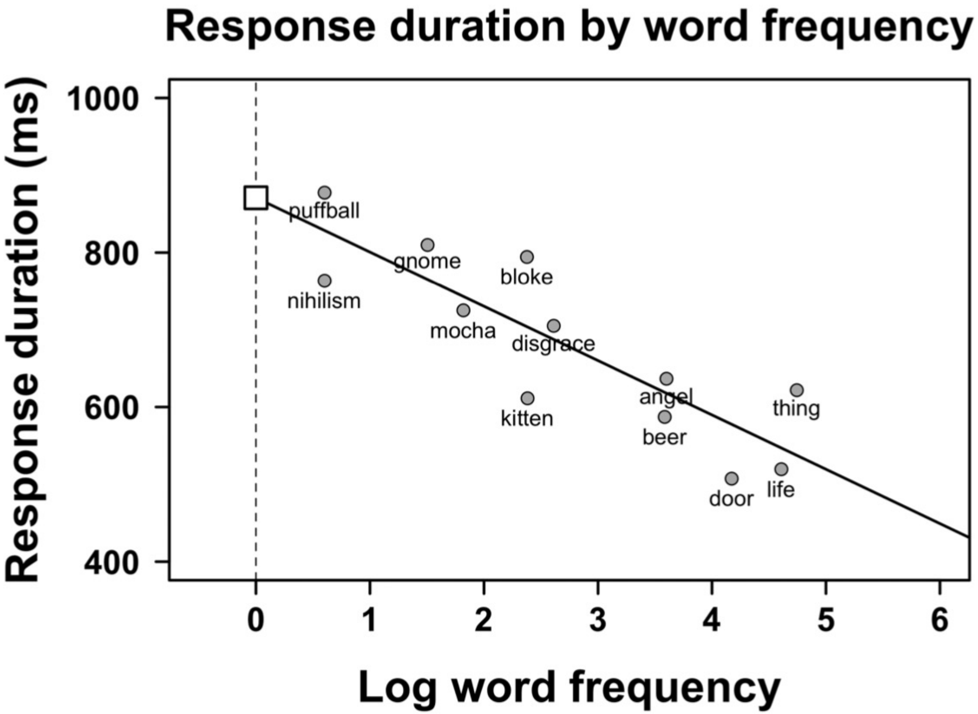
The linear model

Dr. Margriet A. Groen

Outline

- Regression line: intercept and slope.
- Residuals
- Different types of regression
- Assumptions
- Measuring model fit: R^2

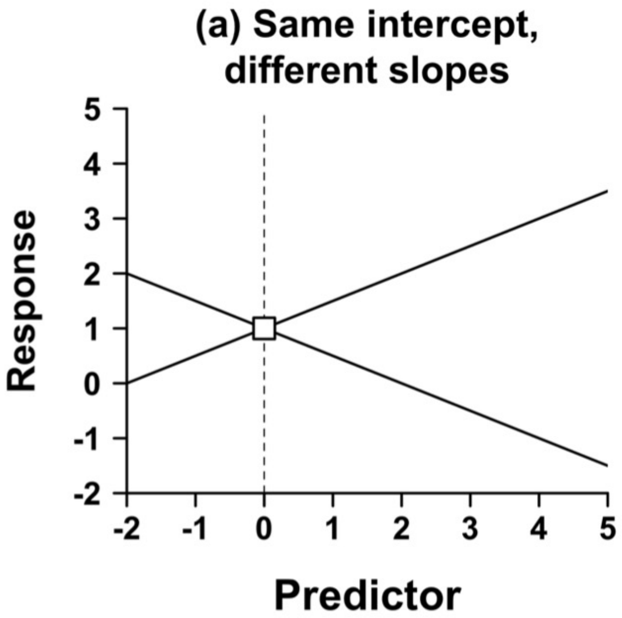
An example: Word frequency effects



Terminology

y	x
response/outcome dependent variable	predictor independent variable explanatory variable regressor

Specifying a line: Slope and intercept



$$slope = \frac{\Delta y}{\Delta x}$$

Regression line

weight

$$y = b_0 + b_1 * x$$

dependent variable

intercept

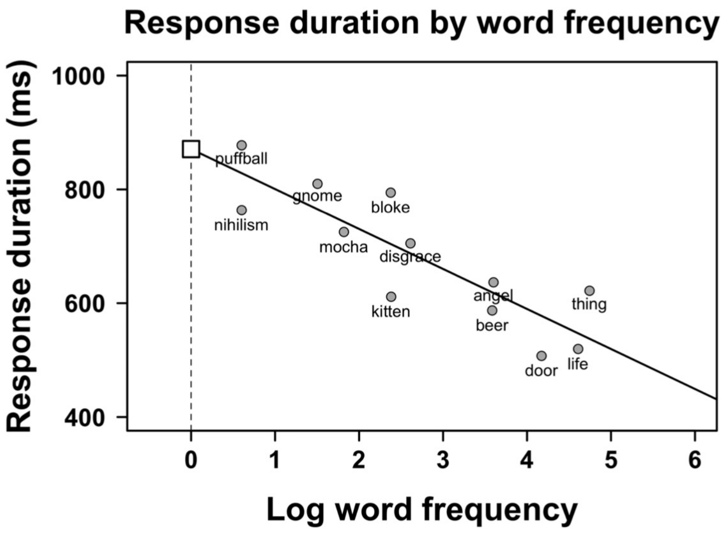
predictor

An example: Word frequency effects (2)

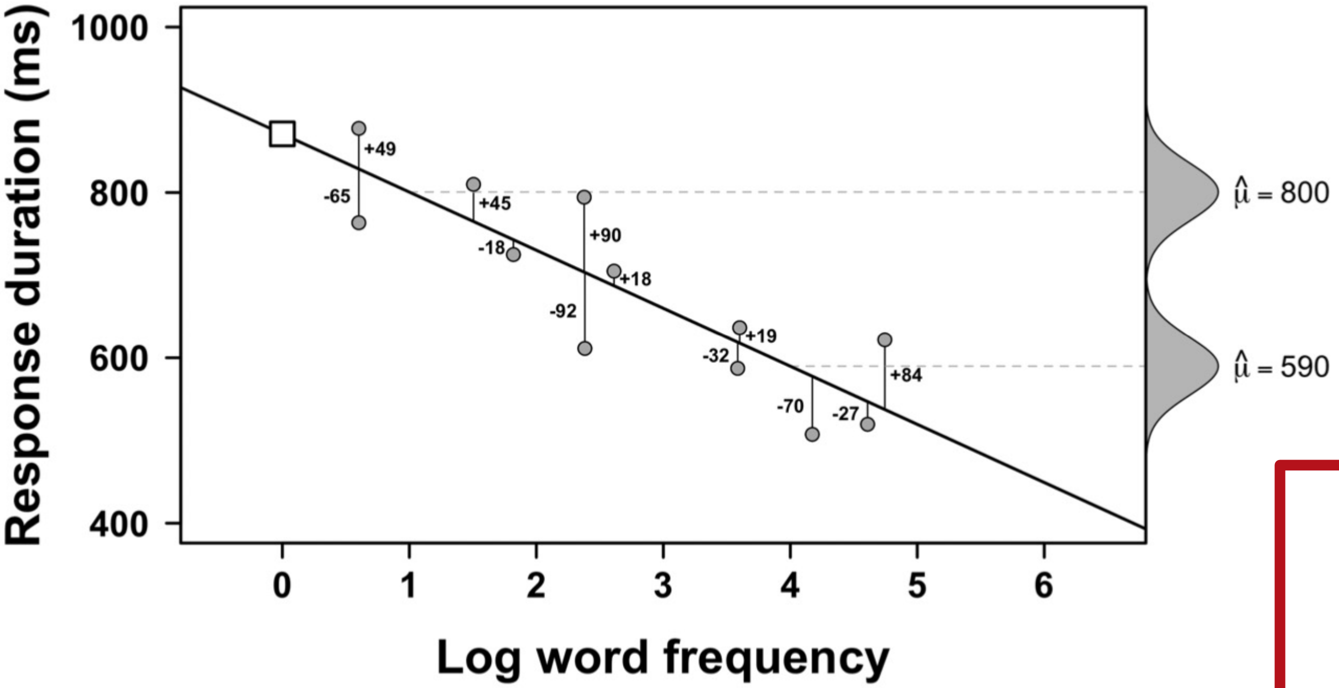
$$y = b_0 + b_1 * x$$

$$response\ duration = 880ms + \left(-70\frac{ms}{freq}\right) * word\ frequency$$

$$response\ duration = 880ms + \left(-70\frac{ms}{freq}\right) * 3\ freq = 670ms$$



Residuals


Response duration (ms)
1000
800
600
400
+49
-65
+45
-18
+90
+18
-92
+19
-32
+84
-70
-27
$\hat{\mu}$ = 800
$\hat{\mu}$ = 590
0
1
2
3
4
5
6
Log word frequency

Regression line (2)

weight

error

$$y = b_0 + b_1 * x + e$$

dependent variable

intercept

predictor

Linear regression

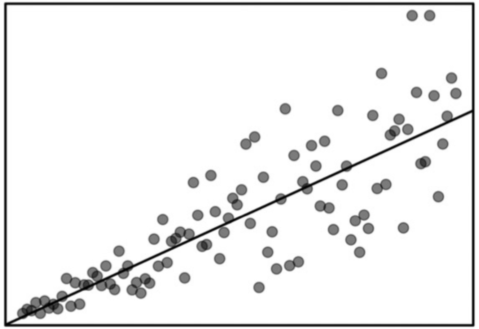
… is a statistical method used to create a linear model

… there are different types:

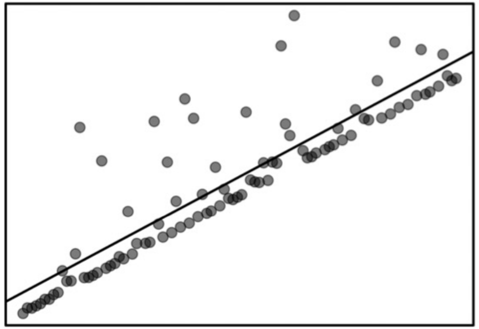
- **Simple linear regression:** models using only one predictor
- **Multiple linear regression:** models using multiple predictors
- **Logistic regression**: models a categorical response variable
- **Multivariate linear regression:** models for multiple response variables

Assumptions

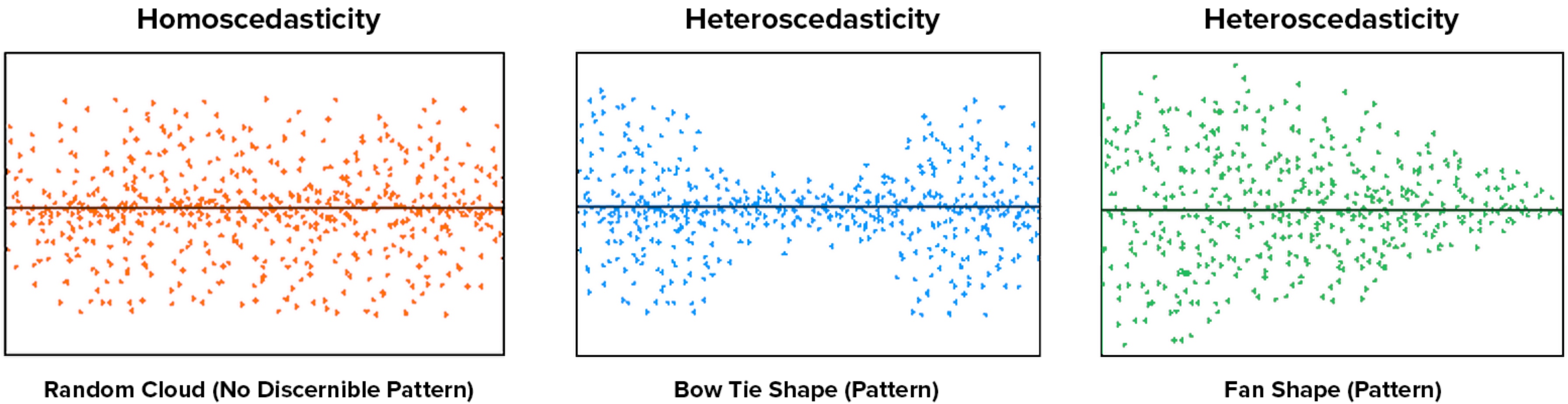
(a) Non-constant variance



(b) Non-normal residuals

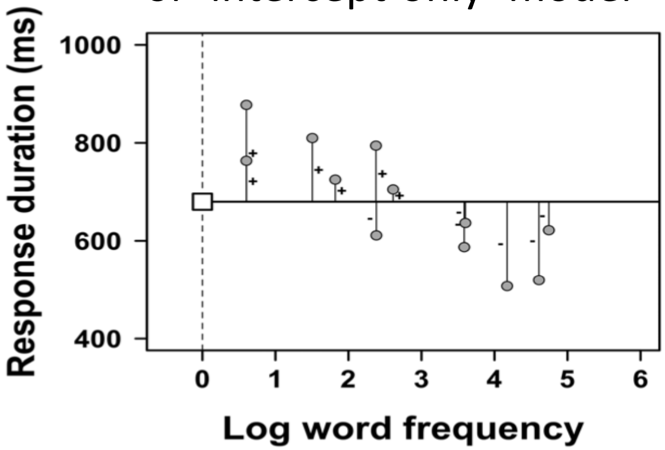


Homoscedasticity



Measuring model fit: The null model

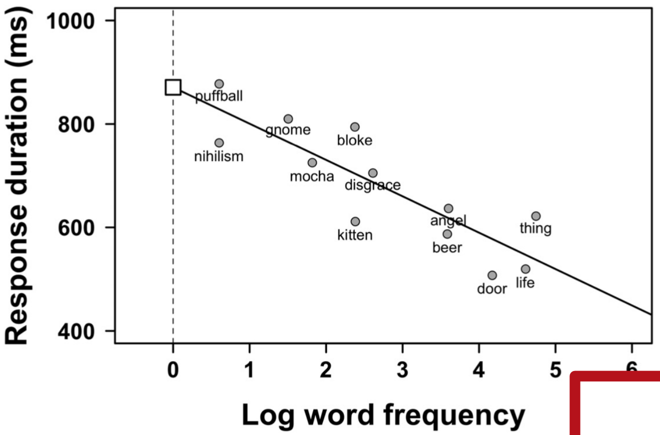
Null model
or ‘intercept-only’ model



$$y = b_0 + b_1 * x + e$$

Model with word frequency predictor

Response duration by word frequency

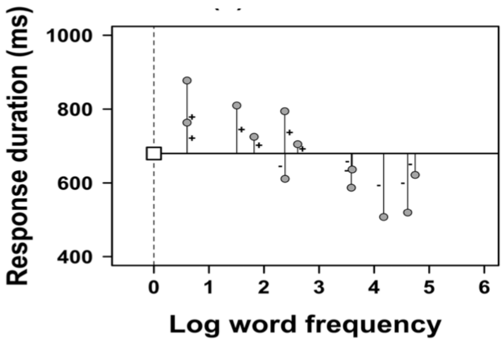


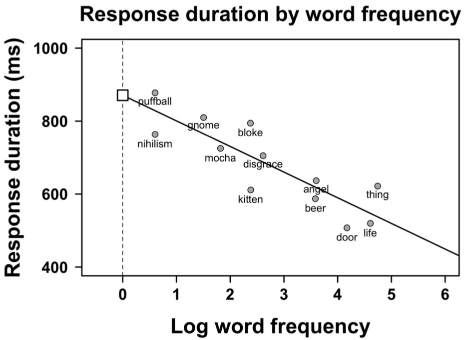
$$y = b_0 + b_1 * x + e$$

Measuring model fit: R-squared

Null model
or 'intercept-only' model

Model with word frequency predictor





sum of
squared errors
(SSE)

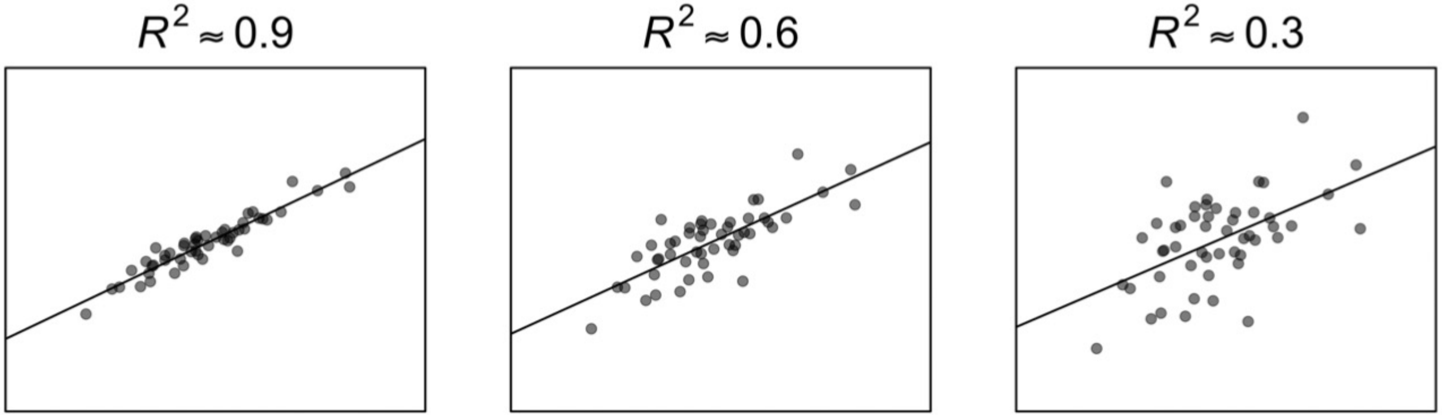
SSE_{null} = 152,767

SSE_{model} = 42,609

$$R^2 = 1 - \frac{SSE_{model}}{SSE_{null}}$$

$$R^2 = 1 - \frac{42,609}{152,767} = 0.72$$

Measuring model fit: R-squared (*cont.*)



R^2 is a measure of effect size.
R^2 values range from 0 to 1.
Values closer to 1 indicate better model fits as well as stronger effects.

Summary

- Mathematical specification of a line: intercept and slope.
- Regression line formula:
- Residuals = observed values – fitted values
- Simple linear regression, multiple linear regression, logistic regression, multivariate regression
- Assumptions: residuals need to be normally distributed and show constant variance (i.e., be homoscedastic)
- R^2 uses the residuals of the null model to standardise the residuals of the main model. This provides an effect size and tells us what proportion of variation in the dependent variable can be accounted for by the predictor in the main model.