Statistics in experimental research

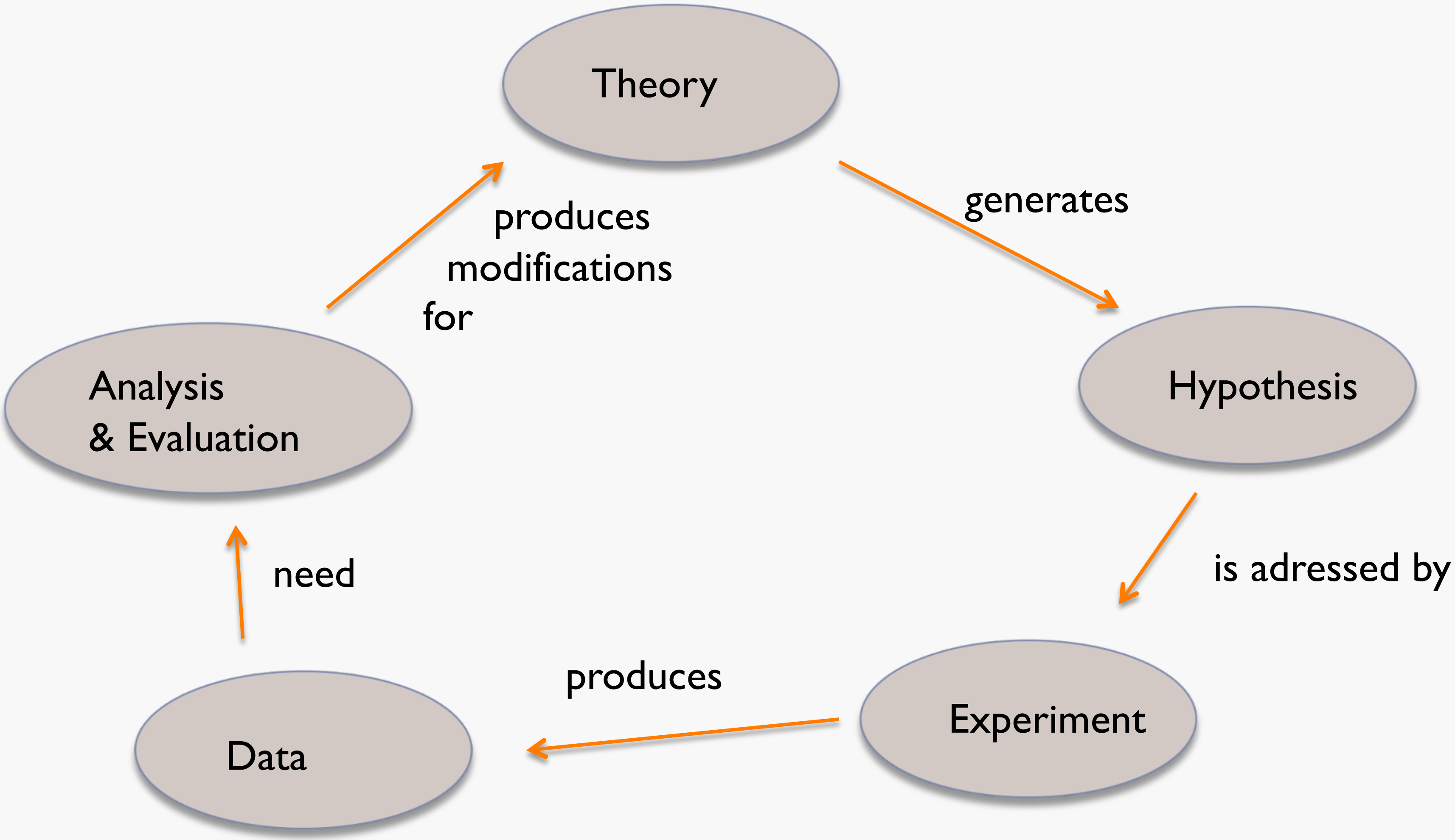
Session 1

Francesca Delogu
delogu@coli.uni-saarland.de

Key concepts

- Experimental variables
- Statistical hypotheses
- Sampling distribution
- Statistical tests (t, chi-square, ANOVA)
- p-value
- α level
- Type I and type II errors

The research cycle



Overview

- Today:
  - Statistical hypothesis testing
    - Sampling, distributions, standardized scores
  - One Sample z-test and t-test
- Thursday:
  - Different data types
  - Two-samples t-test, χ^2 Test
- Friday:
  - Design
  - ANOVAs

Let's start with an example

- Suppose we believe for some reason that caffeine improves cognitive performances of people (e.g., memory, attention)

- How can we find evidence for theory?

- We need a working hypothesis, also called
  - research hypothesis
  - experimental hypothesis
  - alternative hypothesis

The research hypothesis

Example:

- Caffeine improves cognitive performances
  - People are better at recalling a text after a cup of coffee
  - Students are faster to finish their homework after a cup of coffee

- A research hypothesis makes a prediction about the relationship between two (or more) variables

Variables

- Properties of objects or events that can take on different values (hair color, height, length, speed, etc.)
- The goal of an experiment is to measure the effects of one variable on another one
- **Independent variable (IV)**

  the variable manipulated by the experimenter in order to assess whether it has an effect on …
- **Dependent variable (DV)**

  the variable that is measured (the data) in order to assess whether it is influenced by the independent variable

Experimental variables

- What are the experimental variables in our example?

Students are faster to finish their homework after a cup of coffee

The null hypothesis

- The research hypothesis is always tested against a **null hypothesis (H_0)**
- The null hypothesis states that the IV has no influence on the DV
- What is the null hypothesis in our example?

The statistical hypotheses

- Null hypothesis (H_0)
  - Coffee has no influence on time to finish homework
    - Time to finish homework does not differ as a function of whether students have or not a cup of coffee
- Alternative hypothesis (H_1)
  - Coffee has an influence on time to finish homework.
    - Two-tailed hypothesis (neutral with respect to the direction of the effect)
  - Coffee makes students faster to finish homework
    - One-tailed hypothesis (makes a guess on the direction of the effect)

Goal of hypothesis testing

- Decide between H_0 and H_1

- The main goal of hypothesis testing is to assess whether we have enough evidence to *reject* H_0
  - i.e., if the data are inconsistent with H_0

- NB: the starting point of any statistical test is the H_0
  - Statistical tests find the probability of obtaining your data if the null hypothesis were true

  - If this probability is low enough→ reject H_0
  - If the probability is not enough low → fail to reject H_0

Collecting data

Does coffee have an influence on time to finish homework?

- Two options:
  - look at all students and every instance where they finished their homework (impossible)
  - look at subsets of students doing one particular homework with and without coffee

Population *vs.* Sample

Population

Complete set of events we are interested in:
All students
All existing words
All speakers of German
All translations from your system

Sample

the subset of events you look at

- Samples of observations are randomly drawn from a population and used to infer something about the characteristics of the population

A bit of terminology

- **Parameter:** a numerical value summarizing population data (e.g., the population mean μ)
- **Statistic**: a numerical value summarizing sample data (e.g., the sample mean $\overline{X}$)
- **Inferential statistics:** statistical procedures aimed at drawing inferences about parameters from statistics
- **Random sample:** a sample where each member of the population has an equal chance of being selected
- **Sampling error:** variability of a statistic from sample to sample due to chance (the difference between the statistic and the parameter)

Mean, Variance, Standard Deviation

	Population	Sample
Mean	$\mu = \frac{\Sigma X}{N}$	$\overline{X} = \frac{\Sigma X}{N}$
Variance	$\sigma^2 = \frac{\Sigma(X-\mu)^2}{N}$	$s^2 = \frac{\Sigma(X-\overline{X})^2}{N-1}$
Standard Deviation	$\sigma = \sqrt{\sigma^2}$	$s = \sqrt{s^2}$

Variance and Standard Deviation are measures of variability (dispersion), i.e., the degree to which individual data points are distributed around the mean

Case 1- μ and σ known

- Suppose we are interested only in Master students from Coli (our population)
- Students from Coli take on average 30 min to finish homework without any coffee ($\mu = 30$, $\sigma = 4$)
- We want to know whether students are faster (or slower) after a cup of coffee

Statistical hypotheses

H_0: Coffee has no influence on time to finish homework

μ=30 (even with coffee)

H_1: Coffee has an influence on time to finish homework

μ≠30 (if coffee is given)

Data

Sample_Coffee	**Time (min)**
Anna	35
Jim	25
John	28
Mary	25
Peter	19
Carl	31
Judy	18
Bob	30
Liz	26
Betty	23

$Sample_{coffee}$

$\overline{X} = 26$

$S = 5.27$

$Population_{no\_coffee}$

$\mu = 30$

$\sigma = 4$

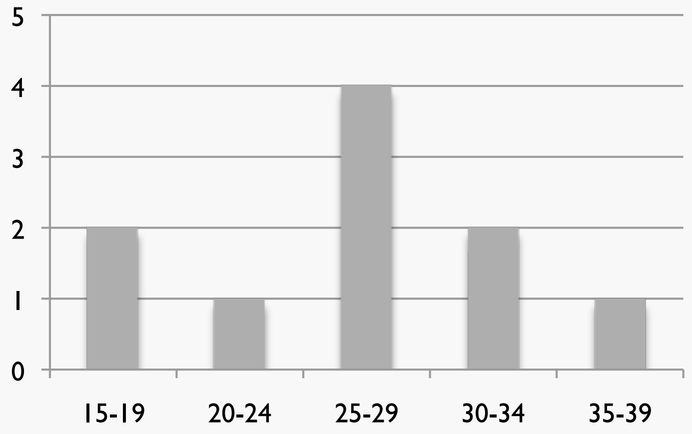
H_0: the deviation from μ is due to sampling error, the sample was drawn from a population with μ=30

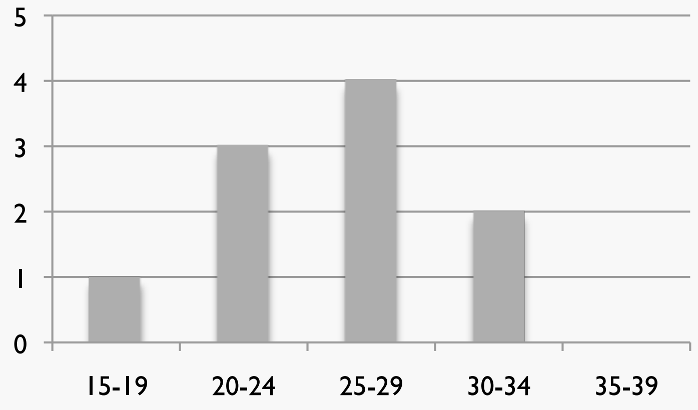
H_1: the deviation from μ is sufficiently large to conclude that the sample was drawn from a population with a different mean

Distribution of data

Sample1	Sample2
35	32
25	28
28	28
25	20
19	29
31	18
18	25
30	20
26	35
23	23
$\overline{X}_1 = 26$	$\overline{X}_2 = 25.6$

histogram





Sampling distribution

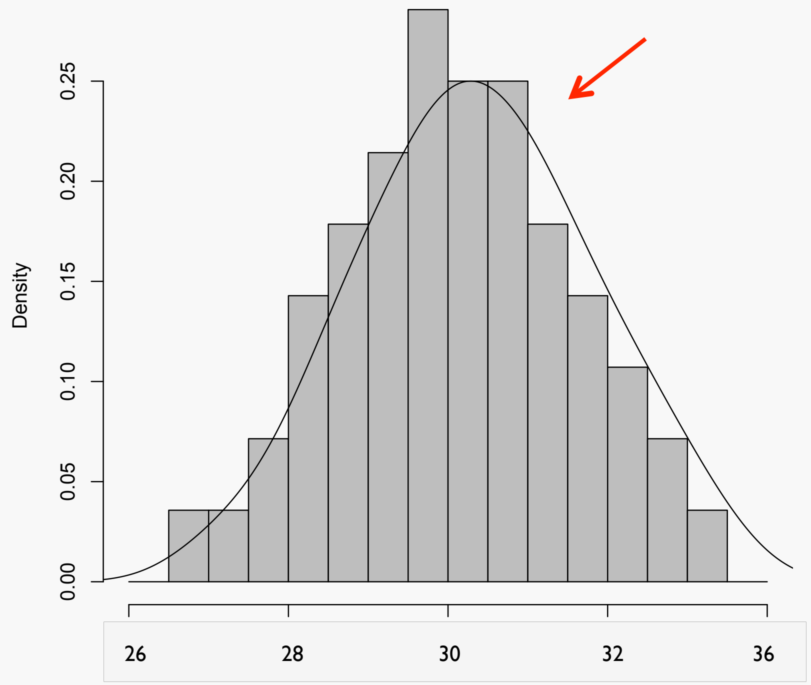
Samples (N=10)	Means
1	26
2	26.5
3	28
4	32
5	30.5
6	33
6	29.6
7	35
8	27
…	…
	$\mu_{\overline{X}} = \mu$

the distribution of a statistic over repeated sampling from a specific population

Sampling distribution of the mean

$$\mu_{pop} = 30$$
$$\sigma_{pop} = 4$$

$$\mu_{\bar{x}} = \mu_{pop} \qquad \sigma_{\bar{x}} = \frac{\sigma_{pop}}{\sqrt{N}}$$



sampling distribution when H_0 is true

- If we know what the sampling distribution looks like when H_0 is true, **we know what sample means are more or less likely to be obtained under H_0**
- We can use this information to assess whether $\bar{X} = 26$ could reasonably have arisen **had we drawn the sample from a population in which μ=30 (i.e., if H_0 is true)**

Central limit theorem

Given a population with mean μ and standard deviation σ:

- The mean of the sampling distribution of the mean is equal to the mean of the source population

$$\mu_{\bar{x}} = \mu$$

- The standard deviation of the sampling distribution of the mean (*standard error*) is equal to the standard deviation of the source population divided by the square root of the sample size

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$$

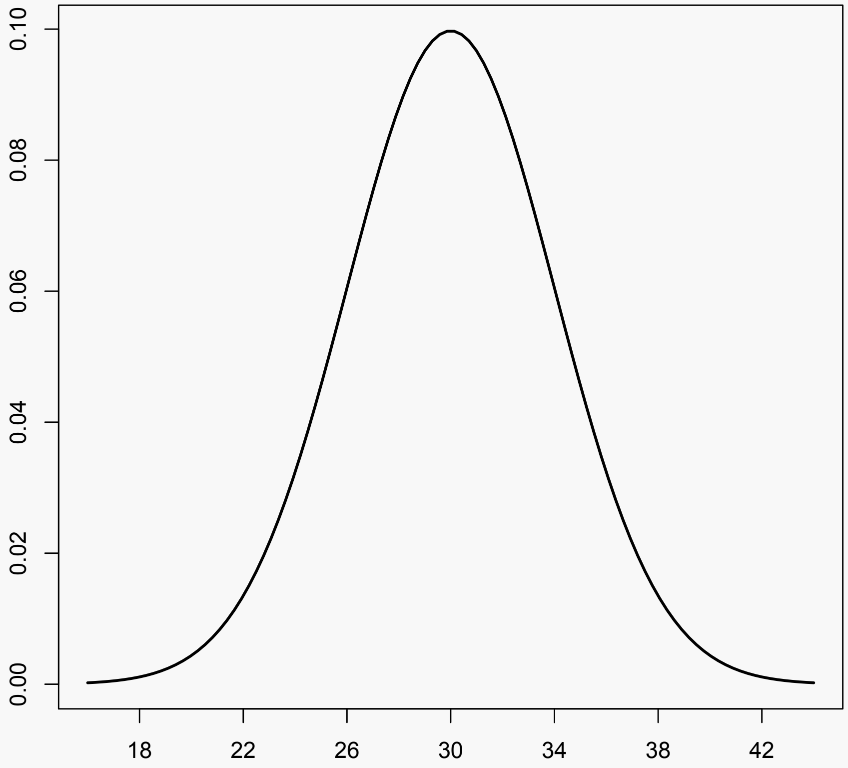
- **The sampling distribution of the mean will approach a normal distribution as the size N of the samples increases**

The normal distribution

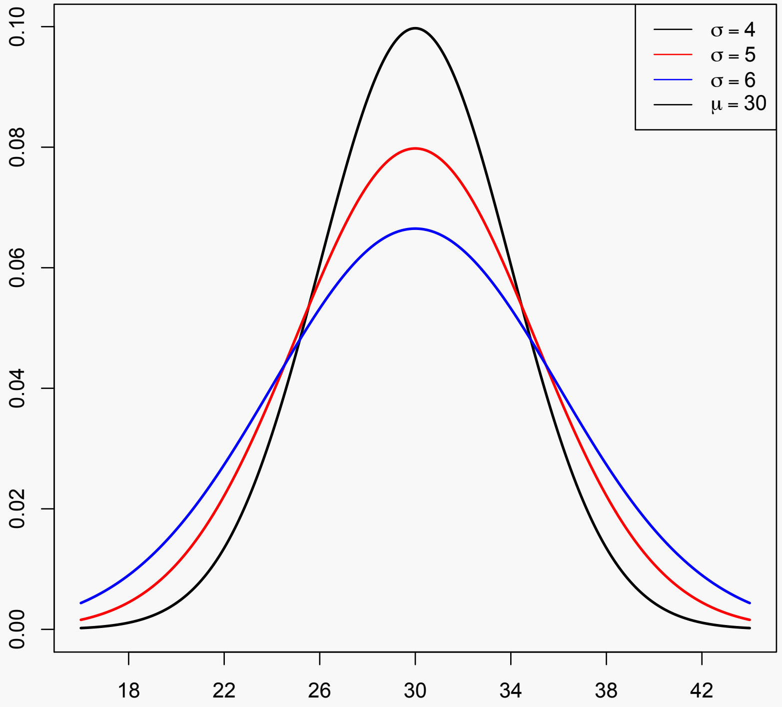
- A probability density function, symmetrical about the mean, bell-shaped, described by μ and σ

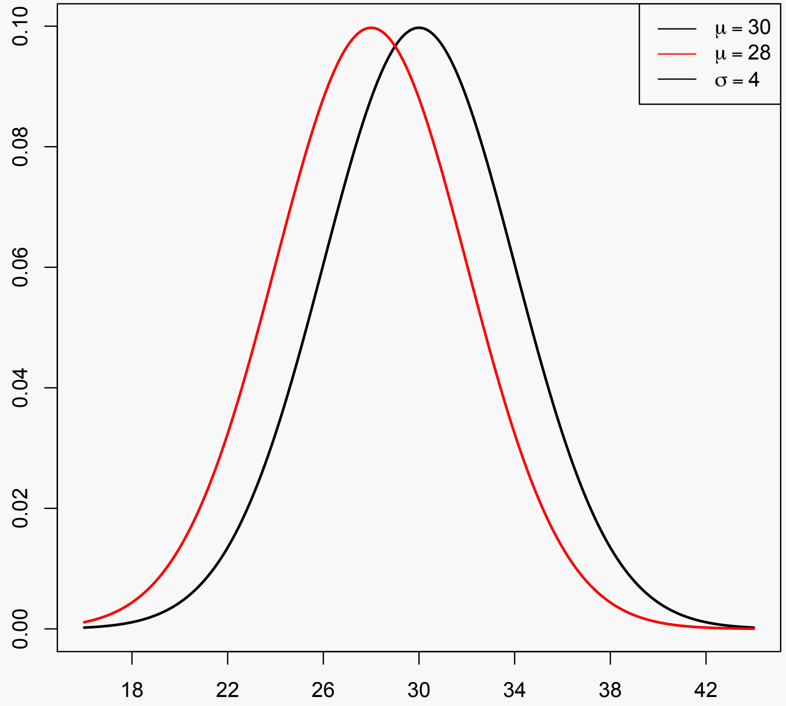
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$-\infty < x < +\infty$$

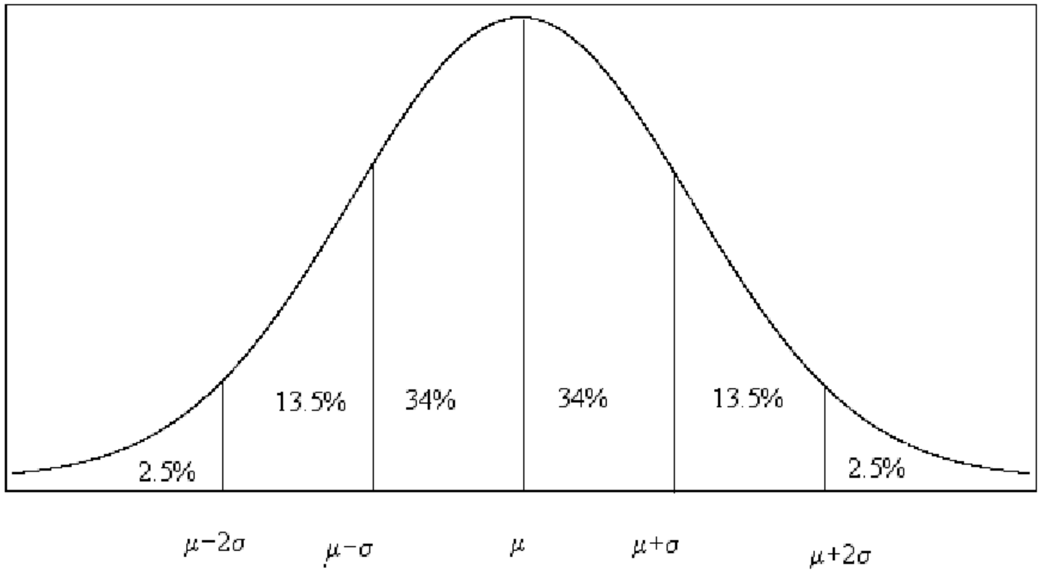


The normal distribution



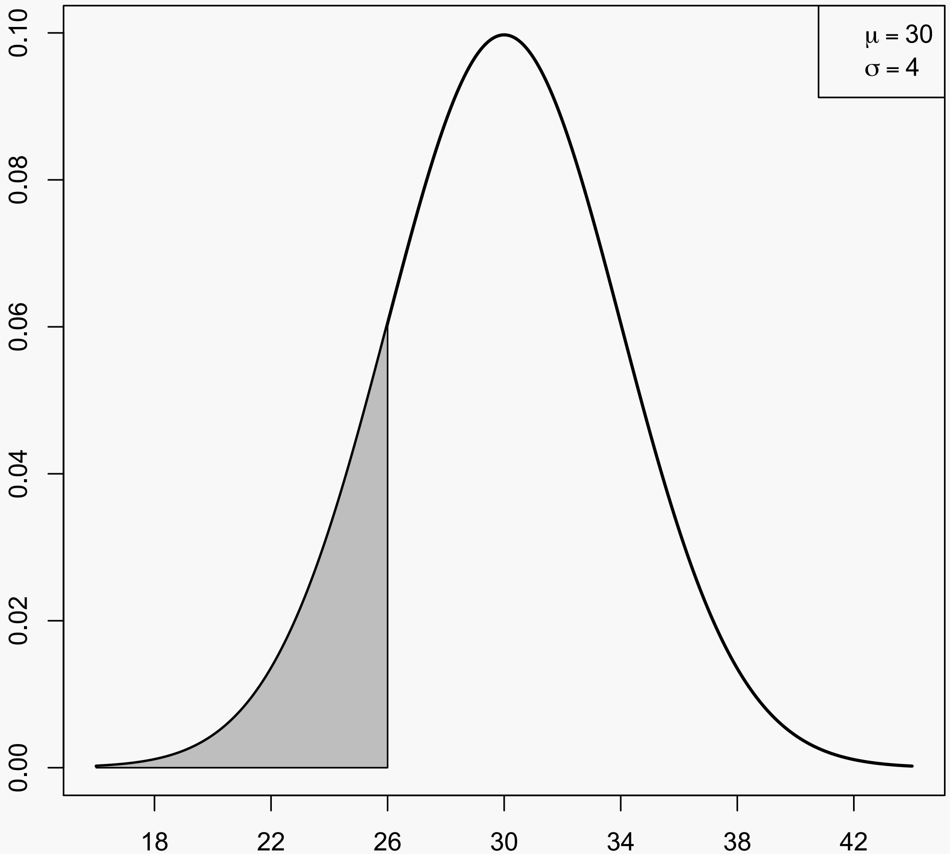


The normal distribution



$$P(\mu - \sigma \le X \le \mu + \sigma) \approx 0.68 \quad P(\mu - 2\sigma \le X \le \mu + 2\sigma) \approx 0.95$$

Standardizing



$X - \mu$:	-12	-8	-4	0	4	8	12
$(X - \mu)/\sigma$:	-3	-2	-1	0	1	2	3

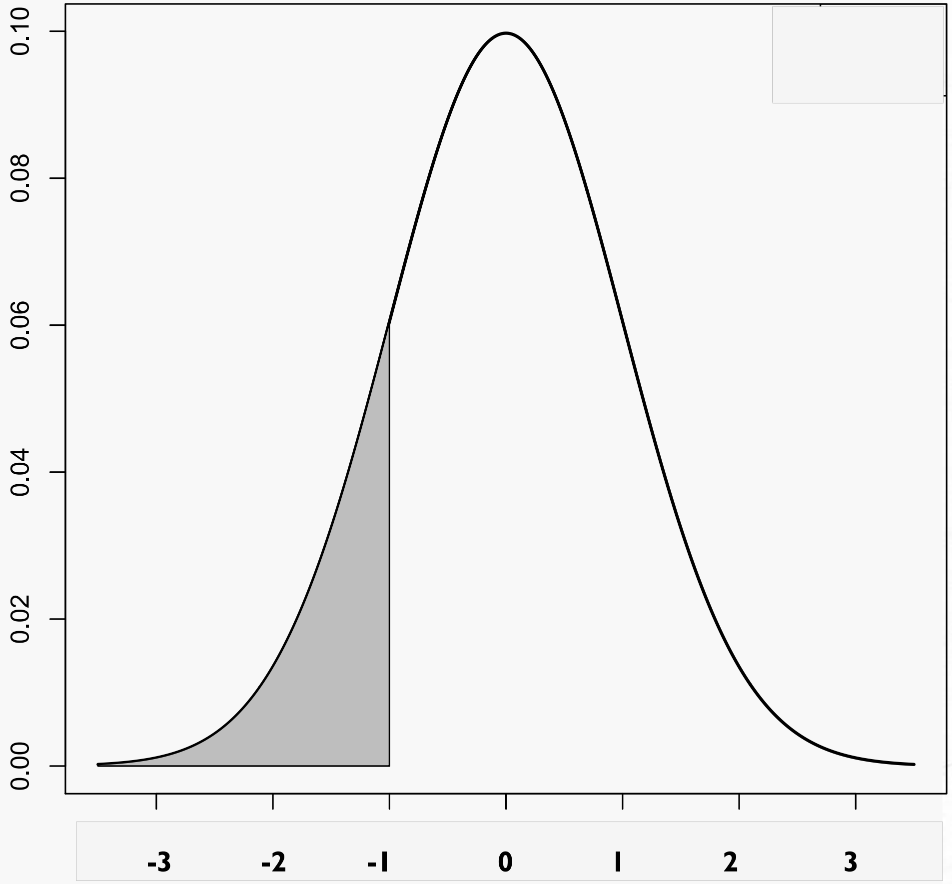
$$P(x \le 26) = ?$$

How many standard deviations is 26 distant from the mean?

$$z = \frac{X - \mu}{\sigma}$$

$$z = \frac{26 - 30}{4} = -1$$

The standard normal distribution



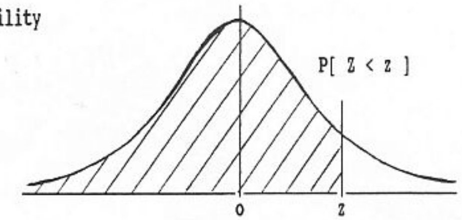
$$P(z \leq -1) = 0.1587$$

STANDARD STATISTICAL TABLES

1. Areas under the Normal Distribution

The table gives the cumulative probability up to the standardised normal value z
i.e.

$$P[\, Z < z \,] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}Z^2)\, dZ$$



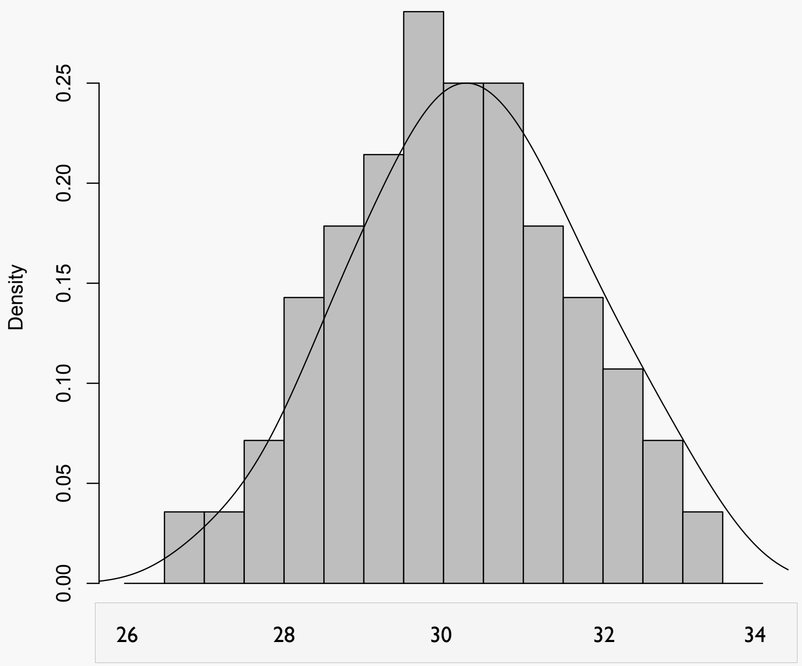
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7854
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8804	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9980	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

z	3.00	3.10	3.20	3.30	3.40	3.50	3.60	3.70	3.80	3.90
P	0.9986	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000

The z test – one sample mean

$$\mu_{\bar{X}} = \mu = 30$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}} = \frac{4}{\sqrt{10}} = 1.26$$



- From the central limit theorem we know the sampling distribution of the mean is normally distributed
- What is the probability of a sample mean ≤ 26min?

$$z = \frac{X - \mu}{\sigma_{\bar{X}}} = \frac{X - \mu}{\frac{\sigma}{\sqrt{N}}} = \frac{26 - 30}{1.26} = -3.17$$

$$P(z \leq -3.17) < 0.001 \text{ (one-tailed)}$$

$$P(z \leq -3.17 \text{ or } \geq 3.17) < 0.002 \text{ (two-tailed)}$$

Recap

- We are testing the effect of coffee on time to finish homework by giving a cup of coffee to a sample of 10 students from Coli and recording the time they take to finish a particular homework
- The average time to finish homework is 26min (S=5.27)
- We know that the mean time for Coli students without influence from coffee is 30min (σ =4)

Recap

H_0: μ =30 (even with coffee)

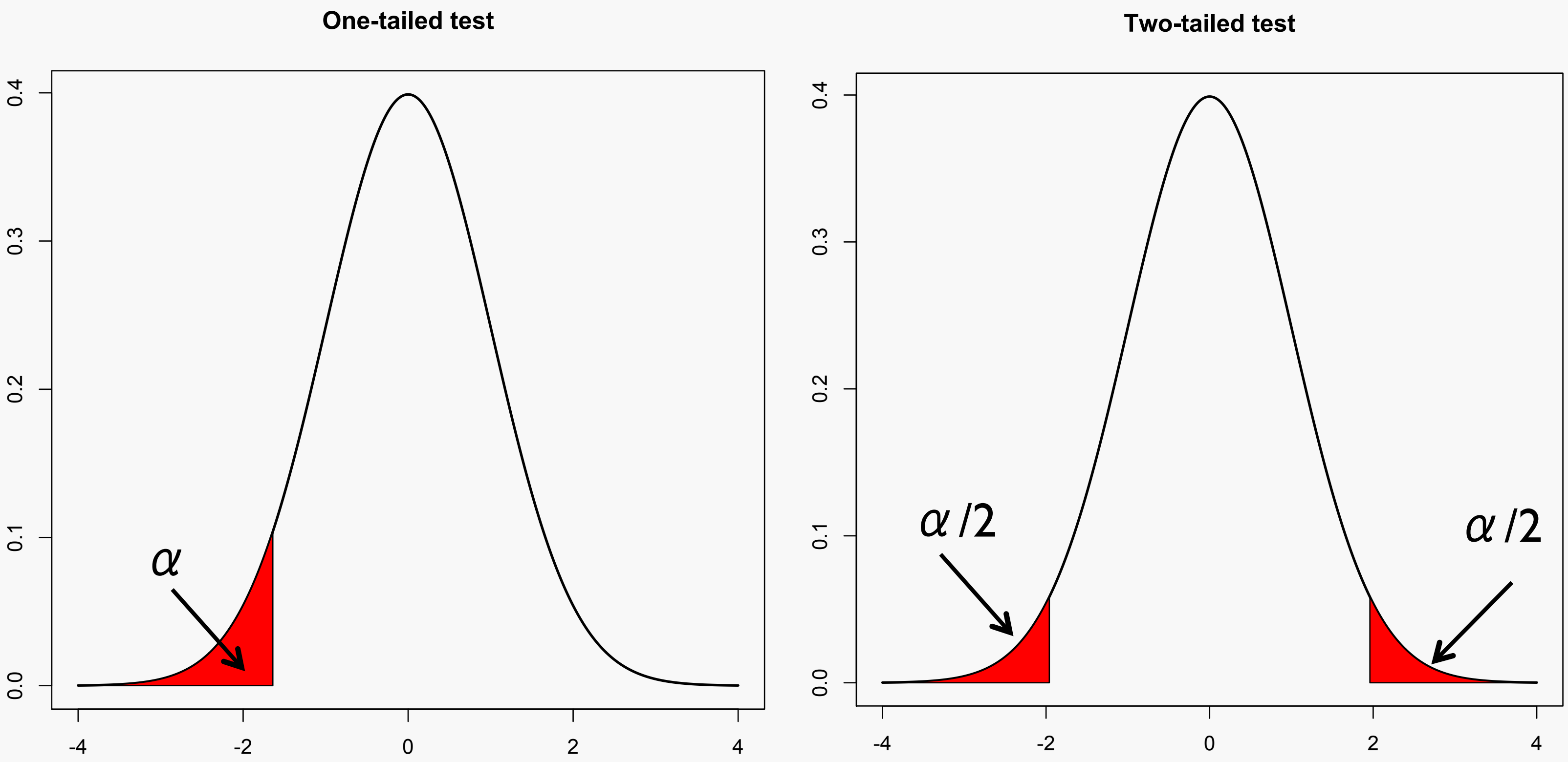
H_1: $\mu \neq 30$ (when coffee is given)

- Assuming that H_0 is true, the probability of obtaining a sample mean **at least as extreme** as 26min is less than 0.002
- Because this probability is very low, it is likely that our sample was drawn from a population with a different mean→ we'll reject the null hypothesis
- What if the probability was 0.08?

Significance level: α

- Alpha (α) is a conventional cutoff value representing the probability with which we are willing to reject H_0 when it is, in fact, correct.

α levels commonly used: 0.05, 0.01



Rejection criterion

- Whenever the probability obtained under H_0 is less than or equal to our predetermined significance level, we will reject H_0

- In our previous example, the probability of obtaining a sample mean at least as extreme as 26min under the null hypothesis ($\mu=30$) was $0.002 < \alpha=0.05$ (two-tailed test)

- Thus we'll reject the null hypothesis!

Exercise

A sample of size 50 is taken from a normal distribution, with a known population standard deviation of 26. The sample mean is 167.02. Use the 0.05 significance level to test the claim that the population mean is greater than 170.

Testing a sample mean when σ is unknown

- We rarely know the standard deviation of the population and usually will have to estimate it by way of the *sample standard deviation* (*S*)
- If we want to test hypotheses when σ is unknown, the nature of the test we'll be using depends on the size of the sample

For N > 30

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{N}}} \approx \frac{\bar{X} - \mu}{\frac{S}{\sqrt{N}}}$$

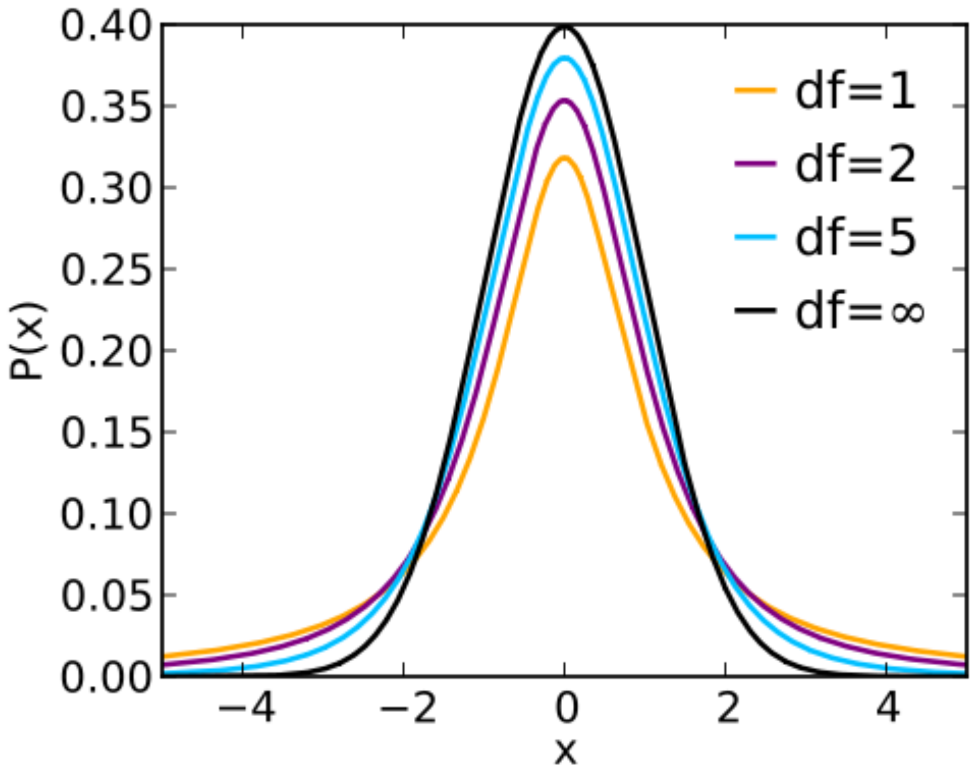
z is normally distributed

For N < 30

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{N}}}$$

t follows the t-Student distribution

The t-distribution



For a one-sample t-test, *df* = N-1

- The t-distribution is the sampling distribution of the *t* statistics
- The shape of the distribution depends on the number of **degrees of freedom** (***df***)
- As **df** go to infinity, the t-distribution converges to the standard normal distribution.

Degrees of freedom

- The number of *df* is a function of both the number of observations and the number of parameters estimated

- It is the number of values in the final calculation of a statistic that are free to vary

- Imagine you have four numbers (a, b, c and d) that must add up to a total of m; you are free to choose the first three numbers at random, but the fourth must be chosen so that it makes the total equal to m - thus your *df* is three.

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{N}}}$$

$$S = \sqrt{\frac{\Sigma(X - \bar{X})^2}{N - 1}} \searrow$$

$$\Sigma(X - \bar{X}) = 0$$

only N-1 of deviations are free to vary because

One sample t-test

$H_0: \mu = 30$ $\qquad \bar{X} = 26$

$H_1: \mu \neq 30$ $\qquad S = 5.27$

Sample_Coffee	Time
Anna	35
Jim	25
John	28
Mary	25
Peter	19
Carl	31
Judy	18
Bob	30
Liz	26
Betty	23

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{N}}} = -2.39$$

t-critical

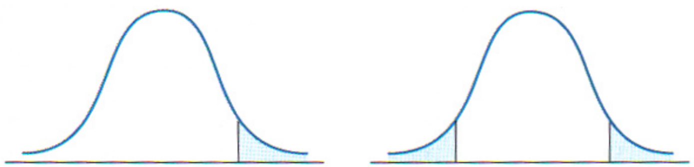
- Once we have computed the t-value, we want to know whether the probability of observing t under the null hypothesis is less than or equal to our level of significance α (0.05)

- i.e., we want to find the value that t must exceed in order for the null hypothesis to be rejected

- This value is called t-critical and depends on
  - The number of df
  - Whether H_1 is one-tailed or two-tailed

TABLE B.2 THE *t* DISTRIBUTION

Table entries are values of *t* corresponding to proportions in one tail or in two tails combined.



One tail (either right or left) | Two tails combined

	PROPORTION IN ONE TAIL					
	0.25	0.10	0.05	0.025	0.01	0.005
	PROPORTION IN TWO TAILS COMBINED					
df	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.319	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
40	0.681	1.303	1.684	2.021	2.423	2.704
60	0.679	1.296	1.671	2.000	2.390	2.660
120	0.677	1.289	1.658	1.980	2.358	2.617
∞	0.674	1.282	1.645	1.960	2.326	2.576

$$t_{\alpha}(9) = 2.262$$

$$t = -2.39$$

$$|t| > t_{\alpha}$$

We'll reject H_0

The result is statistically significant!

$$t(9) = -2.39,\ p < .05$$

Errors

- A significant result does not mean that H_1 is true beyond any doubts doubts
- If α =.05, you have a 5% chance of rejecting H_0 when it is in fact true
  - Type I error (false positive)
- A null result does not mean that H_0 is true
  - Type II error (false negative)

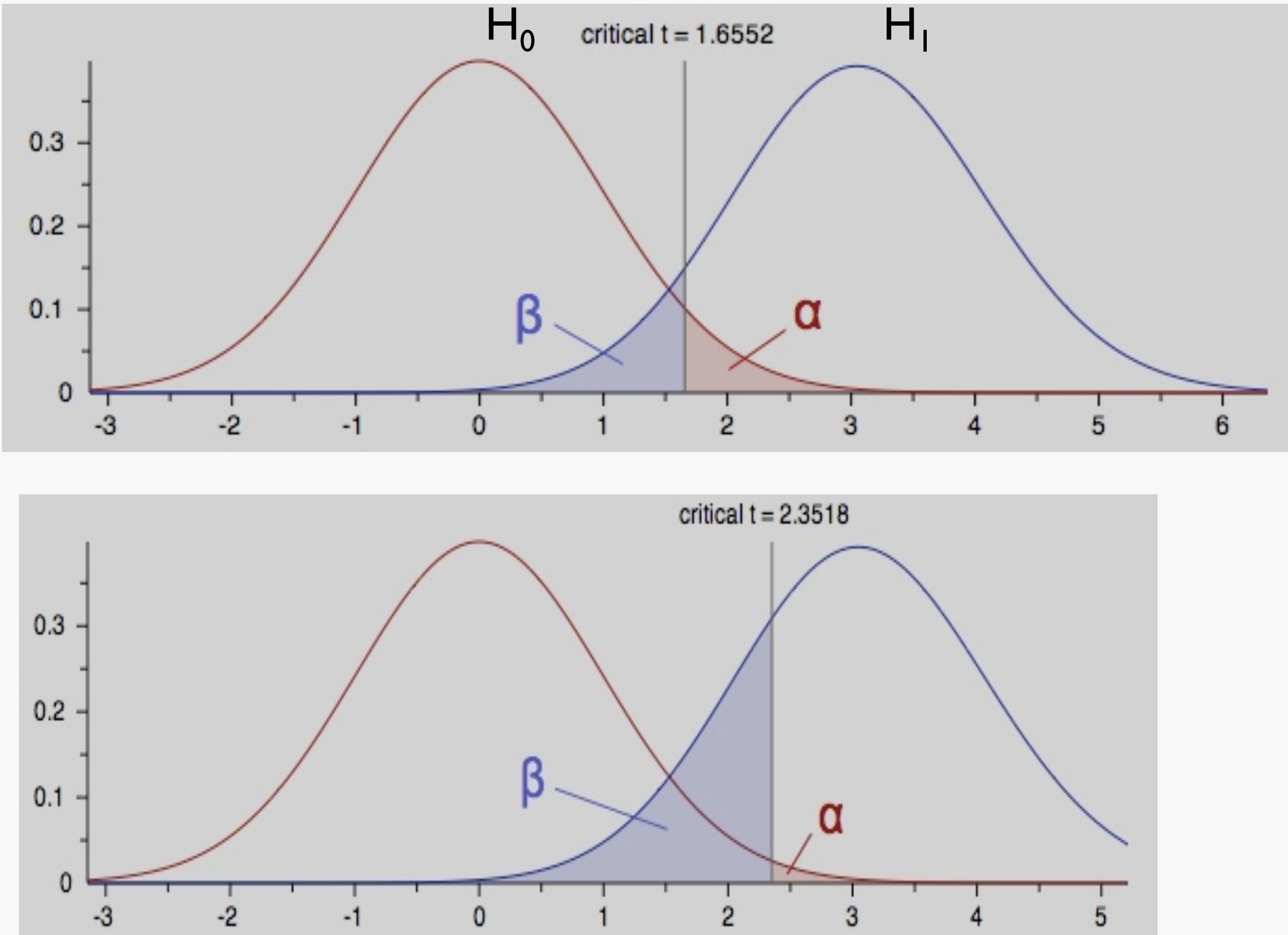
Table of errors

	H_0 True	H_0 False
Reject H_0	Type I Error	Correct Rejection
Fail to Reject H_0	Correct Decision	Type II Error

α is the probability of committing a Type I error

β is the probability of committing a Type II error

Relationship between α and β



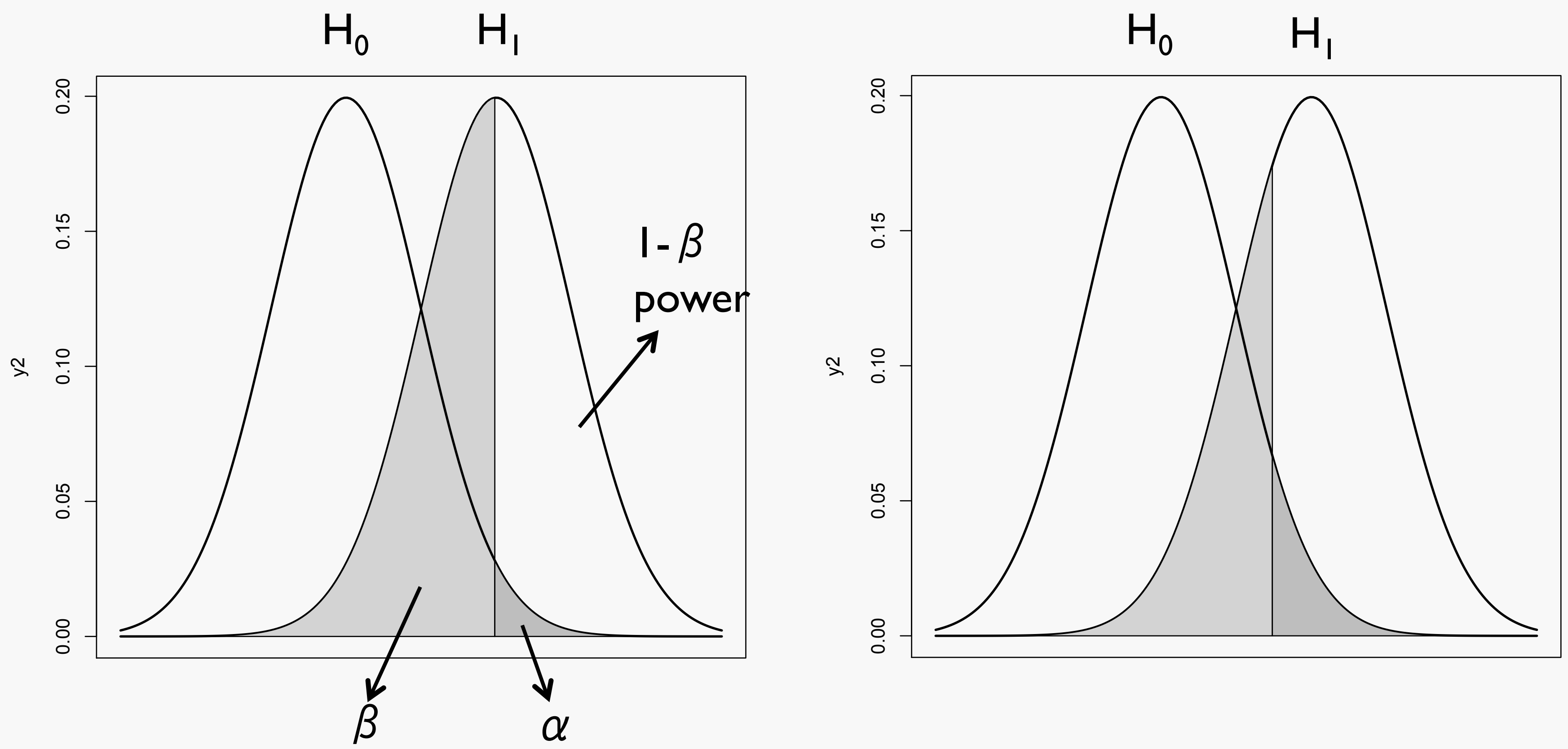
Statistical power

- Power = 1- β

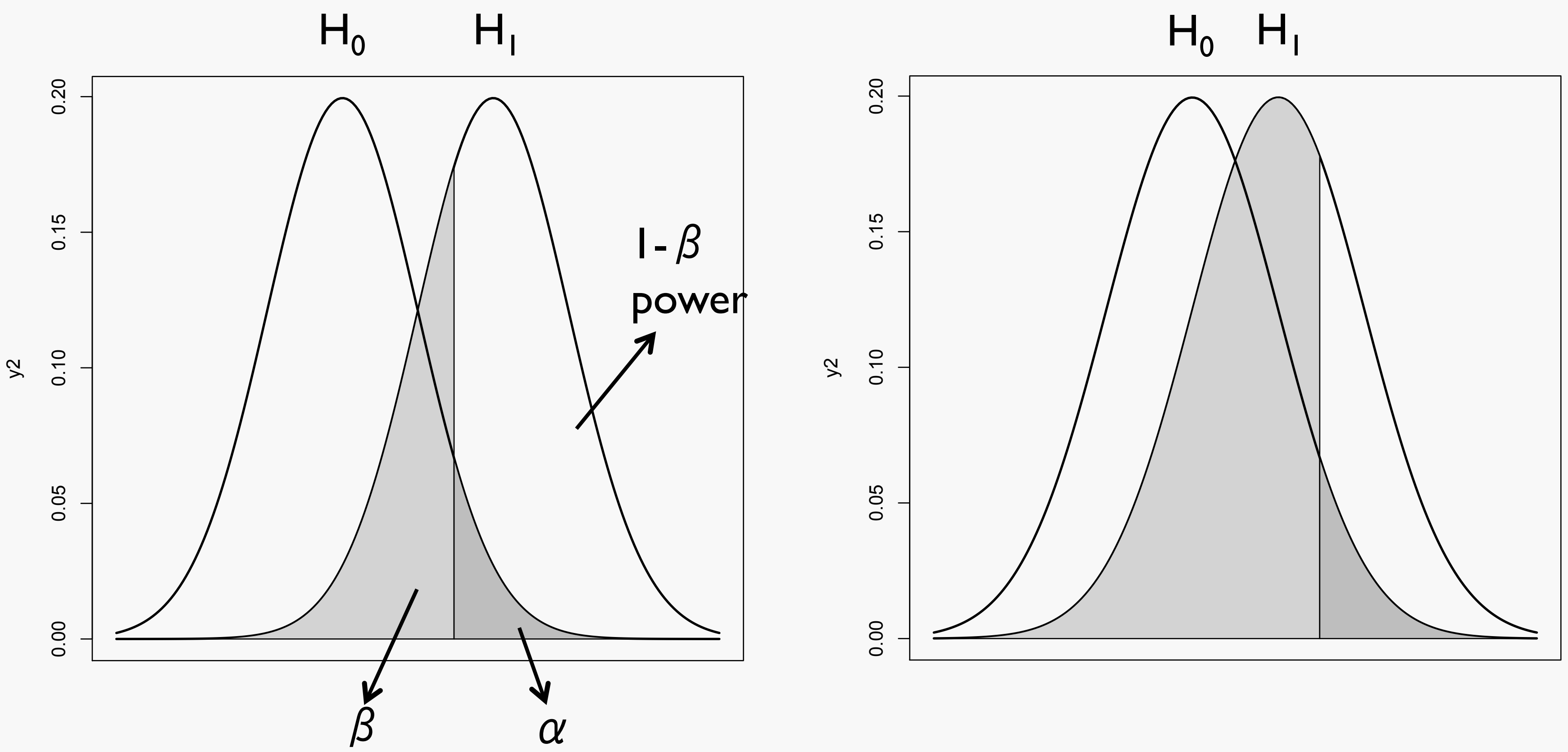
  Power is the probability that the test will reject the null hypothesis when the null hypothesis is actually false

- Power is influenced by:
  - α level
  - Effect size
  - Sample size

Statistical power: alpha level



Statistical power: effect size



Statistical power: variability