VE401 RC Week7

Wang Yangyang

UM-SJTU JI

2022 Spring

Outline

1. **Hypothesis Tests**
   - Fisher's Null Hypothesis Test
   - Neyman-Pearson Decision Theory
   - Null Hypothesis Significance Testing (Horrible!)
2. **Single Sample Tests**
   - Single Sample Tests for the Mean and Variance

Outline

1. **Hypothesis Tests**
   - Fisher's Null Hypothesis Test
   - Neyman-Pearson Decision Theory
   - Null Hypothesis Significance Testing (Horrible!)
2. **Single Sample Tests**
   - Single Sample Tests for the Mean and Variance

Null Hypotheses

Null hypotheses take one of three forms:

$$H_0 : \theta = \theta_0$$
$$H_0 : \theta \leqslant \theta_0$$
$$H_0 : \theta \geqslant \theta_0$$

Our goal is to reject the null hypothesis H_0.
We either

- fail to reject H_0 or
- reject H_0 at the [***p-value***] level of significance.

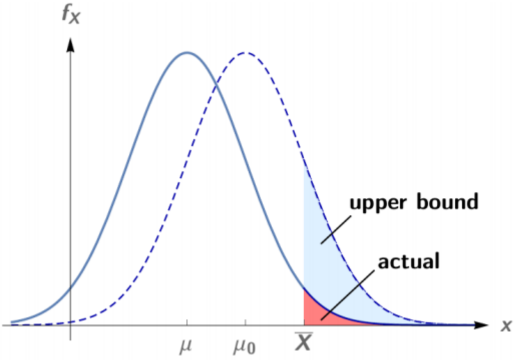
One-Tailed Test

1. Take a random sample of size n and find the value for the estimator $\widehat{\theta}$. Set One-tailed test is the test of a hypothesis of the form
$$H_0 : \theta \leqslant \theta_0$$
or
$$H_0 : \theta \geqslant \theta_0$$
2. Suppose the hypothesis is $H_0 : \theta \leqslant \theta_0$. Next we calculate the significance or P-value of the test, which is the probability of obtaining the measured value of $\widehat{\theta}$ or a larger result if $\theta = \theta_0$, e.g., for $H_0 : \mu \leqslant 26$, the P-value is
$$P[\bar{X} \geqslant \bar{x} \mid \mu = 26]$$

One-Tailed Test

After calculation based on data, we have upper bound of the probability of obtaining the measured value of $\widehat{\theta}$ or a larger result if H_0 is true, denoted as $P[D \mid H_0]$, and

$$P[D \mid H_0] \leqslant P\text{-value}$$



$H_0: \mu \leq \mu_0$

Two-Tailed Test

1. Take a random sample of size n and find the value for the estimator $\widehat{\theta}$. Set One-tailed test is the test of a hypothesis of the form
$$H_0 : \theta = \theta_0$$
2. Next we calculate the significance or P-value of the test, which is the probability of obtaining the measured value of $\widehat{\theta}$ or a larger result if $\theta = \theta_0$. e.g., for $H_0 : \mu = 40$, find
$$P[\bar{X} \geqslant \bar{x} \mid \mu = 40] \quad \text{or} \quad P[\bar{X} \leqslant \bar{x} \mid \mu = 40]$$
based on different situation.
3. After calculation based on data, we need to double the result to get P-value. Then reject H_0 if P-value is small.

Outline

1. **Hypothesis Tests**
   - Fisher's Null Hypothesis Test
   - Neyman-Pearson Decision Theory
   - Null Hypothesis Significance Testing (Horrible!)
2. **Single Sample Tests**
   - Single Sample Tests for the Mean and Variance

Two Types of Error

Given a choice between ***null hypotheses*** H_0 and ***alternative hypotheses*** H_1, there are 4 possible outcomes of the decision-making:

1. We reject H_0 (accept H_1) when H_0 is false.
2. We reject H_0 (accept H_1) even though H_0 is true (***Type I error***).
3. We fail to reject H_0 even though H_1 is true (***Type II error***).
4. We fail to reject H_0 when H_0 is true.

$$\begin{aligned}\alpha := P[\text{ Type I error }] &= P[\text{ reject } H_0 \mid H_0 \text{ true }] \\ &= P[\text{ accept } H_1 \mid H_0 \text{ true }]\end{aligned}$$

$$\begin{aligned}\beta := P[\text{ Type II error }] &= P[\text{ fail to reject } H_0 \mid H_1 \text{ true }] \\ &= P[\text{ accept } H_0 \mid H_1 \text{ true }]\end{aligned}$$

$$\begin{aligned}\text{Power } := 1 - \beta &= P[\text{ reject } H_0 \mid H_1 \text{ true }] \\ &= P[\text{ accept } H_1 \mid H_1 \text{ true }]\end{aligned}$$

Procedure

1. Set up a ***null hypothesis*** H_0 and an ***alternative hypothesis*** H_1.
2. Determine a desirable α and β.
3. Use α and β to determine the appropriate ***sample size*** n.
4. Use α and n to determine the ***critical region***.
5. Obtain sample statistics, and reject H_0 at significance level α and accept H_1 if the test statistic falls into critical region. Otherwise, accept H_0.

Step 1 and 2 are significant before obtaining data.
Step 3 and 4 requires clear calculation and analysis.
The sample size, the critical region must be fixed before data are obtained.

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 not more than α.
- If the value of the test statistic falls into the critical region, then we reject H_0.

Example

Suppose the sample mean $\bar{X}$ follows a normal distribution with unknown mean μ and known variance σ^2, with $H_0 : \mu = \mu_0$.

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

α and the Critical Region

Example

Z follows a standard normal distribution. The critical region is determined by

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

or

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

Remarks

1. We only care about the probability of committing an error if H_0 is falsely rejected.
2. Only H_0 plays a role in the calculation of the critical region. H_1 does not enter into the discussion.

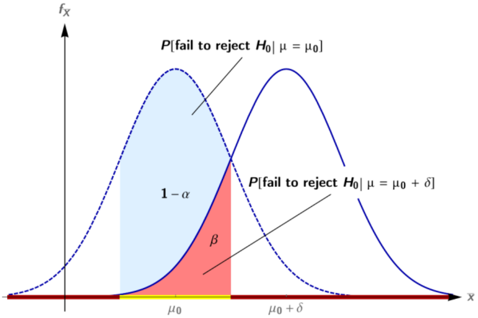
β and the Sample Size

The second type of error concerns failing to reject H_0 even though H_1 is true. Suppose $H_0 : \mu = \mu_0, \quad H_1 : |\mu - \mu_0| \geqslant \delta_0$.
Suppose the true value of mean is $\mu = \mu_0 + \delta, \delta > 0$. The test statistic

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathrm{N}(\delta\sqrt{n}/\sigma, 1)$$

Supposing that α has been fixed, we will fail to reject H_0 if

$$-z_{\alpha/2} \leqslant Z \leqslant z_{\alpha/2}$$



β and the Sample Size

With true mean $\mu = \mu_0 + \delta$, the test statistic $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathrm{N}(\delta\sqrt{n}/\sigma, 1)$.

$$P[\text{ fail to reject } H_0 \mid \mu = \mu_0 + \delta] = \frac{1}{\sqrt{2\pi}} \int_{-z_{\alpha/2}}^{z_{\alpha/2}} e^{-(t - \delta\sqrt{n}/\sigma)^2/2}\, \mathrm{d}t$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-z_{\alpha/2} - \delta\sqrt{n}/\sigma}^{z_{\alpha/2} - \delta\sqrt{n}/\sigma} e^{-t^2/2}\, \mathrm{d}t \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_{\alpha/2} - \delta\sqrt{n}/\sigma} e^{-t^2/2}\, \mathrm{d}t \stackrel{!}{=} \beta,$$

where we set $-z_\beta = z_{\alpha/2} - \delta\sqrt{n}/\sigma$.

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

where $z_{\alpha/2}$ and z_β satisfies that

$$\Phi\left(z_{\alpha/2}\right) = 1 - \alpha/2, \quad \Phi\left(z_\beta\right) = 1 - \beta$$

β and the Sample Size

Remarks

The result applies to ***normal*** case.

1. A desired (small) β can be attained by choosing an appropriate sample size n.
2. Power $= 1 - \beta$. Generally, we would like the probability of rejecting H_0 if the alternative hypothesis is true to be high, i.e., Power $= 1 - \beta$ to be high.

In practice, it may be difficult to perform integral calculations every time to find the probability of failing to reject H_0 as an integral. For this reason, we can refer to ***OC curves*** to read β directly. Each single curve represents a choice of test parameters α and n.

OC Curves

- For normal test, calculate

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

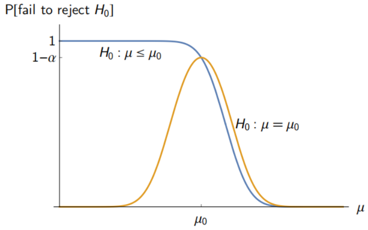
Note. The ***abscissa*** might change corresponding to the distribution of test.

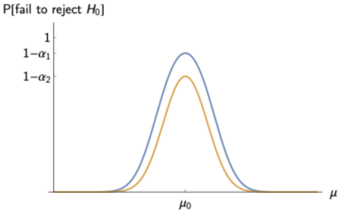
- Look up in OC curve for sample size n.

Remarks

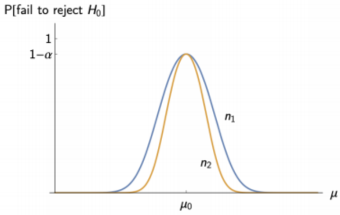
- Be careful with the ***horizontal axis*** for each type of OC curve.
- For the same distribution, OC curves are different for one-sided or two-sided tests.

OC Curves



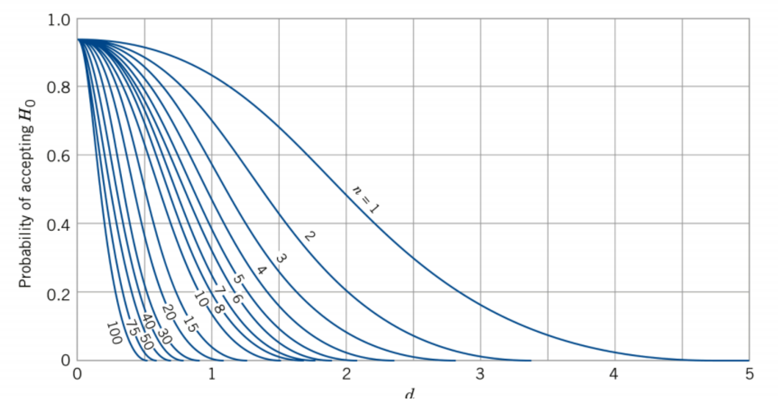


(a) The same null hypothesis for fixed n.



(b) The same null hypothesis for fixed α.

OC Curves



Probability of accepting H_0
1.0
0.8
0.6
0.4
0.2
0
0
1
2
3
4
5
d
n = 1
2
3
4
5
6
7
8
10
15
20
30
40
50
75
100

Relation between CI and CR

We have seen that the two-tailed null hypothesis $H_0 : \mu = \mu_0$ is rejected if

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

This is equivalent to

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

Suppose we would like to estimate the mean μ of a sample $X_1, \ldots, X_n$ of size n.

- *Confidence interval*. Given a sample data with specific values, the *CI* gives an interval for the unknown mean μ.
- *Critical region*. Given a null value μ_0, the critical region gives an interval for sample mean $\bar{X}$ before obtaining

The null hypothesis H_0 is rejected $\Leftrightarrow$ $\bar{X}$ lies in the critical region $\Leftrightarrow$ null value μ_0 lies outside the confidence interval.

Relation between Two Tests

1. ***Neyman-Pearson***: $\bar{X}$ lies in the critical region for α if and only if the null value μ_0 does not lie in a $100(1-\alpha)\%$ two-sided confidence interval for μ.
   Fisher: H_0 is rejected at significance level α if and only if the null value μ_0 does not lie in a $100(1-\alpha)\%$ two-sided confidence interval for μ.
   (This generalizes to one-sided tests and is also true for other distributions.)
2. If the P-value in Fisher's test is no greater than the value of α in Neyman-Pearson's decision process, then H_0 is rejected and H_1 accepted. Otherwise, H_0 is not rejected.

Outline

1. **Hypothesis Tests**
   - Fisher's Null Hypothesis Test
   - Neyman-Pearson Decision Theory
   - Null Hypothesis Significance Testing (Horrible!)

2. **Single Sample Tests**
   - Single Sample Tests for the Mean and Variance

Procedure

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.
   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.
3. In either case, there is *no meaningful discussion* of β, since H_1 is exactly the negation of H_0.

Criticism

- A small P-value does not guarantee that a large probability that H_0 is false. Fisher did not intend for a small P-value to lead to a clear rejection of H_0, but only to serve as ***evidence against*** H_0 if little else is known.
- Rejecting H_0 based on $\alpha = 0.05$ or 0.01 or any other ***P-value*** is arbitrary.
- NHST is actually ***biased against*** failing to reject $\boldsymbol{H}_0$. From a Bayesian point of view, it is far too easy to reject H_0 because $P[H_0]$ does not enter into NHST.
- A ***two-sided test*** such as $H_0 : \theta = \theta_0, H_1 : \theta \neq \theta_0$ is meaningless.
- The ***power*** (β) of the test is not properly defined, since H_1 is just the alternative "not H_0 " rather than referring to a distinct value θ_1.

Outline

1. Hypothesis Tests
   - Fisher's Null Hypothesis Test
   - Neyman-Pearson Decision Theory
   - Null Hypothesis Significance Testing (Horrible!)

2. Single Sample Tests
   - Single Sample Tests for the Mean and Variance

Test for Mean (Variance Known)

Let $X_1, \ldots, X_n$ be a random sample of size n from a ***normal*** distribution with ***unknown*** mean μ and ***known*** variance σ^2. Let μ_0 be a null value of the mean. Then the test statistic is given by

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

We reject at significance level α

- $H_0 : \mu = \mu_0$ if $|Z| > z_{\alpha/2}$,
- $H_0 : \mu \leqslant \mu_0$ if $Z > z_\alpha$,
- $H_0 : \mu \geqslant \mu_0$ if $Z < -z_\alpha$.

OC curve. The abscissa is defined by

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

Test for Mean (Variance Unknown)

Let $X_1, \ldots, X_n$ be a random sample of size n from a ***normal*** distribution with ***unknown*** mean μ and ***unknown*** variance σ^2. Let μ_0 be a null value of the mean. Then the test statistic is given by

$$T_{n-1} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

We reject at significance level α

- $H_0 : \mu = \mu_0$ if $|T_{n-1}| > t_{\alpha/2,n-1}$,
- $H_0 : \mu \leqslant \mu_0$ if $T_{n-1} > t_{\alpha,n-1}$,
- $H_0 : \mu \geqslant \mu_0$ if $T_{n-1} < -t_{\alpha,n-1}$.

OC curve. The abscissa is defined by

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

Comments on T-test

1. The T-distribution may be used for $\frac{\bar{X}-\mu_0}{S/\sqrt{n}}$ when a sample is obtained from a normal population.
2. If a sample is obtained from a non-normal population, then for large to medium sample sizes ($n \geqslant 25$) it can be shown that violating the normality assumption does not significantly change α and β.
3. For small sample sizes, a T-test cannot be used and an alternative (non-parametric) test must be employed.

Comments on Abscissa

Abscissa of OC Curves.

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

where σ is the unknown standard deviation of the random variable. We have three options:

1. If available, we can use prior experiments to insert a rough estimate for σ
2. We can express the difference $\delta = |\mu - \mu_0|$ relative to σ, e.g., prescribing $d = \delta/\sigma < 1$ for a small difference in the mean or $d = \delta/\sigma < 2$ for a moderately large difference.
3. We substitute the sample standard deviation s for σ.

Test for Variance

Let $X_1, \ldots, X_n$ be a random sample of size n from a ***normal*** distribution with ***unknown*** variance σ^2. Let σ_0^2 be a null value of the variance. Then the test statistic is given by

$$\chi^2_{n-1} = \frac{(n-1)S^2}{\sigma_0^2}.$$

We reject at significance level α

- $H_0 : \sigma = \sigma_0$ if $\chi^2_{n-1} \in \left(0, \chi^2_{1-\alpha/2,n-1}\right) \cup \left(\chi^2_{\alpha/2,n-1}, \infty\right)$,
- $H_0 : \sigma \leqslant \sigma_0$ if $\chi^2_{n-1} > \chi^2_{\alpha,n-1}$,
- $H_0 : \sigma \geqslant \sigma_0$ if $\chi^2_{n-1} < \chi^2_{1-\alpha,n-1}$.

OC curve. The abscissa is defined by

$$\lambda = \frac{\sigma}{\sigma_0}$$

Comments on Chi-squared Test

- If the distribution is non-normal, we cannot use Chi-squared test.
- ** Normality of the data must first be tested if we do not know the distribution.

Abscissa of OC Curves.

$$\lambda = \frac{\sigma}{\sigma_0}$$

Note that the OC curves for the left- and right-tailed chi-squared distributions are distinct.

End

Credit to Zhanpeng Zhou (TA of SP21)
Credit to Fan Zhang (TA of SU21)
Credit to Liying Han (TA of SP21)
Credit to Zhenghao Gu (TA of SP20)