Advanced Mathematical Methods

WS 2022/23

6 Hypothesis Testing

Dr. Julie Schnaitmann

Department of Statistics, Econometrics and Empirical Economics

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

WIRTSCHAFTS- UND SOZIALWISSENSCHAFTLICHE FAKULTÄT

Readings

- A. Papoulis and A. U. Pillai. *Probability, Random Variables and Stochastic Processes*.
  Mc Graw Hill, fourth edition, 2002, Chapter 8

Online References

MIT Course on Probabilistic Systems Analysis and Applied Probability (by John Tsitsiklis)

- Lecture 25: Classical Inference III

Hypothesis testing

Testing procedure

step 1: set up a null hypothesis H_0 and an alternative hypothesis H_1

step 2: construct a test statistic t and find its distribution under H_0

step 3: for a significance level α, make a test decision and interpret the outcome.

Hypothesis testing

Test decision: given a specific significance level α

- find the critical value $t_{1-\alpha/2}$ (two-sided test), t_α (left-sided test) or $t_{1-\alpha}$ (right-sided test) and compare it to the test statistic. Reject the null hypothesis if $|t| > |t_{crit}|$.
- calculate the (empirical) p-value and compare it to the significance level α. Reject the null hypothesis if $p \leq \alpha$.
- for two-sided test: Reject H_0 if the hypothetical parameter lies outside of the associated $1 - \alpha$ confidence interval.

Interpretation: has three ingredients

- The null hypothesis H_0 can (not) be rejected at the α significance level.

Two ways of testing

θ unknown parameter in the population:

1. two-sided test:

$$H_0 : \theta = \bar{\theta} \qquad H_1 : \theta \neq \bar{\theta}$$

2. one-sided test: right-sided and left-sided

$$H_0 : \theta \leq \bar{\theta} \qquad H_1 : \theta > \bar{\theta}$$
$$H_0 : \theta \geq \bar{\theta} \qquad H_1 : \theta < \bar{\theta}$$

A testing example

Testing problem:

The price of a cup of cappuccino in 27 cafes in Tübingen and its surroundings is investigated. The average price is given by $\bar{y}_n = 3.10$ €, the standard deviation is given by $\sigma = 0.30$ € and the price of a cup of cappuccino is normally distributed. We want to test whether the average price of a cappuccino μ is significantly different from 3 €.

step 1 State the null and alternative hypothesis for the test.

$$H_0 : \mu = 3 \qquad \text{and} \qquad H_1 : \mu \neq 3$$

step 2 Set up an appropriate test statistic for the null hypothesis. What is the distribution of the test statistic under the null hypothesis?

$$Z = \frac{\bar{Y}_n - 3}{\sigma/\sqrt{n}} \underset{H_0}{\sim} \mathcal{N}(0,1)$$

A testing example

step 3 Calculate the test statistic and the associated p-value. What is the test decision on a 5% (10%) significance level?

test statistic: $z = \frac{3.1-3}{0.30/\sqrt{27}} = 1.73$

$$p = 2\,P(Z > 1.73) = 2\,(1 - P(Z \leq 1.73)) = 2\,(1 - \Phi(1.73))$$
$$= 2\,(1 - 0.9582) = 2 \cdot 0.0418 = 0.0836 > \alpha = 0.05$$

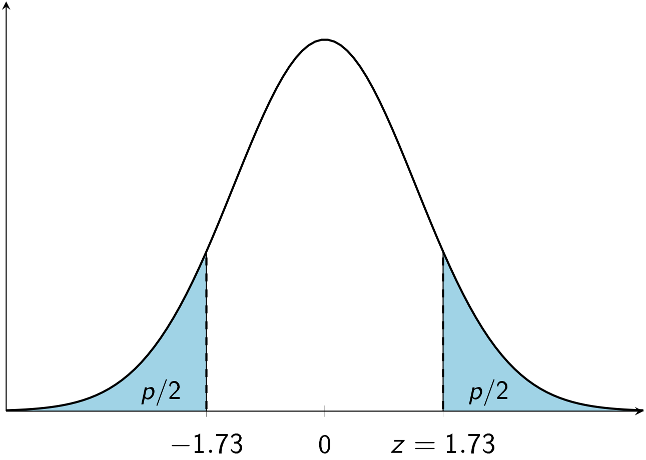
Interpretation:

We cannot reject the null hypothesis $H_0 : \mu = 3$ at the 5% significance level.

But: We can reject the null hypothesis $H_0 : \mu = 3$ at the 10% significance level.

A testing example

Distribution under H_0, test statistic and p-value of a two-sided test



Two types of errors

Two types of errors:

- α - type I error:
  Probability to reject a correct null hypothesis.
- β - type II error:
  Probability not to reject a wrong null hypothesis.

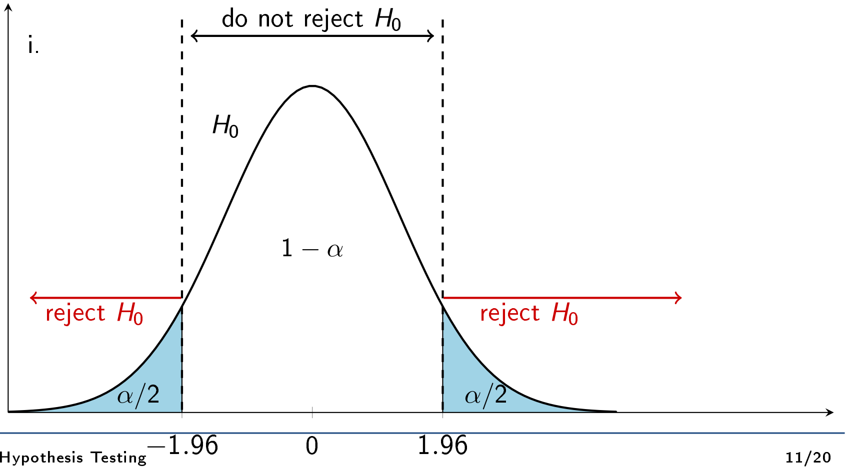
How "good" is the test?

- We control the type I error, the significance level α.
- The goodness of the test is measured by the **power** $= 1 - \beta$, the probability to reject a wrong null hypothesis and make a test decision.

Note: In practice, we do not know whether a given null hypothesis is correct or wrong in the population. We can only observe a finite sample of observations.

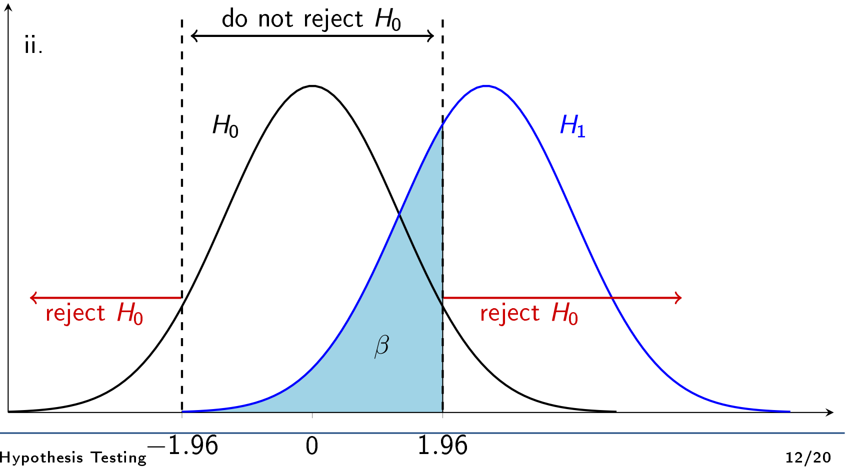
A testing example

Scenario: Sketch the distribution of the test statistic under the null hypothesis that $\mu = 3$ €. Include the rejection and non-rejection region and the type I error. Use a significance level of $\alpha = 5\%$.



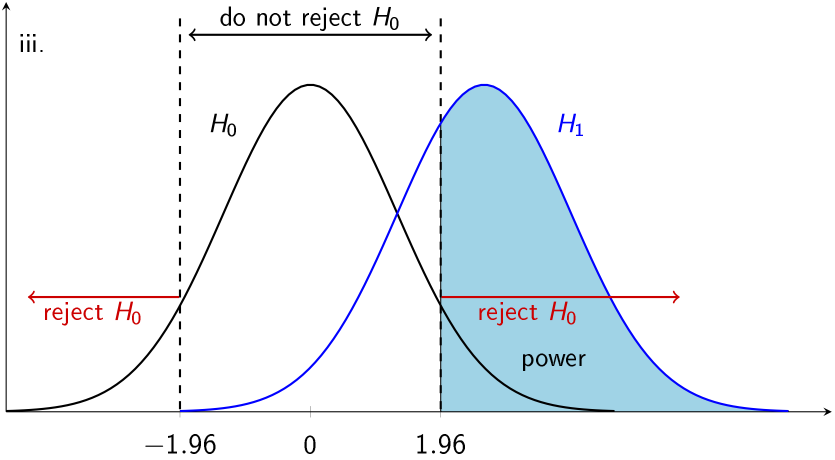
A testing example

Scenario: Now, assume that in reality the true price is given by $\mu = 5$ €. Include a sketch of the distribution under this alternative and the type II error into the plot.



A testing example

What is the power of a test? Where would we include power in the plot above?



Two types of errors

Type I error and significance level α: right-tailed test

- We select α: $P(t > c \mid H_0) = \alpha \quad \rightarrow$ critical value: $c = t_{1-\alpha}$.
- We reject H_0 if $t > t_{1-\alpha}$ (with rejection area $[t_{1-\alpha}, \infty]$) and don't reject H_0 if $t < t_{1-\alpha}$.

Type II error:

- Under H_1, the most likely values of t are on the right of $f(\bar{\theta})$.

$$\underbrace{\beta(\theta)}_{\text{depends on } \theta \text{, the true parameter}} = \int_{-\infty}^{c} f(t, \theta)\, dt$$

$\rightarrow$ can't be controlled!

Power of a test

Trade-off: α vs β

- When α is selected small, the chances to reject H_0 are small. $\rightarrow$ low probability of a type I error.
- However, the probability of a type II error is large: we may fail to reject a incorrect H_0.

Power of a test: $1 - \beta$

- Use the **power** of the test to assess the "goodness" of the test.
- $1 - \beta$ is the probability to reject an null hypothesis given that it is incorrect in population.
- The faster the probability to reject H_0 increases (steeper red line), the better when $\theta \neq \bar{\theta}$ (two-sided test) or $\theta > \bar{\theta}$ (right-sided test) is true in population.

What does significant really mean?

statistical significance

- does not answer the question wether the null hypothesis is wrong or right in population
- does not indicate how (un-) likely the null hypothesis is
- only controlled by maximum probability to run into type I error (α)
- provides no control over probability of type II error (β)

goal: for α given

$\rightarrow$ minimal β

$\rightarrow$ minimal $\alpha + \beta$

$\rightarrow$ maximal $1 - \beta$

Finite sample t-test using normality

estimated parameters $\widehat{\beta_1} \dots \widehat{\beta_K}$

1. **step 1:** define H_0, e.g. $H_0 : \beta_k = \bar{\beta}_k$ and H_1, e.g. $H_1 : \beta_k \neq \bar{\beta}_k$
2. **step 2:** construct test statistic

$$t = \frac{\widehat{\beta_k} - \bar{\beta}_k}{s.e.(\widehat{\beta_k})} \sim t(N-K) \quad \text{under} \quad H_0$$

3. **step 3: test decision** choose significance level α
4. compare test statistic t and critical value or the empirical p-value and the siginifcance level α

Large sample case

estimated parameters $\widehat{\beta}_1 \ldots \widehat{\beta}_K$

1. **step 1:** define H_0, e.g. $H_0 : \beta_k = \bar{\beta}_k$ and H_1, e.g. $H_1 : \beta_k \neq \bar{\beta}_k$
2. **step 2:** construct test statistic and use law of large numbers (LLN) and central limit theorem (CLT)

$$z = \frac{\widehat{\beta}_k - \bar{\beta}_k}{s.e.(\widehat{\beta}_k)} \overset{a}{\sim} \mathcal{N}(0,1) \quad \text{under} \quad H_0$$

3. **step 3: test decision** choose significance level α
4. compare test statistic z and critical value z_{crit} or the empirical p-value and the siginifcance level α

Confidence Interval

Construct a $(1-\alpha)$ confidence interval around $\widehat{\beta}_k$:

In finite sample (+normality assumption), use

$$CI(\widehat{\beta}_k, \alpha) = \left[\widehat{\beta}_k - t_{1-\frac{\alpha}{2}} \cdot s.e.(\widehat{\beta}_k), \widehat{\beta}_k + t_{1-\frac{\alpha}{2}} \cdot s.e.(\widehat{\beta}_k)\right].$$

In large samples, we can also make use of the LLN and CLT and use

$$CI(\widehat{\beta}_k, \alpha) \approx \left[\widehat{\beta}_k - z_{1-\frac{\alpha}{2}} \cdot s.e.(\widehat{\beta}_k), \widehat{\beta}_k + z_{1-\frac{\alpha}{2}} \cdot s.e.(\widehat{\beta}_k)\right]$$

Interpretation: If $\bar{\beta}_k$ is contained in the confidence interval, $H_0 : \beta_k = \bar{\beta}_k$ cannot be rejected on the α significance level.

LLN and CLT

For a random sample $\{X_1, X_2, \ldots, X_n\}$ with finite $E(X_i)$ and $Var(X_i)$ and an appropriately large n, it holds that

Law of large numbers (LLN)

$$\lim_{n\to\infty} P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - E(X_i)\right| > \epsilon\right] = 0 \quad \text{for any } \epsilon > 0.$$

and

Central limit theorem (CLT)

$$\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} X_i - E(X_i)\right] \overset{a}{\sim} \mathcal{N}(0, Var(X_i))$$