11. Introduction to Hypothesis Testing. Large-Sample Test about a Mean

Hypothesis testing

The purpose of hypothesis testing is to determine whether there is enough statistical evidence in favor of a certain belief about a parameter.

Examples:

Is there any statistical evidence in a random sample of potential customers that supports the hypothesis that more than 50% of the potential customers will purchase new products?

Is a new drug effective in curing a certain disease? A sample of patients is randomly selected. Half of them are given the drug where half are given a placebo (sugar pill). The improvement in the patients' conditions is then measured and compared.

Concept of hypothesis testing: two hypotheses

There are two hypotheses about a population parameter(s).

- The *null hypothesis,* denoted by H_0, is a claim about a population parameter that is *initially* assumed to be true.
- The *alternative (research) hypothesis,* H_a, is a competitive claim about the same parameter.

The null hypothesis will be rejected in favor of the alternative if a sample provides a strong evidence that H_0 is false. If the sample do not suggest such evidence, H_0 will not be rejected.

Concept of hypothesis testing: two possible conclusions

Thus we can have two possible conclusions:

- **Reject H_0, accept H_a**
- **Fail to reject H_0, reject H_a**

Concept of hypothesis testing: setup

The form of the null hypothesis is

$$H_0\text{: } population\ parameter = test\ value$$

Here the *test value* is a specific number determined by the problem context.

For the alternative hypothesis we can have *one* of the following three possibilities:

$$H_a\text{: } population\ parameter > test\ value$$

$$H_a\text{: } population\ parameter < test\ value$$

$$H_a\text{: } population\ parameter \neq test\ value$$

Which one will be used depends on the problem.

Concept of hypothesis testing: type I and II errors

Two types of errors are possible when the decision whether to reject the null hypothesis is made.

		The actual state	
		H_0 is true	H_a is true
Our conclusion	Reject H_a	OK	Type II error
	Accept H_a	Type I error	OK

- Probability of Type I error is denoted by α and called the *significance level.*
- Probability of Type II error is denoted by β.

Concept of hypothesis testing: type I and II errors

Two errors are *inversely* related: if we try to reduce one error the other one will increase. Therefore, we have to balance the consequences of Type I and Type II errors.

The accepted practice is to employ the largest α that is tolerable for the problem, and then we try to do our best in minimizing β. The significance level α is always under our control.

Concept of hypothesis testing: test statistic

Once a sample is collected, the typical testing goes like this.

1. Assume the null hypothesis is true.

2. Build a *test statistic* (based on the sample) related to the parameter of interest.

3. Pose the following question. How probable is it to obtain a test statistic value at least as extreme as the observed test statistic if we assume the null hypothesis is true?

4. Make a conclusion based on your answer to that question.

Concept of hypothesis testing: making conclusion

There are two ways how we can answer the question in step 3.

- *The rejection region approach.* Depending on (1) the type of the alternative, (2) distribution of the test statistic under H_0, and (3) chosen significance level α we can construct a region that corresponds to extreme values of test statistics. If our test statistic is in that region we reject H_0, and accept H_a.
- *The p-value approach*. Again depending on (1) the type of the alternative, (2) distribution of the test statistic under H_0, and (3) chosen significance level α we calculate the probability (called p-value) of obtaining a test statistic value at least as extreme as our test statistic, assuming that H_0 is true. If this probability, or p-value, is less than α, then we reject H_0, and accept H_a.

Testing population mean: z-test about μ

The z-test about μ is employed when one wants to check a claim about the population mean, and the sample size n is large, typically, $n \geq 30$.

z-test about μ: setup

The null hypothesis:

$$H_0: \mu = \mu_0$$

Possible alternatives:

$$H_a: \mu > \mu_0$$

$$H_a: \mu < \mu_0$$

$$H_a: \mu \neq \mu_0$$

Which type of alternative and what test value μ_0 will be used is determined by the problem context.

z-test about μ: test statistic

To test the research hypothesis we use the following test statistic

$$z = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

If H_0 is true then by the central limit theorem and the law of large numbers random variable z has approximately standard normal distribution. We use this fact to make our judgment whether the sample is consistent with the null or the alternative.

z-test about μ: rejection region

Three possible expressions for the rejection region:

- If we test $H_a\colon \mu > \mu_0$, then $RR = [z_\alpha, +\infty]$
- If we test $H_a\colon \mu < \mu_0$, then $RR = [-\infty, -z_\alpha]$
- If we test $H_a\colon \mu \neq \mu_0$, then $RR = \left[-\infty, -z_{\alpha/2}\right] \cup [z_{\alpha/2}, +\infty]$

The rejection region tells us which values of the test statistic z are not consistent with the null hypothesis H_0 when it is tested versus a given alternative H_a.

z-test about μ: conclusion based on rejection region

The decision depends on the relationship between the rejection region and the test statistic. The universal rule is

- If the test statistic falls *inside* the rejection region we *accept* H_a, the research claim.
- If the test statistic falls *outside* the rejection region we *reject* H_a.

z-test about μ: p-value method

Instead of the rejection region approach we can calculate p-value. There are three formulas depending on a type of the alternative.

- If we test $H_a: \mu > \mu_0$, then p-value $= P(Z > z)$
- If we test $H_a: \mu < \mu_0$, then p-value $= P(Z < z)$
- If we test $H_a: \mu \neq \mu_0$, then p-value $= 2 \times P(Z > |z|)$

p-value is the probability of obtaining a test statistic value at least as extreme as our observed test statistic, assuming that H_0 is true. If this probability is small it means that our observed test statistic is too extreme to be consistent with H_0.

z-test about μ: conclusion based on p-value

The universal rule is

- **If p-value $< \alpha$, then we *accept* H_a.**
- **If p-value $> \alpha$, then we *reject* H_a.**

Example

Example 1. In a nationwide opinion poll based on a random sample of 240 people, one question is: “How do you rate the ethics of business executives of large companies?” A rating of 3 means “no better or worse than most people”, a rating of 1 is “much better than most people”, and 5 is “much worse than most people” the mean rating is 3.1 and the standard deviation is 0.9. Can we infer at significance level 5% that respondents rated the ethics of executives at different level in comparison to most people?

1. Setup

Let μ be the true (nationwide) mean rating of business executives of large companies.

$$H_0: \mu = 3$$
$$H_a: \mu \neq 3$$

2. Test statistic

Test statistic is given by

$$z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{3.1 - 3}{.9/\sqrt{240}} \approx 1.72$$

3. Rejection region

The rejection region is given by

$$
\begin{aligned}
\boldsymbol{RR} &= \left[-\infty, -\boldsymbol{z}_{\boldsymbol{\alpha/2}}\right] \cup \left[\boldsymbol{z}_{\boldsymbol{\alpha/2}}, +\infty\right] \\
&= \left[-\infty, -\boldsymbol{z}_{.\boldsymbol{025}}\right] \cup \left[\boldsymbol{z}_{.\boldsymbol{025}}, +\infty\right] \\
&= \left[-\infty, -\boldsymbol{1.96}\right] \cup \left[\boldsymbol{1.96}, +\infty\right]
\end{aligned}
$$

4. Conclusion

Since the test statistic falls outside the rejection region we reject H_a, that is, there is no difference in ratings between executives and most people.

3`. p-value

The p-value of the test is

p-value $= 2 \times P(Z > 1.72) = 2 \times .0427 = .0854$

4`. Conclusion based on p-value

Since p-value $=.0854 > \alpha = .05$, we reject H_a, that is, there no difference in ratings between executives and most people.

An important remark: rejection region approach and p-value approach are mathematically equivalent.

z-test about μ: exercise

Exercise 1. According to the U. S. Department of Commerce, the average price for a new home topped $200,000 for the first time in 1999. In November 1999, the average new-home price was $209,700 (Wall Street Journal Interactive Edition, Jan. 7, 2000). The prices of a random sample of 32 new homes sold in November 2000 yielded $\overline{X} = 216981$ and $s = 19805$.

a. What are the appropriate null and alternative hypotheses to test whether the mean price of a new home in November 2000 exceeds $209,700?

b. Compute and interpret the p-value of the test. Do the data provide sufficient evidence to conclude that the mean new-home price in November 2000 exceeded the reported mean price of November 1999?