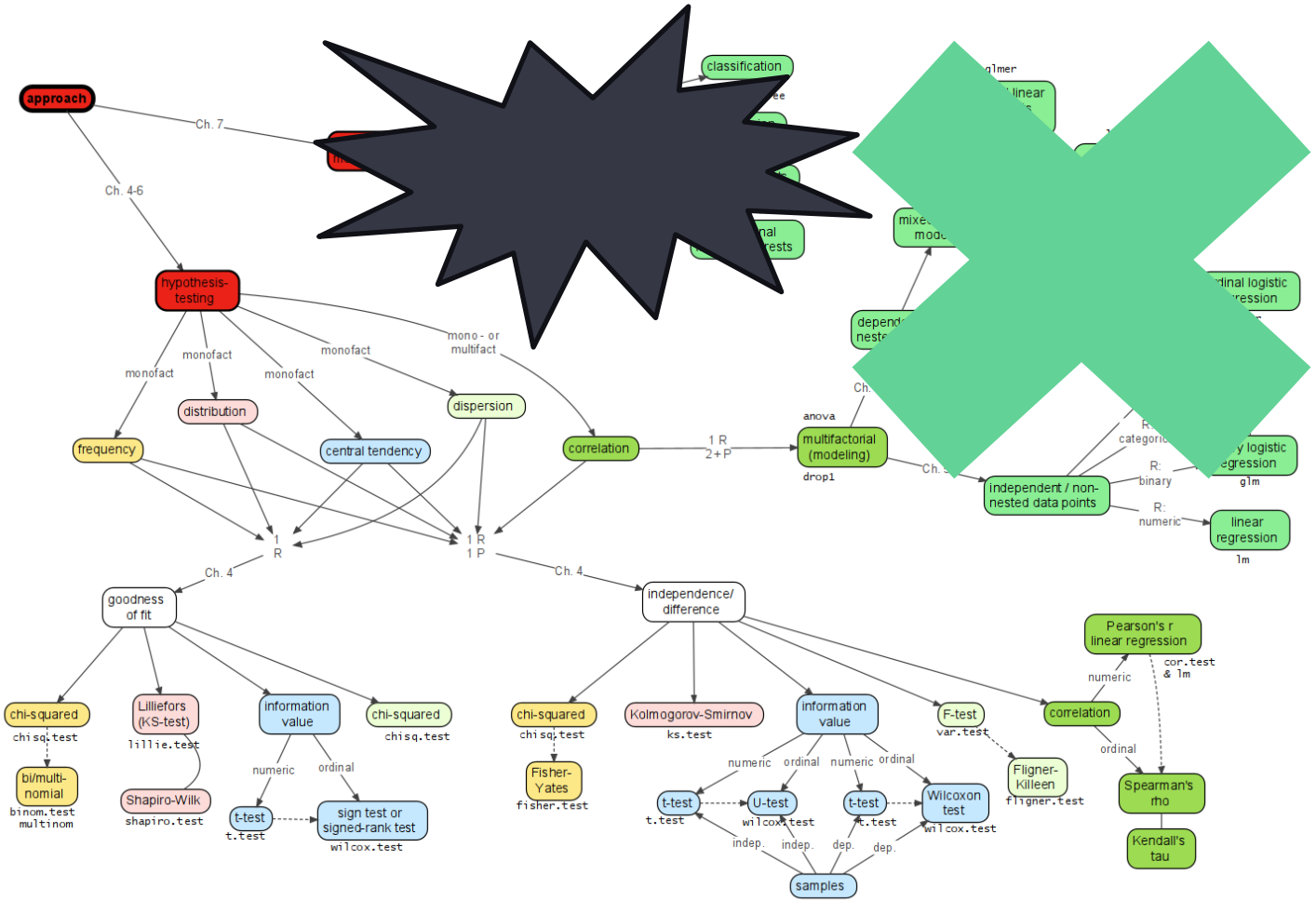
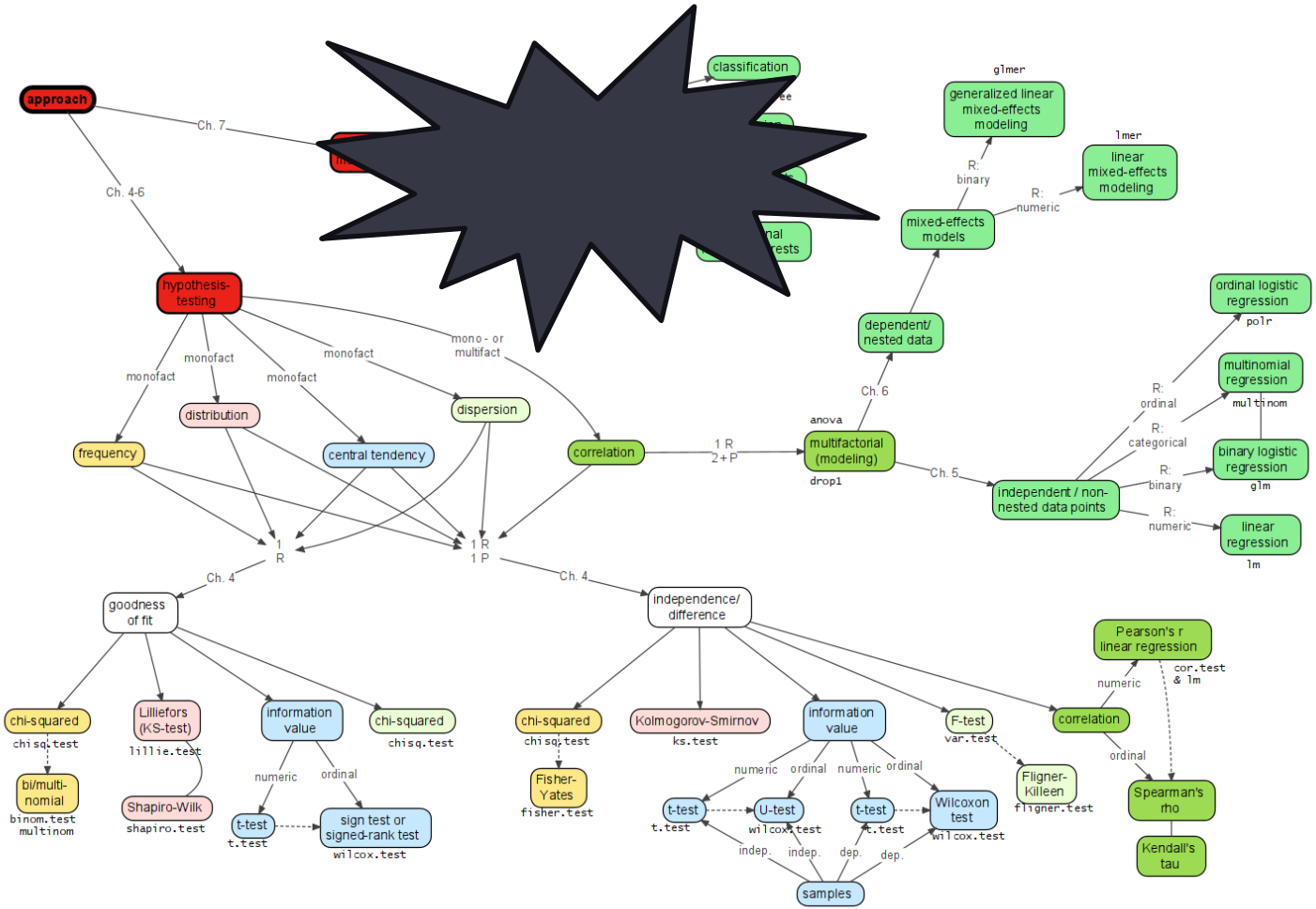
Bayesian statistics, general linear models and mixed effects

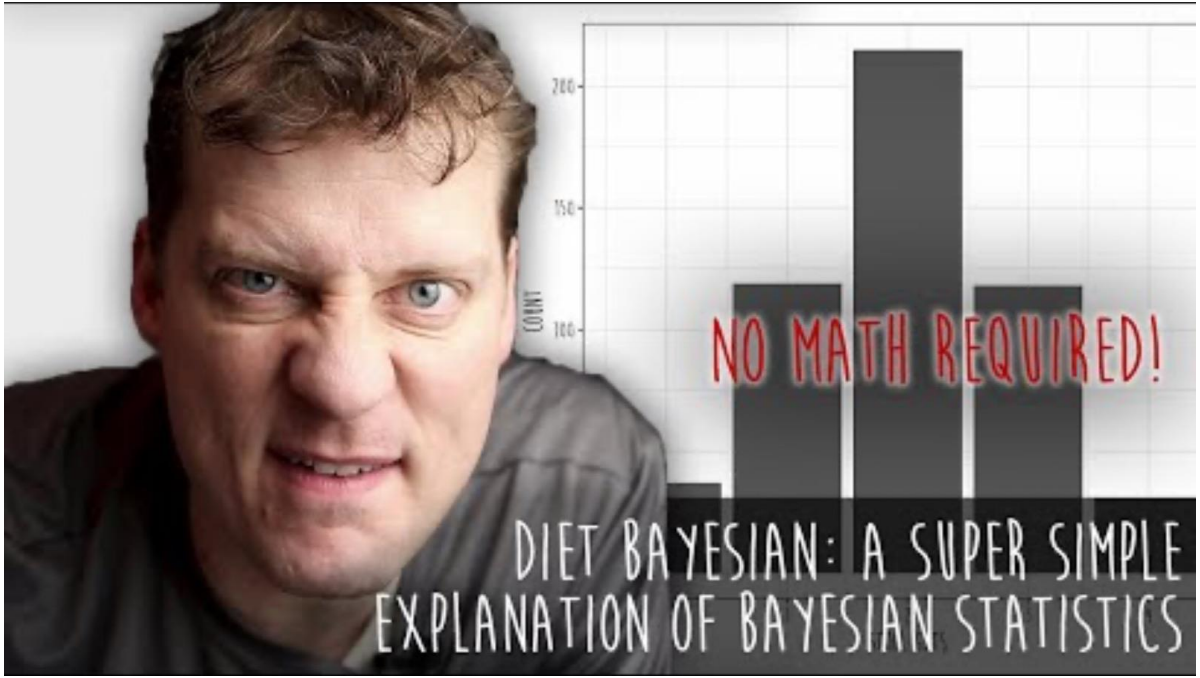

approach
Ch. 7
Ch. 4-6
classification
hypothesis-testing
monofact
monofact
monofact
monofact
mono - or multifact
distribution
dispersion
frequency
central tendency
correlation
1 R
2+P
multifactorial (modeling)
anova
drop1
independent / non-nested data points
R: categoric
R: binary
glm
R: numeric
linear regression
lm
1 R
1 P
Ch. 4
goodness of fit
chi-squared
chisq.test
bi/multi-nomial
binom.test
multinom
Lilliefors (KS-test)
lillie.test
Shapiro-Wilk
shapiro.test
information value
numeric
ordinal
t-test
t.test
sign test or signed-rank test
wilcox.test
chi-squared
chisq.test
Ch. 4
independence/ difference
chi-squared
chisq.test
Fisher-Yates
fisher.test
Kolmogorov-Smirnov
ks.test
information value
numeric
ordinal
numeric
ordinal
t-test
t.test
U-test
wilcox.test
t-test
t.test
Wilcoxon test
wilcox.test
indep.
indep.
dep.
dep.
samples
F-test
var.test
Fligner-Killeen
fligner.test
correlation
Pearson's r linear regression
cor.test & lm
numeric
ordinal
Spearman's rho
Kendall's tau


approach
Ch. 7
classification
Ch. 4-6
hypothesis-testing
monofact
monofact
monofact
monofact
mono - or multifact
frequency
distribution
central tendency
dispersion
correlation
1 R
2+P
anova
multifactorial (modeling)
drop1
Ch. 6
dependent/ nested data
glmer
generalized linear mixed-effects modeling
lmer
linear mixed-effects modeling
R: binary
R: numeric
mixed-effects models
Ch. 5
independent / non-nested data points
R: ordinal
R: categorical
R: binary
R: numeric
ordinal logistic regression
polr
multinomial regression
multinom
binary logistic regression
glm
linear regression
lm
1 R
1 P
Ch. 4
goodness of fit
Ch. 4
independence/ difference
chi-squared
chisq.test
Lilliefors (KS-test)
lillie.test
information value
chi-squared
chisq.test
bi/multi-nomial
binom.test
multinom
Shapiro-Wilk
shapiro.test
numeric
ordinal
t-test
t.test
sign test or signed-rank test
wilcox.test
chi-squared
chisq.test
Kolmogorov-Smirnov
ks.test
information value
F-test
var.test
correlation
Pearson's r linear regression
cor.test & lm
numeric
ordinal
Fisher-Yates
fisher.test
numeric
ordinal
numeric
ordinal
Fligner-Killeen
fligner.test
Spearman's rho
Kendall's tau
t-test
t.test
U-test
wilcox.test
t-test
t.test
Wilcoxon test
wilcox.test
indep.
indep.
dep.
dep.
samples

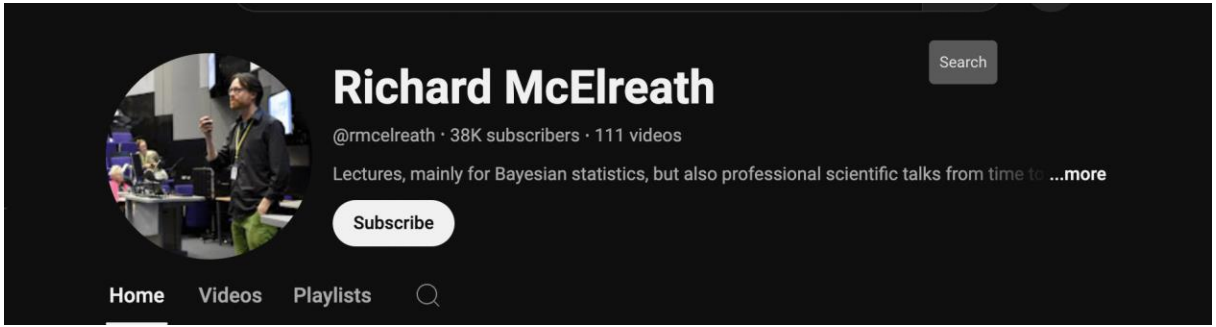
Overview

1. **Bayesian models**
2. Mixed effects
3. Generalised linear models

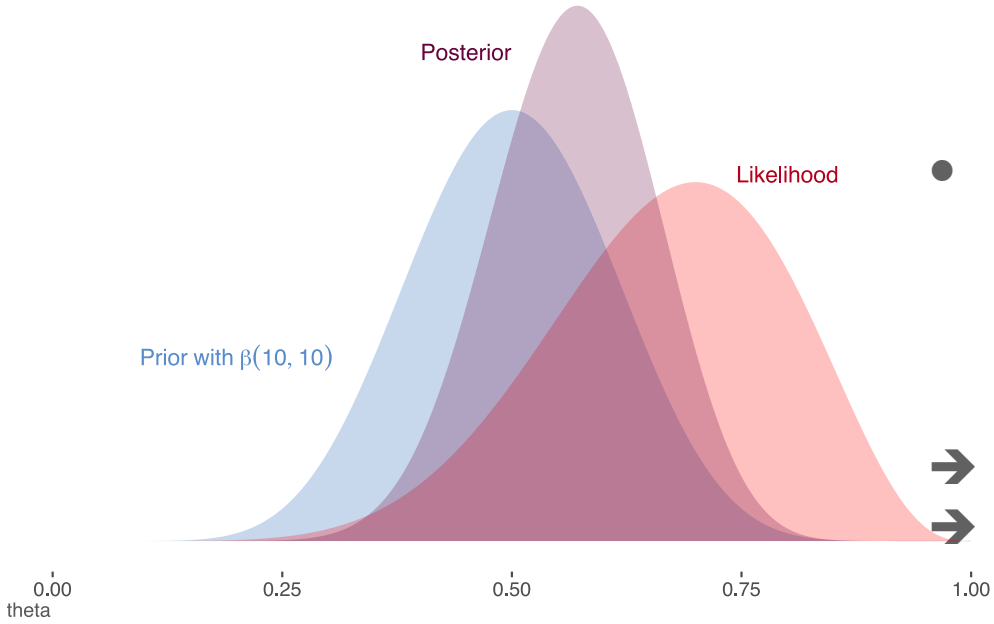
What in the Bayesian



What in the Bayesian?



Bayesian models



- ⇔ frequentists and their p-values
  - Based on…frequency i.e. making statements about a population based on a sample
- Key concepts:
  1. Priors: what you believe before observations
  2. Posteriors: adjustments to beliefs after looking at/adding observations

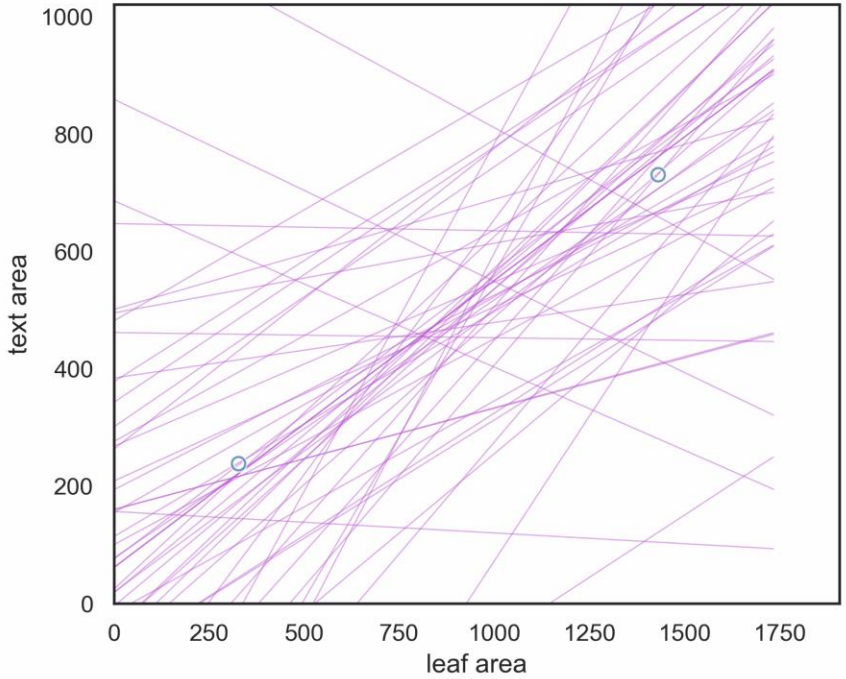
➔ No minimum sample size

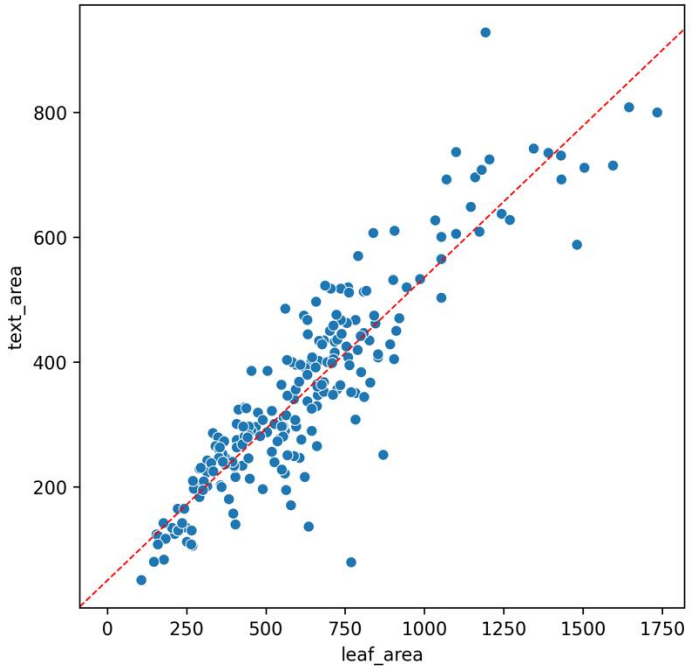
➔ By adding observations, we update prior (= posterior)

➔ Posterior keeps getting closer to actual population

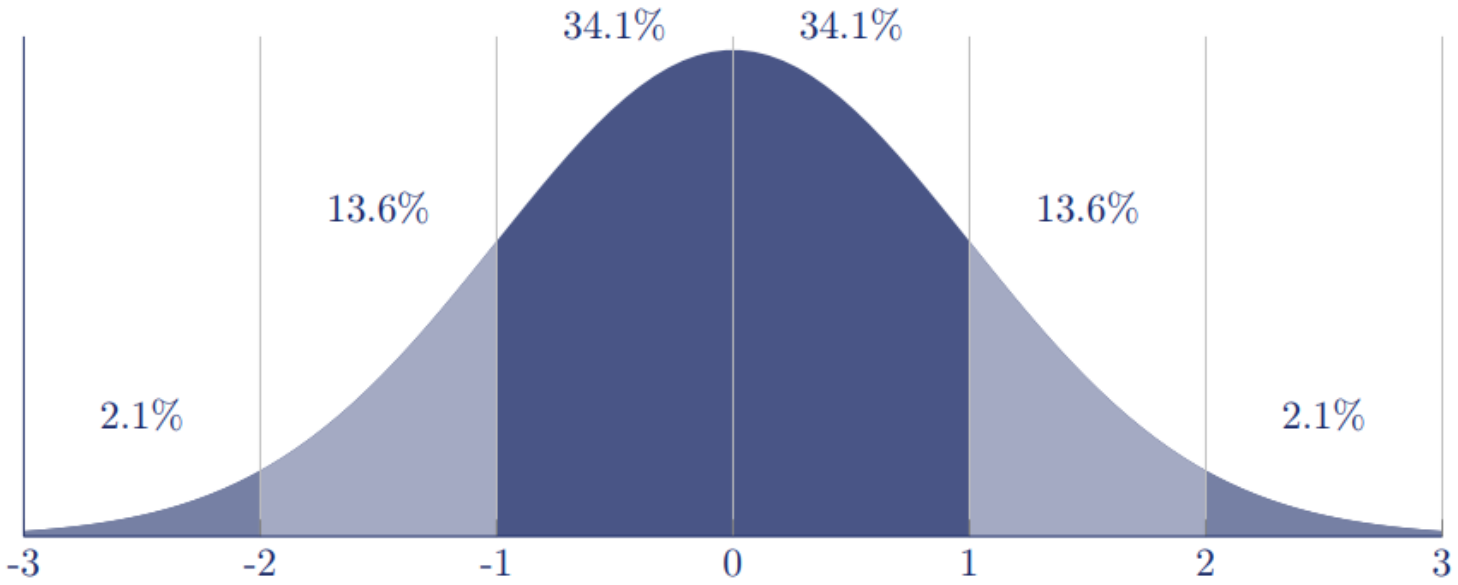
https://m-clark.github.io/bayesian-basics

Bayesian models: an illustration





Sidebar: standard deviations

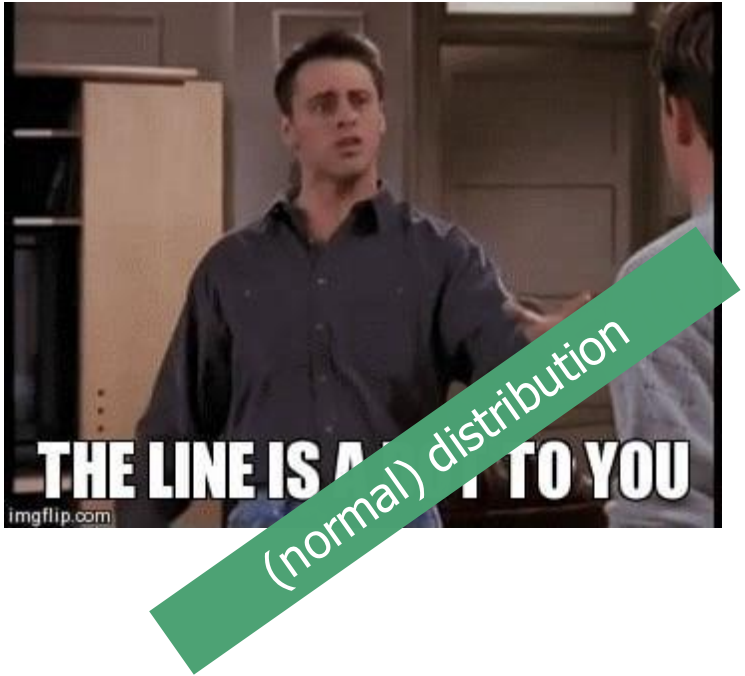


About 68% of values 1 standard deviation from mean

About 95% of values 2 standard deviations from mean

Bayesian models

- Parameter: sigma (& confidence interval)
  - = standard deviation of the posterior
  - Measures how spread out the observations are around the mean
- Important notes
  - “line” = mean value normal distributions of posterior predictions
  - We model posterior predictions because we don’t know what the actual mean of the distribution is
  - Shows (measures) uncertainty and confidence in predictions



Overview

1. Bayesian models
2. **Mixed effects**
3. Generalised linear models

Limitations of (generalised) linear models

- Major assumption of **data independence**
  - Datapoints are not related to each other, but independent
- >< What with dependent data?

- Repeated measurements:
  - E.g. testing the same subject under different circumstances
  - E.g. obtaining multiple responses per subject (e.g. to multiple stimuli)

  ➔ **measurements on the same subject are usually correlated, and this correlation must be taken into account!**

  ➔ **Test before you perform a regression analysis**

Repeated measurements

- Real-world examples?
  - Psycholinguistics: reaction times for one and the same subject on twenty different lexical stimuli (words)
  - Corpus linguistics: emoji count per WhatsApp utterance, for multiple utterances by the same subject
  - Diachronic studies: including measurements of the same subject at different points in time
  - Medicine: applying eye drops to each subject's left eye, then performing tests on both the left and right eye of each subject
  - …

Why should we take repeated measurements into account?

- Difference in the model:
  - Division of variability in the data
  - Model can be more precise

- Risks when ignoring repeated measures:
  - Inaccurate results: imprecise coefficients
  - Often "losing power" (sticking to H_0 even when H_1 is true)
  - Opposite can happen too: overestimating power

Why should we take repeated measurements into account?

Fictitious example:

- Psycholinguistic study on the effect of a night's sleep on memorizing words in a foreign language
  - 5 participants
  - Response = score on a vocabulary test
  - 2 conditions (time of testing):
    - **Day1**: no sleep, taking a test right after memorizing the words for the first time (same day)
    - **Day2**: one night's sleep, retaking the test the day after memorizing the words

Why should we take repeated measurements into account?

subject	day	score
1	day1	10
1	day2	11
2	day1	11
2	day2	12
3	day1	12
3	day2	13
4	day1	13
4	day2	14
5	day1	14
5	day2	10

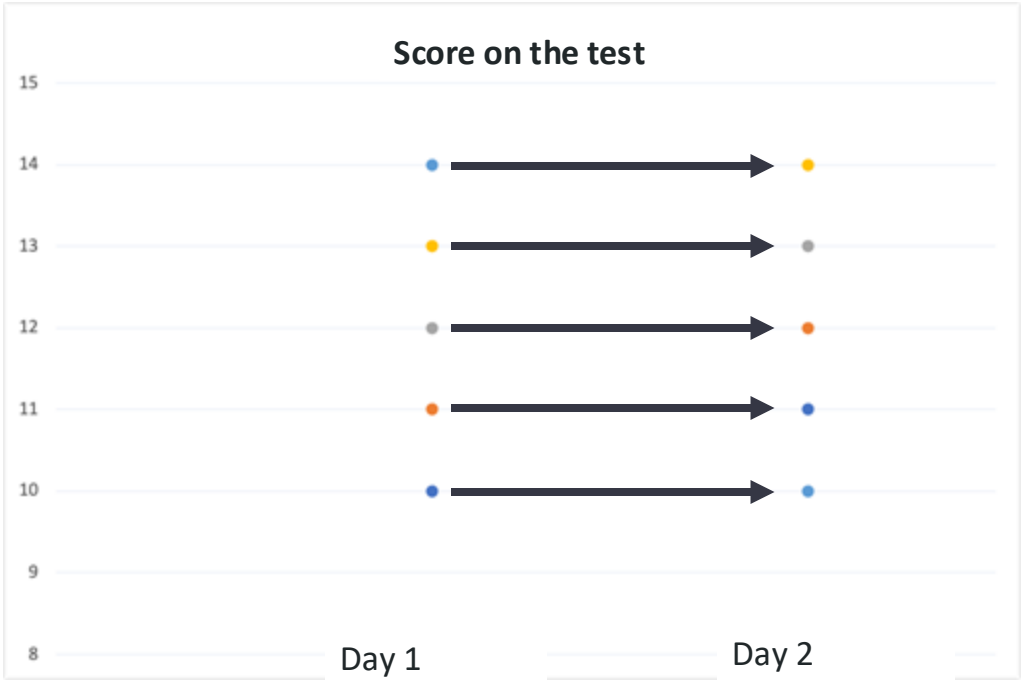
Are these datapoints independent? Why (not)?

Why should we take repeated measurements into account?

day	score
day1	10
day2	11
day1	11
day2	12
day1	12
day2	13
day1	13
day2	14
day1	14
day2	10

Let’s ignore the subject line… i.e. ignore the repeated measurements

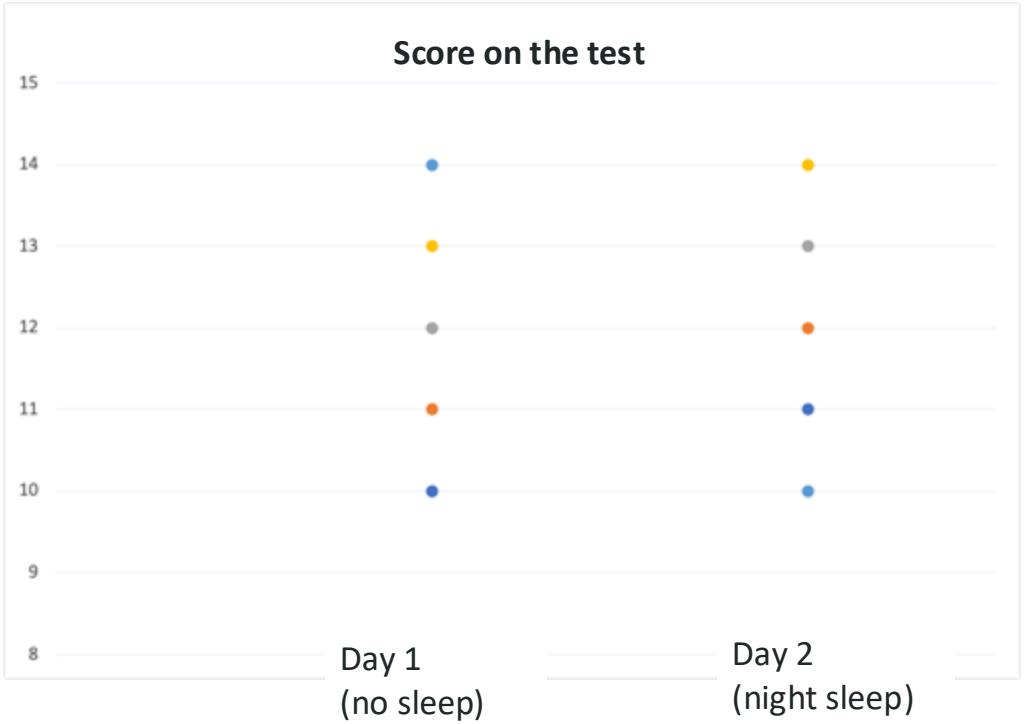
Why should we take repeated measurements into account?



What pattern might we assume, when ignoring repeated measurements (i.e. ignoring the colors of the dots)?

NO sleep effect (no increase/decrease in score on the test)
= H_0

Why should we take repeated measurements into account?



What pattern might we assume, when we DO take into account repeated measurements (i.e. the colors of the dots)?

Why should we take repeated measurements into account?

What pattern might we assume, when we DO take into account repeated measurements (i.e. the colors of the dots)?

General increase in score!
H_1

→ So ignoring data dependence here leads to **losing power!** (sticking to H_0 even though H_1 is more probable)

How to include repeated measures in Python?

- Repeated-measures ANOVA
  - Limitation: only categorical predictors
  - Not treated in this course

- **Mixed effects models / multilevel models**

  = Adaptations of the regression models you (will) know:

  - Linear **mixed effect** model
  - Poisson **mixed effect** model
  - Binary logistic **mixed effect** model

Mixed-effects models

- Include both **fixed** and **random** effects

- **Fixed** effect:
  - Levels cover all possible levels (in the population)
    - E.g. educational track, socio-economic class
  - Rule of thumb:
    - predictor with few levels
    - for which you assume a predictable effect (on theoretical grounds)

Mixed-effects models

- Include both **fixed** and **random** effects

- **Random** effect:
  - Levels do not cover all possible levels (in the population), but are a random sample
    - E.g. subject (participant), item (lexical stimulus)
  - Rule of thumb:
    - variable with many levels
    - for which you assume a rather unpredictable effect
      (no clear theoretical grounds >< e.g. individual variation)

What does the inclusion of a random effect in the model mean?

- Variability gets modeled in a more accurate way
- Different regression lines (or distributions) are fitted for different subjects/items
  - Random **intercepts**: Subject- or item-specific different intercepts
  - Random **slopes**: Subject- or item-specific different slopes

Overview

1. Bayesian models
2. Mixed effects
3. **Generalised linear models**

Limitations of linear models

- Dependent variable:
  - Numeric: interval-/ratio-scaled
  - Covers wide range of values
- Predicted values:
  - *-inf* to *+inf*

>< binary, categorical, frequencies/counts, …

>< predictions make no sense!

(e.g. negative numbers for book dimensions…)

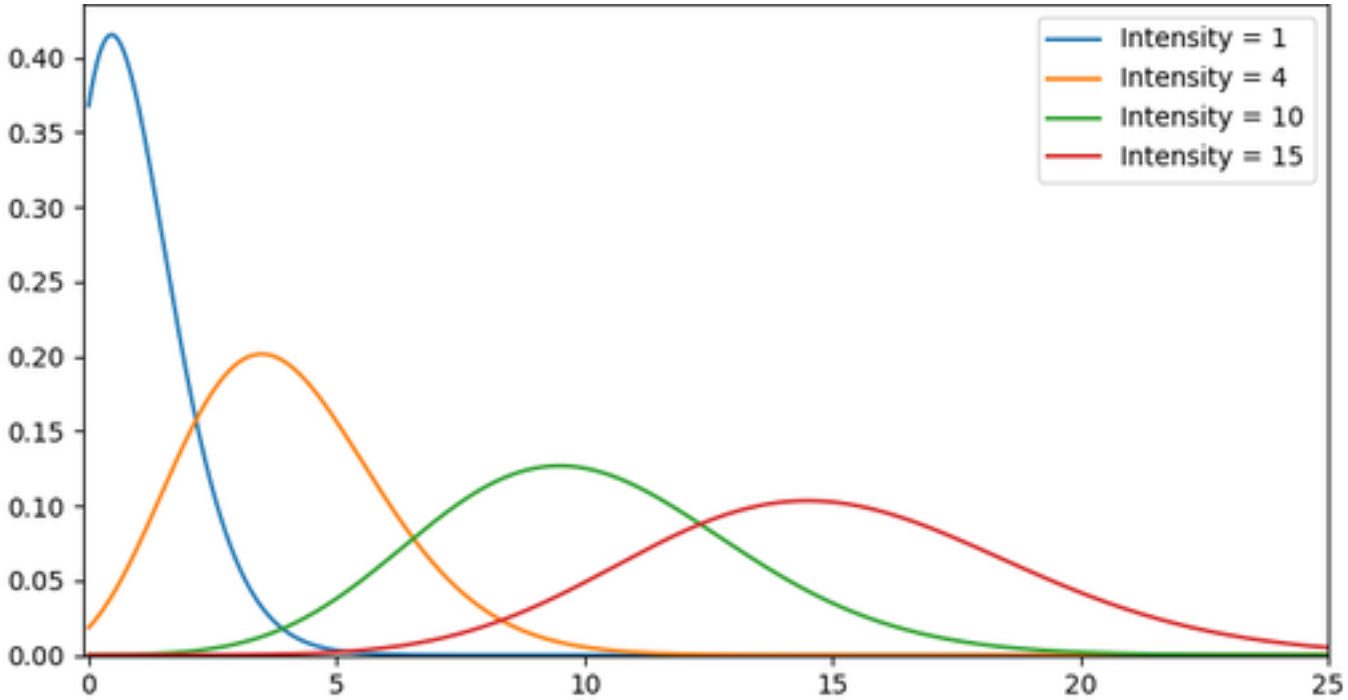
Generalised linear models

- Linear model is applied to these other types of dependent variables
- EXTRA STEP: *link function* transforms the output variable:
  - Predicted range of values transformed:
    (-inf to +inf) → something more appropriate

Types of generalized linear models

Type	Dependent variable	Link function
Binary logistic regression (Gries 5.3)	Binary	Inverse logit function (-inf to +inf) → 0 to 1 (probs)
Poisson/count regression (Gries 5.4.3)	Frequencies / counts	Exponential transformation (-inf to +inf) → 0 to +inf
Other: see Gries (5.3 – 5.4)		

Poisson distribution



Poisson (count) regression

- 'Generalised' case of linear regression
- Dependent variable = counts/freqs (integers, >0, often limited range
- Real-world examples?
  - Traffic: predict nr of accidents at a certain crossroads in a certain timeframe
  - Medicine: predict nr of people entering an emergency room per hour
  - Medicine: predict nr of covid cases per day in a city
  - Catering industry: predict nr of people dining in a restaurant per night
  - Spelling: predict how many errors a person will make on a spelling test
  - Politics: predict how many people of a certain area will cast their vote
  - …

Poisson (count) regression

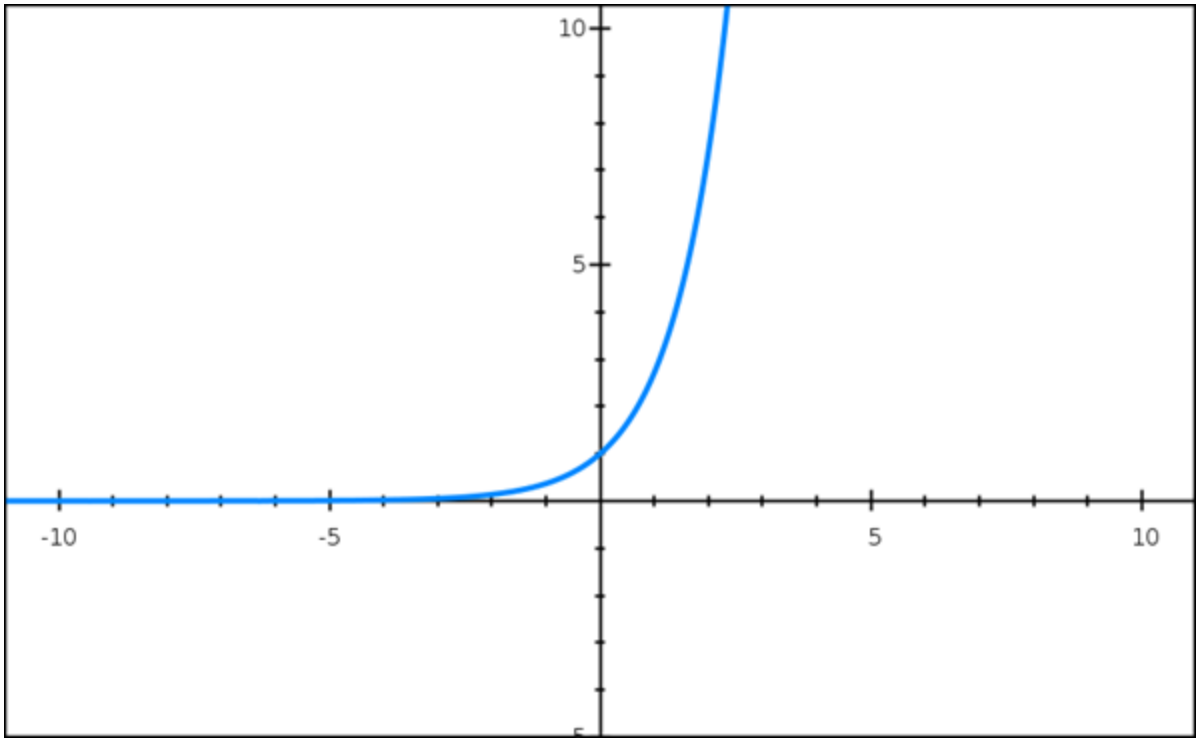
- Predictions from linear model TRANSFORMED to counts/frequencies (integers, >=0) via the exponential function:

$$e^{x}$$

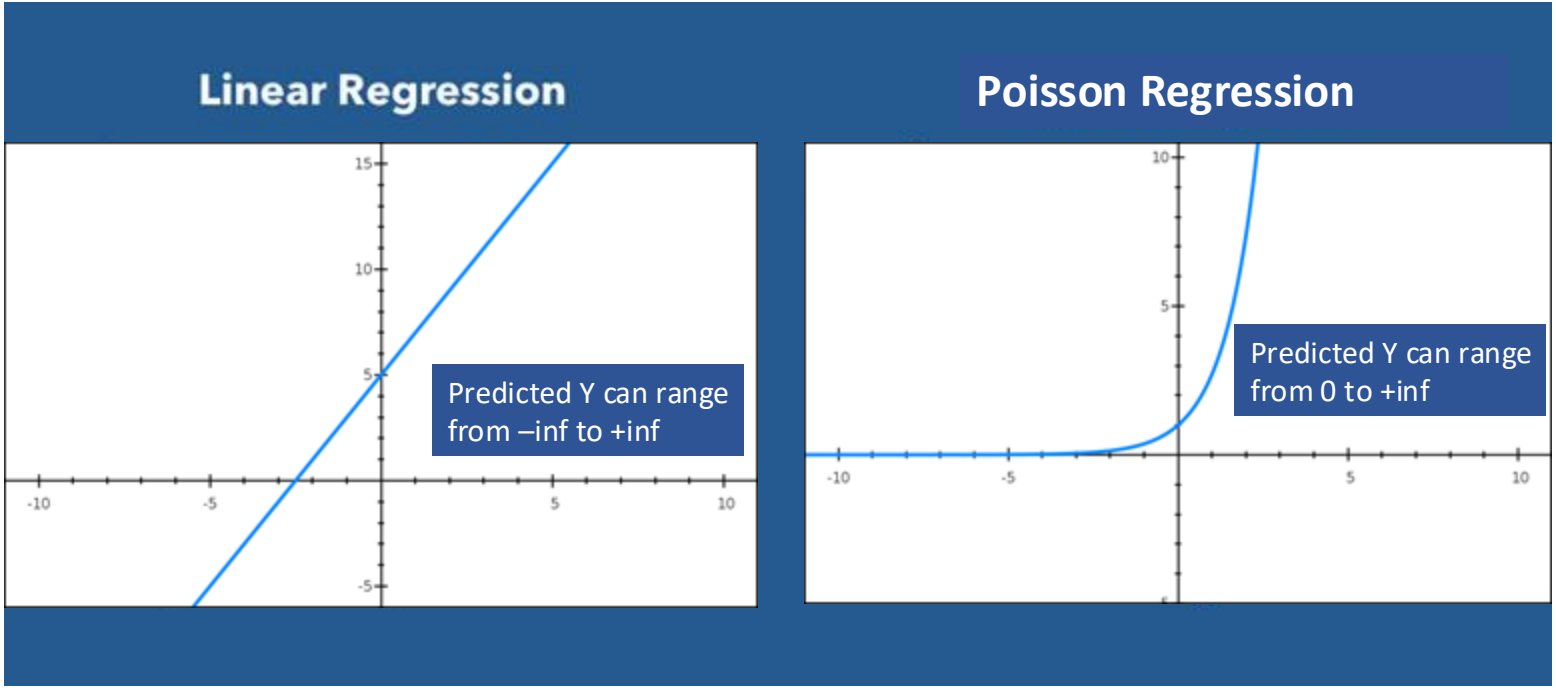
e = 2.7183… (Euler's number)

→ Outcome always positive

Exponential function: $f(x) = e^x$



Linear vs Poisson (count) regression: predicted



Assumptions of Poisson (count) regression

- Data independence
- No overdispersion:
  - = higher variability (dispersion) in the data than expected based on a given distribution (= Poisson)
- No outliers? (disputed!)

Types of generalized linear models

Type	Dependent variable	Link function
Binary logistic regression (Gries 5.3)	Binary	Inverse logit function (-inf to +inf) → 0 to 1 (probs)
Poisson/count regression (Gries 5.4.3)	Frequencies / counts	Exponential transformation (-inf to +inf) → 0 to +inf
Other: see Gries (5.3 – 5.4)		

Binary logistic regression

- ‘Generalised’ case of linear regression
- Dependent variable = binary
- Real-world examples?
  - Spam detection: predict if email is spam / not spam
  - Medicine: predict if a tumor is benign / malignant
  - Finance: predict if a person will get a loan / not
  - Spelling: predict if a person will make a spelling error / not
  - Politics: predict if a candidate will win the election / not
  - Education: predict if a students gets admitted to a university / not
  - …

Binary logistic regression

- Predictions from linear model TRANSFORMED to probalities (no binary outcome) via inverse logit function:

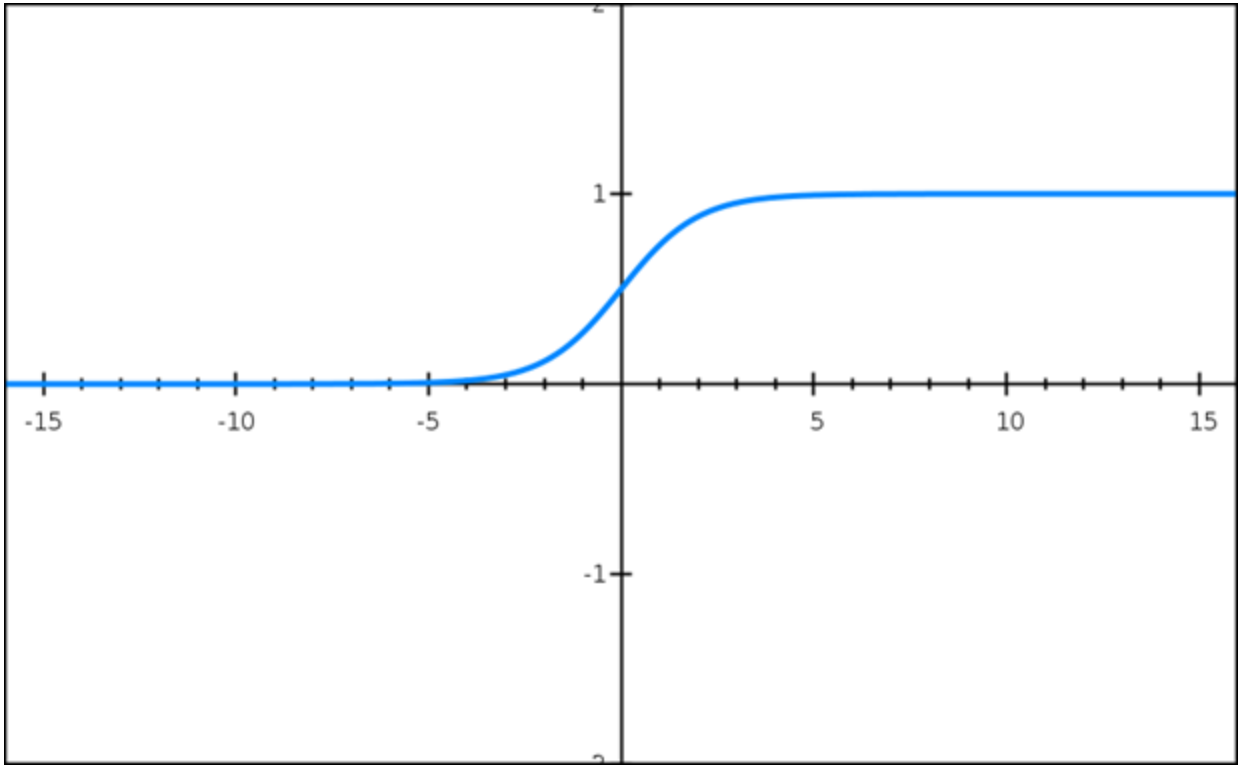
$$\frac{1}{1+e^{-x}} \quad \text{or} \quad \frac{e^{x}}{1+e^{x}}$$

e = 2.7183… (Euler's number)

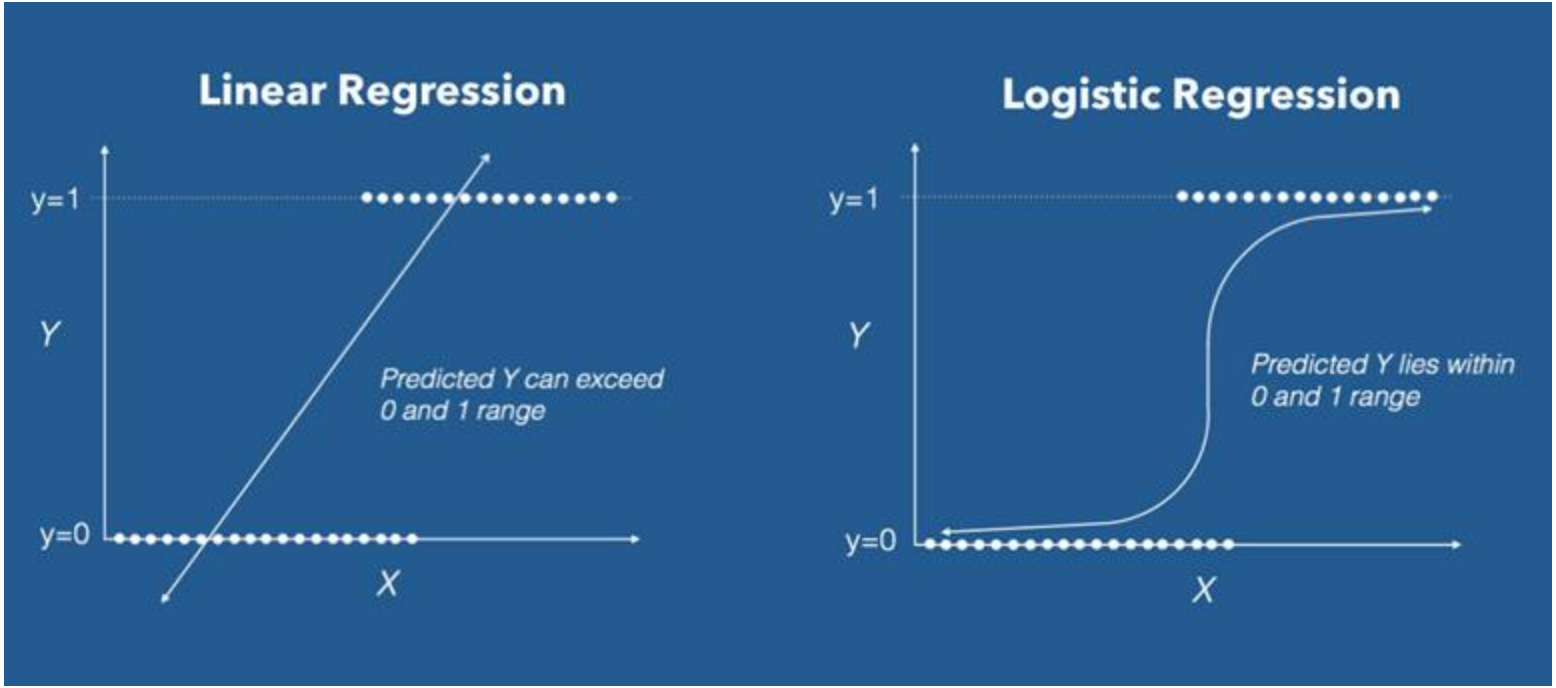
→ Outcome always between 0 and 1 ➔ to be interpreted as probabilities

Inverse logit of x

= sigmoid function/curve



Linear vs logistic regression: predicted



Source: datacamp.com

Assumptions of binary logistic regression

- Data independence
- No overdispersion: (disputed!)
  - = higher variability (dispersion) in the data than expected based on a given distribution (here: binomial)

The ultimate workflow for statistic testing

1. Look at data and mess around with visualisations
2. Postulate dependent and independent variable(s), H1 and H0 for your research
3. Look up assumptions (and H0) relevant test(s)
4. Test the assumptions
5. Report outcome test, measure of effect (and p-value)
6. Interpret results and reject/accept H1/H0
7. Repeat until it makes sense (no results are also results!)