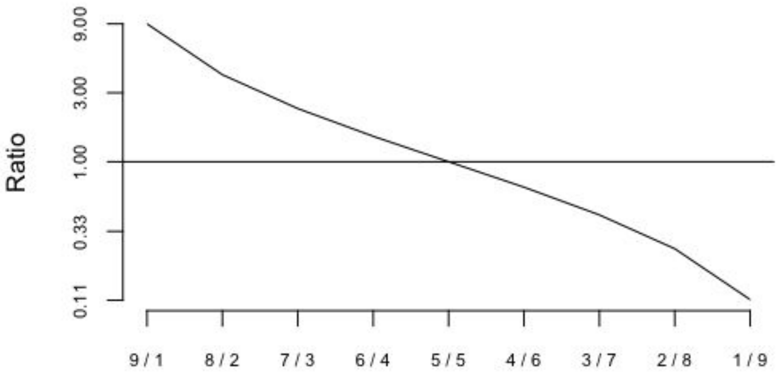
F

=

model / error

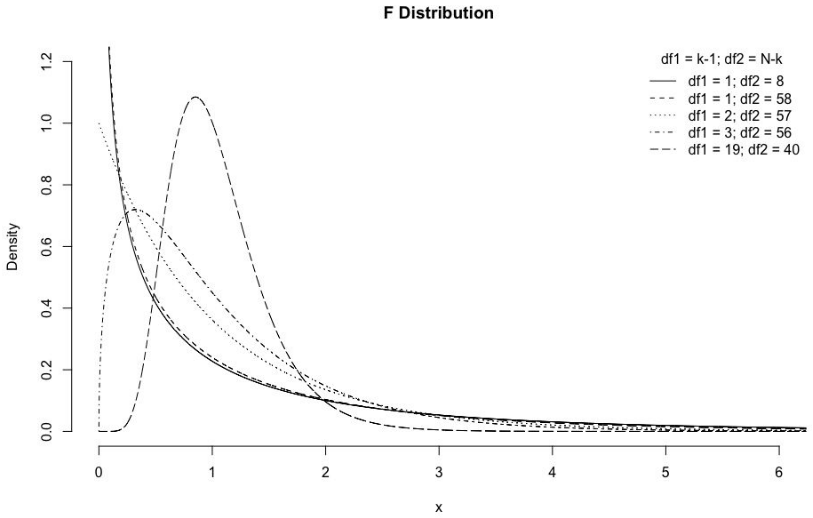
F distribution

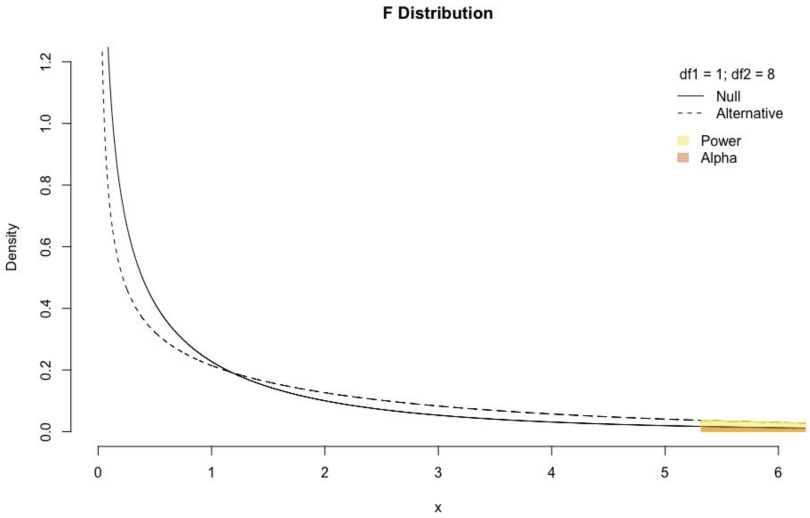




df_{model} = k - 1
df_{error} = N - k

F distribution





$df_{model} = k - 1$
$df_{error} = N - k$

👨‍👩‍👧‍👦 The ANOVA family

Number of independent variables

- 1 one-way
- >1 two-way, three-way, … (factorial)

Type of measurement

- independent (between subject)
- repeated measures (within subject)
- mixed (both)

Number of dependent variables

- 1 ANOVA
- >1 MANOVA

Compare …

- two means (*t*-test, one-way ANOVA)
- several means (one-way ANOVA)
- several means for several independent variables, measured between groups (independent factorial ANOVA)
- several means for several independent variables, measured within groups (repeated measures factorial ANOVA)
- several means for several independent variables, measured between and within groups (mixed-design ANOVA)

🔮 Overview

Topics

Probabilities & distributions
Frequentist inference
Multiple linear regression
| Independent Factorial ANOVA
| Repeated Measures Factorial ANOVA
| ~~Mixed Design ANOVA~~
Nonparametric inference
Bayesian inference

Learning goals

Analyze models with multiple (categorical) independent variables.

Analyze models with repeated measurements for multiple (categorical) independent variables.

Independent Factorial ANOVA

Split Grid by Nicola Rennie (aRt package) 🎨

Student evaluations

De student als consument maakt vrouwelijke docenten extra kwetsbaar

Nieuws | door Frans van Heest

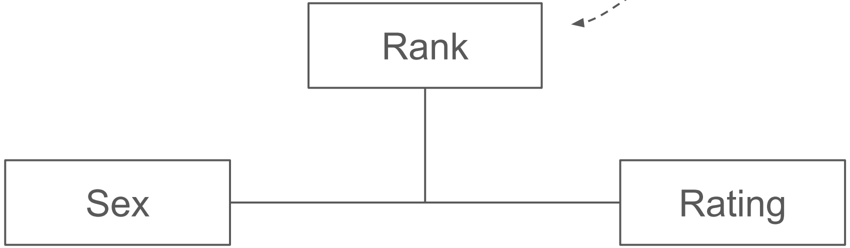
13 september 2023 | Vrouwelijke docenten worden aantoonbaar gediscrimineerd door studentenevaluaties, maar toch blijft het instrument voor veel universiteiten belangrijk om medewerkers te beoordelen. Cursusevaluaties moedigen echter middelmatig onderwijs aan en zijn extra nadelig voor vrouwen.

— ScienceGuide (Sep. 13, 2023)

Q. Is the effect of sex on rating modified by rank?

H. The effect of sex on rating is larger for lower ranked female teachers than for higher ranked female teachers.

E. In the open evals data set, I expect …

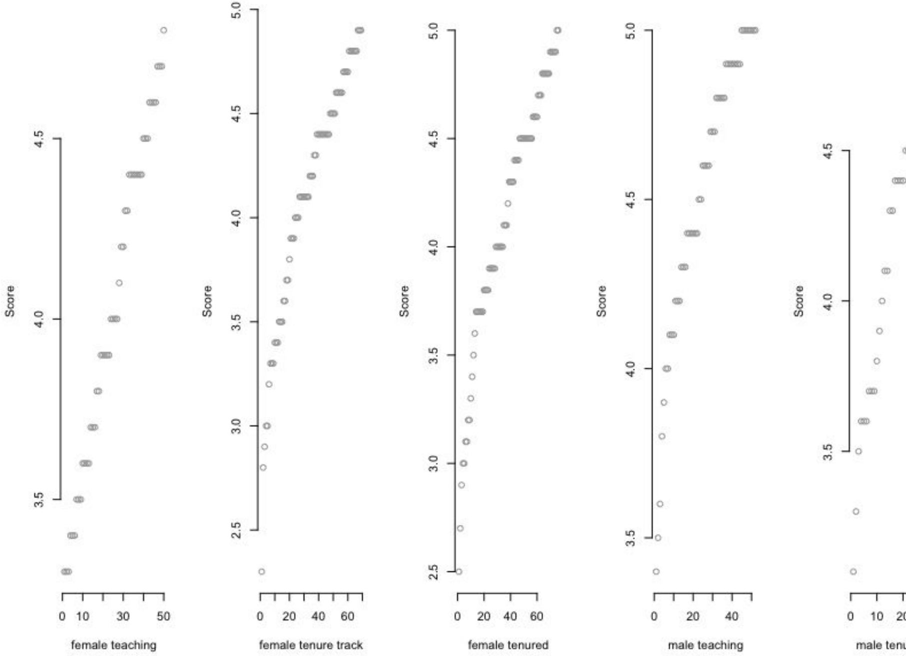


Student evaluations

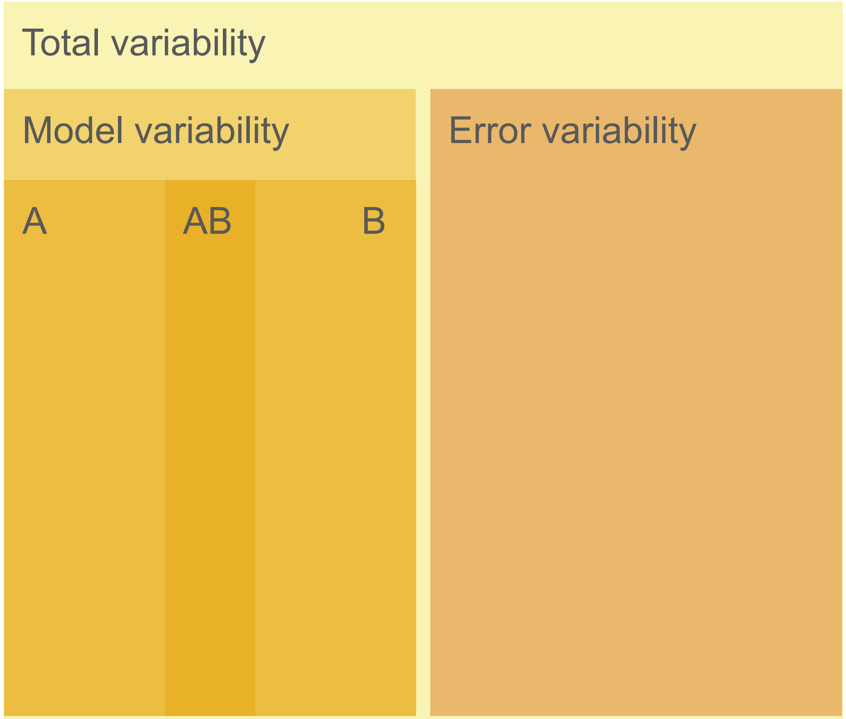
```
library("moderndive")
help(evals)
mod <- score ~ gender + rank + gender :
rank
with(evals, table(gender, rank))
```

	rank		
gender	teaching	tenure track	tenured
female	50	69	76
male	52	39	177

```
> str(evals[, c("score", "gender", "rank")])
tibble [463 x 3] (S3: tbl_df/tbl/data.frame)
 $ score : num [1:463] 3.3 3.3 3.3 3.4 3.4 3.4 3.5 3.5 3.5 3.6 ...
 $ gender: Factor w/ 2 levels "female","male": 1 1 1 1 1 1 1 1 1 1 ...
 $ rank  : Factor w/ 3 levels "teaching","tenure track",..: 1 1 1 1 1 1 1 1 1 1 ...
```



Decomposition of variability



Formulas

Variance	Sum of squares	df	Mean squares	F-ratio
Model	$SS_{\text{model}} = \sum n_k(\bar{X}_k - \bar{X})^2$	$k_{model} - 1$	$\frac{SS_{\text{model}}}{df_{\text{model}}}$	$\frac{MS_{\text{model}}}{MS_{\text{error}}}$
A	$SS_{\text{A}} = \sum n_k(\bar{X}_k - \bar{X})^2$	$k_A - 1$	$\frac{SS_{\text{A}}}{df_{\text{A}}}$	$\frac{MS_{\text{A}}}{MS_{\text{error}}}$
B	$SS_{\text{B}} = \sum n_k(\bar{X}_k - \bar{X})^2$	$k_B - 1$	$\frac{SS_{\text{B}}}{df_{\text{B}}}$	$\frac{MS_{\text{B}}}{MS_{\text{error}}}$
AB	$SS_{A \times B} = SS_{\text{model}} - SS_{\text{A}} - SS_{\text{B}}$	$df_A \times df_B$	$\frac{SS_{\text{AB}}}{df_{\text{AB}}}$	$\frac{MS_{\text{AB}}}{MS_{\text{error}}}$
Error	$SS_{\text{error}} = \sum s_k^2(n_k - 1)$	$N - k_{model}$	$\frac{SS_{\text{error}}}{df_{\text{error}}}$	
Total	$SS_{\text{total}} = SS_{\text{model}} + SS_{\text{error}}$	$N - 1$	$\frac{SS_{\text{total}}}{df_{\text{total}}}$	

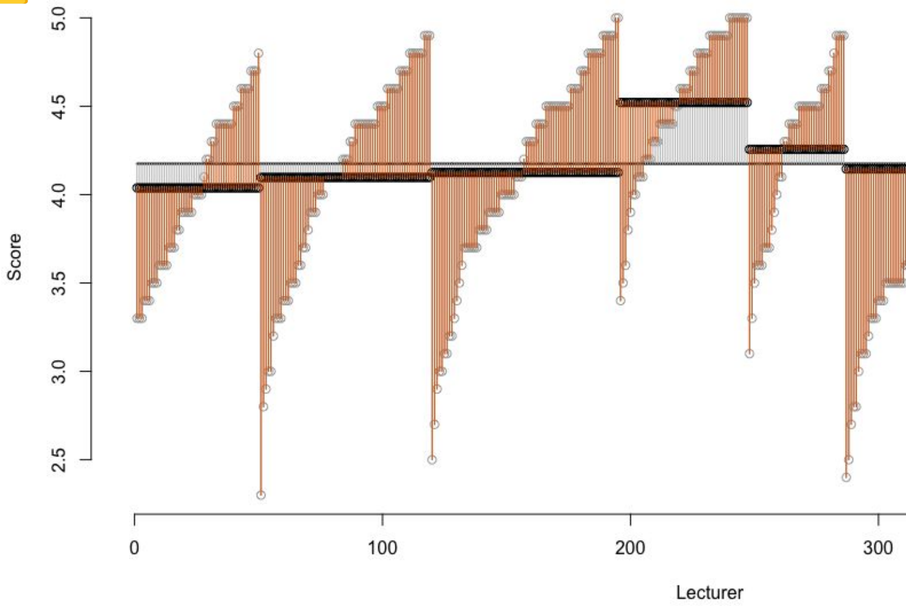
Unexplained variance

Formulas

Variance	Sum of squares	df	Mean squares	F-ratio
Model	$SS_{\text{model}} = \sum n_k(\bar{X}_k - \bar{X})^2$	$k_{model} - 1$	$\frac{SS_{\text{model}}}{df_{\text{model}}}$	$\frac{MS_{\text{model}}}{MS_{\text{error}}}$
A	$SS_{\text{A}} = \sum n_k(\bar{X}_k - \bar{X})^2$	$k_A - 1$	$\frac{SS_{\text{A}}}{df_{\text{A}}}$	$\frac{MS_{\text{A}}}{MS_{\text{error}}}$
B	$SS_{\text{B}} = \sum n_k(\bar{X}_k - \bar{X})^2$	$k_B - 1$	$\frac{SS_{\text{B}}}{df_{\text{B}}}$	$\frac{MS_{\text{B}}}{MS_{\text{error}}}$
AB	$SS_{A \times B} = SS_{\text{model}} - SS_{\text{A}} - SS_{\text{B}}$	$df_A \times df_B$	$\frac{SS_{\text{AB}}}{df_{\text{AB}}}$	$\frac{MS_{\text{AB}}}{MS_{\text{error}}}$
Error	$SS_{\text{error}} = \sum s_k^2(n_k - 1)$	$N - k_{model}$	$\frac{SS_{\text{error}}}{df_{\text{error}}}$	
Total	$SS_{\text{total}} = SS_{\text{model}} + SS_{\text{error}}$	$N - 1$	$\frac{SS_{\text{total}}}{df_{\text{total}}}$	

```r
var_score_k1 <- var(subset(x = evals, gender == "female" & rank == "teaching", select = "score", drop = TRUE))
var_score_k2 <- var(subset(x = evals, gender == "female" & rank == "tenure track", select = "score", drop = TRUE))
var_score_k3 <- var(subset(x = evals, gender == "female" & rank == "tenured", select = "score", drop = TRUE))
var_score_k4 <- var(subset(x = evals, gender == "male" & rank == "teaching", select = "score", drop = TRUE))
var_score_k5 <- var(subset(x = evals, gender == "male" & rank == "tenure track", select = "score", drop = TRUE))
var_score_k6 <- var(subset(x = evals, gender == "male" & rank == "tenured", select = "score", drop = TRUE))
n_k1 <- table(evals$gender, evals$rank)[["female", "teaching"]]
n_k2 <- table(evals$gender, evals$rank)[["female", "tenure track"]]
n_k3 <- table(evals$gender, evals$rank)[["female", "tenured"]]
n_k4 <- table(evals$gender, evals$rank)[["male", "teaching"]]
n_k5 <- table(evals$gender, evals$rank)[["male", "tenure track"]]
n_k6 <- table(evals$gender, evals$rank)[["male", "tenured"]]
ss_error_k1 <- var_score_k1 * (n_k1 - 1)
ss_error_k2 <- var_score_k2 * (n_k2 - 1)
ss_error_k3 <- var_score_k3 * (n_k3 - 1)
ss_error_k4 <- var_score_k4 * (n_k4 - 1)
ss_error_k5 <- var_score_k5 * (n_k5 - 1)
ss_error_k6 <- var_score_k6 * (n_k6 - 1)
ss_error <- sum(ss_error_k1, ss_error_k2, ss_error_k3, ss_error_k4, ss_error_k5, ss_error_k6)
n <- length(evals$score)
k_model <- 6
df_error <- n - k_model
ms_error <- ss_error / df_error
```

$MS_{error} = 0.2810958$



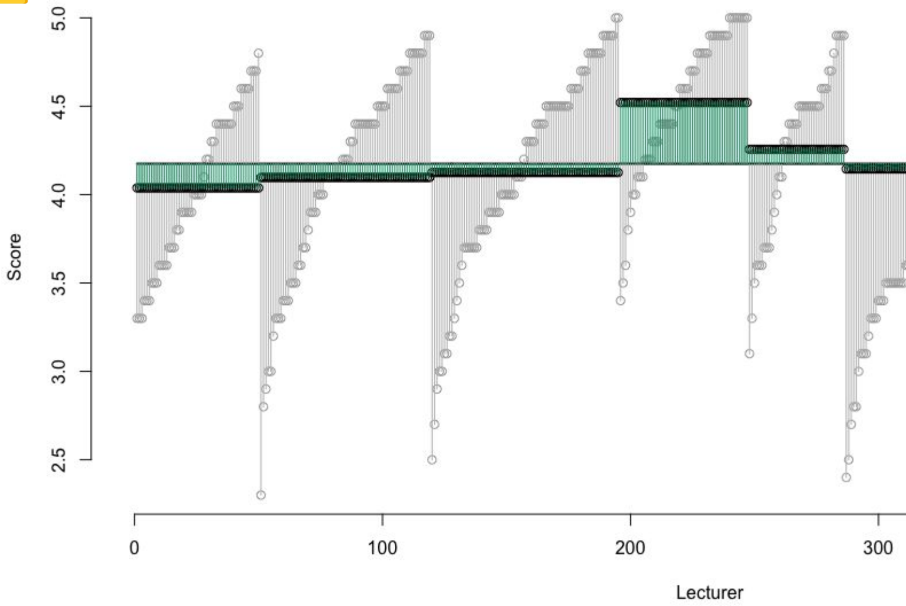
Explained variance (full model)

Formulas

Variance	Sum of squares	df	Mean squares	F-ratio
Model	$SS_{\text{model}} = \sum n_k(\bar{X}_k - \bar{X})^2$	$k_{model} - 1$	$\frac{SS_{\text{model}}}{df_{\text{model}}}$	$\frac{MS_{\text{model}}}{MS_{\text{error}}}$
A	$SS_{\text{A}} = \sum n_k(\bar{X}_k - \bar{X})^2$	$k_A - 1$	$\frac{SS_{\text{A}}}{df_{\text{A}}}$	$\frac{MS_{\text{A}}}{MS_{\text{error}}}$
B	$SS_{\text{B}} = \sum n_k(\bar{X}_k - \bar{X})^2$	$k_B - 1$	$\frac{SS_{\text{B}}}{df_{\text{B}}}$	$\frac{MS_{\text{B}}}{MS_{\text{error}}}$
AB	$SS_{A \times B} = SS_{\text{model}} - SS_{\text{A}} - SS_{\text{B}}$	$df_A \times df_B$	$\frac{SS_{\text{AB}}}{df_{\text{AB}}}$	$\frac{MS_{\text{AB}}}{MS_{\text{error}}}$
Error	$SS_{\text{error}} = \sum s_k^2(n_k - 1)$	$N - k_{model}$	$\frac{SS_{\text{error}}}{df_{\text{error}}}$	
Total	$SS_{\text{total}} = SS_{\text{model}} + SS_{\text{error}}$	$N - 1$	$\frac{SS_{\text{total}}}{df_{\text{total}}}$	

```r
mean_score <- mean(evals$score)
mean_score_k1 <- mean(subset(x = evals, gender == "female" & rank == "teaching", select = "score", drop = TRUE))
mean_score_k2 <- mean(subset(x = evals, gender == "female" & rank == "tenure track", select = "score", drop = TRUE))
mean_score_k3 <- mean(subset(x = evals, gender == "female" & rank == "tenured", select = "score", drop = TRUE))
mean_score_k4 <- mean(subset(x = evals, gender == "male" & rank == "teaching", select = "score", drop = TRUE))
mean_score_k5 <- mean(subset(x = evals, gender == "male" & rank == "tenure track", select = "score", drop = TRUE))
mean_score_k6 <- mean(subset(x = evals, gender == "male" & rank == "tenured", select = "score", drop = TRUE))
ss_model_k1 <- n_k1 * (mean_score_k1 - mean_score)^2
ss_model_k2 <- n_k2 * (mean_score_k2 - mean_score)^2
ss_model_k3 <- n_k3 * (mean_score_k3 - mean_score)^2
ss_model_k4 <- n_k4 * (mean_score_k4 - mean_score)^2
ss_model_k5 <- n_k5 * (mean_score_k5 - mean_score)^2
ss_model_k6 <- n_k6 * (mean_score_k6 - mean_score)^2
ss_model <- sum(ss_model_k1, ss_model_k2, ss_model_k3, ss_model_k4, ss_model_k5, ss_model_k6)
df_model <- k_model - 1
ms_model <- ss_model / df_model
ms_model / ms_error
```

$MS_{\text{model}} = 1.638715$



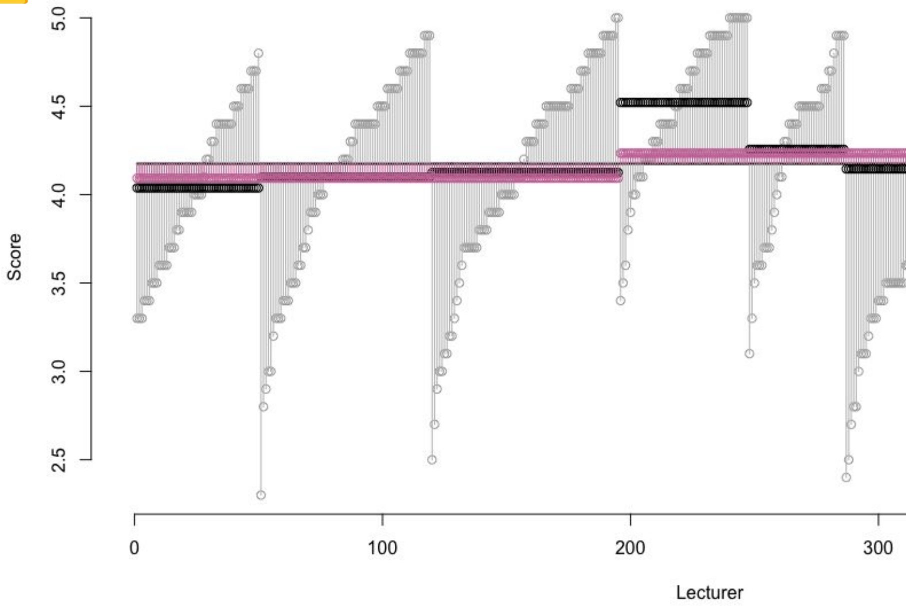
Explained variance (gender)

Formulas

Variance	Sum of squares	df	Mean squares	F-ratio
Model	$SS_{model} = \sum n_k(\bar{X}_k - \bar{X})^2$	$k_{model} - 1$	$\frac{SS_{model}}{df_{model}}$	$\frac{MS_{model}}{MS_{error}}$
A	$SS_A = \sum n_k(\bar{X}_k - \bar{X})^2$	$k_A - 1$	$\frac{SS_A}{df_A}$	$\frac{MS_A}{MS_{error}}$
B	$SS_B = \sum n_k(\bar{X}_k - \bar{X})^2$	$k_B - 1$	$\frac{SS_B}{df_B}$	$\frac{MS_B}{MS_{error}}$
AB	$SS_{A\times B} = SS_{model} - SS_A - SS_B$	$df_A \times df_B$	$\frac{SS_{AB}}{df_{AB}}$	$\frac{MS_{AB}}{MS_{error}}$
Error	$SS_{error} = \sum s_k^2(n_k - 1)$	$N - k_{model}$	$\frac{SS_{error}}{df_{error}}$	
Total	$SS_{total} = SS_{model} + SS_{error}$	$N - 1$	$\frac{SS_{total}}{df_{total}}$	

```r
mean_score_female <- mean(subset(x = evals, subset = gender == "female", select = "score", drop = TRUE))
mean_score_male <- mean(subset(x = evals, subset = gender == "male", select = "score", drop = TRUE))
n_female <- table(evals$gender)[["female"]]
n_male <- table(evals$gender)[["male"]]
ss_female <- n_female * (mean_score_female - mean_score)^2
ss_male <- n_male * (mean_score_male - mean_score)^2
ss_gender <- sum(ss_female, ss_male)
k_gender <- 2
df_gender <- k_gender - 1
ms_gender <- ss_gender / df_gender
ms_gender / ms_error
```

$MS_{gender} = 2.260213$



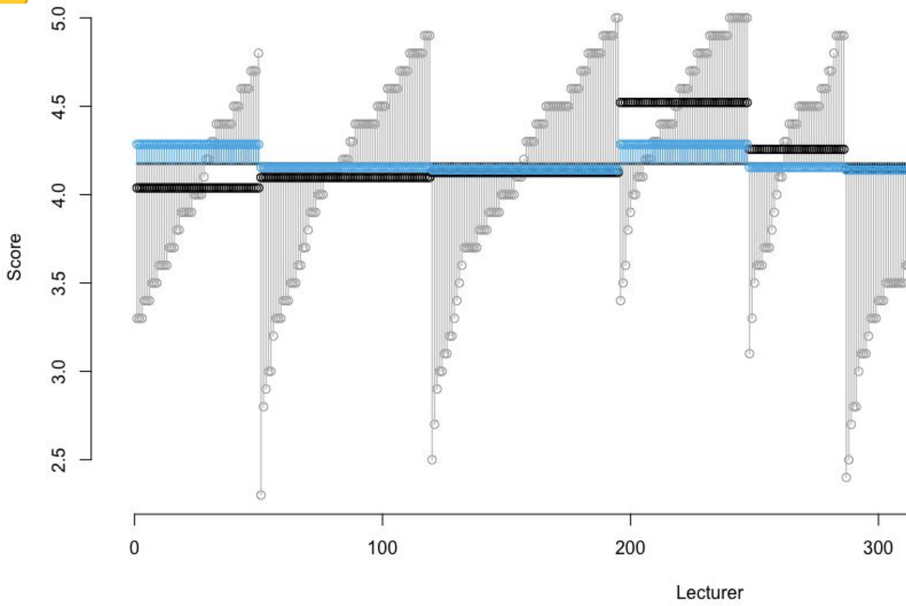
Explained variance (rank)

Formulas

Variance	Sum of squares	df	Mean squares	F-ratio
Model	$SS_{\text{model}} = \sum n_k(\bar{X}_k - \bar{X})^2$	$k_{model} - 1$	$\frac{SS_{\text{model}}}{df_{\text{model}}}$	$\frac{MS_{\text{model}}}{MS_{\text{error}}}$
A	$SS_{\text{A}} = \sum n_k(\bar{X}_k - \bar{X})^2$	$k_A - 1$	$\frac{SS_{\text{A}}}{df_{\text{A}}}$	$\frac{MS_{\text{A}}}{MS_{\text{error}}}$
B	$SS_{\text{B}} = \sum n_k(\bar{X}_k - \bar{X})^2$	$k_B - 1$	$\frac{SS_{\text{B}}}{df_{\text{B}}}$	$\frac{MS_{\text{B}}}{MS_{\text{error}}}$
AB	$SS_{A \times B} = SS_{\text{model}} - SS_{\text{A}} - SS_{\text{B}}$	$df_A \times df_B$	$\frac{SS_{\text{AB}}}{df_{\text{AB}}}$	$\frac{MS_{\text{AB}}}{MS_{\text{error}}}$
Error	$SS_{\text{error}} = \sum s_k^2(n_k - 1)$	$N - k_{model}$	$\frac{SS_{\text{error}}}{df_{\text{error}}}$	
Total	$SS_{\text{total}} = SS_{\text{model}} + SS_{\text{error}}$	$N - 1$	$\frac{SS_{\text{total}}}{df_{\text{total}}}$	

```r
mean_score_teaching <- mean(subset(x = evals, subset = rank == "teaching", select = "score", drop = TRUE))
mean_score_tenure_track <- mean(subset(x = evals, subset = rank == "tenure track", select = "score", drop = TRUE))
mean_score_tenured <- mean(subset(x = evals, subset = rank == "tenured", select = "score", drop = TRUE))
n_teaching <- table(evals$rank)[["teaching"]]
n_tenure_track <- table(evals$rank)[["tenure track"]]
n_tenured <- table(evals$rank)[["tenured"]]
ss_teaching <- n_teaching * (mean_score_teaching - mean_score)^2
ss_tenure_track <- n_tenure_track * (mean_score_tenure_track - mean_score)^2
ss_tenured <- n_tenured * (mean_score_tenured - mean_score)^2
ss_rank <- sum(ss_teaching, ss_tenure_track, ss_tenured)
k_rank <- 3
df_rank <- k_rank - 1
ms_rank <- ss_rank / df_rank
ms_rank / ms_error
```

$MS_{rank} = 0.7945727$

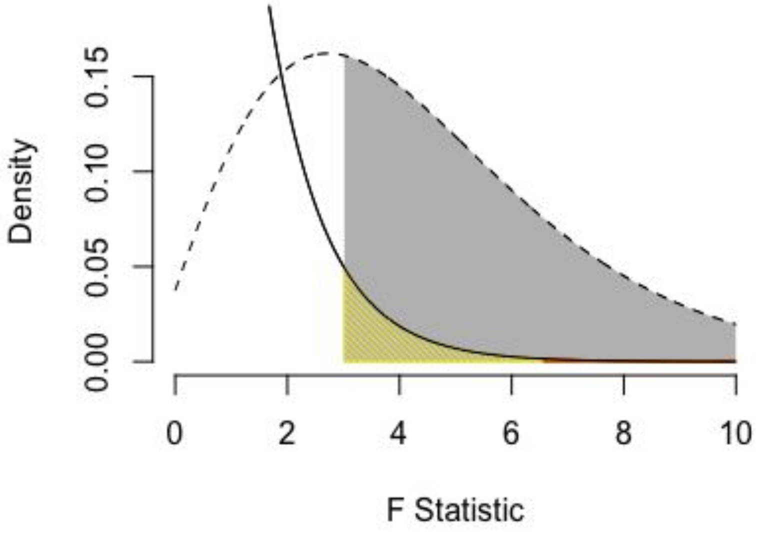


Explained variance (gender × rank)

Formulas

Variance	Sum of squares	df	Mean squares	F-ratio
Model	$SS_{\text{model}} = \sum n_k(\bar{X}_k - \bar{X})^2$	$k_{model} - 1$	$\frac{SS_{\text{model}}}{df_{\text{model}}}$	$\frac{MS_{\text{model}}}{MS_{\text{error}}}$
A	$SS_A = \sum n_k(\bar{X}_k - \bar{X})^2$	$k_A - 1$	$\frac{SS_A}{df_A}$	$\frac{MS_A}{MS_{\text{error}}}$
B	$SS_B = \sum n_k(\bar{X}_k - \bar{X})^2$	$k_B - 1$	$\frac{SS_B}{df_B}$	$\frac{MS_B}{MS_{\text{error}}}$
AB	$SS_{A \times B} = SS_{\text{model}} - SS_A - SS_B$	$df_A \times df_B$	$\frac{SS_{AB}}{df_{AB}}$	$\frac{MS_{AB}}{MS_{\text{error}}}$
Error	$SS_{\text{error}} = \sum s_k^2(n_k - 1)$	$N - k_{model}$	$\frac{SS_{\text{error}}}{df_{\text{error}}}$	
Total	$SS_{\text{total}} = SS_{\text{model}} + SS_{\text{error}}$	$N - 1$	$\frac{SS_{\text{total}}}{df_{\text{total}}}$	

⚠️ $F_{\text{gender×rank}}$ = 7.727288 (6.572717)



⚠️ (Un)balanced data

```
library("ez")
ez::ezANOVA(data = evals,
            dv = score,
            wid = ID,
            between = c(gender, rank),
            type = 2,
            return_aov = TRUE)
```

```
with(evals, table(gender, rank))  #
balanced?
aov(); anova()  # type I
car::Anova()  # type II/III (type III
requires contrasts)
ez::ezANOVA()  # type I/II/III (default
is II)
```

Type I, II and III

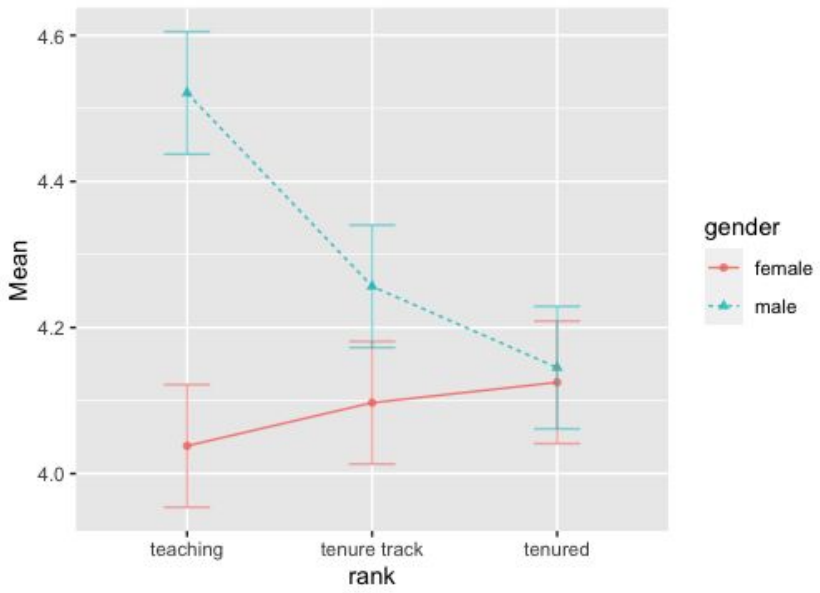
- I (sequential): SS(A), SS(B | A), SS(AB | A, B)
- II (hierarchical): SS(A | B), SS(B | A), SS(AB | A, B)
- III (unique): SS(A | B, AB), SS(B | A, AB), SS(AB | A, B)

Type I, II or III?

- Balanced? I/II/III
- Highest-order interaction of interest? I/II/III
- Unbalanced, no significant interaction? II
- Confused? Check robustness and consult a statistician.
- SPSS? III

👀

```
ez::ezPlot(data = evals,
           x = .(gender),
           split = .(rank),
           dv = .(score),
           wid = .(ID),
           between = .(gender, rank)
)
```



Multiple comparisons

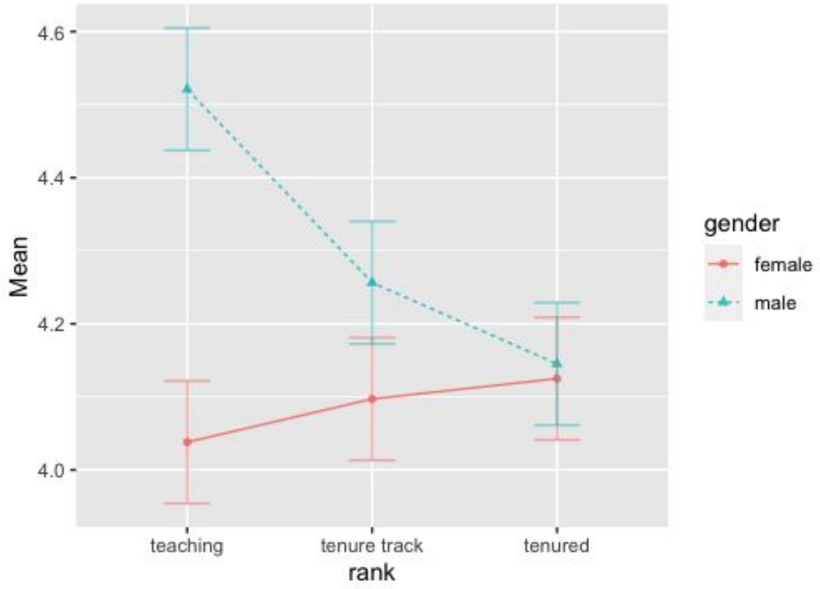
🔎 Contrast

- Planned comparison
- Theoretical interest
- High power
- High precision

🔎 Post hoc

- Unplanned comparisons
- Explore all differences
- *Adjust t value for inflated type I error?*

💡 [Multiple comparisons problem]



Repeated Measures Factorial ANOVA

Split Grid by Nicola Rennie (aRt package) 🎨

Sleep

Problem: 😳

Drug A: 💊 → 😴

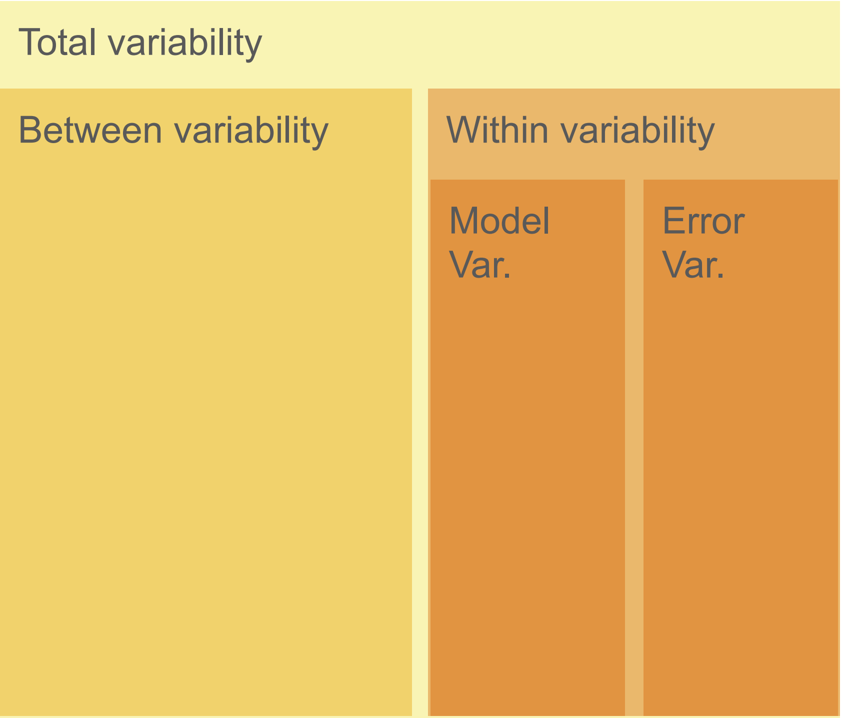
Drug B: 🍬 → 😴

Q. Which drug improves sleep the most (length)?

H. No hypothesis, let's explore!

E. In the sleep data set, …

Decomposition of variability



Formulas

Variance	Sum of Squares	df	Mean Squares	F-ratio
Between	$SS_{between} = SS_{total} - SS_{within}$	$DF_{total} - DF_{within}$	$\frac{SS_{between}}{DF_{between}}$	
Within	$SS_{within} = \sum s_i^2 (n_i - 1)$	$(n_i - 1)n$	$\frac{SS_{within}}{DF_{within}}$	
• Model	$SS_{model} = \sum n_k (\bar{X}_k - \bar{X})^2$	$k - 1$	$\frac{SS_{model}}{DF_{model}}$	$\frac{MS_{model}}{MS_{error}}$
• Error	$SS_{error} = SS_{within} - SS_{model}$	$(n-1)(k-1)$	$\frac{SS_{error}}{DF_{error}}$	
Total	$SS_{total} = s^2_{grand}(N - 1)$	$N - 1$	$\frac{SS_{total}}{DF_{total}}$	

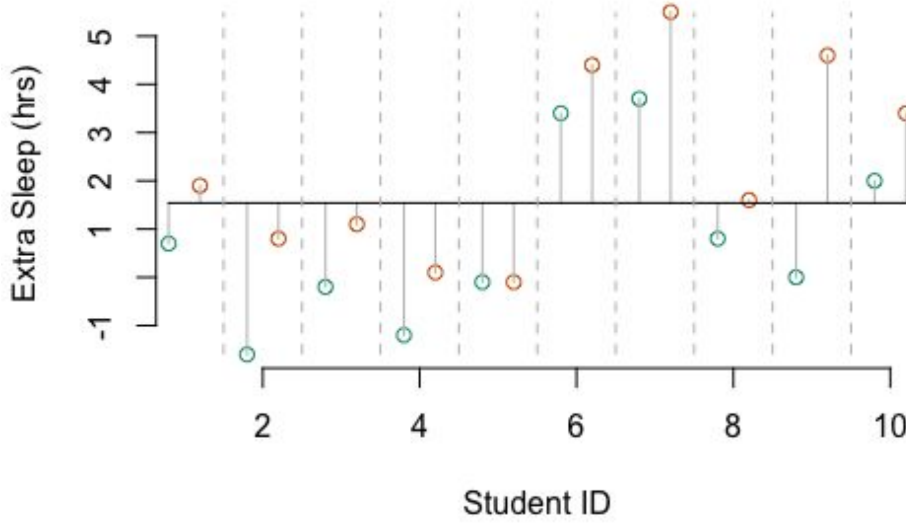
Where n_i is the number of observations per person and k is the number of conditions. These two are equal for a one-way repeated ANOVA. Furthermore n is the number of subjects per condition and N is the total number of data points $n \times k$.

Total

Formulas

Variance	Sum of Squares	df	Mean Squares	F-ratio
Between	$SS_{between} = SS_{total} - SS_{within}$	$DF_{total} - DF_{within}$	$\frac{SS_{between}}{DF_{between}}$	
Within	$SS_{within} = \sum s_i^2(n_i - 1)$	$(n_i - 1)n$	$\frac{SS_{within}}{DF_{within}}$	
• Model	$SS_{model} = \sum n_k(\bar{X}_k - \bar{X})^2$	$k - 1$	$\frac{SS_{model}}{DF_{model}}$	$\frac{MS_{model}}{MS_{error}}$
• Error	$SS_{error} = SS_{within} - SS_{model}$	$(n-1)(k-1)$	$\frac{SS_{error}}{DF_{error}}$	
Total	$SS_{total} = s^2_{grand}(N - 1)$	$N - 1$	$\frac{SS_{total}}{DF_{total}}$	

Where n_i is the number of observations per person and k is the number of conditions. These two are equal for a one-way repeated ANOVA. Furthermore n is the number of subjects per condition and N is the total number of data points $n \times k$.

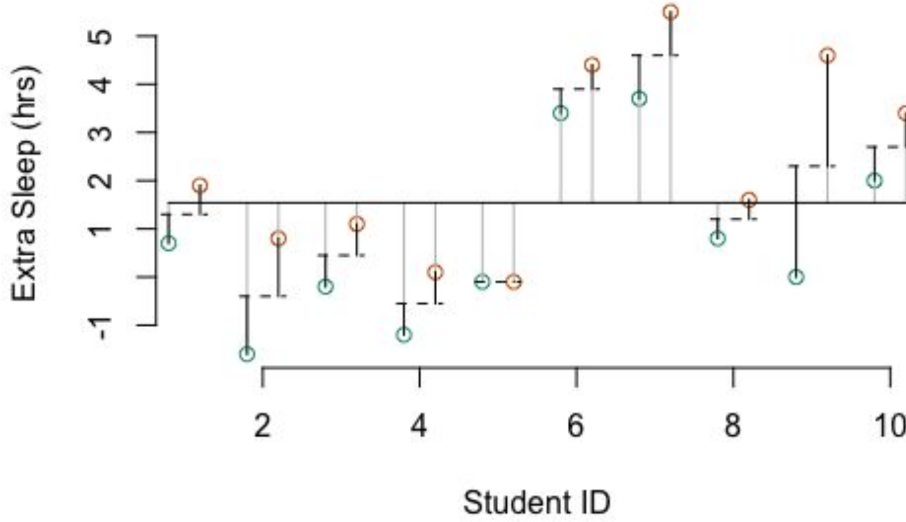


Within

Formulas

Variance	Sum of Squares	df	Mean Squares	F-ratio
Between	$SS_{between} = SS_{total} - SS_{within}$	$DF_{total} - DF_{within}$	$\frac{SS_{between}}{DF_{between}}$	
Within	$SS_{within} = \sum s_i^2(n_i - 1)$	$(n_i - 1)n$	$\frac{SS_{within}}{DF_{within}}$	
• Model	$SS_{model} = \sum n_k(\bar{X}_k - \bar{X})^2$	$k - 1$	$\frac{SS_{model}}{DF_{model}}$	$\frac{MS_{model}}{MS_{error}}$
• Error	$SS_{error} = SS_{within} - SS_{model}$	$(n-1)(k-1)$	$\frac{SS_{error}}{DF_{error}}$	
Total	$SS_{total} = s^2_{grand}(N-1)$	$N - 1$	$\frac{SS_{total}}{DF_{total}}$	

Where n_i is the number of observations per person and k is the number of conditions. These two are equal for a one-way repeated ANOVA. Furthermore n is the number of subjects per condition and N is the total number of data points $n \times k$.

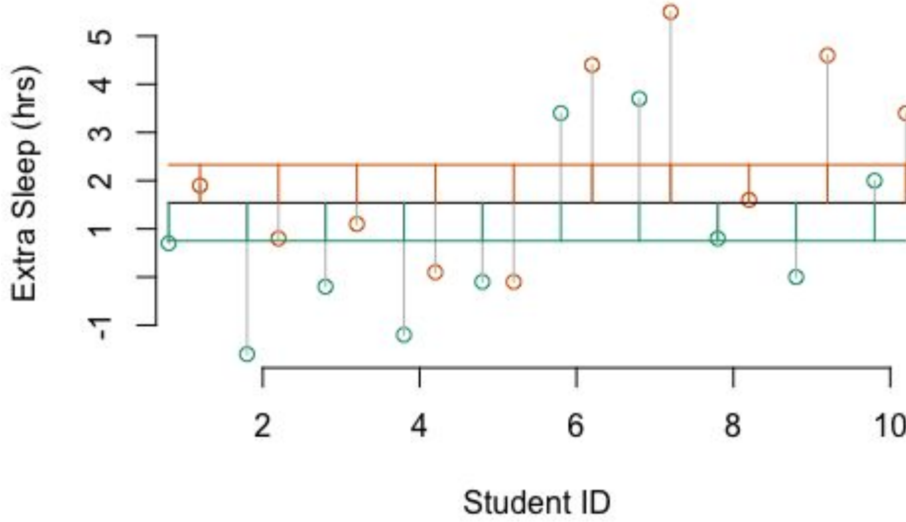


Model

Formulas

Variance	Sum of Squares	df	Mean Squares	F-ratio
Between	$SS_{between} = SS_{total} - SS_{within}$	$DF_{total} - DF_{within}$	$\frac{SS_{between}}{DF_{between}}$	
Within	$SS_{within} = \sum s_i^2 (n_i - 1)$	$(n_i - 1)n$	$\frac{SS_{within}}{DF_{within}}$	
• Model	$SS_{model} = \sum n_k(\bar{X}_k - \bar{X})^2$	$k - 1$	$\frac{SS_{model}}{DF_{model}}$	$\frac{MS_{model}}{MS_{error}}$
• Error	$SS_{error} = SS_{within} - SS_{model}$	$(n - 1)(k - 1)$	$\frac{SS_{error}}{DF_{error}}$	
Total	$SS_{total} = s^2_{grand}(N - 1)$	$N - 1$	$\frac{SS_{total}}{DF_{total}}$	

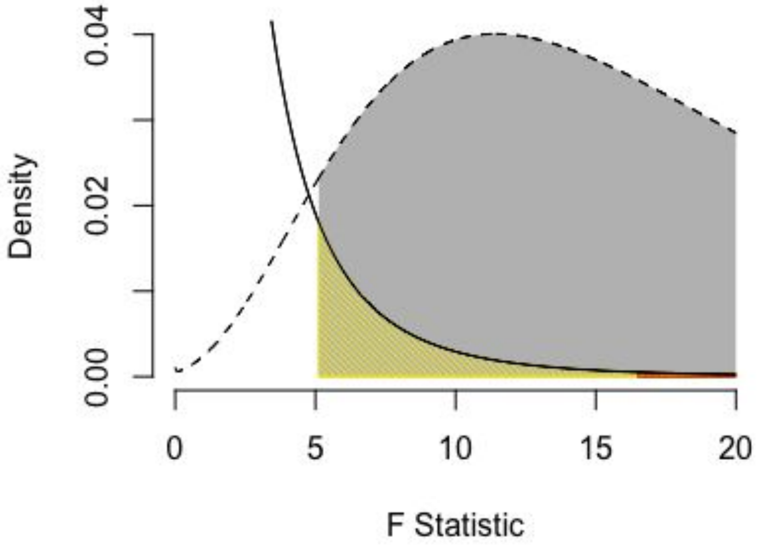
Where n_i is the number of observations per person and k is the number of conditions. These two are equal for a one-way repeated ANOVA. Furthermore n is the number of subjects per condition and N is the total number of data points $n \times k$.



F

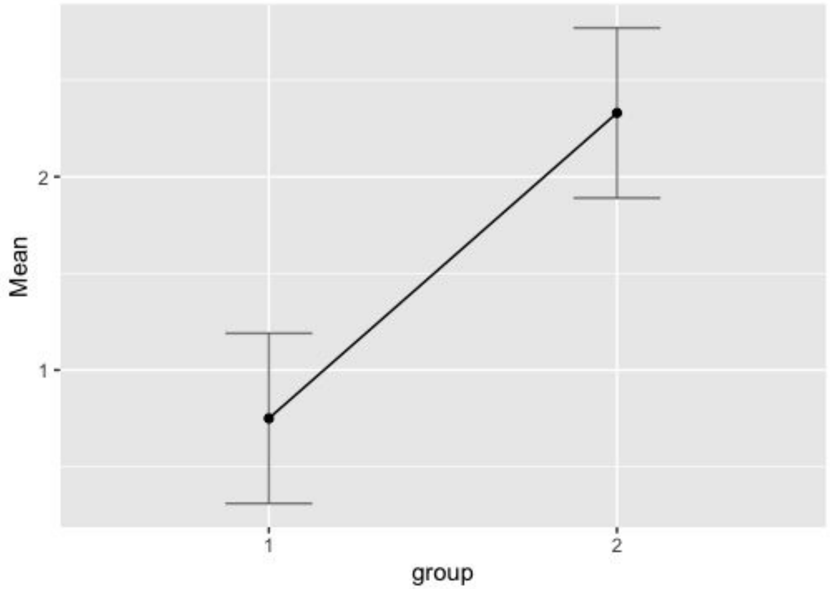
```
ez::ezANOVA(data = sleep,
            dv = extra,
            wid = ID,
            within = group,
            type = 2,
            return_aov = TRUE
)
```

F = 16.50088



👀

```
ez::ezPlot(data = sleep,
           x = .(group),
           dv = .(extra),
           wid = .(ID),
           within = .(group)
)
```



🧘 Cooling Down

Split Grid by Nicola Rennie (aRt package) 🎨

How t F R^2 all of these related?

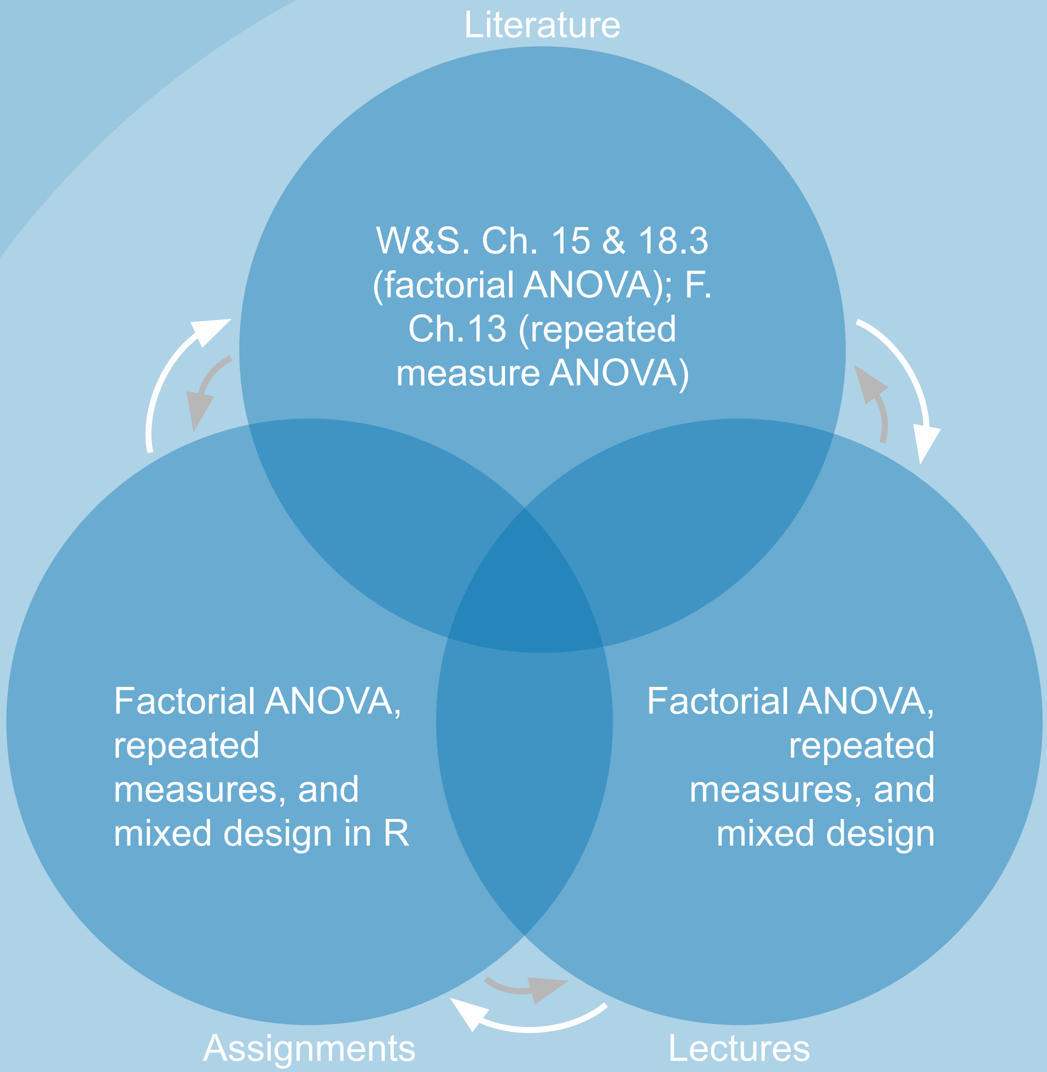
$F = t^2$

```
?sleep
mod <- extra ~ group
summary(aov(mod, sleep))[[1]]["F
value"][1, ]
t.test(mod, sleep)$statistic^2
```

$F = (R^2 / (1 - R^2)) \times (\mathrm{df}_{\mathrm{error}} / \mathrm{df}_{\mathrm{model}})$

```
?iris
mod <- Petal.Length ~ Sepal.Length +
Sepal.Width
fit <- summary(lm(formula = mod, data =
iris))
fit$fstatistic["value"]
(fit$r.squared / (1 - fit$r.squared)) *
(fit$fstatistic["dendf"] /
fit$fstatistic["numdf"])
```

🥇 Nail it


Literature
W&S. Ch. 15 & 18.3 (factorial ANOVA); F. Ch.13 (repeated measure ANOVA)
Factorial ANOVA, repeated measures, and mixed design in R
Factorial ANOVA, repeated measures, and mixed design
Assignments
Lectures

🥇 Exam(ple) question

Om de interactie tussen twee verschillende behandelingen te bepalen, voert een onderzoeker een independent factorial ANOVA uit. Ze berekent de Sum of Squares van het model (SS_{model} = 433), van behandeling A (SS_A = 99) en van behandeling B (SS_B = 157).

Geef de Sum of Squares van de interactie tussen behandeling A en B.

Take-home assignments

Weekly assignment

Pub quiz

Create an *informative* four-choice question about the content of today's lecture.

An informative question has a large spread in responses across answer options.

Clarify answer options (which are (in)correct and why).

Illustration adapted from Snippets.com

🔮 Overview

Topics

Probabilities & distributions
Frequentist inference
Multiple linear regression
Factorial ANOVA
| Nonparametric inference
Bayesian inference

Illustration by Jennifer Cheuk

🗃️ Look here!

Explore how individual data points affect ANOVA results (Seeing Theory).

Gain a true understanding of Type I to III sums of squares (Danielle Navarro).

⚠️ Don’t look here!

Create a challenge for another student. Also create a solution and make it a fun challenge!

Additional challenge: create an exam(ple) question and see how many of your fellow students can solve it within 5 minutes.

Share your attempt.

Hints (select and copy/paste the invisible text below to reveal it)

0.

1.

2.

3.

✍️ Colophon

Slides

alexandersavi.nl/teaching/

License

Statistical Reasoning by Alexander Savi is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. An Open Educational Resource. Approved for Free Cultural Works.