

# Introduction to Statistical Modeling

## Introduction

Joris Vankerschaver

# Problem setting

- 26 observations from class of 2021-22 (19 female and 7 male) + 1 professor (**27 total**)
- Measurement of right **palm width** and **height** (both in cm).
- Random sample? From which population?
- Sources of bias, error?

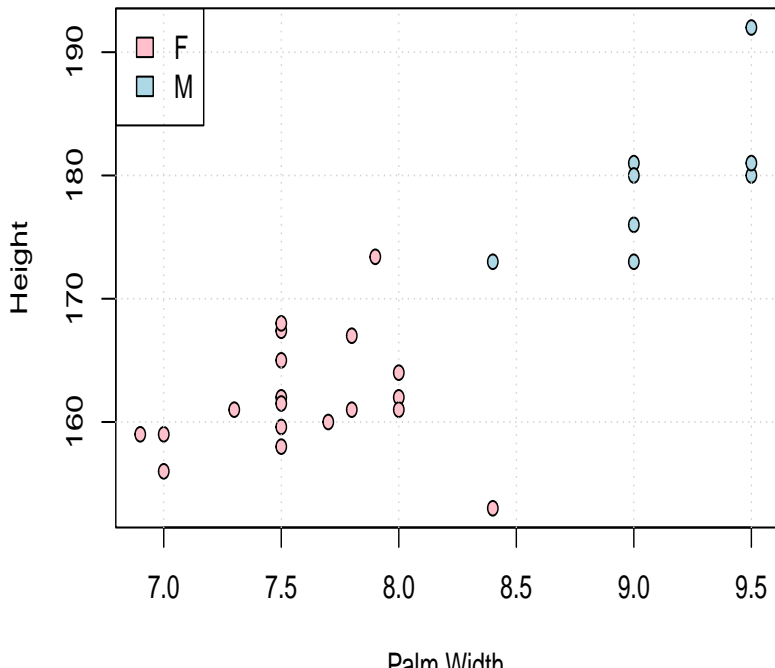
## **i** Research questions

- Is there an association between height and palm width?
- Can we predict a person's height from their palm width?
- If yes, how confident are we in these results?

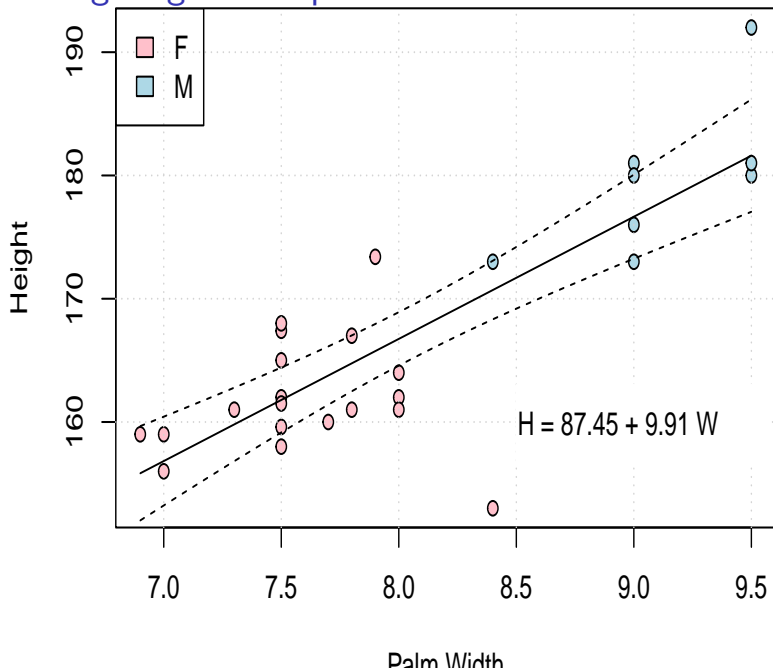
# Simple and multiple linear regression

- In this lecture, we build a **simple linear regression** model.
- Simple regression: effect on height of a single predictor (palm width)
- Multiple regression: multiple predictors (palm width, gender, year, ...)

## The raw data



## Associating height with palm width



# Via R

```
m <- lm(Height ~ Palm.width, data=heights)
summary(m)
```

Call:

```
lm(formula = Height ~ Palm.width, data = heights)
```

Residuals:

Min	1Q	Median	3Q	Max
-17.7055	-3.1967	-0.2853	3.1882	10.3919

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	87.450	10.814	8.087	1.93e-08 ***
Palm.width	9.911	1.338	7.408	9.26e-08 ***

---

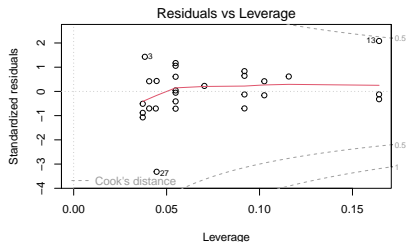
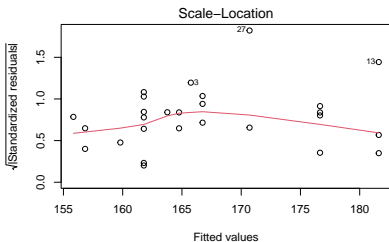
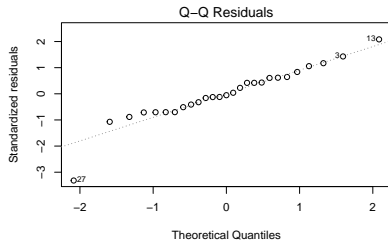
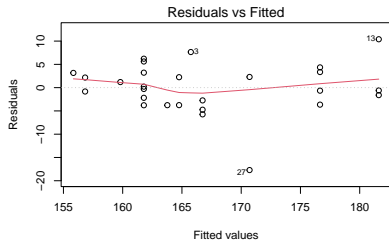
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.459 on 25 degrees of freedom

Multiple R-squared: 0.687, Adjusted R-squared: 0.6745

F-statistic: 54.88 on 1 and 25 DF, p-value: 9.264e-08

# Model diagnostics



## Predicting height from palm width

- Model:  $E(H|W = w) = 87.45 + 9.91 \times w$ .
- Predicted expected height of a person with palm width 8.75cm:

$$E(H|W = 8.75) = 87.45 + 9.91 \times 8.75 = 174.17 \text{ cm}$$

Regression coefficients:

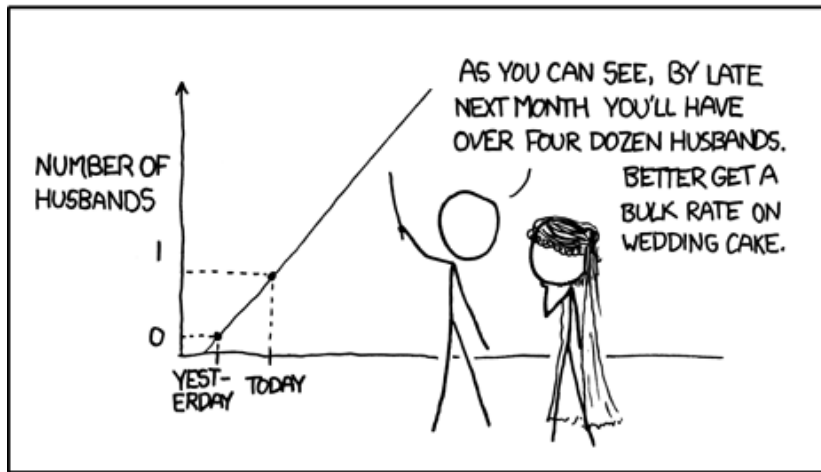
- **Intercept** (87.45cm): height of a hypothetical student with palms that are 0 cm wide. Often makes more sense after mean-centering.
- **Slope** (9.91): each extra cm in palm width is associated with an increase of 9.91 cm in height.



## Be careful with extrapolating

Predicting outside the range of the data can yield misleading results.

### MY HOBBY: EXTRAPOLATING



Source: XKCD

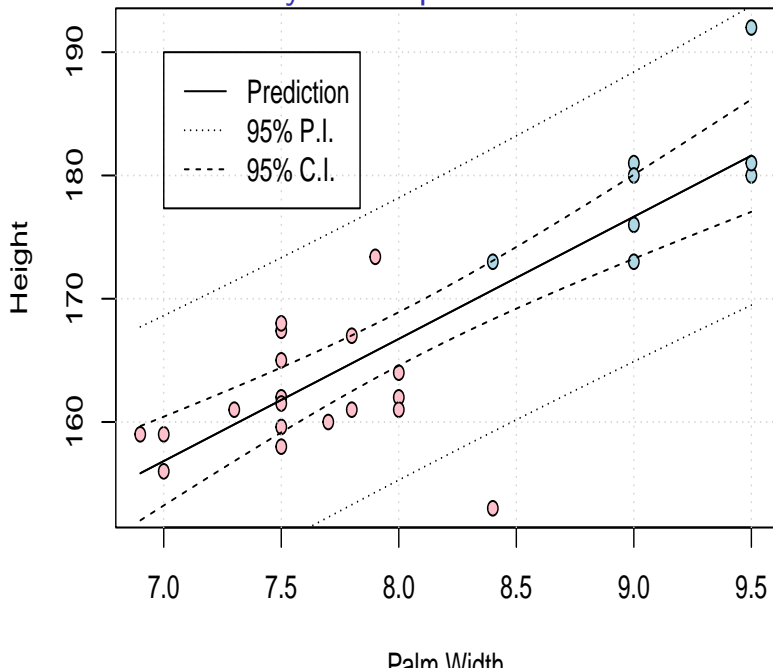
# What is the uncertainty in our prediction?

Assuming that our model is good, how accurate are the predictions from it?

For prediction  $E(H|W = 8.75) = 174.17$  cm:

- 95% confidence interval:  $[171.27, 177.08]$ . Uncertainty in **average prediction**.
- 95% prediction interval:  $[162.56, 185.79]$ . Uncertainty in **individual predictions**.

## What is the uncertainty in our prediction?



## Association between predictor and outcome

The regression slope  $\beta = 9.91$  measures the strength of the association between palm width and height.

- If close to 0: no association
- If different from 0: some degree of association

How do we test whether  $\beta$  is 0?

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	87.449557	10.813838	8.086819	1.928912e-08
Palm.width	9.911427	1.337903	7.408178	9.263523e-08