Lecture #11

Hypothesis Testing: One Sample

BMIR Lecture Series on Probability and Statistics

Ching-Han Hsu, Ph.D.
Department of Biomedical Engineering
and Environmental Sciences
National Tsing Hua University

Statistical Hypotheses: Motivations

- We have discussed about how to construct a confidence interval estimate of a parameter from sample data.
- In many applications, we need to decide which of two competing claims or statements about some parameter is true.
- The statement is called **hypothesis**.
- The decision making procedure is called **hypothesis testing**.

Definition

A **statistical hypothesis** is a claim or assertion about the parameters of one or more populations.

Statistical Hypotheses:Example I

Example

Suppose that we are measuring the burning rate of a solid propellant for some air crew escape system. We are interested in deciding whether or not the mean burning rate is 50 centermeters per second. We may express this formally as:

$$H_0 : \mu = 50 \text{ cm/sec} \quad (1)$$
$$H_1 : \mu \neq 50 \text{ cm/sec} \quad (2)$$

- The statement Eq.(1) is called the **null hypothesis**.
- The statement Eq.(2) is called the **alternative hypothesis**.

Statistical Hypotheses:Example II

- Since the alternative statement specifies the value of μ can be either greater or less than 50 cm/sec, this is called a **two-sided alternative hypothesis**.
- We can also formulate a **one-sided alternative hypothesis**

$$H_0 : \mu = 50 \text{ cm/sec} \quad (3)$$

$$H_1 : \mu < 50 \text{ cm/sec} \quad (4)$$

or,

$$H_0 : \mu = 50 \text{ cm/sec} \quad (5)$$

$$H_1 : \mu > 50 \text{ cm/sec} \quad (6)$$

- Note that hypotheses are always statements of the **population (or distribution) parameter** under study, not statements about sample.

Statistical Hypotheses

The value of the population parameter (ex. $\mu = 50$ cm/sec) is usually specified in one of three ways:

1. It may result from past experience, knowledge, or experiment. The object of hypothesis testing is usually to determine whether the parameter value has been changed.
2. The value may be determined from some theory or model. The objective is to verify the theory or model.
3. The value results from external considerations, such as design or specification. The objective is conformance testing.

Test of Statistical Hypotheses

- We want to test:

$$H_0 : \mu = 50 \text{ cm/sec}$$
$$H_1 : \mu \neq 50 \text{ cm/sec}$$

- Suppose that a sample of $n = 10$ specimens is tested and that the sample mean $\bar{x}$ of burning rate is computed.
- The sample mean is an estimate of the true population mean, i.e. the population parameter of interest.
- If $\bar{x}$ is close to the hypothesized true mean ($\mu = 50$ cm/sec in this example, H_0 is accepted.
- Otherwise, H_0 is rejected, i.e., the alternative hypothesis H_1 is true.
- The sample mean $\bar{x}$ is the **test statistic** in this example.

Test of Statistical Hypotheses

- How to implement the task:

$$H_0 : \mu = 50 \text{ cm/sec}; \quad H_1 : \mu \neq 50 \text{ cm/sec}$$

- Suppose that if $48.5 \leq \bar{x} \leq 51.5$, we will not reject the null hypothesis H_0.
- If either $\bar{x} \leq 48.5$ or $\bar{x} \geq 51.5$, we will reject the null hypothesis in favor of the alternative hypothesis H_1.
- $48.5 \leq \bar{x} \leq 51.5$ is called the **acceptance region**.
- $\bar{x} \leq 48.5$ or $\bar{x} \geq 51.5$ is the **critical region** of the test.
- The boundaries between acceptance and critical regions are called the **critical values** (48.5 and 51.5 in this example).

Test of Statistical Hypotheses

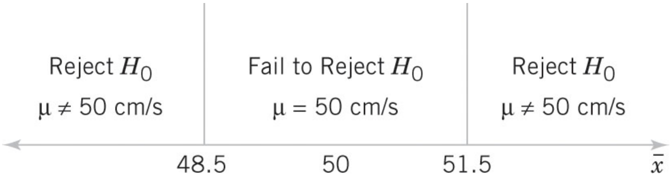

Figure 1: Decision criteria for testing $H_0 : \mu = 50$ cm/sec versus $H_1 : \mu \neq 50$ cm/sec.

Errors in Hypothesis Testing

- Rejecting the null hypothesis H_0 when it it true is defined as **type I error**.
- Failing to reject the null hypothesis H_0 when it it false is defined as **type II error**.
- In testing any hypothesis, there are four different situations determining whether final decision is correct or in error:

Table 1: Decisions in Hypothesis Testing

Decision	H_0 is True	H_1 is True
Fall to Reject H_0	no error	type II error
Reject H_0	type I error	no error

Ching-Han Hsu, Ph.D.

Type I Error: I

- The probability of making a type I error is denoted by α:

$$\begin{aligned} \alpha &= P(\text{type I error}) \\ &= P(\text{reject } H_0 \text{ when } H_0 \text{ is true}) \end{aligned} \quad (7)$$

  The probability of type I error α is also called the **significant level**, the **α-error** or the **size of test**.

- In the pollutant example, if the true mean $\mu = 50$ cm/sec and the standard deviation $\sigma = 2.5$, then the sample mean is normal with standard deviation $\sigma/\sqrt{n} = 2.5/\sqrt{10} = 0.79$.

Type I Error: II

- The corresponding probability of type I error is

$$
\begin{aligned}
\alpha &= P(\bar{X} < 48.5; \mu = 50) + P(\bar{X} > 51.5; \mu = 50) \\
&\left(z_1 = \frac{48.5 - 50}{0.79} = -1.90,\ z_2 = \frac{51.5 - 50}{0.79} = 1.90 \right) \\
&= P(Z < -1.90) + P(Z > 1.90) \\
&= 0.0287 + 0.0287 = 0.0574
\end{aligned}
$$

- This implies that 5.74% of all random samples would lead to rejection of the hypothesis $H_0 : \mu = 50$ cm/sec when the true mean burning rate is really 50 cm/sec.
- The more significant the test is, the more restrictive the test becomes and the higher the type I error will be.

Type I Error: III

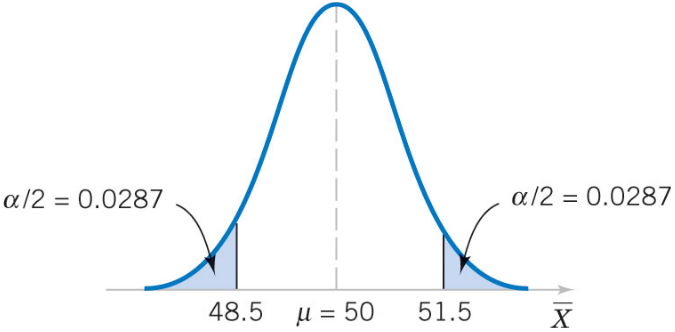

Figure 2: The critical region for $H_0 : \mu = 50$ cm/sec versus $H_1 : \mu \neq 50$ cm/sec and $n = 10$.

Type I Error: IV

- We can reduce α by increasing the sample size. If $n = 16$, then $\sigma/\sqrt{n} = 0.625$. We have

$$
\begin{aligned}
z_1 &= \frac{48.5 - 50}{0.625} = -2.40 \\
z_2 &= \frac{51.5 - 50}{0.625} = 2.40 \\
\alpha &= P(Z < -2.40) + P(Z > 2.40) \\
&= 0.0082 + 0.0082 = 0.0164.
\end{aligned}
$$

Type II Error: I

- The probability of making a type II error is denoted by β:

$$\begin{aligned} \beta &= P(\text{type II error}) \\ &= P(\text{fail to reject } H_0 \text{ when } H_0 \text{ is false}) \end{aligned} \tag{8}$$

- To calculate β (sometimes called the **β-error**), we must have a specific alternative hypothesis, i.e, we must have a particular value of μ.
- A type II error will be committed if the sample mean $\bar{X}$ falls between 48.5 and 51.5, when $\mu = 52$ and $n = 10$.

Type II Error: II

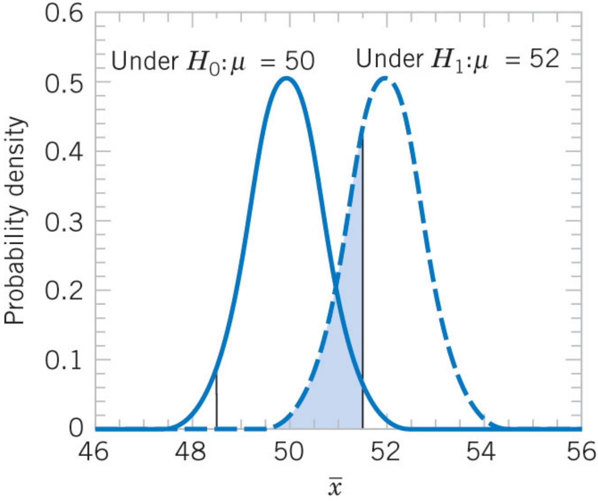

Figure 3: The probability of type II error when $\mu = 52$ and $n = 10$.

Type II Error: III

- The corresponding probability of type II error is

$$
\begin{aligned}
\beta &= P(48.5 \leq \bar{X} \leq 51.5; \mu = 52) \\
z_1 &= \frac{48.5 - 52}{0.79} = -4.43 \\
z_2 &= \frac{51.5 - 52}{0.79} = -0.63 \\
\beta &= P(-4.43 \leq Z \leq -0.63) \\
&= P(Z \leq -0.63) - P(Z \leq -4.43) \\
&= 0.2643 - 0.0000 = 0.2643
\end{aligned}
$$

- If we are testing $H_0 : \mu = 50$ against $H_1 : \mu \neq 50$ with $n = 10$, and the true value of the mean is $\mu = 52$, the probability that we **fail to reject** the false null hypothesis is 0.2643.

Type II Error: IV

- If the true mean is $\mu = 50.5$, then the value of β is

$$\beta = P(48.5 \le \bar{X} \le 51.5; \mu = 50.5)$$
$$z_1 = \frac{48.5 - 50.5}{0.79} = -2.53$$
$$z_2 = \frac{51.5 - 50.5}{0.79} = 1.27$$
$$\beta = P(-2.53 \le Z \le 1.27)$$
$$= P(Z \le 1.27) - P(Z \le -2.53)$$
$$= 0.8980 - 0.0057 = 0.8923$$

- As the true mean approaches the hypothesized value, the probability of making type II error increase rapidly.

Type II Error: V

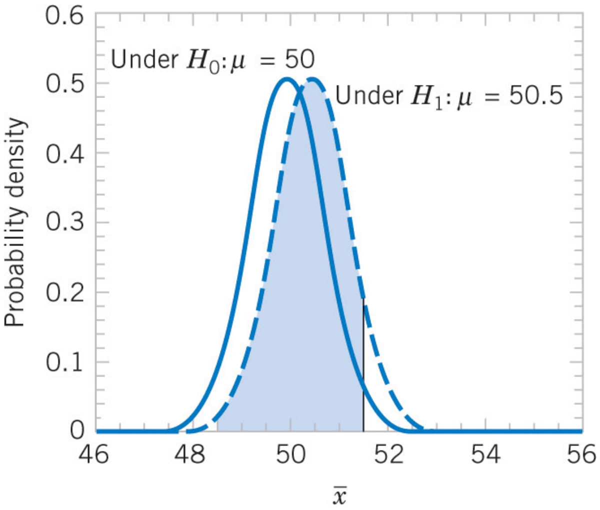

Figure 4: The probability of type II error when $\mu = 50.5$ and $n = 10$.

Type II Error: VI

- When $n = 16$, then the standard deviation of $\bar{X}$ becomes $\sigma/\sqrt{n} = 2.5/\sqrt{16} = 0.625$. We have

$$
\begin{aligned}
z_1 &= \frac{48.5 - 52}{0.625} = -5.60 \\
z_2 &= \frac{51.5 - 52}{0.625} = -0.80 \\
\beta &= P(-5.60 \leq Z \leq -0.80) \\
&= P(Z \leq -0.80) - P(Z \leq -5.60) \\
&= 0.2119 - 0.0000 = 0.2119.
\end{aligned}
$$

- Increasing the sample size results in a decrease in the probability of type II error.

Results and Discussions of This Section

- The results from this section and a few other similar calculations are summarized in the following table:

Acceptance Region	Sample Size	α	β at $\mu = 52$	β at $\mu = 50.5$
$48.5 < \bar{x} < 51.5$	10	0.0576	0.2643	0.8923
$48 < \bar{x} < 52$	10	0.0114	0.5000	0.9705
$48.81 < \bar{x} < 51.19$	16	0.0576	0.0966	0.8606
$48.42 < \bar{x} < 51.58$	16	0.0114	0.2515	0.9578

Results and Discussions of This Section

- The size of the critical region, and consequently the probability of a type I error α, can always be reduced by appropriate selection of the critical values.
- Type I and type II errors are related. When sample size n is fixed, a decrease in the probability of one type of error always results in an increase in the probability of the other.
- An increase in sample size will generally reduce both α and β, provided that the critical values are held constant.
- When the null hypothesis is false, β increases as the true value of the parameter approaches the value hypothesized in the null hypothesis.

Weak and Strong Conclusions

- Since we can directly control the probability of wrongly rejecting H_0, we always think of *rejection* of the null hypothesis H_0 as a **strong conclusion**.
- The probability of type II error β is not a constant, but depends on the true value of the parameter and the sample size. It is customary to think of the decision to *accept* H_0 as a **weak conclusion**, unless we know that β is acceptably small.
- Rather than saying we "**accept** H_0", we prefer to the terminology **"fail to reject H_0"**.

Power of Statistical Test

Definition

The **power** of a statistical test is the probability of rejecting the null hypothesis H_0 when the alternative hypothesis is true.

- The power is computed as $1 - \beta$.
- Power can be interpreted as the probability of correctly rejecting a false null hypothesis.
- In previous example, when the true mean is $\mu = 52$. When $n = 10$, $\beta = 0.2643$ and the power of the test is $1 - \beta = 1 - 0.2643 = 0.7357$.
- When $\mu = 50.5$ and $n = 10$, $\beta = 0.8923$ and the power is $1 - \beta = 1 - 0.8923 = 0.1077$.
- Power is a very descriptive and concise measure of the sensitivity of a statistical test, where by sensitivity we mean the ability of the test to detect differences.

Motivations

- One way to report the results of a hypothesis test is to state that the null hypothesis was or was not rejected at a specified α-value or level of significance.
- For previous example, we can say that $H_0 : \mu = 50$ was rejected at the 0.05 level of significant.
- However, we have no idea about whether the computed value of the test statistic was just barely in the rejection region or whether it was very far into this region.
- This approach may be unsatisfactory because we might be uncomfortable with the risks implied by $\alpha = 0.05$.

P-Value I

A P**-value** conveys how much information about the strength of evidence *against* H_0, and allows an individual decision maker to draw a conclusion at any specified significant level α.

Definition

The P-value is the smallest level of significance which H_0 would be rejected when a specified test procedure is used on a given data set. Once the P-value has been determined, the conclusion at any particular level α from comparing the P-value to α:

1. P-value $\leq \alpha \rightarrow$ reject H_0 at level α.
2. P-value $> \alpha \rightarrow$ do not reject H_0 at level α.

P-Value II

- Consider the two-sided hypothesis test for burning

$$H_0 : \mu = 50$$
$$H_1 : \mu \neq 50$$

with $n = 16$ and $\sigma = 2.5$.

- Suppose that the observed sample mean is $\bar{x} = 51.3$ cm/sec.
- We set the critical values at 51.3 and 48.7.
- The P-value of the test is the α associated with this critical region.

$$\begin{aligned} P\text{-value} &= 1 - P(48.7 < \bar{X} < 51.3) \\ &= 1 - P\left(\frac{48.7 - 50}{2.5/\sqrt{16}} < Z < \frac{51.3 - 50}{2.5/\sqrt{16}}\right) \\ &= 1 - P(-2.08 < Z < 2.08) = 1 - 0.962 \\ &= 0.038 \end{aligned}$$

P-Value III

- Recall that at 0.05 level of significance, the critical values of Z are -1.96 and 1.96.
- The interval derived from P-value contains the interval from 0.05 level of significance.
- $p = 0.038 < \alpha = 0.05$ in this example.
- This null hypothesis would be rejected for any $0.038 \leq \alpha$ (such as 0.05); but not rejected for a smaller significant level $0.038 > \alpha$, such as 0.01.

Connection to CI

- Suppose that the parameter θ is of interest.
- $[l, u]$ is the a $100(1-\alpha)\%$ confident interval for θ.
- The test of size α of the hypothesis:

$$H_0 : \theta = \theta_0 \tag{9}$$

$$H_1 : \theta \neq \theta_0 \tag{10}$$

  will lead to the rejection of H_0 if and only if θ is NOT in the $100(1-\alpha)\%$ confident interval $[l, u]$.
- In previous propellant problem, we have $\bar{x} = 51.3$, $\sigma = 2.5$ and $n = 16$. The 95% two-sided CI on μ is $51.3 \pm 2.5/\sqrt{16}$ which is $50.075 \le \mu \le 52.525$.
- Because $\mu = 50$ is not included in the interval $50.075 \le \mu \le 52.525$, the null hypothesis $H_0 : \mu = 50$ is rejected.

Practical Issues

Consider the propellant example. We are testing

$$H_0 : \mu = 50 \text{ cm/sec} \quad (11)$$
$$H_1 : \mu \neq 50 \text{ cm/sec} \quad (12)$$

with known $\sigma = 2.5$. Suppose that the sample mean is $\bar{x} = 50.5$ with various sample sizes n:

Sample Size n	P-value When $\bar{x} = 50.5$	Power (at $\alpha = 0.05$) When True $\mu = 50.5$
10	0.527	0.097
25	0.317	0.170
50	0.157	0.293
100	0.046	0.516
400	6.3×10^{-5}	0.979
1000	2.5×10^{-10}	1.000

Hypothesis Setting

- We first consider the hypothesis testing about the mean μ of a single normal distribution with know variance σ^2.
- We assume that a random sample $X_1, \ldots, X_n$ has been taken from the population.
- The sample mean $\bar{X}$ is an **unbiased point estimator** of μ with variance σ^2/n.
- Suppose that we wish to test the hypotheses:

$$H_0 : \mu = \mu_0 \quad (13)$$

$$H_1 : \mu \neq \mu_0 \quad (14)$$

where μ_0 is a specified constant.

Test Statistics

- The $\bar{X}$ has a normal distribution with mean μ_0 and variance σ^2/n.
- The test procedure for $H_0 : \mu = \mu_0$ uses the **test statistic**:

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

- Z_0 is the standard normal distribution.
- If $H_0 : \mu = \mu_0$ is true, the probability is $1 - \alpha$ that the test statistic Z_0 falls between $-z_{\alpha/2}$ and $z_{\alpha/2}$.
- $z_{\alpha/2}$ is the $100\alpha/2$ percentage point.
- If the observed value of the test statistic z_0 is either $z_0 > z_{\alpha/2}$ or $z_0 < -z_{\alpha/2}$, we should reject H_0.
- We shall **fail to reject** H_0, if $-z_{\alpha/2} \leq z_0 \leq z_{\alpha/2}$.

Statistical Hypotheses: Example I

Example

Suppose that we are measuring the burning rate of a solid propellant for some air crew escape system. Specifications require that the mean burning rate must be 50 cm/sec. We know that the standard deviation of burning rate is $\sigma = 2$ cm/sec. The experiment decides to specify a type I error probability or level of significant of $\alpha = 0.05$, selects a random sample of $n = 25$, and obtains a sample average burning rate of $\bar{x} = 51.3$ cm/sec. What conclusion should be drawn?

1. The parameter of interest is μ, the mean burning rate.
2. $H_0 : \mu = 50$ cm/sec.
3. $H_1 : \mu \neq 50$ cm/sec.
4. $\alpha = 0.05$.

Statistical Hypotheses: Example II

5. The test statistic:

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{51.3 - 50}{2/\sqrt{25}} = 3.25$$

6. Reject H_0 if $z_0 < -1.96$ or $z_0 > 1.96$.
7. Conclusion: Since $z_0 = 3.25 > 1.96$, we shall reject $H_0 : \mu = 50$ cm/sec at $\alpha = 0.05$ level of significance.

Note that there is a strong evidence that the mean burning rate exceeds 50 cm/sec.

One-Sided Hypotheses

Theorem

Suppose that we specify the hypotheses as

$$H_0 : \mu = \mu_0 \tag{15}$$
$$H_1 : \mu > \mu_0 \tag{16}$$

We place the critical region in the upper tail of the standard normal distribution and reject H_0 is the computed value of z_0 is too large, i.e., $z_0 > z_\alpha$. Similarly, to test

$$H_0 : \mu = \mu_0 \tag{17}$$
$$H_1 : \mu < \mu_0 \tag{18}$$

we would calculate the test statistic Z_0 and reject H_0 is the computed value of z_0 is too small, i.e., $z_0 < -z_\alpha$.

Test on Mean of Normal Dist., Variance Known

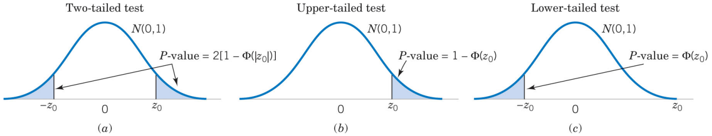

Figure 5: The distribution of Z_0 when $H_0 : \mu = \mu_0$ is true, with critical region for (a) the two-sided alternative $H_1 : \mu \neq \mu_0$, (b) the one-sided alternative $H_1 : \mu > \mu_0$, and (c) the one-sided alternative $H_1 : \mu < \mu_0$.

P-value

- If z_0 is the computed value of the test statistics, the P value is

$$P = \begin{cases} 2[1-\Phi(|z_0|)] & H_0: \mu = \mu_0,\ H_1: \mu \neq \mu_0 \\ 1-\Phi(z_0) & H_0: \mu = \mu_0,\ H_1: \mu > \mu_0 \\ \Phi(z_0) & H_0: \mu = \mu_0,\ H_1: \mu < \mu_0 \end{cases}$$

Hypothesis Testing Example: Two Sided I

Example

A manufacturer of sport equipment has developed a new synthetic fishing line that the company claims has a mean breaking strength of 8 kilograms with a standard deviation of 0.5 kilogram. Test the hypothesis that $\mu = 8$ kilograms against the alternative that $\mu \neq 8$ kilograms if a random sample of size 50 lines is test and found to have a mean breaking strength of 7.8 kilograms. Use a 0.01 level of significance.

- $H_0 : \mu = 8$ kilograms.
- $H_1 : \mu \neq 8$ kilograms.
- $\alpha = 0.01$.
- Critical region: $z < -2.575$ and $z > 2.575$, where $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$.

Hypothesis Testing Example: Two Sided II

- Given $\mu_0 = 8$, $\bar{x} = 7.8$, $n = 50$ and $\sigma = 0.5$,

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{7.8 - 8.0}{0.5/\sqrt{50}} = -2.83$$

- Since $z = -2.83 < -z_{\alpha/2} = -2.575$, H_0 is rejected.
- The rejection of H_0 concludes that the average breaking strength is not equal to 8 kilograms, but is less than 8 kilograms.
- The P-value corresponds to $z = 2.02$ is

$$P = P(|Z| > 2.83) = 2P(Z < -2.83) = 0.0046$$

- $P = 0.0046 < \alpha = 0.01$ also leads to the rejection of H_0 at the 0.01 level of significance.

Hypothesis Testing Example: Upper Tail I

Example

A random sample of 100 recorded deaths in the US during the pass year shown an average life span of $\bar{x} = 71.8$ years. Assume a population standard deviation of $\sigma = 8.9$ years, does this seem to indicate that the mean life span today is greater than 70 years old? Use a 0.05 level of significance.

- $H_0 : \mu = 70$ years.
- $H_1 : \mu > 70$ years.
- $\alpha = 0.05$.
- Critical region: $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = z > 1.645$.
- Given $\mu_0 = 70$, $\bar{x} = 71.8$, $n = 100$ and $\sigma = 8.9$,

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{71.8 - 70}{8.9/\sqrt{100}} = 2.02$$

Hypothesis Testing Example: Upper Tail II

- Since $z = 2.02 > z_\alpha = 1.645$, H_0 is rejected.
- The rejection of H_0 concludes that the mean life span today is greater than 70 years.
- The P-value corresponds to $z = 2.02$ is

$$P = P(Z > 2.02) = 0.0217$$

- $P = 0.0217 < \alpha = 0.05$ also leads to the rejection of H_0 at the 0.05 level of significance.

Test on Mean of Normal Dist., Variance Unknown I

- Let $X_1, \ldots, X_n$ be a random sample from a normal population with unknown μ and σ^2.
- The random variable $T_0 = \frac{\bar{X} - \mu}{s/\sqrt{n}}$ is a Student t-distribution with $n-1$ degree of freedom.

Theorem

For the two-sided alternative hypothesis:

$$H_0 : \mu = \mu_0 \tag{19}$$

$$H_1 : \mu \neq \mu_0 \tag{20}$$

we reject H_0 at significant level α when the computed t-statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

exceeds $t_{\alpha/2,n-1}$ or is less than $-t_{\alpha/2,n-1}$.

Test on Mean of Normal Dist., Variance Unknown II

Theorem

For the one-sided alternative hypothesis:

- *Null Hypothesis:* $H_0 : \mu = \mu_0$
- *Test statistic:* $T_0 = \frac{\bar{X}-\mu}{s/\sqrt{n}}$
- *Upper Tail:*

$$H_1 : \mu > \mu_0, \ t_0 > t_{\alpha,n-1}$$

- *Lower Tail:*

$$H_1 : \mu < \mu_0, \ t_0 < -t_{\alpha,n-1}$$

Test on Mean of Normal Dist., Variance Unknown III

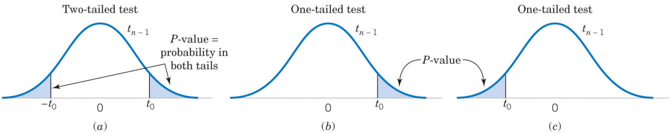

Figure 6: The reference distribution for $H_0 : \mu = \mu_0$ with critical region for (a) $H_1 : \mu \neq \mu_0$, (b) $H_1 : \mu > \mu_0$, (c) $H_1 : \mu < \mu_0$.

Hypothesis Testing Example: Two-Sided, Unknown Var I

Example

The Edison Electric Institute has published figures on the number of kilowatts hours used annually by various home appliance. It is claimed that a vacuum cleaner uses an average of 46 kilowatt hours per year. If a random sample of 12 homes included in a planned study indicates that vacuum cleaners use an average of 42 kilowatt hours per year with a standard deviation of 11.9 kilowatt hours, does this suggest at the 0.05 level of significance that vacuum cleaners use, on average, less than 46 kilowatt hours annually? Assume the population of kilowatt hours to be normal.

- $H_0 : \mu = 46$ kilowatt hours.
- $H_1 : \mu < 46$ kilowatt hours.

Hypothesis Testing Example: Two-Sided, Unknown Var II

- $\alpha = 0.05$.
- Critical region: $\frac{\bar{x}-\mu_0}{s/\sqrt{n}} = t < -1.796$.
- Given $\mu_0 = 46$, $\bar{x} = 42$, $n = 12$ and $s = 11.9$,

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{42 - 46}{11.9/\sqrt{12}} = -1.16$$

- Since $t = -1.16 > t_{0.05,11} = -1.796$, H_0 is not rejected.
- The average number of kilowatt hours used annually by home vacuum cleaner is **NOT** significantly less than 46.

Test on Variance of Normal Distribution I

- Let $X_1, \ldots, X_n$ be a random sample from a normal population.
- We wish to test the hypothesis that the variance of this normal population σ^2 equals a specified value, say σ_0^2.
- To test

$$H_0 : \sigma^2 = \sigma_0^2 \qquad (21)$$

$$H_1 : \sigma^2 \neq \sigma_0^2 \qquad (22)$$

we will use the test statistic:

$$X_0^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

Test on Variance of Normal Distribution II

- The null hypothesis $H_0 : \sigma^2 = \sigma_0^2$ will be rejected if $\chi_0^2 > \chi^2_{\alpha/2,n-1}$ or if $\chi_0^2 < \chi^2_{1-\alpha/2,n-1}$, where $\chi^2_{\alpha/2,n-1}$ and $\chi^2_{1-\alpha/2,n-1}$ are the upper and lower $100\alpha/2$ percentage point of the chi-squared distribution with $n-1$ degree of freedom.
- For one-sided hypothesis (upper tail):

$$H_0 : \sigma^2 = \sigma_0^2 \tag{23}$$
$$H_1 : \sigma^2 > \sigma_0^2 \tag{24}$$

  we would reject H_0 if $\chi_0^2 > \chi^2_{\alpha,n-1}$.
- For one-sided hypothesis (lower tail):

$$H_0 : \sigma^2 = \sigma_0^2 \tag{25}$$
$$H_1 : \sigma^2 < \sigma_0^2 \tag{26}$$

  we would reject H_0 if $\chi_0^2 < \chi^2_{1-\alpha,n-1}$.

Test on Variance of Normal Distribution III

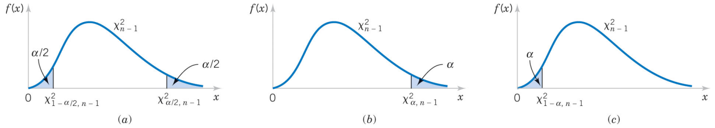

Figure 7: Reference distribution for the test of $H_0 : \sigma^2 = \sigma_0^2$ with critical region for (a) $H_1 : \sigma^2 \neq \sigma_0^2$, (b) $H_1 : \sigma^2 > \sigma_0^2$, (c) $H_1 : \sigma^2 < \sigma_0^2$.

Test on Variance of Normal Distribution I

Example

An automatic filling machine is used to fill bottles with liquid detergent. A random sample of 20 bottles results in a sample variance of fill volume of $s^2 = 0.0153$. If the variance of fill volume exceeds 0.01, an unacceptable portion of bottle will be under-filled or overfilled. Is there any evidence in the sample data to suggest that the manufacturer has a problem with under-filled or overfilled bottles? Use $\alpha = 0.05$, and assume that fill volume has a normal distribution.

- $H_0 : \sigma^2 = 0.01$.
- $H_1 : \sigma^2 > 0.01$.
- $\alpha = 0.05$.

Test on Variance of Normal Distribution II

- The test statistic:

$$\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(20-1)0.0153}{0.01} = 29.07$$

- Critical region: $\chi^2_{0.05,19} = 30.14$
- Since $\chi_0^2 = 29.07 < 30.14 = \chi^2_{0.05,19}$, we fail to reject H_0.
- We conclude that there is no strong evidence that the variance of fill volume exceeds 0.01.
- Given $\chi^2_{0.1,19} = 27.2$ and $\chi^2_{0.05,19} = 30.14$ and $27.2 < 29.07 < 30.14$, we can conclude that the P-value of this example is in the interval $0.05 < P < 0.1$. ($P > \alpha$, fail to reject H_0).
- The actual P-value is 0.0649.

Test on a Single Proportion: Small Sample I

- We consider the problem of testing the hypothesis that the proportion of success, p, in a binomial experiment equals some specified value, p_0.
- We are testing the null hypothesis and the alternative hypothesis which may be one or two-sided alternatives:

$$p < p_0, \;\; p > p_0, \;\; p \neq p_0$$

- We will use the test statistic:

$$\hat{p} = \frac{X}{n}$$

where X is a binomial random variable.
- Values of X that are far from the mean $\mu = np_0$ will lead to rejection of the null hypothesis.

Test on a Single Proportion: Small Sample II

- Since binomial distribution is discrete, it is unlikely that a critical region can be established whose size is exactly equal to the prescribed value of α.
- In dealing with **small sample**, it is preferable to base the discussion on P-value.
- To test the lower-tail hypothesis:

$$H_0 : p = p_0 \tag{27}$$

$$H_1 : p < p_0 \tag{28}$$

we use the binomial distribution to compute the P-value:

$$P = P(X \leq x \text{ when } p = p_0)$$

The value of X is the number of success in n trials. If the P-value is than or equal to the α, the test is significant at the α level and we reject H_0 in favor of H_1.

Test on a Single Proportion: Small Sample III

- Similarly, to test the upper-tail hypothesis:

$$H_0 : p = p_0 \tag{29}$$
$$H_1 : p > p_0 \tag{30}$$

at the α-level of significance, we compute

$$P = P(X \geq x \text{ when } p = p_0)$$

and reject H_0 in favor of H_1 if this P-value is less than or equal to α.

- To test the two-sided hypothesis:

$$H_0 : p = p_0 \tag{31}$$
$$H_1 : p \neq p_0 \tag{32}$$

at the α-level of significance, we compute

$$P = 2P(X \leq x, \text{ when } p = p_0), \quad \text{if } x < np_0$$

Test on a Single Proportion: Small Sample IV

or

$$P = 2P(X \geq x, \text{ when } p = p_0), \quad \text{if } x > np_0$$

and reject H_0 in favor of H_1 if this P-value is less than or equal to α.

Test on a Single Proportion: Small Sample I

Example

A builder claims that heat pumps are installed in 70% of all homes being constructed today in the city of Hsinchu, Taiwan. Would you agree with this claims if a random number survey of new homes in this city showed that 8 out 15 had heat pumps installed? Use a 0.10 level of significance.

- $H_0 : p = 0.7$
- $H_1 : p \neq 0.7$
- $\alpha = 0.10$
- Test Statistic: Binomial random variable X with $p = 0.7$ and $n = 15$.

Test on a Single Proportion: Small Sample II

- Given $x = 8$ and $np_0 = 15 \times 0.7 = 10.5$, the computed P-value is

$$\begin{aligned} P &= 2P(X \leq 8,\ \text{when } p = 0.7) \\ &= 2\sum_{x=0}^{8} (0.7)^x (0.3)^{15-x} \\ &= 0.2622 > 0.10 = \alpha \end{aligned}$$

- Do not reject H_0. There is insufficient reason to doubt builder's claim.

Test on a Single Proportion: Large Sample

- For large n, approximations procedures are required.
- When the hypothesized value p_0 is very close to 0 or 1, the Poisson distribution with parameter $\lambda = np_0$ can be used.
- If P_0 is not too close to either 0 or 1, and np_0 and $n(1 - p_0)$ are greater than 5, we can use normal approximation.
- The z-value for testing $p = p_0$ is given by

$$z = \frac{x - np_0}{\sqrt{np_0q_0}} = \frac{\hat{p} - p_0}{\sqrt{p_0q_0/n}}$$

which is a value of the standard normal variable Z.

Test on a Single Proportion: Large Sample

- For a two-sided test at the α-level of significance, the critical region is $z < -z_{\alpha/2}$ and $z > z_{\alpha/2}$.
- For the one-sided alternative $p < p_0$, the critical region is $z < -z_{\alpha}$.
- For the one-sided alternative $p > p_0$, the critical region is $z > z_{\alpha}$.

Test on a Single Proportion: Large Sample I

Example

A commonly prescribed drug for relieving nervous tension is believed to be only 60% effective. Experimental results with a new drug administered to a random sample of 100 adults who were suffering from nervous tension show that 70 received relief. Is this sufficient evidence to conclude that the new drug is superior to the one commonly prescribed? Use a 0.05 level of significance.

- $H_0 : p = 0.6$
- $H_1 : p > 0.6$
- $\alpha = 0.05$
- Critical region $z > 1.645$

Test on a Single Proportion: Large Sample II

- Given $x = 70$, $n = 100$, $\hat{p} = 0.7$, and

$$\begin{aligned} z &= \frac{0.7 - 0.6}{\sqrt{(0.6)(0.4)/100}} = 2.04 \\ P &= P(Z > 2.04) \\ &< 0.0207 \\ &< 0.05 = \alpha \end{aligned}$$

- Reject H_0 and conclude the new is superior.