VE401 Recitation Class Note8

Hypothesis Testing

Chen Siyi
siyi.chen_chicy@sjtu.edu.cn

Z-test:

To review the three general hypothesis testing methods , we take Z-test for the mean as examples.

Let $X_1, . . . , X_n$ be a random sample of size n from a normal distribution with variance σ^2, and let $\overline{X}$ denote the sample mean. Let μ be the unknown population mean and μ_0 a null value of that mean. Then any test based on the standard normal distributed statistic below is called a Z-test:

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

1 Fisher's Null Hypothesis Testing

Goal:

Find statistical evidence that allows us to **reject** the null hypothesis H_0.

Null Hypothesis:

1. $\theta = \theta_0$
2. $\theta \leq \theta_0$
3. $\theta \geq \theta_0$

Steps:

1. Set H_0 as what you wish to reject
2. Gather data for a random sample.
3. Calculate P-value for the data.
4. If P-value is **small** enough, reject H_0 at the [P-value] level of significance.

1.1 P-value

Definition:

$$P[D|H_0] \leq P\text{-value}$$

Interpretation:

D represents the statistical data.

But do notice here, take Z-test as example, D does not means obtaining this specific value of $\overline{x}$. D can be understood as the case "value of $\overline{X}$ being $\overline{x}$ or worse ones for supporting H_0 true". Simply means $\overline{X}$ being $\overline{x}$ or even further from what H_0 expects.

Example:

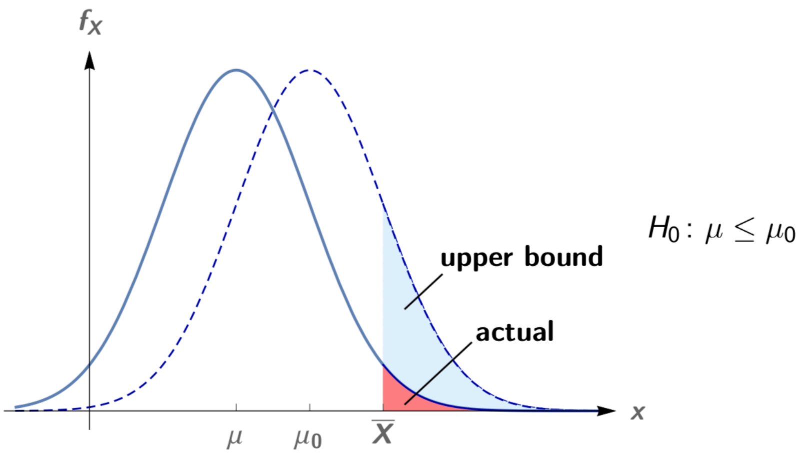
Z-test, $H_0 : \mu \leq \mu_0$, then

$$P\text{-value} = max\{P[\overline{X} \geq \overline{x}|\mu \leq \mu_0]\} = P[\overline{X} \geq \overline{x}|\mu = \mu_0]$$

$$P\text{-value} = P[\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \geq \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}] = P[Z \geq \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}]$$

Rejecting H_0:

Reject H_0 if P-value is small. Then we say we reject H_0 at the [P-value] level of significance.



1.2 One-tailed and Two-tailed Test

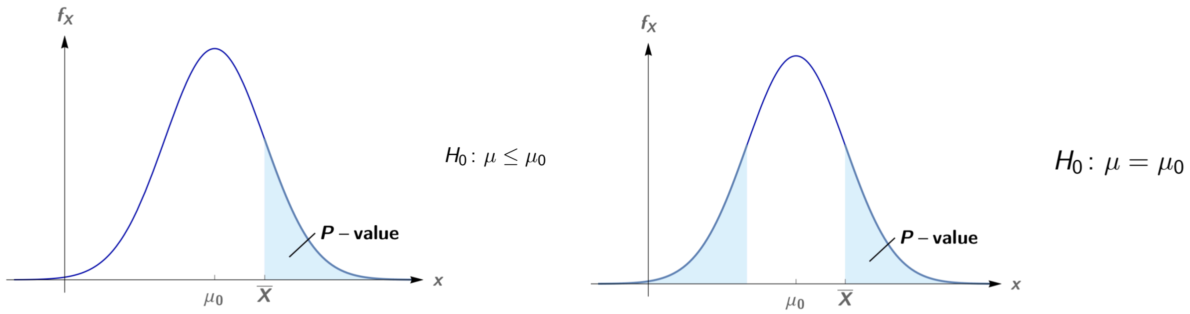
One-tailed:

$H_0 : \theta \leq \theta_0$ or $H_0 : \theta \geq \theta_0$

Two-tailed:

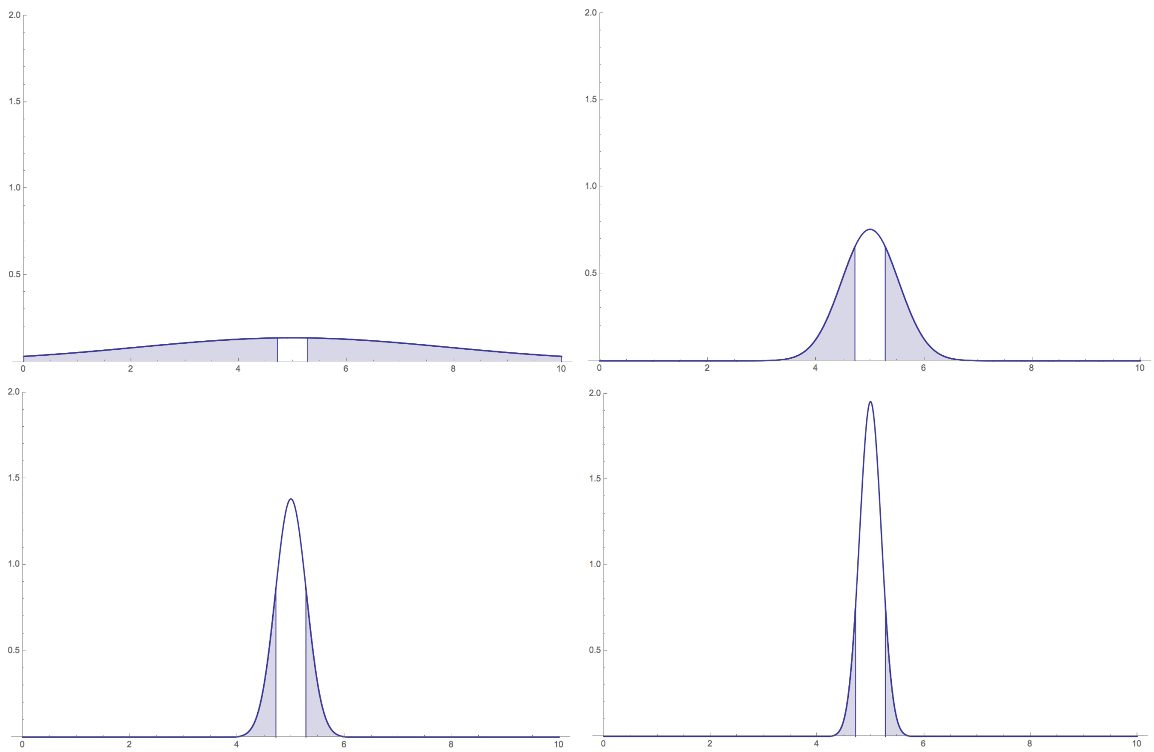
$H_0 : \theta = \theta_0$

Be careful with the different P-value for one-tailed and two-tailed test. For two-tailed test, there is a double.



1.3 Remarks

1. Need multiple, independent significant tests.
2. $P[H_0|D]$ is wanted instead of $P[D|H_0]$.
3. Two-tailed tests are pointless. H_0 can always be rejected if the sample size n is chosen large enough.



Demo 3.1.

2 Neyman–Pearson Decision Theory

Goal:

Seek to reject H_0, in which case we accept H_1.

1. H_0: null hypothesis
2. H_1: research hypothesis, or alternative hypothesis.

Steps:

(i) Select appropriate hypotheses H_1 and H_0 and a test statistic;

(ii) Fix α and β for the test;

(iii) Use α and β to determine the appropriate the sample size;

(iv) Use α and the sample size to determine the critical region;

(v) Obtain the sample statistic; if the test statistic falls into the critical region, reject H0 at significance level α and accept H_1. Otherwise, accept H_0.

2.1 Type I, Type II Errors and Power

	Actual situation	
Decision	H_0 true	H_1 true
Reject H_0	Type I error (probability α)	Correct decision
Fail to reject H_0	Correct decision	Type II error (probability β)

1. $\alpha := P[\text{ Type I error }] = P[\text{ reject } H_0 \mid H_0 \text{ true }] = P[\text{ accept } H_1 \mid H_0 \text{ true }]$
2. $\beta := P[\text{ Type II error }] = P[\text{ fail to reject } H_0 \mid H_1 \text{ true }] = P[\text{ accept } H_0 \mid H_1 \text{ true }]$
3. Power $:= 1 - \beta = P[\text{ reject } H_0 \mid H_1 \text{ true }] = P[\text{ accept } H_1 \mid H_1 \text{ true }]$

2.2 α and the Critical Region

Definition:
If H_0 is true, then the probability of the test statistic's values falling into the critical region is $\leq \alpha$.

Rejecting H_0:
If the value of the test statistic falls **into** the critical region, then we **reject** H_0.

Example:
Z-test, $H_0 : \mu = \mu_0, \quad (H_1 : |\mu - \mu_0| \geq \delta)$, then

$$\alpha = P[\overline{X} \text{ in the critical region } | \mu = \mu_0]$$

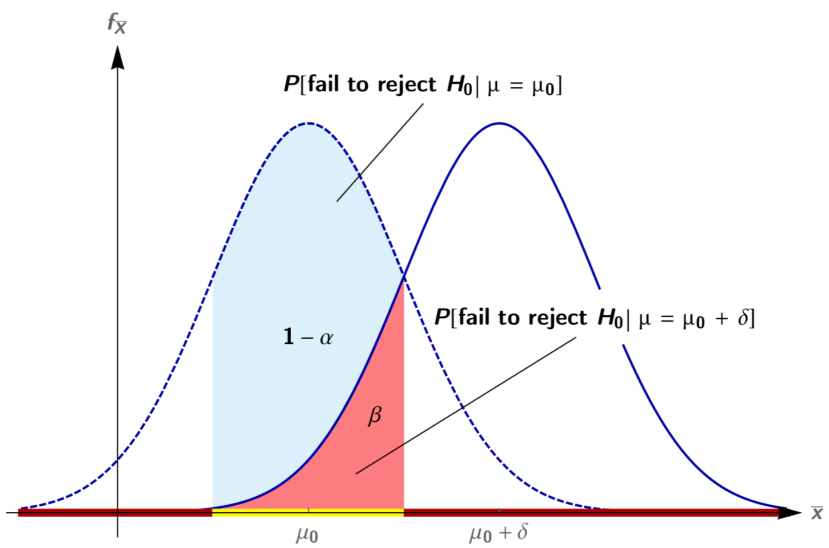
Notice $P[\frac{|\overline{X}-\mu_0|}{\sigma/\sqrt{n}} > z_{\alpha/2}] = \alpha$
So the critical region is set as:

$$\bar{x} \neq \mu_0 \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Comment:
The critical region depends on H_0, α, and always the sample size n.
But it has no relation to H_1.

2.3 β and the Sample Size

β is decided by H_1, α, and sample size n. (Of course also related to the distribution of the statistic you use.)



Intuitively, you can calculate β by integeration.

Example:

Z-test, $H_0 : \mu = \mu_0, \quad H_1 : |\mu - \mu_0| \geq \delta$, then

$$\beta = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-z_\beta} e^{-t^2/2} dt$$

$$-z_\beta \approx z_{\alpha/2} - \delta\sqrt{n}/\sigma$$

$$n \approx \frac{\left(z_{\alpha/2} + z_\beta\right)^2 \sigma^2}{\delta^2}$$

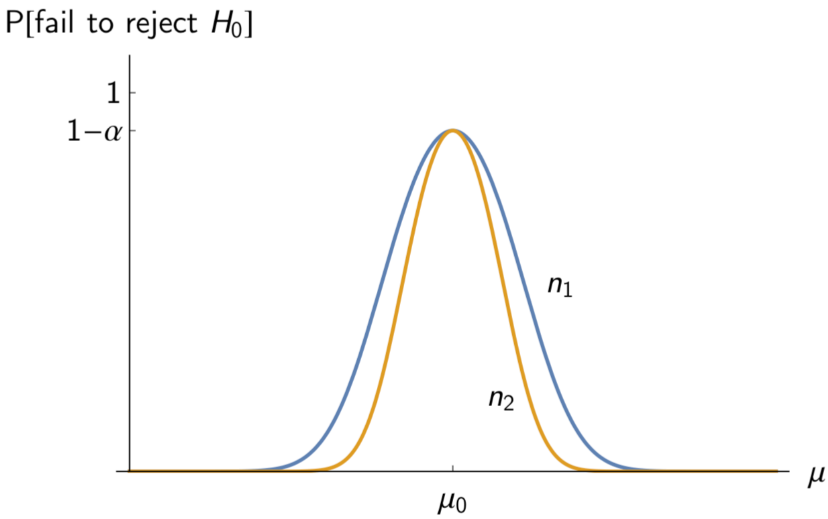
When using a statistic of a typical distribution, read n from OC curves is another efficient way.

2.4 OC Curves

α, β, n, H_1, when any two of them are fixed, then the left two will have a relationship described by a ”function”, where you need a parameter representing the effect of H_1.

See an example to understand...

First try to understand this graph you see in class by yourself.

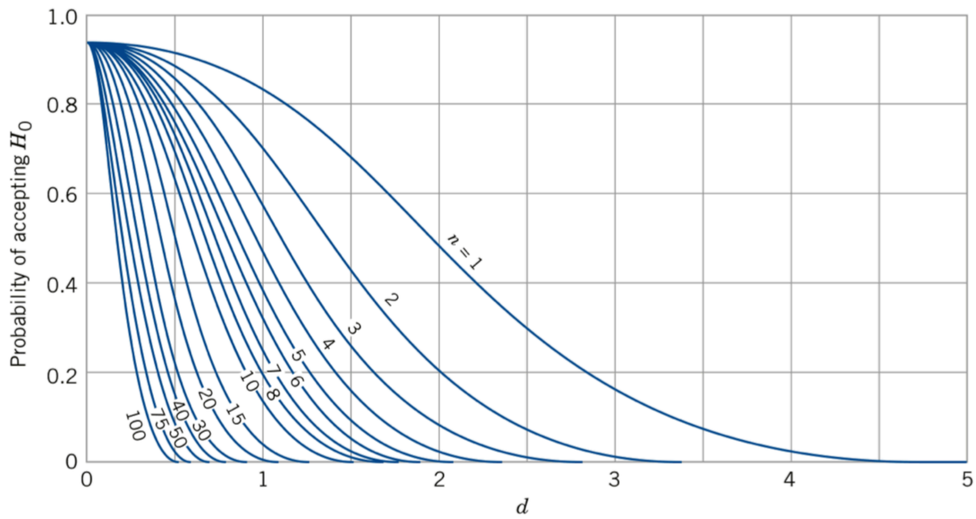


We prefer standardized curves.

Always, in one graph, we fix α, and choose several n. The horizontal ordinate uses a standardized parameter representing the effect of H_1, and the vertical coordinate is $1 - \beta$.

Example for two-tailed Z-test, where the abscissa is standardized as:

$$d = \frac{|\mu - \mu_0|}{\sigma}$$



Basic Reading:

1. Known H_0, H_1, α, β. Determine the needed n.
2. Known H_0, H_1, α, n. Obtain β.
3. Known H_0, α, β, n. Make inferences on H_1.

Question: Draw OC curves

Consider Z-test for μ, with any given α, n, H_0 and H_1, do you know how to draw the OC curve for that specific case?

Demo 3.2.

3 Null Hypothesis Significance Testing

Steps:

1. Two hypotheses, H_0 and H_1 are set up, but H_1 is always the logical negation of H_0
2. Then either a "hypothesis test" is performed, whereby a critical region for given α is defined, the test statistic is evaluated and H_0 is either rejected or accepted.
3. Alternatively (and more commonly), the test statistic is evaluated immediately, a P-value is found, and H_0 is either rejected or accepted based on that value.
4. In either case, there is no meaningful discussion of β, since H_1 is exactly the negation of H_0.