CV 510$^0_1$
Modeling, Uncertainty, and Data for Engineers
(July – Nov 2025)

Dr. Prakash S Badal

1

Parameter estimation

- Next remaining classes:
  - Estimation: MoM, MLE
  - Inference: Confidence interval (CI)
  - Inference: Hypothesis testing, p-value, power
  - Goodness of fit (GoF): χ^2, KS
- Risk & reliability (Q2C8)
  - Introduction
  - Risk Evaluation
- Extreme value analysis (Q2C7)
  - Peak over threshold
  - Return period and design life

2

2

Module Overview

B.2

Inferences

B.2.1

Confidence interval (CI)
- Interval estimation
- Mean, stdev

Q1C3.5

B.2.2

Hypothesis testing (HT)
- p-value
- Null/alternative hypothesis
- Type I/II error
- Power

Q1C3.9

B.2.3

Goodness-of-Fit (GoF).
- Likelihood ratio
- KS test

Q1C5.4

3

Flow

- Null/alternative hypothesis
- Type I/II error
- Hypothesis testing method
- Error calculations
- Power of test
- p-value
- GoF

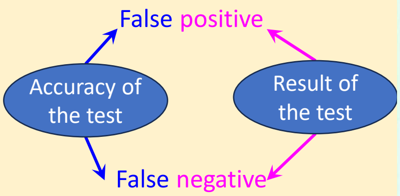
1. Revise H0-H1 definitions
2. Emphasize the justice example; contrast the same errors with safety alarm at the airport where Type II is more important.
3. Power of test.
4. p-value
5. Show xkcd 882 for p-value, then introduce formally.
6. GoF
    1. KS test
    2. AIC

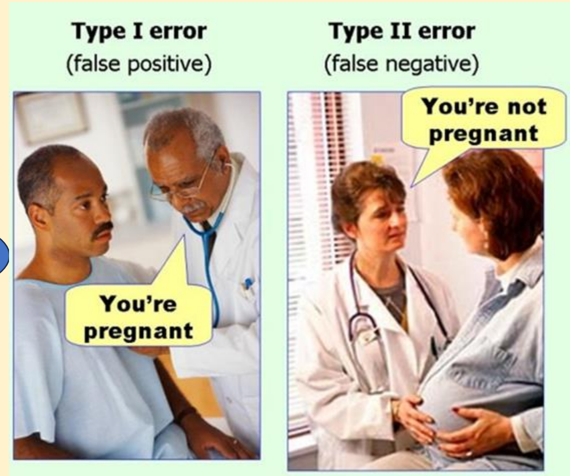
4

4

Spot the error*

*: not an example from real life





People are generally not pregnant!

Default	You are not pregnant	H_0
Alternative	You are pregnant	H_1

5

5

Justice system in a democracy

- How can someone be arrested if they really are presumed innocent?
- Why is a defendant pronounced "Not Guilty" instead of innocent?
- Why do citizens put up with a system that allows criminals to go free on technicalities?

6

6

Common challenges in justice, research (life?)

- A strong urge to believe in unusual event/discovery
  - A charged /medicine discovery/a new concrete mix
- Huge repercussions of making a wrong decision
  - Two types of wrong decisions

You acted on your belief, BUT The belief turned out to be FALSE	You did NOT act on your belief, and the belief turned out to be TRUE
Innocent punished/side-effects death of patients/unsafe buildings	guilty going free/suffering patients/wasted resources

- No sure-shot way to prove the unusual event
- Limited data

7

7

Hypothesis testing is the solution!

8

8

Hypothesis testing (& justice system)

No numerical values in courts, but they share four common features:

1. **The alternative hypothesis:**
   This is why a *criminal is arrested* (or a research undertaken)
   denoted by H_1 or H_A.
2. **The null hypothesis:**
   The *presumption of innocence* (the solution won't work!)
   H_0 is the logical opposite of H_1.
3. **A standard of justice:**
   Absolute proof IMPOSSIBLE, hence a *reasonable doubt* (sig. level)
   Reject the null hypothesis beyond a reasonable doubt.
4. **A data sample:**
   Evaluation of *partial information* (test score)
   The "whole truth and nothing but the truth" is often impossible.

9

Errors in Hypothesis testing (& justice system)

$$\text{Type I error} \equiv \left\{ \begin{matrix} \text{an innocent person goes to jail} \\ + \\ \text{a guilty person walks free} \end{matrix} \right\}$$

HT (& justice system) puts a lot of emphasis on
avoiding Type I error.

Counter example: Airport security
Innocent person and alarm goes off (Type I error)
Guilty person and alarm goes not go off (Type II error)

Sometimes, Type II error is worse!

10

Spot the error type

According to a report from the United States Environmental Protection Agency, burning one gallon of gasoline typically emits about 8.9 kg of CO_2. A fuel company wants to test a new type of gasoline designed to have lower CO_2 emissions. Here are their hypotheses:

$H_0: \mu = 8.9$ kg
$H_a: \mu < 8.9$ kg

(where μ is the mean amount of CO_2 emitted by burning one gallon of this new gasoline).

Under which of the following conditions would the company commit a Type I error?

Choose 1 answer:

(A) The mean amount of CO_2 emitted by the new fuel is actually 8.9 kg, and they fail to conclude it is lower than 8.9 kg.

(B) The mean amount of CO_2 emitted by the new fuel is actually lower than 8.9 kg, and they conclude it is lower than 8.9 kg.

(C) The mean amount of CO_2 emitted by the new fuel is actually 8.9 kg, and they conclude it is lower than 8.9 kg.

(D) The mean amount of CO_2 emitted by the new fuel is actually lower than 8.9 kg, and they fail to conclude it is lower than 8.9 kg.

A large university is curious if they should build another cafeteria. They plan to survey a sample of their students to see if there is strong evidence that the proportion interested in a meal plan is higher than 40%, in which case they will consider building a new cafeteria.

Let p represent the proportion of students interested in a meal plan. Here are the hypotheses they'll use:

$H_0: p \leq 0.40$
$H_a: p > 0.40$

What would be the consequence of a Type II error in this context?

Choose 1 answer:

(A) They don't consider building a new cafeteria when they should.

(B) They don't consider building a new cafeteria when they shouldn't.

(C) They consider building a new cafeteria when they shouldn't.

(D) They consider building a new cafeteria when they should.

Spot the error type

A quality control expert wants to test the null hypothesis that a new solar panel is no more effective than the older model.

What would be the consequence of a Type I error in this context?

Choose 1 answer:

(A) They don't conclude the new panel is more effective when it actually is not more effective.

(B) They don't conclude the new panel is more effective when it actually is more effective.

(C) They conclude the new panel is more effective when it actually is more effective.

(D) They conclude the new panel is more effective when it actually is not more effective.

According to a report from the United States Environmental Protection Agency, burning one gallon of gasoline typically emits about 8.9 kg of CO_2. A fuel company wants to test a new type of gasoline designed to have lower CO_2 emissions. Here are their hypotheses:

$H_0: \mu = 8.9$ kg
$H_a: \mu < 8.9$ kg

(where μ is the mean amount of CO_2 emitted by burning one gallon of this new gasoline).

Which of the following would be a Type II error in this setting?

Choose 1 answer:

(A) The mean amount of CO_2 emitted by the new fuel is actually 8.9 kg, and they conclude it is lower than 8.9 kg.

(B) The mean amount of CO_2 emitted by the new fuel is actually lower than 8.9 kg, and they fail to conclude it is lower than 8.9 kg.

(C) The mean amount of CO_2 emitted by the new fuel is actually 8.9 kg, and they fail to conclude it is lower than 8.9 kg.

(D) The mean amount of CO_2 emitted by the new fuel is actually lower than 8.9 kg, and they conclude it is lower than 8.9 kg.

Example 1

- Consider bars made of IS 1786-compliant Fe500 grade steel
  - Let's focus on the mean yield strength
  - We are interested in deciding whether or not the mean yield strength is 500 MPa

Formally,

$$H_0: \mu = 500 \text{ MPa} \quad \text{versus} \quad H_1: \mu \neq 500 \text{ MPa}$$

Four situations may arise:

Decision	H_0 is True	H_0 is False
Fail to reject H_0	No error	Type II error (β)
Reject H_0	Type I error (α)	No error

$$\Pr(\text{Type I error}) = \alpha \qquad \Pr(\text{Type II error}) = \beta$$

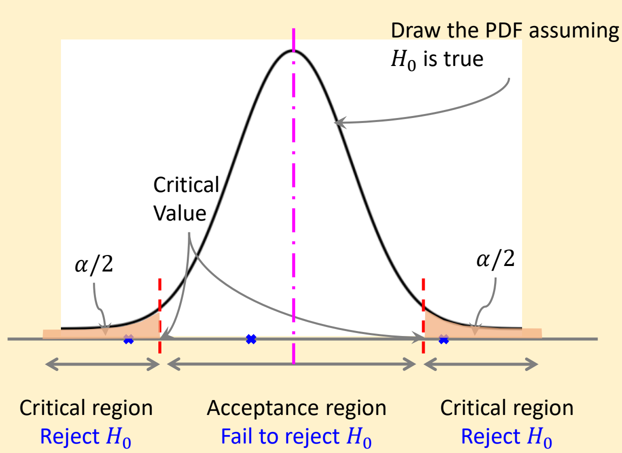
13

Example 1

$H_0: \mu = 500$ MPa
$H_1: \mu \neq 500$ MPa

Checklist:

1. How to draw PDF
2. Critical value
3. Acceptance region
4. Critical region
5. Sig. level
6. Areas
7. When reject/fail to reject H_0



14

Type I error

Probability of Type I Error, α

- $\alpha = \Pr(\text{Type I error})$

$$= \Pr(\text{reject } H_0 \text{ when } H_0 \text{ is true})$$

aka the significance level or the α-error

We can reduce α by

- by widening the acceptance region
- by increasing the sample size n

Probability of Type I Error, α

- Consider Fe500 bar again; suppose $\sigma_{\text{Fe500}} = 25\text{ MPa}$
- Suppose we tested a sample with $n = 10$ specimens
- If a client accepts bars with average $f_y \in [485\text{ MPa}, 515\text{ MPa}]$
- Find the probability of Type I error?

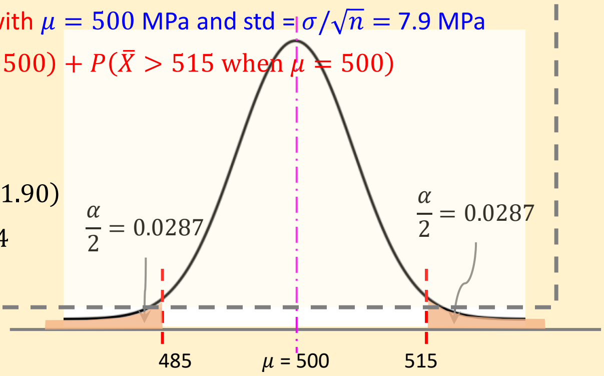
- Type I error will occur when $\bar{x} > 515$ or $\bar{x} < 485$ for $\mu = 500\text{ MPa}$.
- CLT ⇒ $\bar{x}$ is approx. normal with $\mu = 500$ MPa and std $= \sigma/\sqrt{n} = 7.9$ MPa
- $\alpha = P(\bar{X} < 485 \text{ when } \mu = 500) + P(\bar{X} > 515 \text{ when } \mu = 500)$

$Z(\bar{X} = 485) = -1.90$

$Z(\bar{X} = 515) = 1.90$

$\alpha = P(Z < -1.90) + P(Z > 1.90)$

$= 0.0287 + 0.0287 = 0.0574$



Standard Normal Probabilities

Z-table

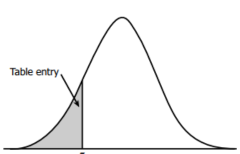


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Type II error

Probability of Type II Error, β

- $\beta = \Pr(\text{Type II error})$

$= \Pr(\text{fail to reject } H_0 \text{ when } H_0 \text{ is false})$

aka the β-error

- To calculate β we must have a specific alternative hypothesis,

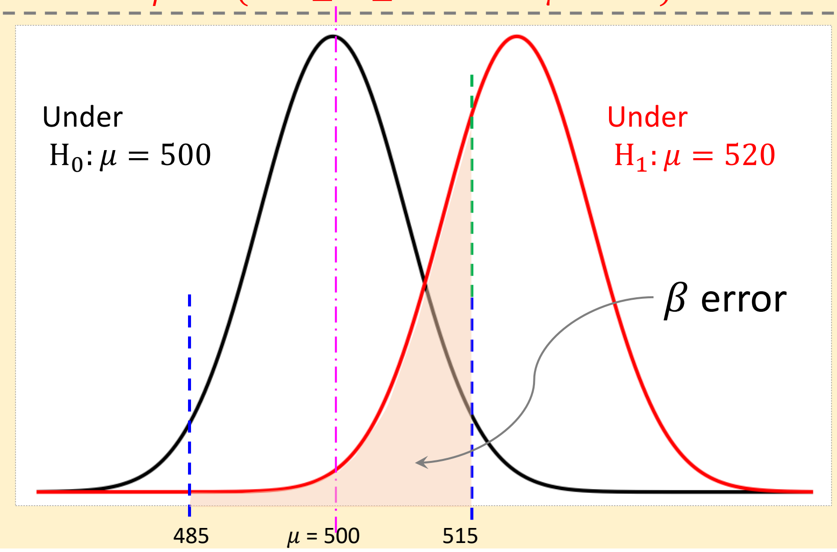
For example: choose a specific alternative value, e.g., 520 MPa,

$H_0: \mu = 500 \text{ MPa}$ versus $H_1: \mu = 520 \text{ MPa}$

Type II error

Type II error will be committed if $\bar{x} \in [485,515]$ when $\mu = 520$

$\beta = P(485 \leq \bar{x} \leq 515 \text{ when } \mu = 520)$



21

Probability of Type II Error, β

- Fe500; $\sigma_{\mathrm{Fe500}} = 25$ Mpa; we tested a sample with $n = 10$ specimens
- If an alternative hypothesis exists that $\mu_{F_y} = 520$ MPa
- Find the probability of Type II error?

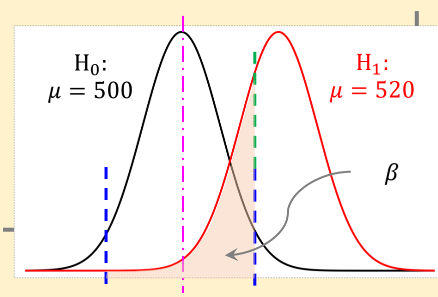
- A type II error will occur when $\bar{x} \in [485, 515]$ for $\mu = 520$ MPa.
- CLT ⇒ $\bar{x} \approx \mathrm{N}(\mu = 500$ Mpa; std = 7.9 MPa)
- $\beta = P(\bar{X} \in [485, 515] \text{ when } \mu = 520)$

$Z(\bar{X} = 485) = -4.43$

$Z(\bar{X} = 515) = -0.63$

- $\beta = \Pr(Z \leq -0.63) - \Pr(Z \leq -4.43)$

$= 0.2643 - 0.0000 = 0.2643$



22

Power of a statistical test

- The power of a statistical test

$$\Pr(\text{rejecting } H_0 \text{ when } H_1 \text{ is true})$$

$$Power = 1 - \beta$$

interpreted as the probability of correctly rejecting a false null hypothesis

23

Power: steel strength example

- In the steel strength problem, we found that $\beta = 0.2643$ when the true mean is $\mu = 520$, what is the power of this test

$$1 - \beta = 0.7357$$

if the true mean is really 520, this test will correctly reject H_0: $\mu = 500$ MPa and "detect" this difference 73.57% of the time

24

24

p-value

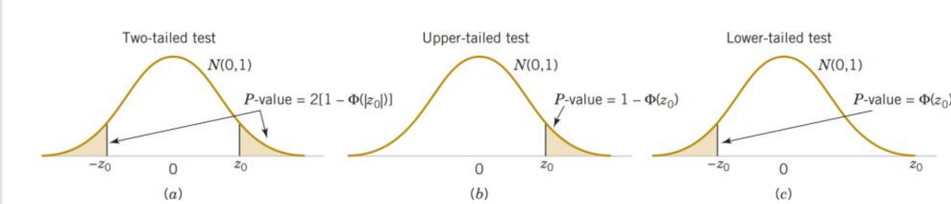
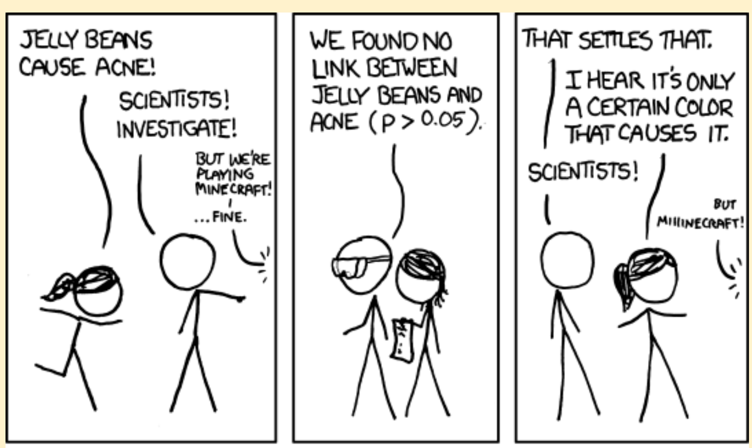


FIGURE 9-10 The P-value for a z-test. (a) The two-sided alternative $H_1 : \mu \neq \mu_0$. (b) The one-sided alternative $H_1 : \mu > \mu_0$. (c) The one-sided alternative $H_1 : \mu < \mu_0$.

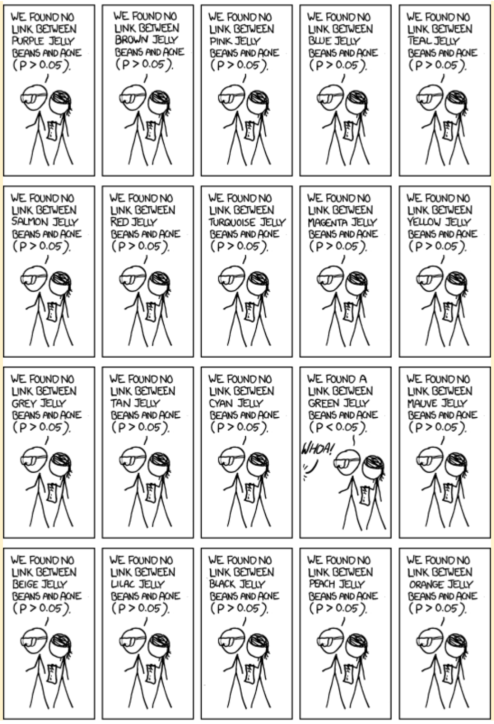
25

25

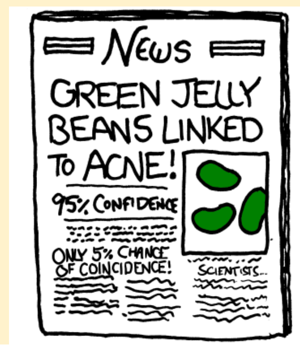


https://xkcd.com/882

26



https://xkcd.com/882



27

Questions, comments, or concerns?

28

28