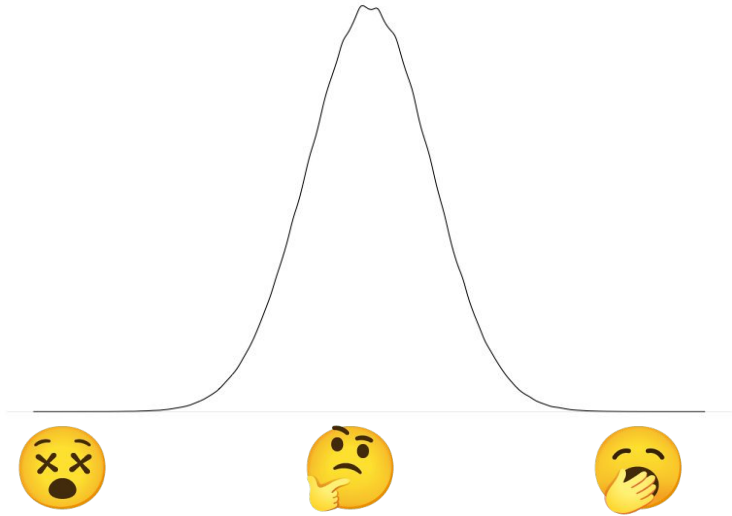
Philosophy of Science and

Statistical Reasoning

Moderation

But first, …

📢



Student Puzzlement Scale

> In het begin voel je je als iemand die in de woestijn staat. Waar je ook om je heen kijkt, er is alleen maar zand. Geen idee waar je naartoe moet. Maar dan langzaamaan begint er toch iets te dagen, en dan zie je toch licht aan het eind van de tunnel – dat is weer een andere beeldspraak.
>
> — Professor Einmahl, uitzonderingsstatisticus, in NRC

Previously, on statistical reasoning

```
Residuals:
     Min       1Q   Median       3Q      Max
-1.25582 -0.46922 -0.05741  0.45530  1.75599

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.52476    0.56344  -4.481 1.48e-05 ***
Sepal.Length  1.77559    0.06441  27.569  < 2e-16 ***
Sepal.Width  -1.33862    0.12236 -10.940  < 2e-16 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6465 on 147 degrees of freedom
Multiple R-squared:  0.8677,    Adjusted R-squared:  0.8659
F-statistic:   482 on 2 and 147 DF,  p-value: < 2.2e-16
```

Ordinary least squares (OLS) Explained Visually.

Compute t-statistic for β_1 (same procedure as for the mean): t = (1.776 − 0) / 0.064 = 27.569

Pub quiz

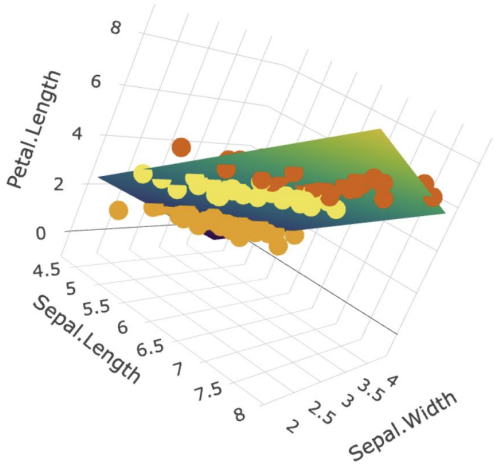
What will we learn today?

Topics

Statistical reasoning
Empirical cycle
Probability distributions
Frequentist inference
Sample / sampling distribution
Central limit theorem
Normal distribution
P-value
Type I/II errors
Effect size
Confidence interval
Power
Test statistics
Linear regression
t-Test
Moderation
F-distribution
Nonparametric inference
ANOVA
Bayesian inference

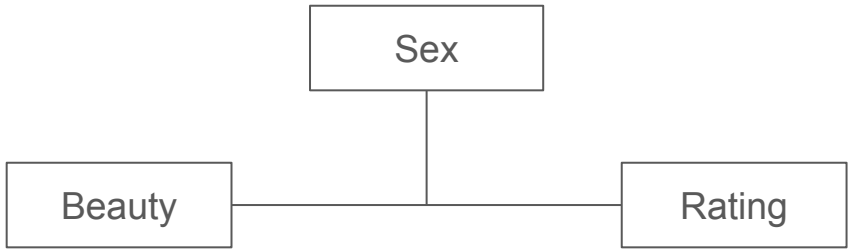
Questions

How can we determine if the relation between two variables depends on a third variable?



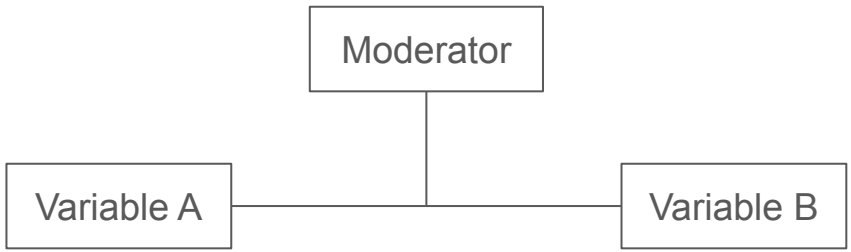
Estimating relationships between variables

> Instructors who are viewed as better looking receive higher instructional ratings, [...]. This impact exists within university departments and even within particular courses, and is larger for male than for female instructors.
>
> Disentangling whether this outcome represents productivity or discrimination is, as with the issue generally, probably impossible.
> — Hamermesh & Parker, 2005 🔒; NBER



Moderation / interaction

> In statistics and regression analysis, moderation (also known as effect modification) occurs when the relationship between two variables depends on a third variable. The third variable is referred to as the moderator variable [...]. — Wikipedia



Moderation

🤦‍♀️

library(“moderndive”)
help(evals)

```
> str(evals)
tibble [463 × 16] (S3: tbl_df/tbl/data.frame)
 $ ID          : int [1:463] 117 227 409 116 120 250 111 124 125 92 ...
 $ prof_ID     : int [1:463] 20 42 83 20 20 48 20 21 21 17 ...
 $ score       : num [1:463] 3.3 3.3 3.3 3.4 3.4 3.4 3.5 3.5 3.5 3.6 ...
 $ age         : int [1:463] 57 39 47 57 57 50 57 52 52 56 ...
 $ bty_avg     : num [1:463] 4.33 8.17 6.67 4.33 4.33 ...
 $ gender      : Factor w/ 2 levels "female","male": 1 1 1 1 1 1 1 1 1 1 ...
 $ ethnicity   : Factor w/ 2 levels "minority","not minority": 2 2 2 2 2 2 2 2 2 2 ...
 $ language    : Factor w/ 2 levels "english","non-english": 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 ...
 $ pic_outfit  : Factor w/ 2 levels "formal","not formal": 2 2 2 2 2 2 2 2 2 2 ...
 $ pic_color   : Factor w/ 2 levels "black&white",..: 2 2 1 2 2 2 2 2 2 2 ...
 $ cls_did_eval: int [1:463] 8 22 16 14 12 18 17 31 17 34 ...
 $ cls_students: int [1:463] 19 24 21 20 15 28 28 36 19 49 ...
 $ cls_level   : Factor w/ 2 levels "lower","upper": 2 1 1 2 2 2 2 2 2 2 ...
 $ mean_gender : num [1:463] 4.09 4.09 4.09 4.09 4.09 ...
 $ mean_rank   : num [1:463] 4.28 4.28 4.28 4.28 4.28 ...
```

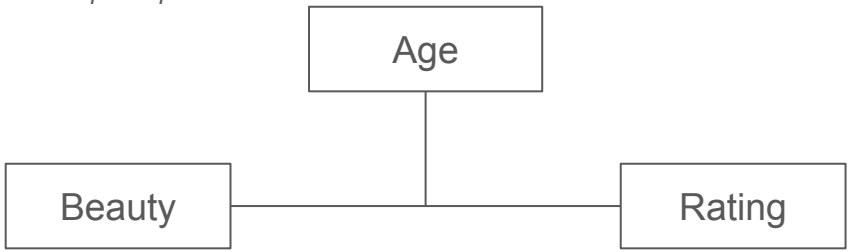
mod <- score ~ bty_avg + age + bty_avg : age

Q. Is the effect of beauty on instructional rating modified by age?

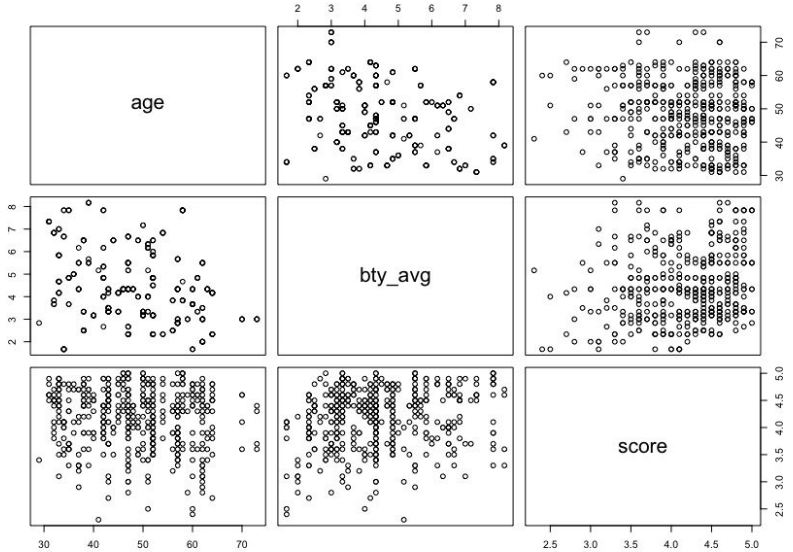
H. 💭 *What’s your hypothesis?*

E. (*Your hypothesis in terms of your operationalization.*)

$$Rating_i = \beta_o + \beta_1 Beauty_i + \beta_2 Age_i + \beta_3 Beauty_i \times Age_i + e_i$$



Student evaluations



Linear regression w/ interaction term

🤦

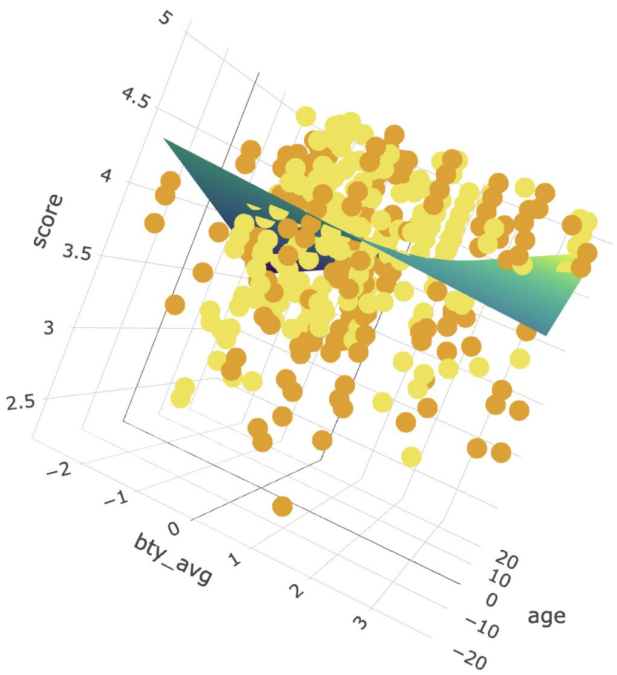
```
# mean centering
dat <- evals
dat$bty_avg <- dat$bty_avg -
mean(dat$bty_avg) # 4.4
dat$age <- dat$age - mean(dat$age) # 48.4

fit <- lm(formula = mod, data = dat)
summary(fit)
```

💭 To mean center or not to mean center? See last paragraph of the Discussion section for practical advice.

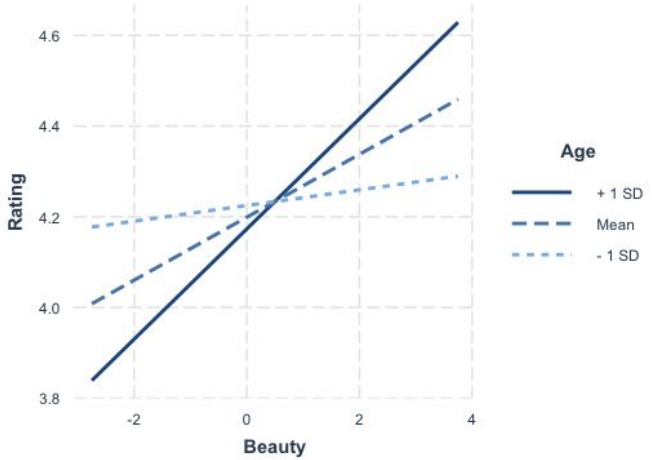
```
Residuals:
    Min      1Q  Median      3Q     Max
-1.9410 -0.3517  0.1231  0.4040  1.0066

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.198930   0.025603 164.001  < 2e-16 ***
age         -0.002636   0.002638  -0.999 0.318201
bty_avg      0.069389   0.017107   4.056 5.86e-05 ***
age:bty_avg  0.005318   0.001580   3.366 0.000827 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5287 on 459 degrees of freedom
Multiple R-squared:  0.06096,   Adjusted R-squared:  0.05482
F-statistic: 9.933 on 3 and 459 DF,  p-value: 2.349e-06
```

$$Rating_i = 4.20 + 0.07\ Beauty_i - 0.00\ Age_i + 0.01\ Beauty_i \times Age_i + e_i$$

Visualize interaction




```
library("interactions")
interactions::interact_plot(model = fit, pred =
bty_avg, modx = age, data = dat)
```



Simple slopes analysis & Johnson–Neyman interval

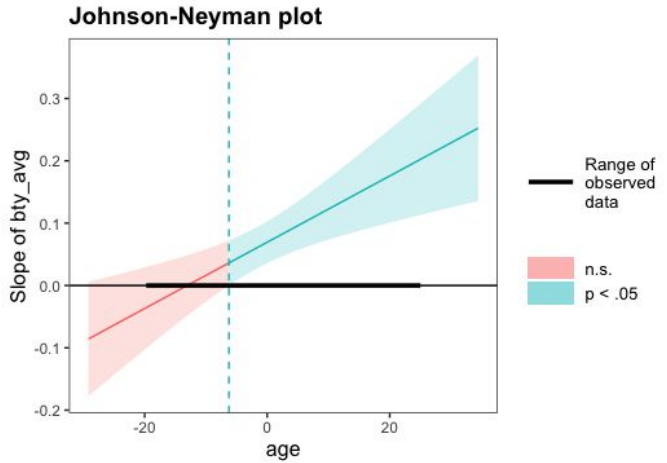
👩‍💻

```
library("sandwich")
interactions::sim_slopes(fit, pred = bty_avg,
modx = age)
```

```
SIMPLE SLOPES ANALYSIS

Slope of bty_avg when age = -9.802742e+00 (- 1 SD):

  Est.   S.E.   t val.      p
------ ------ -------- ------
  0.02   0.02     0.81   0.42

Slope of bty_avg when age =  1.930589e-14 (Mean):

  Est.   S.E.   t val.      p
------ ------ -------- ------
  0.07   0.02     4.06   0.00

Slope of bty_avg when age =  9.802742e+00 (+ 1 SD):

  Est.   S.E.   t val.      p
------ ------ -------- ------
  0.12   0.02     4.91   0.00
```

👩‍💻

```
interactions::johnson_neyman(fit, pred =
bty_avg, modx = age, alpha = .05)
```



Model evaluation

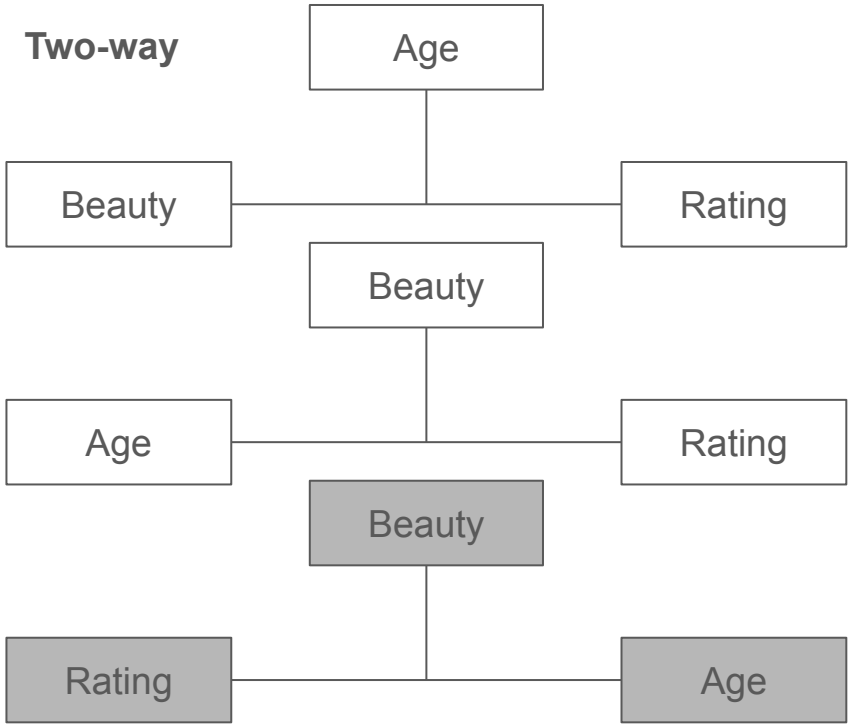
See previous lecture

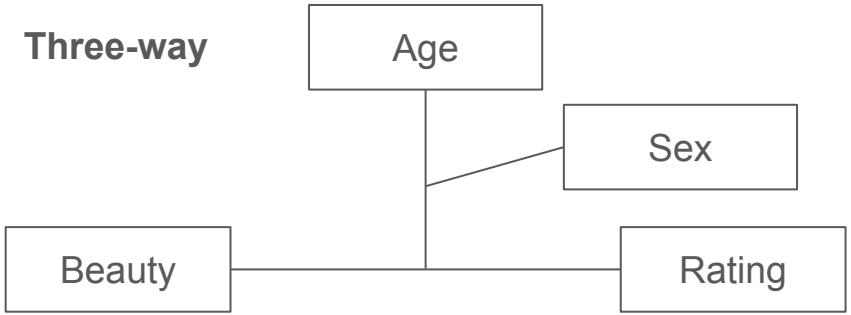
p-values; R^2

```
Residuals:
    Min      1Q  Median      3Q     Max
-1.9410 -0.3517  0.1231  0.4040  1.0066

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.198930   0.025603 164.001  < 2e-16 ***
age         -0.002636   0.002638  -0.999 0.318201
bty_avg      0.069389   0.017107   4.056 5.86e-05 ***
age:bty_avg  0.005318   0.001580   3.366 0.000827 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5287 on 459 degrees of freedom
Multiple R-squared:  0.06096,   Adjusted R-squared:  0.05482
F-statistic: 9.933 on 3 and 459 DF,  p-value: 2.349e-06
```

🧠 Students don’t know what’s best for their own learning (The Conversation)

Higher-order interactions





$$Rating_i = \beta_o + \beta_1 Beauty_i + \beta_2 Age_i + \beta_3 Sex + \beta_4 Beauty_i \times Age_i + \beta_5 Beauty_i \times Sex_i + \beta_6 Age_i \times Sex_i + \beta_7 Beauty_i \times Age_i \times Sex_i + e_i$$

👩‍💻

score ~ bty_avg * age * gender

Cooling down

What did we learn?

Topics

Statistical reasoning
Empirical cycle
Probability distributions
Frequentist inference
Sample / sampling distribution
Central limit theorem
Normal distribution
P-value
Type I/II errors
Effect size
Confidence interval
Power
Test statistics
Linear regression
t-Test
Moderation
F-distribution
Nonparametric inference
ANOVA
Bayesian inference

Illustration by Jennifer Cheuk

Take-home assignments

📅 Weekly assignment

Q2/3 ask you to use a different method to determine multicollinearity.

🍻 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

Colophon

Slides
alexandersavi.nl/teaching/