STAT22000 Autumn 2013 Lecture 21

Yibi Huang

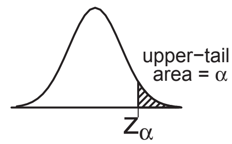
November 15, 2013

6.3 Use and Abuse of Tests
6.4 Power and Inference as a Decision

Notation z_α

Let z_α be the value that the area to the right of z_α under the standard normal curve is α. I.e.,

$P(Z > z_\alpha) = \alpha$ or



For a one-sided test $\begin{cases} H_0: & \mu = \mu_0 \\ H_a: & \mu > \mu_0 \end{cases}$, the P-value $< \alpha$ if and only if the z-statistic $z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha$.

Similarly, for a two-sided test $\begin{cases} H_0: & \mu = \mu_0 \\ H_a: & \mu \neq \mu_0 \end{cases}$, the P-value $< \alpha$ if and only if the z-statistic $z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} > z_{\alpha/2}$.

Another Way to Look at Hypothesis Testing

For a one-sided test,

$$\begin{cases} H_0 & : \mu = \mu_0 \\ H_a & : \mu > \mu_0 \end{cases}$$

the H_0 is rejected at level α if the test statistic $z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha$, or equivalently if the sample mean

$$\overline{X} > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}.$$

For a two-sided test,

$$\begin{cases} H_0 & : \mu = \mu_0 \\ H_a & : \mu \neq \mu_0 \end{cases}$$

the H_0 is rejected at level α) if the absolute value of the z-statistic $|z| = \frac{|\overline{X}-\mu_0|}{\sigma/\sqrt{n}} > z_{\alpha/2}$, or equivalently if

$$|\overline{X} - \mu_0| > z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Confidence Interval Revisit

Recall in Lecture 18, we show the confidence interval for the population μ at confidence level $(1-\alpha)$ is

$$\overline{X} \pm z^* \frac{\sigma}{\sqrt{n}}$$

in which z^* is the **critical value** that a standard Normal density has area $(1-\alpha)$ between $-z^*$ and z^*.

With the z_α notation, the z^* above is simply $z_{\alpha/2}$. So the confidence interval for the population μ at confidence level $(1-\alpha)$ is

$$\overline{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}.$$

Duality of Confidence Intervals and Two-Sided Tests

In a two-sided test,

$$H_0 : \mu = \mu_0$$
$$H_a : \mu \neq \mu_0$$

the null hypothesis is rejected at level α exactly when the value of μ_0 falls outside a $(1-\alpha)100\%$ confidence interval for μ.

Reason. The null hypothesis is NOT rejected at level α if and only if

$$\frac{|\overline{X}-\mu_0|}{\sigma/\sqrt{n}} \leq z_{\alpha/2} \iff -z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \overline{X} - \mu_0 \leq z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$
$$\iff \overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

I.e. μ_0 is in the interval,

$$\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

which is exactly the $(1-\alpha)$ confidence level for μ.

6.3 Use and Abuse of Tests

- **Not rejecting H_0 $\neq$ Accept H_0**: just not enough evidence to conclude
- **Significance $\neq$ Importance**
  That is because statistical significance doesn't tell you about the magnitude of the effect, only that there is one.
  An effect could be too small to be relevant. And with a large enough sample size n, significance can be reached even for the tiniest effect.
  Example. A drug for lung cancer is found to increase the 5-year survival rate from 30% to 31% (with P-value < 0.001). Is this a great news for patients with lung cancer?
- **Interpreting Effect Size: It's All About Context**
  There is no consensus on how big an effect has to be in order to be considered meaningful. In some cases, effects that may appear to be trivial can be very important.

Type I Error & Type II Error

The hypothesis H_0 can be either **true** or **false**.
We can **reject** or **not reject** H_0
The following outcomes are possible when conducting a test:

	Our Decision	
Reality	Not Reject H_0	Reject H_0
H_0 is true	Correct decision	Type I Error
H_0 is false	Type II Error	Correct decision

Example:

	The suspect is actually ...	
Criminal Trial	innocent	guilty
Verdict of not guilty (Not reject H_0)	Correct Decision	The guilty goes free (Type I Error)
Verdict of guilty (Reject H_0)	Convict the innocent (Type II Error)	Correct Decision

Power = Ability to Detect the Alternative Hypothesis

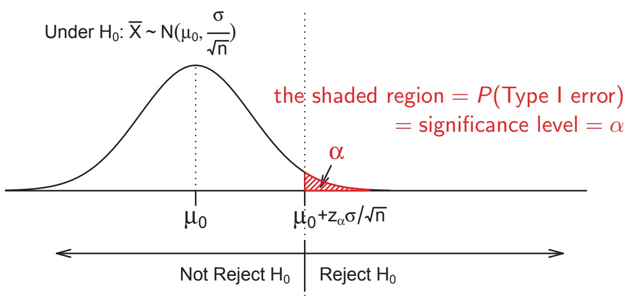
The **power** is the probability of the test to reject the null H_0 when the alternative H_a is true.

$$\begin{aligned} \text{Power} &= P(\text{reject } H_0 | H_a \text{ is true}) \\ &= 1 - P(\text{type II error}) \\ &= \text{Ability of a test to distinguish } H_a \text{ from } H_0 \end{aligned}$$

Power calculation requires knowledge of distribution of test statistics under H_a. Simply knowing $\mu \neq \mu_0$ is not enough. So we hypothesize a value for $\mu = \mu_a$ under H_a.

Power of a Test

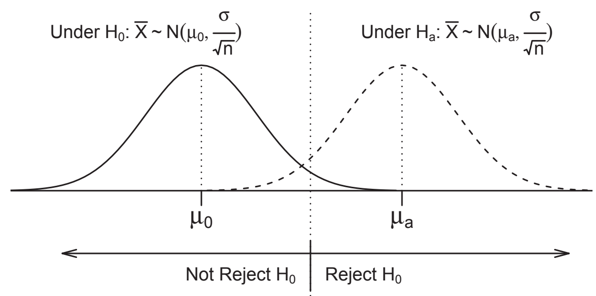
Consider the one-sided test $H_0: \mu = \mu_0$ v.s. $H_a: \mu > \mu_0$.



H_0 is rejected at level α if the test statistic $z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha$, or equivalently if the sample mean $\overline{X} > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$.

Power of a Test

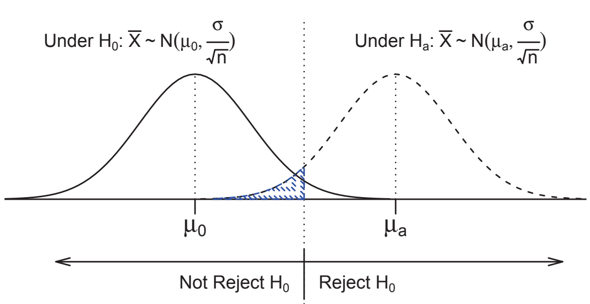
Consider the one-sided test $H_0: \mu = \mu_0$ v.s. $H_a: \mu > \mu_0$.



H_0 is rejected at level α if the test statistic $z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha$, or equivalently if the sample mean $\overline{X} > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$.

Power of a Test

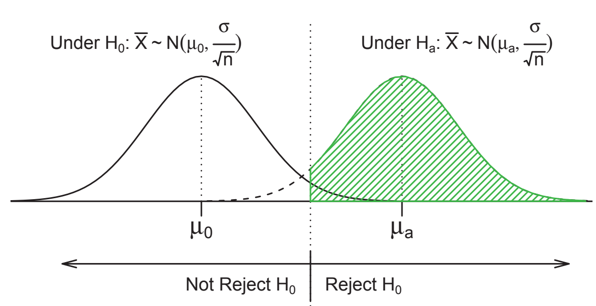
Consider the one-sided test $H_0: \mu = \mu_0$ v.s. $H_a: \mu > \mu_0$.



$$\begin{aligned} \text{Area of the blue shaded region} &= P(\text{fail to reject } H_0 | H_a \text{ is true}) \\ &= P(\text{Type II error}) \end{aligned}$$

Power of a Test

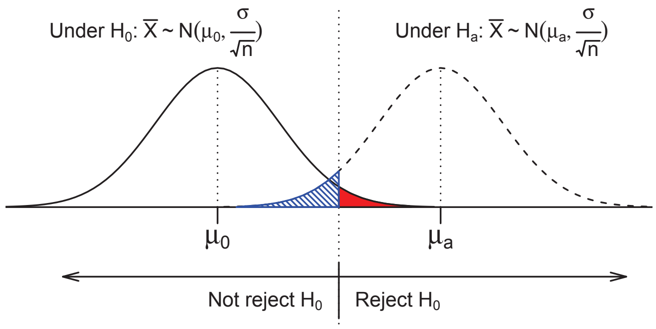
Consider the one-sided test $H_0: \mu = \mu_0$ v.s. $H_a: \mu > \mu_0$.



$$\begin{aligned} \text{Area of the green shaded region} &= P(H_0 \text{is rejected} | H_a \text{ is true}) \\ &= \textbf{power of the test} \\ &= 1 - P(\text{Type II error}) \end{aligned}$$

Trade-off Between Type I and Type II Errors

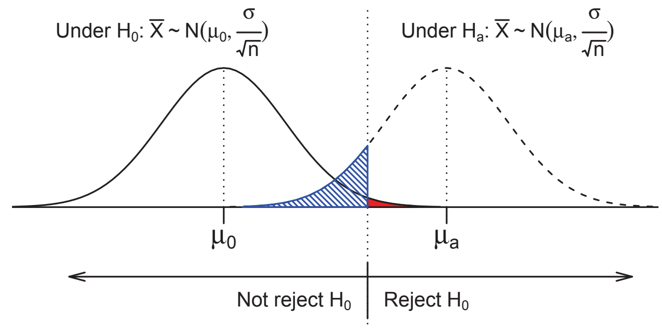
We wish to minimize both the chances to make type I error and type II error. Unfortunately, there is a trade-off.



Whenever we reduce P(Type I error) (area of the red filled region), P(Type II error) (area of the blue shaded region) is increased, and vice versa.

Trade-off Between Type I and Type II Errors

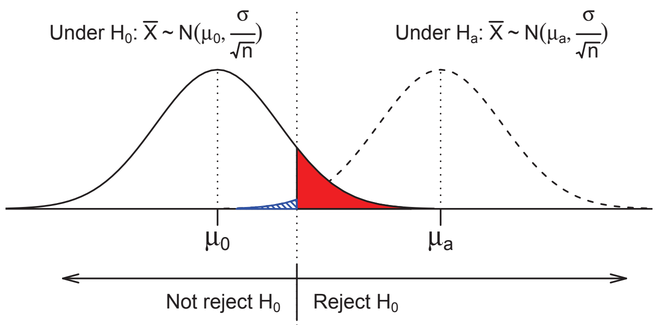
We wish to minimize both the chances to make type I error and type II error. Unfortunately, there is a trade-off.



Whenever we reduce P(Type I error) (area of the red filled region), P(Type II error) (area of the blue shaded region) is increased, and vice versa.

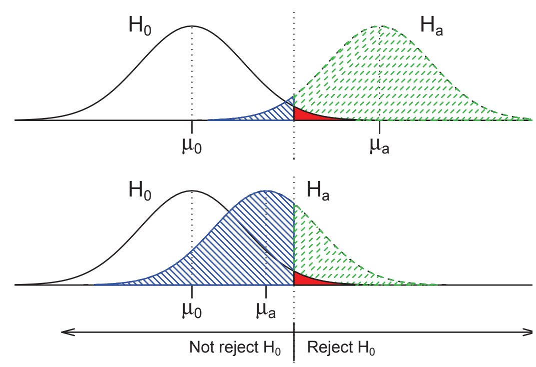
Trade-off Between Type I and Type II Errors

We wish to minimize both the chances to make type I error and type II error. Unfortunately, there is a trade-off.



Whenever we reduce P(Type I error) (area of the red filled region), P(Type II error) (area of the blue shaded region) is increased, and vice versa.

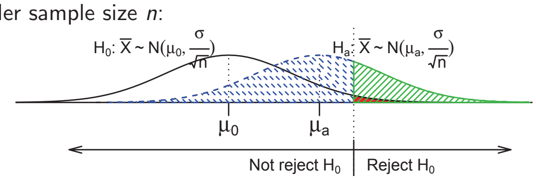
The Smaller the Difference $\mu_a - \mu_0$, The Less the Power



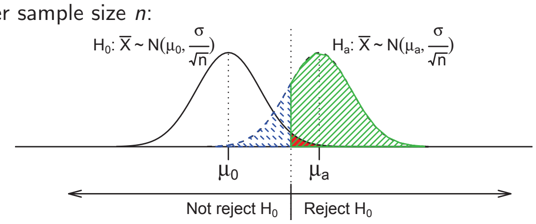
If μ_a is closer to μ_0, the two densities overlap more. The significance level (red region) can be maintained at α, at the cost of **increased Type II error** (blue region) and **less power** (red + green region).

The Larger the Sample Size, the More the Power

Smaller sample size n:



Larger sample size n:



The larger the sample size n, the more concentrate the distribution of the sample mean $\overline{X}$, and the larger the power.

Ways to Increase the Power

- **Increase** α. A 5% test of significance will have a greater chance of rejecting H_0 than a 1% test because the strength of evidence required for rejection is less.
- Consider a particular μ_a that is **farther away from** μ_0. Values of μ that are in H_a but lie close to the hypothesized value μ_0 are harder to detect than values of μ that are far from μ_0.
- **Increase the sample size** n. More data will provide more information about μ so we have a better chance of distinguishing values of μ.
- **Decrease** σ**.** Improving the precision of measurements and restricting attention to a subpopulation are two common ways to decrease σ.

Example: Does Exercise Make Strong Bones?

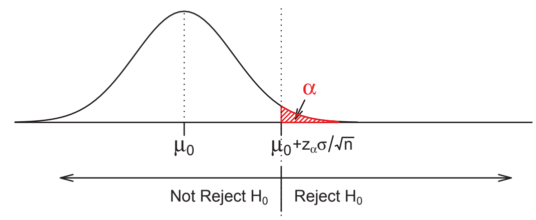
Can a 6-month exercise program increase the total body bone mineral content (TBBMC) of young women? A team of researchers is planning a study to examine this question.

- Based on the results of a previous study, they are willing to assume that $\sigma = 2$ for the percent change in TBBMC over the 6-month period.
- A 1% change in TBBMC would be considered important
- What is the chance of detecting a change of 1% or larger with a sample size of 25? Use $\alpha = 0.05$.

Example – Power Computation (1)

Let μ denote the mean percent change. Under $H_0 : \mu = \mu_0 = 0$, we have

$$\overline{X} \sim N(\mu_0, \frac{\sigma}{\sqrt{n}}) = N(0, \frac{2}{\sqrt{25}}) = N(0, 0.4)$$

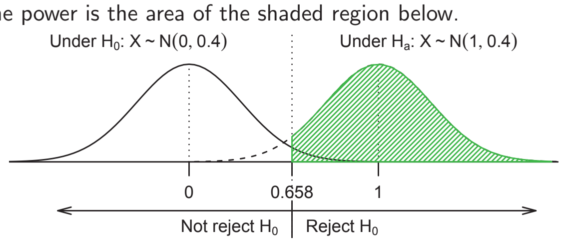


For a one-sided test, we will reject H_0 at level $\alpha = 0.05$ if

$$\overline{X} > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} = 0 + 1.645 \frac{2}{\sqrt{25}} = 0.658.$$

Example – Power Computation (2)

- As a 1% change in TBBMC would be considered important, we consider the power of the test at $H_a : \mu_a = 1$.
- Under $H_a : \overline{X} \sim N(\mu_a, \sigma/\sqrt{n}) = N(1, 2/\sqrt{25}) = N(1, 0.4)$, so the power is the area of the shaded region below.



- The cutoff 0.658 has a z-score of $\frac{0.658 - 1}{0.4} = -0.855$. From the normal table we see the area corresponds to $z = -0.855$ is about 0.1963. So the power is $1 - 0.1963 = 0.8037$.