Statistical Tests

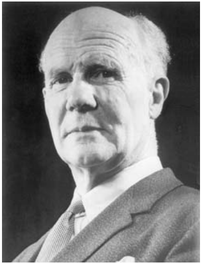
Basics: Jerzy Neyman (1894–1981) and Egon S. Pearson (1895–1980)

Example: is the filling capacity of bottles exactly 500ml as argued?

- It is a decision problem between two hypotheses.
- We speak of a test problem.
- These hypotheses can be expressed using the parameter of interest (here μ).

Jerzy Neyman (1894–1981)

Egon S. Pearson (1895–1980)



Let $X \sim F_\vartheta$, $\vartheta \in \Theta$. Θ_0 and Θ_1 are disjunct and $\Theta_0 \cup \Theta_1 = \Theta$.

The question regarding ϑ is expressed in terms two hypotheses H_0 (null hypothesis) and H_1 alternative hypothesis:

$$H_0: \ \vartheta \in \Theta_0 \quad \text{versus} \quad H_1: \ \vartheta \in \Theta_1 \ .$$

Action: On the basis of the sample $x_1, \ldots, x_n$ we should make a decision whether if $\mu = 500\, ml$ is fulfilled or not.

$\Rightarrow \vartheta = \mu$ - expectation (expected capacity) $\Theta_0 = \{500\ ml\}$.

Such a decision rule is called a statistical test. $\Theta_1 = \mathbb{R} \setminus \{500\}$.

Here we shortly sketch how does a test look like:

Example: filling capacity of bottles $\theta \in \mathbb{R}\setminus\{500\} \Leftrightarrow \theta \neq 500$

$$H_0 : \mu = 500ml \quad \text{vs} \quad H_1 : \mu \neq 500ml.$$

20 bottles ⇒ 495, 502, 495 ...
$\bar{x} = 499$ ⇒ H_0 prob. correct
20 bottles ⇒ 480; 580; ...
$\bar{x} = 492$ ⇒ H_0 prob wrong.

Starting point: Estimator for μ, here $\hat{\mu} = \bar{x}$

Consideration

Reject H_0, when $|\bar{x} - 500|$ is large enough, e. g. when $|\bar{x} - 500| > c$ for some constant $c > 0 \rightsquigarrow$ i.e. reject H_0 if we have sufficient evidence against H_0.

?

Decision rule

possible.

$|\bar{x} - 500| > c \quad \rightsquigarrow \quad$ reject H_0 (accept H_1),

$|\bar{x} - 500| \leq c \quad \rightsquigarrow \quad$ do not reject H_0 (H_0 CANNOT be accpeted).

Note: If the hypothesis is an inequality, then it must be chosen as H_1. A hypothesis with equality is always in H_0.

One sample Z-test

- We assume
  - $G \sim \mathrm{N}(\mu; \sigma)$ with known σ
  - random sample $X_1, \ldots, X_n$
- Different pairs of hypotheses

simplifying assumption

we try to detect deviations in both direction

$\mu = 500$ ml

$\Rightarrow$ underfilling and overfill.

two-sided

$$H_0 : \mu = \mu_0 \qquad H_1 : \mu \neq \mu_0$$

one-sided test

brewery

right-sided test	$H_0 : \mu \leq \mu_0$	$H_1 : \mu > \mu_0$
left-sided test	$H_0 : \mu \geq \mu_0$	$H_1 : \mu < \mu_0$

consumer.

- $X_1, \ldots, X_{25}$ with X_i = capacity of the i-th bottle $\sim \mathrm{N}(\mu; 1.5^2)$
  null hypotheses $H_0 : \mu = 500$, i.e. $\mu_0 = 500$

Possible errors: rejection of H_0, if H_0 is correct and non-rejection of H_0, if H_0 is wrong:

	Decision	**Reality**	
		rain	no rain
predict „rain"	take umbrella	correct decision	error
predict "no rain"	do not take umbrella	error	correct decision

$\{H_1\}$: decide to reject H_0
$\{H_0\}$: decide not to reject H_0

μ is really 500ml.

μ is 495

decision	Reality	
	H_0 is correct	H_0 is not correct (H_1 is correct)
H_0 is not rejected: $\{H_0\}$ $\bar{X}$ is close $\mu_0=500$	correct decision $\{H_0\}\|H_0$ $P(\{H_0\}\|H_0) = 1-\alpha$	error of the 2nd type $\{H_0\}\|H_1$ $P(\{H_0\}\|H_1) = \beta$ $\bar{x} = 501$ ⇒ we cannot detect the deviation from 500.
H_0 is rejected: $\{H_1\}$	error of the 1st type $\{H_1\}\|H_0$ $P(\{H_1\}\|H_0) = \alpha$ $\bar{x} = 490$ ml	correct decision $\{H_1\}\|H_1$ $P(\{H_1\}\|H_1) = 1-\beta$ $\bar{x} = 490$. - far from 500

⇒ just unlucky choice of bottles

Level of significance α:
the highest allowed probability of the type 1 error.

$\alpha = 0.05$ ⇒ we fix the prob of making the error of type I.

β - power of test.

Test statistic (test function)

- for testing pusposes we aggregate the information in the sample to a test statistic

$$V = V(X_1, ..., X_n).$$

compute a value (a RV) from the sample which is specific for the test.

- The functional form of V depends on the test/hypotheses, ect.
- The distribution of V under H_0 should be known (at least asymptotically), i.e. if H_0 is correct

$$F(v|\mu_0)$$

as for CI we need its distribution.

- Using this distribution split the set of possible values of the test statistics into
  - rejection area - if v takes a value here then H_0 is rejected ?
  - non-rejection area - if v takes a value here then H_0 is NOT rejected ?

Universität Augsburg

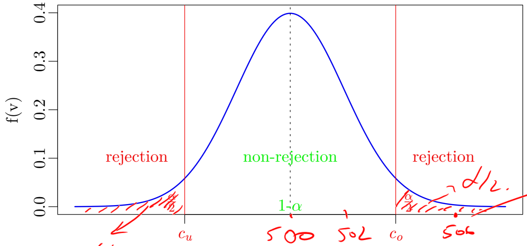
Critical values

$H_0: \mu = 500 \Rightarrow V \in$ rejection area (we reject H_0)
$\Rightarrow$ Type 1 error $\Rightarrow$ prob is α !!!

The rejection area (or critical values) is determined in such was, that the probability of getting a test statistic in this area (assuming that H_0 is correct) is not higher then the given α

$$P(V \in \text{ rejection area of } H_0 \mid \mu_0) \leq \alpha$$

$$P(V \in \text{ non-rejection area of } H_0 \mid \mu_0) \geq 1 - \alpha$$



$\bar{x} - 500$ large
$\Rightarrow$ reject H_0

$\Rightarrow \bar{x} - 500$ is small
H_0 is not rejected

Universität Augsburg

- two-sided test

$$H_0 \ : \ \mu = \mu_0 \qquad H_1 \ : \ \mu \neq \mu_0$$

- non-rejection area $[c_u; c_o]$

$$P(c_u \leq V \leq c_o \mid \mu_0) = 1 - \alpha \quad = 0.95$$

- rejection area $B = (-\infty; c_u) \cup (c_0; \infty)$

probability of lying in the rejection area.

$$P(V \leq c_u \mid \mu_0) + P(V \geq c_o \mid \mu_0) = \alpha/2 + \alpha/2 = \alpha$$

- test statistic

$$V = \sqrt{n}\frac{|\bar{x} - \mu_0|}{\sigma}$$

as on slide 174.

with normalization by $\sigma/\sqrt{n}$.

the same standardisation as for CI.

Universität Augsburg

- Relying on the symmetry of the normal distribution:

prob. that V lies in the rejection area (prob of Type I error)

$$\begin{aligned} P(|V| > c) &= P(V > c \text{ or } -V > c) = P(V > c \text{ or } V < -c) \\ &= P(V > c) + P(V < -c) = 2 \cdot P(V > c) \\ &= 2 \cdot [1 - P(V \leq c)] = 2 \cdot [1 - \Phi(c)] \overset{!}{=} \alpha \iff \end{aligned}$$

prob. of type I error.

$$\Phi(c) = 1 - \frac{\alpha}{2} \iff c = z_{1-\frac{\alpha}{2}}$$

$1-\alpha/2$ -quantile of $N(0,1)$ as for CI.

H_0 is rejected if $|v| > z_{1-\frac{\alpha}{2}}$.

- $B = (-\infty; -z_{1-\frac{\alpha}{2}}) \cup (z_{1-\frac{\alpha}{2}}; \infty)$ is the rejection area.

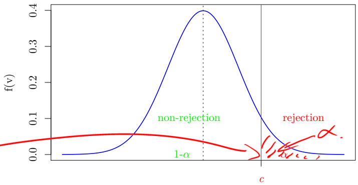
$v = \sqrt{n}\dfrac{\bar{x} - 500}{\sigma} \in B \Rightarrow$ reject H_0 $\Rightarrow$ filling capacity is not 500

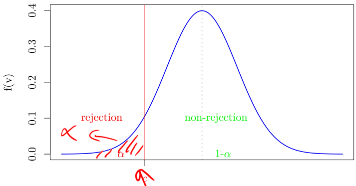
$\notin B \Rightarrow$ H_0 is not rejected.

Similarly for one sided tests

H_0 is rejected if $v \in B$ with

$$\left.\begin{array}{lll} B = \left(-\infty; -z_{1-\frac{\alpha}{2}}\right) \cup \left(z_{1-\frac{\alpha}{2}}; \infty\right) & & \text{case a)} \\ B = \left(-\infty; -z_{1-\alpha}\right) & & \text{case b)} \\ B = \left(z_{1-\alpha}; \infty\right) & & \text{case c)} \end{array}\right\} \quad (1)$$





$H_0: \mu > 500$ ml $\quad H_1: \mu < 500$ ml.

$\bar{x} << 500$ ml then reject

$H_0: \mu < 500 \quad H_1: \mu > 500.$

$\bar{x} >> 500$ ml $\Rightarrow$ reject

Example

$X_1, \ldots, X_{25}$ with $X_i \sim \mathrm{N}(\mu; 1.5^2)$ and $\bar{x} = 499.28$

Test $H_0 : \mu = 500, H_1 : \mu \neq 500$ with $\alpha = 0.01$

1. $\alpha = 0.01$ → value of the test statistic.
2. $v = \frac{499.28 - 500}{\underbrace{1.5}_{=\sigma}} \cdot \sqrt{\underbrace{25}_{=n}} = -2.4$
3. $\mathrm{N}(0; 1) : \; z_{1-\frac{\alpha}{2}} = z_{1-0.005} = z_{0.995} = 2.576$
   $\Rightarrow B = (-\infty; -2.576) \cup (2.576; \infty)$ → rejection area.
4. $v \notin B \Rightarrow H_0$ not rejected

$\leadsto$ With significance level of 1% we cannot prove that the capacity deviates from the stated capacity.

$n = 100 \qquad v = -4.8 \in B \Rightarrow H_0$ is rejected.
⇒ increased information ⇒ the distance between 499,28 and 500 becomes "large enough".

p-value

- If you change α you have to run the test again $\rightsquigarrow$ stat software do not ask for α, but compute the p-value which allows you to run the test for any α
- p-value is the largest level of significance for which H_0 is still not rejected
- The smaller the p-value is, the more evidence the sample contains against H_0
- decision rule

$$\boxed{\begin{array}{ll} \alpha > \alpha' & \Rightarrow H_0 \text{ rejected} \\ \alpha \leq \alpha' & \Rightarrow H_0 \text{ not rejected} \end{array}}$$

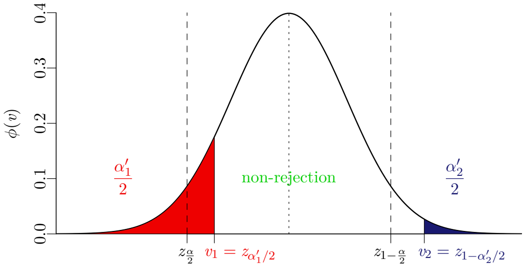
Case a: two-sided Z-test

The smallest value of α, for which H_0 is still not rejected, satisfies

$$\begin{cases} -z_{1-\frac{\alpha'}{2}} = z_{\frac{\alpha'}{2}} = v, & \text{if } v < 0 \Rightarrow \\ z_{1-\frac{\alpha'}{2}} = -z_{\frac{\alpha'}{2}} = v, & \text{if } v > 0 \Rightarrow \end{cases}$$

one equation in α'
solve to get
p-value

We are looking for such a value of α', that $\Phi(v) = \frac{\alpha'}{2}$ (if $v < 0$) or $\Phi(v) = 1 - \frac{\alpha'}{2}$ (if $v > 0$).



Universität Augsburg

income = $\beta_0 + \beta_1 \cdot$ field of studies + u_t

$H_0: \beta_1 = 0$ ⇒ field has no impact on income.

$\alpha = 10\%$ you will tend to reject H_0 even if field of studies have no impact.

⇒ "weak evidence".

Example

Using the quantiles of $\mathcal{N}(0,1)$ we obtain

$$\Phi(-2.4) = 1 - \Phi(2.4) = 1 - 0.9918 = 0.0082 = \frac{\alpha'}{2}.$$

Thus $\alpha' = 0.0164$

Let $\alpha = 0.01$. Since $\alpha' > \alpha$, we cannot reject H_0 at $\alpha = 0.01$.
Let $\alpha = 0.05$. Since $\alpha' < \alpha$, we reject H_0 at $\alpha = 0.05$.

⇝ For all $\alpha' < \alpha$ we can reject H_0.

⇒ α is in fact central for the decision.

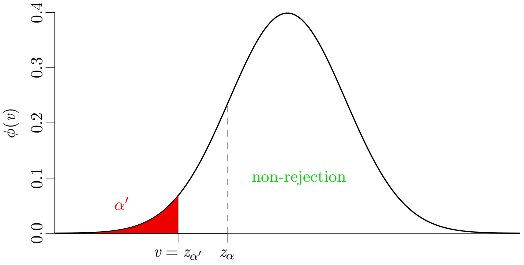
H_0: someone is ill. ⇒ afraid to reject H_0 erroneously
⇒ α very small 0,001
⇒ prob of treating someone as healthy if he is in fact ill is very small

Case b: left-sided Z-test

The smaller value of α, for which H_0 is still not rejected, satisfies

$$-z_{1-\alpha'} = z_{\alpha'} = v$$

We are looking for such value of α', that $\Phi(v) = \alpha'$.



Note:

In Case a) we can also run the test using the confidence intervals: We compute for given α the symmetric CI $[v_u; v_o]$ centered at $\bar{x}$ and reject $H_0: \ \mu = \mu_0$ vs. $H_1: \ \mu \neq \mu_0$ if $\mu_0 \notin [v_u; v_o]$.

Here: CI for μ :

1. $1 - \alpha = 1 - 0.01 = 0.99$
2. $\mathrm{N}(0; 1): \ c = z_{1-\frac{\alpha}{2}} = z_{1-\frac{0.01}{2}} = z_{0.995} = 2.576$
3. $\bar{x} = 499.28$
4. $\frac{\sigma c}{\sqrt{n}} = \frac{1.5 \cdot 2.576}{\sqrt{25}} = 0.77$
5. $[499.28 - 0.77; 499.28 + 0.77] = [498.51; 500.05]$

$\mu_0 = 500 \in [498.51; 500.05] \Rightarrow H_0$ cannot be rejected

Other tests

- Test of $H_0 : \mu = \mu_0$ with $X_i \sim N(\mu, \sigma^2)$, but σ^2 is unknown
  - Estimate σ^2 by s^2
  - $V = \sqrt{n}\frac{\bar{X}-\mu_0}{s} \sim t_{n-1}$
  - The rejection area for a two-sided test is

$n > 50 \Rightarrow z_{1-\alpha/2}$

$$B = (-\infty; -t_{n-1;1-\alpha/2}) \cup (t_{n-1;1-\alpha/2}; +\infty)$$

- Test of $H_0 : \mu = \mu_0$ if the distribution is unknown (asymptotic Z-test)
  - Rely on the CLT
  - $V = \sqrt{n}\frac{\bar{X}-\mu_0}{s} \overset{approx}{\sim} \mathcal{N}(0, 1)$
  - The rejection areas as for the simple Z-test

Other tests

- Test of $H_0 : p = p_0$, with $X_i \sim B(n,p)$
  - Check if $5 \leq \sum x_i \leq n-5$
  - Estimate $\hat{p} = \bar{x}$
  - Compute the test statistic $v = \sqrt{n}\frac{\bar{x}-p_0}{\sqrt{p_0(1-p_0)}} \overset{asymp}{\sim} \mathcal{N}(0,1)$
  - Follow the idea of the asymptotic Z-test
- Test of $H_0 : \sigma^2 = \sigma_0^2$, with $X_i \sim N(\mu, \sigma^2)$
  - Estimate σ^2 by s^2
  - $V = \frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{n-1}$
  - The rejection area for a two-sided test is

$$B = (0; \chi^2_{n-1;\alpha/2}) \cup (\chi^2_{n-1;1-\alpha/2}; +\infty)$$

Example

$X_1, \ldots, X_{2000} \sim \mathrm{B}(1;p)$ mit

$$X_i = \begin{cases} 1, & \text{falls } i\text{th person voted for the party A} \\ 0, & \text{else} \end{cases}$$

$$\sum_{i=1}^{2000} x_i = 108$$

Test $H_0: \ p \leq 0.05$ vs. $H_1: \ p > 0.05$ with $\alpha = 2\,\%$

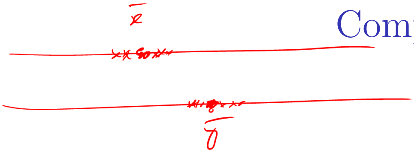
Asymptotic Z-test, Case (c); $5 \leq \sum x_i \leq n-5$: $\ 5 \leq 108 \leq 2000 - 5$

1. $\alpha = 0.02$
2. $v = \frac{\frac{108}{2000} - 0.05}{\sqrt{0.05 \cdot (1-0.05)}} \sqrt{2000} = 0.82$
3. $\mathrm{N}(0;1): \ z_{1-\alpha} = z_{0.98} = 2.05 \Rightarrow B = (2.05; \infty)$
4. $v \notin B \Rightarrow H_0$ not rejected

$\Rightarrow$ The party can not prove to get more then 5% of votes.

Two-sample tests

- Given
  - two independent samples $X_1, \dots, X_{n_1}$ and $Y_1, \dots, Y_{n_2}$ with
  - sample sizes n_1 and.n_2
  - expectations $E(X_i) = \mu_1$ and $E(Y_i) = \mu_2$
  - variances $Var(X_i) = \sigma_1^2$ and $Var(Y_i) = \sigma_2^2$
  - means $\bar{X}$ and $\bar{Y}$
  - sample variances S_1^2 and S_2^2
- Object of interest:
  - Comparison of means/expectations $\mu_1 \lesseqqgtr \mu_2$
  - Comparison of variances $\sigma_1^2 \lesseqqgtr \sigma_2^2$



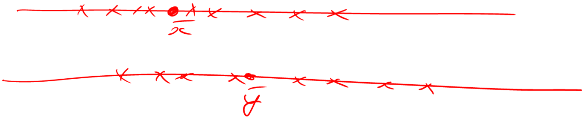
Comparison of means

Hypotheses

$$\boxed{\begin{array}{lll} \text{a)} & H_0: \ \mu_1 = \mu_2 & H_1: \ \mu_1 \neq \mu_2 \\ \text{b)} & H_0: \ \mu_1 \geq \mu_2 & H_1: \ \mu_1 < \mu_2 \\ \text{c)} & H_0: \ \mu_1 \leq \mu_2 & H_1: \ \mu_1 > \mu_2 \end{array}} \tag{2}$$

Estimator for $\mu_1 - \mu_2$: $\quad \bar{X} - \bar{Y}$

in contrary to $\bar{x} - \mu_0$ for a single sample test



Two-sample Z-test

If the variance σ_1^2 and σ_2^2 are known, then

$$Var(\bar{X} - \bar{Y}) = \underbrace{\frac{\sigma_1^2}{n_1}}_{= Var(\bar{X})} + \underbrace{\frac{\sigma_2^2}{n_2}}_{= Var(\bar{Y})}.$$

Thus the test statistics is

$$V = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

Under H_0 $(\mu_1 = \mu_2)$ and for Gaussian samples it holds

$$V \sim N(0, 1).$$ $\Rightarrow$ test like the t-test

Universität Augsburg

Two-sample t-test

If the variances σ_1^2 and σ_2^2 are unknown, but $\sigma_1^2 = \sigma_2^2 = \sigma^2$, then estimate σ^2 with

$$\tilde{\sigma}^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

with (under H_0)

$$\frac{(n_1 + n_2 - 2)\tilde{\sigma}^2}{\sigma^2} \sim \chi^2(n_1 + n_2 - 2).$$

The test statistic is

$$V = \frac{\bar{X} - \bar{Y}}{\sqrt{\tilde{\sigma}^2 \frac{n_1 + n_2}{n_1 n_2}}}.$$

Under H_0 $(\mu_1 = \mu_2)$ it holds

$$V \sim t_{n_1 + n_2 - 2}.$$

Asymptotic two-sample Z-test

If the variances σ_1^2 and σ_2^2 are unknown and arbitrary, then the test statistic is

$$V = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}.$$

Under H_0 $(\mu_1 = \mu_2)$ and from the CLT it holds

$$V \overset{\text{approx.}}{\sim} N(0,1).$$

The rejection area is given in all three situation by

$$B = (-\infty; -x_{1-\frac{\alpha}{2}}) \cup (x_{1-\frac{\alpha}{2}}; \infty) \qquad \text{in case a)}$$

$$B = (-\infty; -x_{1-\alpha}) \qquad \text{in case b)}$$

$$B = (x_{1-\alpha}; \infty) \qquad \text{in case c)}$$

with the corresponding quantiles defined by the above distributions of the test statistics.

	Assumption	test statistics V	Distr. of V under H_0
1.	$X_i \sim N(\mu_1;\sigma_1^2)$ $Y_i \sim N(\mu_2;\sigma_2^2)$ σ_1^2 and σ_2^2 known	$\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	$V \sim N(0;1)$
2.	$X_i \sim N(\mu_1;\sigma_1^2)$ $Y_i \sim N(\mu_2;\sigma_2^2)$ σ_1^2 and σ_2^2 unknown but $\sigma_1^2 = \sigma_2^2$	$\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\cdot\frac{n_1+n_2}{n_1 n_2}}}$	$V \sim t_{n_1+n_2-2}$
3.	$X_i \sim N(\mu_1;\sigma_1^2)$ $Y_i \sim N(\mu_2;\sigma_2^2)$ σ_1^2 and σ_2^2 unknown, but $\sigma_1^2 \neq \sigma_2^2$, $n_1 = n_2 = n$	$\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\cdot\frac{n_1+n_2}{n_1 n_2}}}$	$V \overset{\text{approx.}}{\sim} t_{(n-1)\left[1+\frac{2}{s_1^2/s_2^2+s_2^2/s_1^2}\right]}$
4.	$X_i \sim B(1;p_1)$ $Y_i \sim B(1;p_2)$ $5 \leq \sum x_i \leq n_1 - 5$ $5 \leq \sum y_i \leq n_2 - 5$	$\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{(\sum X_i+\sum Y_i)(n_1+n_2-\sum X_i-\sum Y_i)}{(n_1+n_2)n_1 n_2}}}$	$V \overset{\text{approx.}}{\sim} N(0;1)$
5.	X_i, Y_i arbitr. distr. $n_1 > 30;\ n_2 > 30$ σ_1^2, σ_2^2 unknown	$\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$	$V \overset{\text{approx.}}{\sim} N(0;1)$
6.	X_i, Y_i arbitr. distr. $n_1 > 30;\ n_2 > 30$ σ_1^2, σ_2^2 unknown	$\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	$V \overset{\text{approx.}}{\sim} N(0;1)$

1. two-sample Z-test; 2./3. two-sample t-test;
4./5./6. approx. Z-test

Example: has the expected return of an asset increased after the announcement of the acquisition?

Let X_1 be the return before the announcement and X_2 the return after. Assume $X_i \sim N(\mu_i, \sigma_i)$ and X_1 and X_2 are independent. D

$$H_0 : \mu_1 \geq \mu_2 \qquad vs \qquad H_1 : \mu_1 < \mu_2$$

$n_1 = 115$, $\bar{x}_1 = 6.5$, $s_1 = 0.4$, $n_2 = 110$, $\bar{x}_2 = 8.14$ and $s_2 = 0.78$. Thus

$$v = \frac{6.5 - 8.14}{\sqrt{\frac{0.4^2}{115} + \frac{0.78^2}{110}}} = -19.71.$$

We obtain $z_{0.99} = 2.3263$ and $B = (-\infty; -2.3263)$. Since $v < -2.3263$ we reject H_0 and conclude that the expected return is significantly larger after the announcement.

p-value $= \alpha' = 2 \cdot 10^{-86}$

$\Rightarrow$ very strong evidence in favour of H_1

Example:

$$X_1, \ldots, X_{80} \sim \mathrm{B}(1; p_1)$$

with

$$X_i = \begin{cases} 1, & \text{if the } i\text{th product is defective} \\ 0, & \text{else} \end{cases}, \qquad \sum_{i=1}^{80} x_i = 20$$

$$Y_1, \ldots, Y_{100} \sim \mathrm{B}(1; p_2)$$

with

$$Y_i = \begin{cases} 1, & \text{if the } i\text{th product is defective} \\ 0, & \text{else} \end{cases}, \qquad \sum_{i=1}^{100} y_i = 50$$

Can we argue that the probability of being defective is higher for Type 1 products than for Type 2 products?

1. $\alpha = 0.1$
2. $\bar{x} = \frac{20}{80} = 0.25; \quad \bar{y} = \frac{50}{100} = 0.5; \quad v = \frac{0.25-0.5}{\sqrt{\frac{(20+50)(80+100-20-50)}{(80+100)\cdot 80\cdot 100}}} = -3.42$
3. $\mathrm{N}(0;1): \quad z_{1-\alpha} = z_{0.9} = 1.282 \Rightarrow B = (-\infty; -1.282)$
4. $v \in B \Rightarrow H_0$ rejected, i.e. $p_1 < p_2$ is confirmed

Test for correlation/dependence

Assumption: let (X, Y) follow a 2–dim. normal distribution

$$E(X) = \mu_x \ , \ Var(X) \ = \ \sigma_x^2 \, ,$$

$$E(Y) = \mu_y \ , \ Var(Y) \ = \ \sigma_y^2 \, .$$

Let

$$\rho = Corr(X, Y) \ := \ \frac{Cov(X, Y)}{\sigma_x \, \sigma_y} \ = \ \frac{E\big((X - \mu_x)\,(Y - \mu_y)\big)}{\sigma_x \, \sigma_y} \, .$$

Note: if the samples are normal, then zero correlation implies independence.

no linear dependence.

no functional relationship between variables (no square, sin)

The estimator for ρ is

$$r_{XY} = \hat{\rho} = \frac{s_{XY}}{s_X\, s_Y} = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2 \sum\limits_{i=1}^{n}(y_i - \bar{y})^2}}$$

Note: similar tests can be derived the contingency tables and for the rank correlation of Spearman.

ranks.

to test monotone dependence.

$$r_{SP} = \frac{\sum_{i=1}^{n}\left(R(x_i) - \bar{R}\right)\left(R(y_i) - \bar{R}\right)}{\sqrt{\sum_{i=1}^{n}\left(R(x_i) - \bar{R}\right)^2 \sum_{i=1}^{n}\left(R(y_i) - \bar{R}\right)^2}}$$

with $\bar{R} = \frac{n+1}{2}$

Test problem

- $H_0 : \rho = 0$ (i.e. X and Y are uncorrelated / independent assuming normality) vs

  → equality hypothesis ⇒ H_0.
- $H_1 : \rho \neq 0$ (i.e. X and Y are dependent)

- Test statistics: $v = \sqrt{n-2}\frac{\hat{\rho}}{\sqrt{1-\hat{\rho}^2}}$.
- Under H_0 it holds $V \overset{approx}{\sim} \mathcal{N}(0,1)$ (or t_{n-2} for small samples)
- Rejection area

  or $-z_{1-\alpha/2}$

$$B = (-\infty; -t_{n-2;1-\frac{\alpha}{2}}) \cup (t_{n-2;1-\frac{\alpha}{2}}; \infty)$$

Note: Similarly with r_{SP} with "no monotone dependence"

$$v = \sqrt{n-2}\,\frac{r_{sp}}{\sqrt{1-r_{sp}^2}}.$$

Example: test with $\alpha = 0.05$ if the is a significant correlation between the body height of fathers (Y) and sons at the age of 5 (X)?

x_i	109	114	116	105	114	116	114	108	108	122	117	115
115	112	122	113	108	109	115	108	118	110	113	111	116
y_i	167	176	186	175	175	182	180	172	185	186	183	178
175	175	180	181	172	179	170	172	172	172	176	180	182

It holds $n = 25$, $\bar{x} = 113.12$, $s_X = 4.352$, $\bar{y} = 177.24$, $s_Y = 5.206$ and $s_{XY} = 10.05333$. Thus $\hat{\rho} = 0.44365$ and

$$v = \sqrt{23}\,\frac{0.44365}{\sqrt{1 - 0.44365^2}} = 2.374\,.$$

For $\alpha = 0.05$ it holds $t_{23;0.975} = 2.069$ and

$$B = (-\infty; -2.069) \cup (2.069; +\infty)\,.$$

Since $v \in B$, we conclude that $H_0 : \rho = 0$ can be rejected.

$H_0: \rho = 0 \Rightarrow H_1: \rho \neq 0 \Rightarrow$ there is a significant correlation between the body heights!

Universität Augsburg

Kolmogorov-Smirnov Goodness-of-Fit Test

Requirement: an independent random sample $X_1, \dots, X_n$ with $X_i \sim F$ for $i = 1, \dots, n$

the true unknown distribution of data

target distribution.

Testing problem:

$$H_0 : F = F_0 \qquad \text{against} \qquad H_1 : F \neq F_0,$$

where F_0 is a given and known distribution, e.g. $N(\mu_0, \sigma_0^2)$.

$H_0: \mu = \mu_0 \Rightarrow$ test for scalar quantities.

$H_0: F = F_0$ is a test for functions!

Universität Augsburg

$H_0: \mu = \mu_0 \xrightarrow{\text{estimator}} |\bar{x} - \mu_0|$

Idea of the test: Comparison of the empirical distribution function with F_0.

$F \longrightarrow$ estimated by ecdf $\hat{F}$.

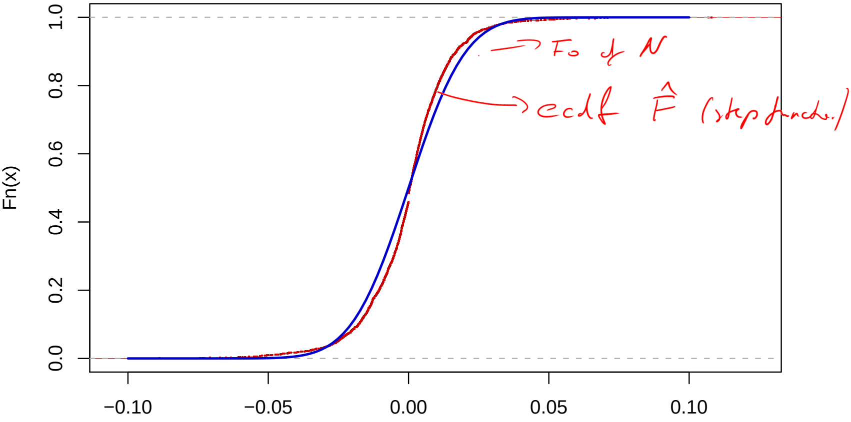
compare $\hat{F}$ with F_0. function

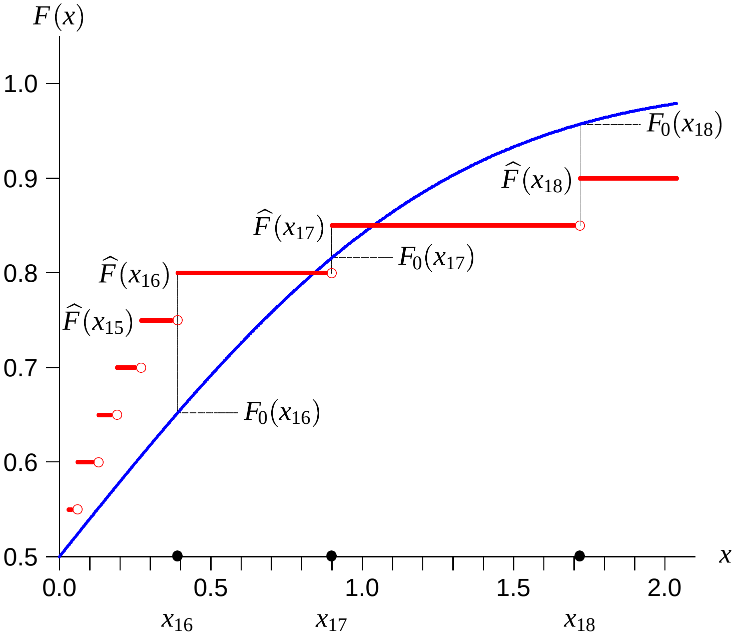
Distribution function:

$$F_0(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy$$

- $0 \leq F_0(x) \leq 1$;
- $F_0(x)$ is a non-decreasing function;
- $F_0(x)$ is right-continuous.

Example: Distribution function of the normal distribution (blue) and the empirical distribution function (red) for DAX returns





$F(x)$
1.0
0.9
0.8
0.7
0.6
0.5
$F_0(x_{18})$
$\widehat{F}(x_{18})$
$\widehat{F}(x_{17})$
$F_0(x_{17})$
$\widehat{F}(x_{16})$
$\widehat{F}(x_{15})$
$F_0(x_{16})$
x
0.0
0.5
1.0
1.5
2.0
x_{16}
x_{17}
x_{18}

largest vertical distance between the N( ) and ecdf.

Test statistic: $D = \max_{x\in\mathbb{R}} |\hat{F}(x) - F_0(x)|$

The distribution of D under H_0 is a non-standard distribution and is independent from F_0 if F_0 is continuous!

⇒ not N, t, χ^2, F ...

Decision: using the *p-value*-approach.

⇒ we cannot get the critical values c (or rejection area in a simple way ⇒ MC)

In practice: Let F_0 be continuous and $x_1 \le x_2 \le \ldots \le x_T$.

$$\rightsquigarrow \quad D = \max_{1\le t\le T} \left\{ \hat{F}(x_t) - F_0(x_t), F_0(x_t) - \hat{F}(x_{t-1}) \right\},$$

where $\hat{F}(x_0) := 0$.

Now: F_0 is a non-predetermined distribution, but a class of distributions, e.g. $N(\cdot,\cdot)$.
Testing problem:

$$H_0 : F \in \mathcal{F}_0 := \left\{ F_0\left(\frac{x-\mu}{\sigma}\right) : \mu \in \mathbb{R}, \sigma > 0 \right\}$$

$$H_1 : F \in \mathcal{D} - \mathcal{F}_0\,,$$

where F_0 is known (e. g. $F_0 = \Phi$).

Modified Kolmogorov-Smirnov Test

Test statistic: $D^* = \max_{x\in\mathbb{R}} \left| \hat{F}(x) - \Phi\big((x-\hat{\mu})/\hat{\sigma}\big) \right|$

If $D^* > c^*$, then H_0 is rejected.

Example:

For the DAX-Index with $F_0 = \mathcal{N}(5.7854 \cdot 10^{-6}; 2.5551 \cdot 10^{-4})$ we get

```
> ks.test(rdax, "pnorm", mean=mean(rdax), sd= sd(rdax))

        One-sample Kolmogorov-Smirnov test

data:  rdax
D = 0.0736, p-value = 1.067e-12
alternative hypothesis: two-sided
```

$<< \alpha = 0.05 \Rightarrow H_0$ rejected

$\Rightarrow$ The returns are not normally distributed.

Data clearly not normal $\Rightarrow$ use bootstrap to set rejection area

Power of a test

- Parametric test:

$$H_0: \ \vartheta \in \Theta_0 \quad \text{vs} \quad H_1: \ \vartheta \in \Theta_1$$

  with $\Theta_0 \cup \Theta_1 = \Theta \subseteq \mathbb{R}$.

- Performance measures of a test:
  a) Prob. of type I error should not exceed α.
  b) Prob. of type II error should be a small as possible.

Prob. of rejection H_0 depending on the true value of the parameter

$$G(\mu) = P(V \in \text{rejection area} H_0|\mu) = P(\{H_1\}|\mu)$$

→ true expectation (any value).

- Is $\vartheta \in \Theta_0$ so we made a wrong decision $(\{H_1\}|H_0)$.
- The power function is in this case the prob. of type I error: ⇒ "size of test"

$\boxed{G(\mu_0) = \alpha}$ ⇐ $$G(\mu) = P(\{H_1\}|\mu) \leq \alpha \text{ for all } \mu \in \Theta_0$$ ⇒ H_0 is correct, for example $\mu = \mu_0$. $\mu = \mu_0$

- Is $\vartheta \in \Theta_1$ so we made the correct decision $(\{H_1\}|H_1)$.
- The prob. of type II error :

$$G(\mu) = P(\{H_1\}|\mu) \leq 1 - \beta \text{ for all } \mu \in \Theta_1$$ $\mu \neq \mu_0$.

⇒ H_0 is wrong.

$= P(\text{choose } H_1 \mid \text{if } H_1 \text{ is correct})$

$= 1 - \underbrace{P(\text{choose } H_0 \mid \text{if } H_1 \text{ is correct})}_{\text{prob of type II error}}$ ⇒ $G(\mu)$ contains probs of both erroneous decision

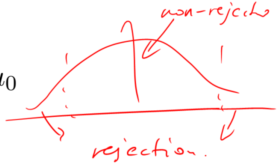
Universität Augsburg

Power function of the test for the mean

Assumption: α and n are fixed, normal distribution and σ^2 is known.

$$
\begin{aligned}
G(\mu) &= P(V \in \text{ rejection area } H_0|\mu) \\
&= P(\{H_1\}|\mu) \\
&= 1 - P(V \in \text{ non-rejection area } H_0|\mu) \\
&= 1 - P(\{H_0\}|\mu)
\end{aligned}
$$

two-sided test



$$H_0 \;:\; \mu = \mu_0 \text{ vs. } H_1 \;:\; \mu \neq \mu_0$$

H_0 is correct only if $\mu = \mu_0$.

$= 1 - P(\text{non-rejection})$

$$
\begin{aligned}
G(\mu) =& 1 - P(-z_{1-\alpha/2} \leq V \leq z_{1-\alpha/2} \mid \mu) \\
=& 1 - P\left(-z_{1-\alpha/2} \leq \frac{\bar{X} - \mu_0}{\sigma}\sqrt{n} \leq z_{1-\alpha/2} \mid \mu\right) \\
=& 1 - P\left(-z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq (\bar{X} - \mu_0) \leq z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \mid \mu\right) \\
=& 1 - P\left(-z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} + \mu_0 - \mu \leq (\bar{X} - \mu) \leq z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} + \mu_0 - \mu \mid \mu\right) \\
=& 1 - P\left(-z_{1-\alpha/2} + \frac{\mu_0 - \mu}{\sigma}\sqrt{n} \leq \underbrace{\frac{\bar{X} - \mu}{\sigma}\sqrt{n}}_{\sim N(0,1)} \leq z_{1-\alpha/2} + \frac{\mu_0 - \mu}{\sigma}\sqrt{n} \mid \mu\right)
\end{aligned}
$$

not $N(0,1)$ because $E(\bar{X}) = \mu$ and not μ_0.

Since μ is the true mean, it holds $\frac{\bar{X}-\mu}{\sigma}\sqrt{n} \sim N(0,1)$.

$$
\begin{aligned}
G(\mu) &= 1 - \Big[P\Big(\frac{\bar{X}-\mu}{\sigma}\sqrt{n} \leq z_{1-\alpha/2} + \frac{\mu_0-\mu}{\sigma}\sqrt{n}\Big) \\
&\quad -\Big[P\Big(-z_{1-\alpha/2} + \frac{\mu_0-\mu}{\sigma}\sqrt{n} \leq \frac{\bar{X}-\mu}{\sigma}\sqrt{n}\Big)\Big] \\
&= 1 - \Big[\Phi\Big(z_{1-\alpha/2} + \frac{\mu_0-\mu}{\sigma}\sqrt{n}\Big) - \Phi\Big(-z_{1-\alpha/2} + \frac{\mu_0-\mu}{\sigma}\sqrt{n}\Big)\Big]
\end{aligned}
$$

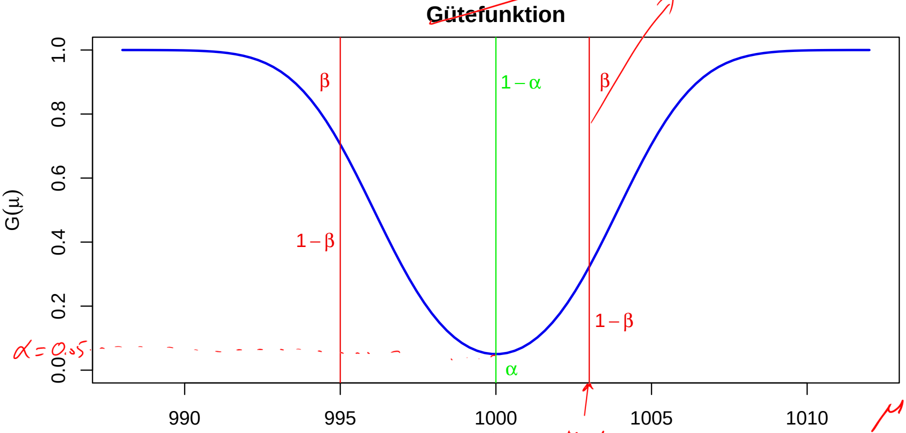
$$
G(\mu) = \begin{cases} \alpha = P(\{H_1\}|\mu), & \text{for } \mu = \mu_0 \\ 1-\beta(\mu) = P(\{H_1\}|\mu), & \text{for } \mu \neq \mu_0 \end{cases}
$$

$H_0: \mu = 1000 \quad H_1: \mu \neq 1000$

$G(\mu_0) = G(1000)$

Powerfunction

prob of type II error is 70%!



$\mu = 1003$

$\parallel$

μ_0. $\quad H_1: \mu \neq 1000$ is correct.

$G(1003) = 0{,}3 \Rightarrow$ We reject the wrong H_0 hypothesis with probability of 30%

Universität Augsburg

Example: target filling capacity $\mu_0 = 1000$. Let $\sigma = 10$, $\alpha = 0.05$, $n = 25$. What is the prob. of type II error, if the true filling capacity is $\mu = 1002$?

$$
\begin{aligned}
G(1002) &= 1 - \Big[\Phi\Big(1.96 + \frac{1000 - 1002}{10}\sqrt{25}\Big) \\
&\quad -\Phi\Big(-1.96 + \frac{1000 - 1002}{10}\sqrt{25}\Big)\Big] = 0.170066 = 1 - \beta
\end{aligned}
$$

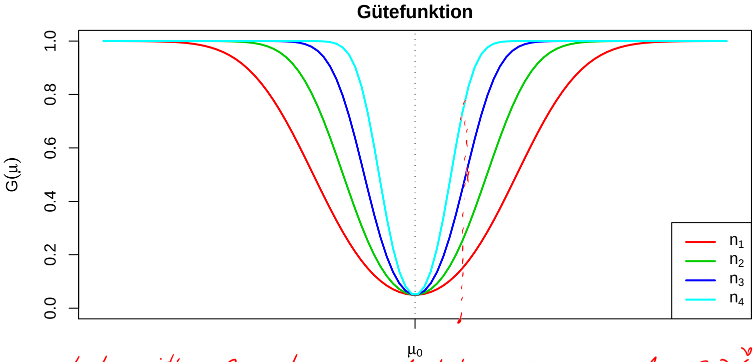
$$P(\{H_0\}|\mu = 1002) = \beta = 1 - G(1002) = 0.83$$

We will not detect the deviation of 2ml from the target capacity of 1000ml in 83% of the cases!!!

Let $\mu = 989$. Then $G(989) = 0.9998$ and $\beta = 0.0002$.

If the true capacity is $\mu = 989$, then we will NOT detect it only in 0.02% of all samples of size $n = 25$

Power function for different sample sizes $n_1 < n_2 < n_3 < n_4$

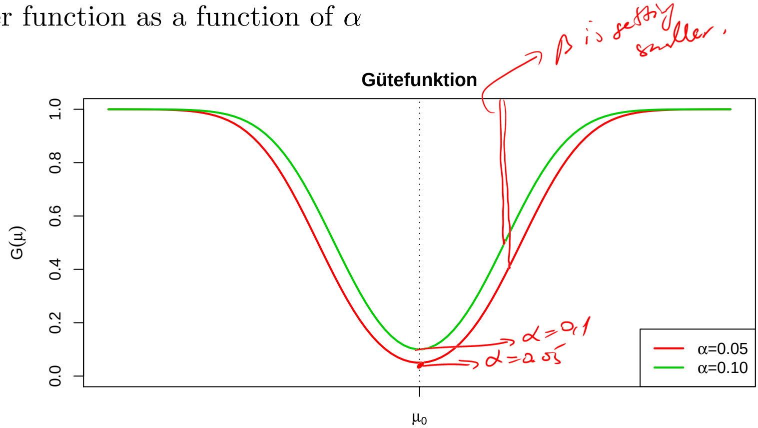


the test with n_4 has prob of type 2 error of 20%
but with n_1 - - 80%.

$\leadsto$ increasing the sample size reduces the prob. of type II error (ceteris paribus) .

Universität Augsburg

Power function as a function of α



$\rightsquigarrow$ increasing the prob. of type I error reduces the prob. of type II error (ceteris paribus)

$\rightsquigarrow$ both probabilities cannot be reduced simultaneously!!!