Lecture 06. Hypothesis Testing (Chapters 9 and 10)

Ping Yu

HKU Business School
The University of Hong Kong

Plan of This Lecture

- Hypothesis Testing: One Population
  - Concepts of Hypothesis Testing
  - One Normal Mean, Known Population Variance
  - One Normal Mean, Unknown Population Variance
  - One Proportion, Large Samples
  - Assessing the Power of a Test
  - One Normal Variance
- Hypothesis Testing: Two Populations
  - Matched Pair: Two Means
  - Independent Samples: Two Normal Means, Known Population Variances
  - Independent Samples: Two Normal Means, Unknown Equal Population Variances
  - Independent Samples: Two Normal Means, Unknown Unequal Population Variances
  - Independent Samples: Two Proportions, Large Samples
  - Independent Samples: Two Normal Variances

Hypothesis Testing: One Population

Concepts of Hypothesis Testing: Null and Alternative Hypotheses

- In hypothesis testing, we first state two alternatives/options/hypotheses that cover all possible outcomes, and then select one of them using statistics computed from random samples.
  - What is the difference between estimation and hypothesis testing?
- A hypothesis is a claim (assumption) about a population parameter.
- The null hypothesis is the maintained hypothesis unless there is strong evidence against it.
- The alternative hypothesis is the complement of the null (hypothesis), i.e., "rejecting the null" means "accepting the alternative".
- H_0 and H_1 are asymmetric: H_0 is the default state of the world, and we focus on using data to reject H_0.
- The textbook uses the term "fail to reject the null" instead of "accept the null" since the null need not be correct (even if we cannot reject it) but only because we do not have sufficient evidence to reject it; anyway, we will use these two terms interchangeably.

Simple and Composite Hypotheses

- The specification of null and alternative hypotheses depends on the problem.

Example

To test whether the mean package weight of a ready-to-eat cereal is 16 ounces, we can set our null hypothesis as

$$H_0 : \mu = 16,$$

and the alternative hypothesis can be

$$H_1 : \mu > 16 \text{ or } H_1 : \mu \neq 16.$$

If the company wants to avoid legal action and/or customer dissatisfaction, then it can set $H_0 : \mu \leq 16$ vs. $H_1 : \mu > 16$.

- The hypothesis like $\mu = 16$, which specifies a single value of μ, is called the simple hypothesis.
- Either of the two alternative hypotheses in the above example includes more than one values of μ, so is called the composite hypothesis.
- Among which, $\mu > 16$ is a one-sided (composite) hypothesis, and $\mu \neq 16$ is a two-sided (composite) hypothesis.

Testing Procedure

- Define a test statistic; if its value has a small probability to occur under H_0, then we will reject H_0; otherwise, accept H_0.
  - How to construct the test statistic from the observed data?
  - To study the probability of its realized value, we need to derive the distribution of the test statistic.
  - Also, how small is small? $10\%, 5\%$ or 1%?
- Answering the above three questions determines a test.
- Analog: in criminal jury trial,

$$H_0 : \text{you are innocent vs. } H_1 : \text{you are guilty;}$$

  evidences cannot happen if you are innocent, but they indeed happened, so you must be guilty.
- Rejecting H_0 is a strong statement, but accepting H_0 is not. [why?]
  - This is also why the textbook uses the terms "accept H_1" and "fail to reject H_0".
- So if seek strong evidence in favor of a particular outcome, we should define that outcome as the alternative hypothesis.

Type I and II Errors

- Because the test statistic is random, the decision rule based on the test statistic is random, i.e., we have some chance to make mistakes.
- A false rejection of H_0 (rejecting H_0 when H_0 is true) is called a Type I error.
  - Usually, restrict the Type I error rate $P\left(\text{Reject } H_0 | H_0 \text{ is true}\right) = \alpha$ to be small, i.e., the probability of convicting an innocent should be small.
  - α is called the significance level of the test.
- A false acceptance of H_0 (accepting H_0 when H_1 is true) is called a Type II error.
  - $\pi := 1 - P\left(\text{Accept } H_0 | H_1 \text{ is true}\right) =: 1 - \beta$ is called the power of the test, which is the probability of correctly rejecting H_0, i.e., convicting a gulity.
  - Minimizing the probability of the Type II error, β, is equivalent to maximizing the power, i.e., trying our best to convict every guilty person.
  - Type I error is more serious than Type II error because the former involves declaring a scientific finding that is not correct. This is why we must restrict the Type I error rate to be small.

continue

Table 9.1 States of Nature and Decisions on the Null Hypothesis, with Probabilities of Making the Decisions, Given the States of Nature

Decisions on Null Hypothesis	States of Nature: Null Hypothesis Is True	States of Nature: Null Hypothesis Is False
Fail to reject H_0	Correct decision Probability $= 1 - \alpha$	Type II error Probability $= \beta$
Reject H_0	Type I error Probability $= \alpha$ (α is called the significance level)	Correct decision Probability $= 1 - \beta$ ($1 - \beta$ is called the power of the test)

- We fail to reject H_0 either because H_0 is true or we have committed a Type II error.
- α and β cannot be minimized simultaneously, so there is a trade-off between α and β. [see the next slide]
- Usually, we fix α and try to minimize β (or equivalently, maximize π).

Trade-off Between Type-I and Type-II Errors

- Suppose we have only one data point z in hand and we know $z \sim N(\mu, 1)$.
- We want to test H_0: $\mu = 0$ against H_1: $\mu = 3$.
- A natural test is to reject H_0 if z is large, e.g., $z > c$ for some $c > 0$.
- $\alpha = P(z > c|\mu = 0) = 1 - \Phi(c)$, which is a decreasing function of c, where $\Phi(\cdot)$ is the distribution function of the standard normal.
- $\beta = P(z \leq c|\mu = 3) = P(z - 3 \leq c - 3) = \Phi(c - 3)$ which is an increasing function of c. [figure here]
- There is not a direct linear substitution between α and β.
- What is the difference in the two legal systems with H_0 and H_1 switched?

- Fix $\alpha = 0.05$, then c is chosen such that $1 - \Phi(c) = 0.05$, i.e., $c = 1.645$.
- Now, $\pi = 1 - \Phi(1.645 - 3)$.

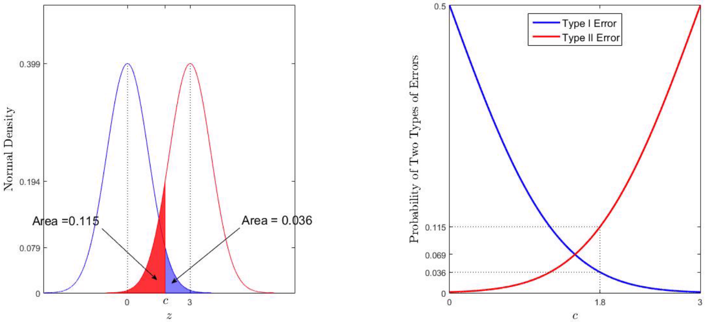
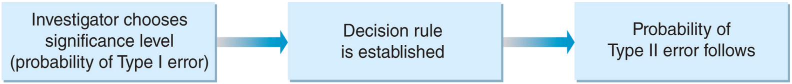


Figure: Trade-Off Between Type-I and Type-II Errors

- The left panel illustrates α and β when $c = 1.8$, and the right panel illustrates these probabilities as a function of c.

Example Continued

- In testing $H_0 : \mu \leq 16$ vs. $H_1 : \mu > 16$, suppose our test statistic is $\bar{x}$, and the decision rule is "reject H_0 if $\bar{x} > 16.13$".
- $\alpha = P\left(\bar{x} > 16.13|\mu\right)$ with $\mu \leq 16$ and $\beta = P\left(\bar{x} \leq 16.13|\mu\right)$ with $\mu > 16$.
- Obviously, α and β are functions of μ, i.e., they should be written as $\alpha\left(\mu\right)$ and $\beta\left(\mu\right)$.
  - $\sup_{\mu\leq 16}\alpha\left(\mu\right)$ is called the size of the test, and $\sup_{\mu\leq 16}\alpha\left(\mu\right)$ is restricted to be no greater than the significance level α.
  - $\pi\left(\mu\right) := 1-\beta\left(\mu\right)$ is called the power function of the test, and is the target of maximization (uniformly over $\mu > 16$ although maybe impossible).
- Figure 9.1: Consequences of Fixing the Significance Level of a Test:



Summary

- One hypothesis testing problem includes the following steps.

0. specify the null and alternative.
1. construct the test statistic.
2. derive the distribution of the test statistic under the null.
3. specify a level of significance.
4. determine the decision rule by finding the critical value.
5. study the power of the test.

- The difficult step is Step 2 which has been studied in Lecture 5.

One Normal Mean, Known Population Variance

- Data: $\{x_i\}_{i=1}^n$, where $x_i \sim N\left(\mu, \sigma^2\right)$, and n is the sample size.
- σ^2 is known from the historical data and is assumed to be maintained, and we only want to know whether the mean of the new data meets a standard, μ_0.
- $H_0 : \mu = \mu_0$ vs. $H_1 : \mu > \mu_0$
  - In the previous example, let $\mu_0 = 16.1$ ounces to meet the industry regulation with label weight 16 ounces.
  - $\mu > \mu_0$ is chosen as H_1. (why?)

continue

- Test Statistic:
$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}},$$
which follows the $N(0,1)$ distribution under H_0.
- Decision Rule: reject H_0 if $z > z_\alpha$.
  - Why? Under H_1, z tends to be large (the mean of $\bar{x}$ is μ, greater than μ_0, and $\sigma/\sqrt{n} > 0$ is fixed), and $P(z > z_\alpha) = \alpha$ (i.e., z has only α probability to be larger than z_α) under H_0.
- Caution on the notation: When we state the distribution of z, z is a random variable; in the decision rule, z is the realized value of z.[1]
- $z > z_\alpha \Longleftrightarrow \bar{x} > \bar{x}_c = \mu_0 + z_\alpha \sigma / \sqrt{n}$, where $\bar{x}_c (> \mu_0)$ is called the critical value for the decision.[2]

[1] The textbook uses Z to denote the random variable, and z its realized value, but the authors seem not consistent in their notations.

[2] The critical value is defined relative to the test statistic: if z is the test statistic, then z_α is its critical value.

A Numerical Example

- Suppose $n = 25$, $\sigma = 0.4$ and $\alpha = 0.05$ (so $z_\alpha = 1.645$).
- The decision rule is to reject H_0 if

$$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - 16.1}{0.4/\sqrt{25}} > 1.645$$

$$\Longleftrightarrow \bar{x} > \bar{x}_c = \mu_0 + z_\alpha \sigma/\sqrt{n} = 16.1 + 1.645 \times \left(0.4/\sqrt{25}\right) = 16.232.$$

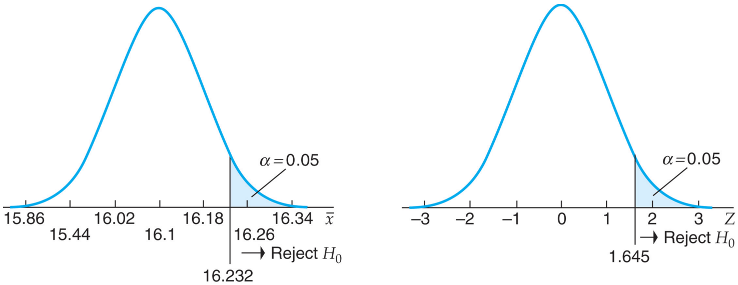
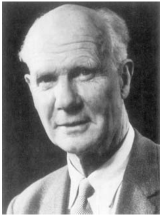


Figure: Normal Densities Showing Both z and $\bar{x}$ Values for the Decision Rule to Test $H_0 : \mu = 16.1$ vs. $H_1 : \mu > 16.1$

History of the Neyman-Pearson Approach and the *p*-Value Approach

Jerzy Neyman (1894-1981), Berkeley



Egon Pearson (1895-1980), UCL

Ronald A. Fisher (1890-1962), UCL

- The rejection/acceptance dichotomy is associated with the Neyman-Pearson approach to hypothesis testing; *p*-value is associated with R.A. Fisher.

p-Value

- There is a shortcoming in the above testing procedure: for a different α, we need to repeat the test. The p-value can avoid this problem.
- If α is made smaller and smaller, there will be a point where H_0 cannot be rejected anymore. [refer to the figure in the last slide, where $\bar{x} = 16.3$ (or $z = 2.5$)]
- The reason is that, by lowering α, we need stronger evidences to reject H_0, and the current evidence becomes not enough.
- The smallest α at which H_0 is still rejected, is called the p-value of the hypothesis test. [Do you expect the p-value $< .05$ or $> .05$ when $\bar{x} = 16.3$?]
- The p-value is the α at which one is indifferent between rejecting and not rejecting the null hypothesis.
- Alternatively, the p-value is the probability of observing a test statistic as extreme as or more extreme than what we obtained if H_0 is true.
  - $p = P(Z > z)$,[3] where $Z \sim N(0,1)$, the null distribution of z.
- A null hypothesis is rejected if and only if the corresponding p-value is smaller than α:

$$\alpha = P(Z > z_\alpha), \text{ so } z > z_\alpha \Longleftrightarrow p < \alpha.$$

[3] Note that $P(Z > z) = P(Z \geq z)$ since the probability of a single point for $N(0,1)$ is 0.

More on the p-Value

- In the above example,

$$p = P(Z > 2.5) = 1 - \Phi(2.5) = 0.0062 < 0.05.$$

- Note that $p = P(Z > z) = 1 - \Phi(z)$ is actually a random variable because z is random, and the observed p-value is one of its realized value.
- A small p-value is evidence against H_0 because one would reject H_0 even at small α's.
- P-values are more informative than tests at fixed α's because you can choose your own α.
- Caveat: the p-value should not be interpreted as the probability that either hypothesis is true. For example, p is NOT the probability "that H_0 is true." Rather, p is a measure of the strength of information against H_0.

Composite Null and Alternative Hypotheses

- $H_0 : \mu \leq \mu_0$ vs. $H_1 : \mu > \mu_0$
- The data, test statistic, decision rule and p-value are exactly the same as in the previous test.
  - Why is z_α the appropriate critical value to guarantee the size of the test to be α? [Exercise]

- $H_0 : \mu \geq \mu_0$ vs. $H_1 : \mu < \mu_0$
  - e.g., from the regulator's perspective in the cereal example, $H_0 : \mu = 16$ (or $H_0 : \mu \geq 16$) vs. $H_1 : \mu < 16$
- The data and test statistic are exactly the same as in the previous test.
- Decision Rule: reject H_0 if $z < -z_\alpha$, or equivalently, $\bar{x} < \bar{x}_c = \mu_0 - z_\alpha \sigma / \sqrt{n}$.
  - e.g., in the setup of the previous numerical example,

$$\bar{x}_c = 16 - 1.645 \times \left(0.4/\sqrt{25}\right) = 13.868.$$

- The p-value is now $P(Z < z)$, where $Z \sim N(0,1)$.

Two-Sided Alternative Hypotheses

- $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$
  - e.g., the diameter of an automobile engine piston cannot be too large or too small.
- The data and test statistic are exactly the same as in the previous test.
- Decision Rule: reject H_0 if $|z| > z_{\alpha/2}$, or equivalently, $\bar{x} < \mu_0 - z_{\alpha/2}\sigma/\sqrt{n}$ or $\bar{x} > \mu_0 + z_{\alpha/2}\sigma/\sqrt{n}$.
  - Note that $z_{\alpha/2} > z_\alpha$, e.g., $z_{0.05/2} = 1.96$.
- The p-value is now $P\left(|Z| > |z|\right)$, where $Z \sim N\left(0,1\right)$.

Acceptance Interval and Critical Interval (Pages 264-266)

- The acceptance interval is the interval where $\bar{x}$ occurs such that H_0 cannot be rejected.
- This concept can be applied to any test, but we discuss it here to aid understanding the acceptance interval on Page 264.
- Specifically, the acceptance interval for the above test is

$$\left[\mu_0 - z_{\alpha/2}\sigma/\sqrt{n}, \mu_0 + z_{\alpha/2}\sigma/\sqrt{n}\right] =: \text{AI} \left(\bar{x} \in \text{AI} \Longleftrightarrow \left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right| \le z_{\alpha/2}\right).$$

- The acceptance interval provides an operating rule for process-monitoring to determine if product standards continue to be achieved over time.
- In US industries, $z_{\alpha/2} = 3$, which results in the so-called Six Sigma methodology.
- Often, the process is adjusted so that σ is small, and the resulting acceptance interval is called the control interval, which is plotted over time and is called the control chart (or more specifically, X-bar chart for $\bar{x}$).

- The critical interval is the interval where $\bar{x}$ occurs such that H_0 is rejected, i.e., it is the complement of the acceptance interval.

Example 6.6: Cereal Package Weights

- A random sample of five packages is collected every 30 minutes, $\mu = 448.8$, and the implied σ from $451 - 448.8 = 3 \times \frac{\sigma}{\sqrt{5}}$ is 1.64.

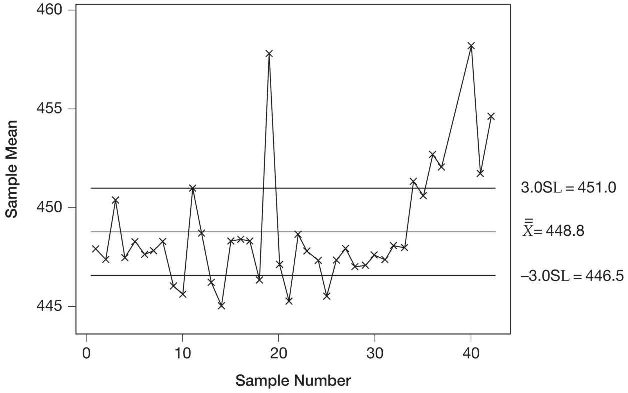


Figure: X-Bar Chart For Cereal-Package Weight

One Normal Mean, Unknown Population Variance

- The hypotheses are exactly the same as in the known population variance case:
  - (i) $H_0 : \mu = \mu_0$ or $H_0 : \mu \leq \mu_0$ vs. $H_1 : \mu > \mu_0$
  - (ii) $H_0 : \mu = \mu_0$ or $H_0 : \mu \geq \mu_0$ vs. $H_1 : \mu < \mu_0$
  - (iii) $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$
- Test Statistic: the t-statistic

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

  which follows the Student's t_{n-1} distribution under H_0. [see the next four slides for the definition and history of the t distribution]
- Decision Rule: reject H_0 if $t > t_{n-1,\alpha}$ in (i), if $t < -t_{n-1,\alpha}$ in (ii), and $|t| > t_{n-1,\alpha/2}$ in (iii).
  - The corresponding decision rule based on $\bar{x}$ is the same as before except replacing z_α by $t_{n-1,\alpha}$, and σ by s.
- The p-value is $P(T > t)$ in (i), $P(T < t)$ in (ii), and $P(|T| > |t|)$ in (iii), where $T \sim t_{n-1}$.

t Distribution (Section 7.3)

- If Z is a standardized normal r.v.,

$$Z \sim N(0,1),$$

and the r.v. X has a χ^2 (chi-square) distribution with v degrees of freedom,

$$X \sim \chi^2_v, \text{ [see the next slide for review]}$$

independent of Z, then

$$\frac{Z}{\sqrt{X/v}} = \frac{\text{standard normal variable}}{\sqrt{\text{independent chi-square variable}/df}} \sim t_v,$$

a t-distribution with v degrees of freedom.

continue

- If $Z_1, \cdots, Z_v$ are i.i.d. such that $Z_i \sim N(0,1)$, $i = 1, \cdots, v$, then
$$X = \sum_{i=1}^{v} Z_i^2 \sim \chi_v^2.$$
- Note that
$$\sum_{i=1}^{v} Z_i^2 \Big/ v \to E\left[Z_i^2\right] = 1 \text{ as } v \to \infty$$
by the LLN, so
$$t_v \to N(0,1) \text{ as } v \to \infty.$$
- Recall that $E\left[Z_i^2\right] = Var(Z_i) + E[Z_i]^2 = 1 + 0^2 = 1$.
- In practice, when $v \geq 20$, the approximation is good enough.

continue

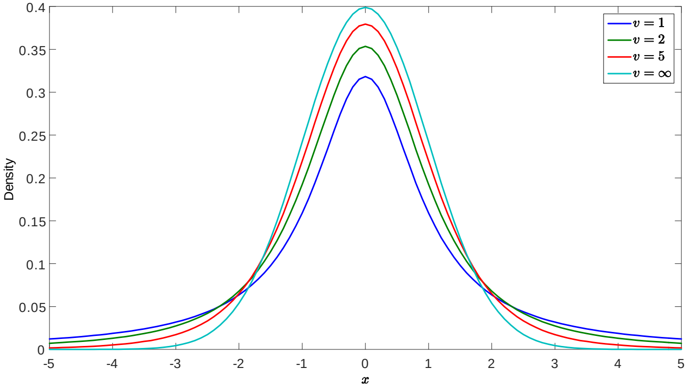


Figure: Density of the t_v Distribution with $v = 1, 2, 5, \infty$

- Compared to $N(0,1)$, the t-distribution is also symmetric, but has a heavier tail, which implies the upper αth quantile of the t_{n-1} distribution $t_{n-1,\alpha} > z_\alpha$.

History of the t Test

William S. Gosset (1876-1937)

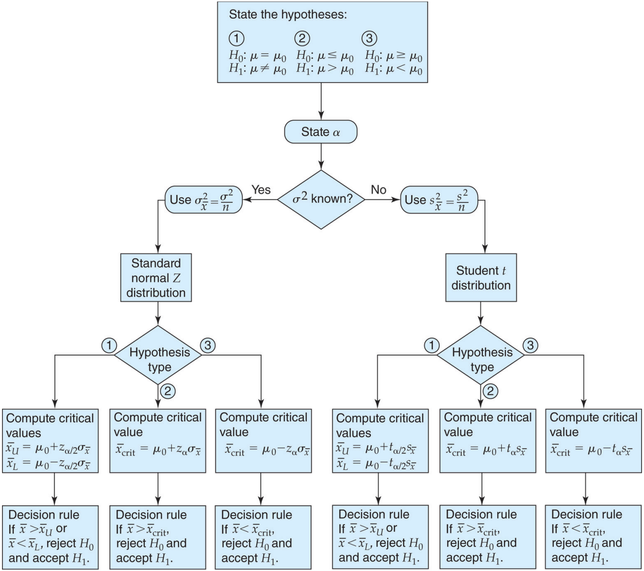
- The t-test is named after Gosset (1908), “The probable error of a mean”. At the time, Gosset worked at Guiness Brewery, which prohibited its employees from publishing in order to prevent the possible loss of trade secrets. To circumvent this barrier, Gosset published under the pseudonym “Student”. Consequently, this famous distribution is known as the Student’s t rather than Gosset’s t! The name “t” was popularized by R.A. Fisher.

Why the t-Statistic Follows the t-Distribution Under H_0?

- Note that

$$
\begin{aligned}
t &= \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{(\bar{x}-\mu_0)/\sqrt{\sigma^2/n}}{\sqrt{\frac{s^2}{n}\Big/\frac{\sigma^2}{n}}} \\
&= \frac{(\bar{x}-\mu_0)/sd(\bar{x})}{\sqrt{\frac{(n-1)s^2}{\sigma^2}\Big/(n-1)}} \\
&\sim \frac{N(0,1)}{\sqrt{\chi^2_{n-1}\Big/(n-1)}} = t_{n-1},
\end{aligned}
$$

  where $N(0,1)$ and χ^2_{n-1} are independent [proof not required].
- When the σ in $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ is replaced by its estimator s, the null distribution changes from $N(0,1)$ to t_{n-1}.
- When $n\to\infty$, the two null distributions coincide [(*) because s is consistent to σ], but when n is small, e.g., $n\leq 10$, the t_n-distribution differs greatly from the normal distribution.


State the hypotheses:
① $H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$
② $H_0: \mu \leq \mu_0$ $H_1: \mu > \mu_0$
③ $H_0: \mu \geq \mu_0$ $H_1: \mu < \mu_0$
State α
σ^2 known?
Yes
No
Use $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$
Use $s_{\bar{x}}^2 = \frac{s^2}{n}$
Standard normal Z distribution
Student t distribution
Hypothesis type
Compute critical values $\bar{x}_U = \mu_0 + z_{\alpha/2}\sigma_{\bar{x}}$ $\bar{x}_L = \mu_0 - z_{\alpha/2}\sigma_{\bar{x}}$
Compute critical value $\bar{x}_{crit} = \mu_0 + z_\alpha \sigma_{\bar{x}}$
Compute critical value $\bar{x}_{crit} = \mu_0 - z_\alpha \sigma_{\bar{x}}$
Compute critical values $\bar{x}_U = \mu_0 + t_{\alpha/2} s_{\bar{x}}$ $\bar{x}_L = \mu_0 - t_{\alpha/2} s_{\bar{x}}$
Compute critical value $\bar{x}_{crit} = \mu_0 + t_\alpha s_{\bar{x}}$
Compute critical value $\bar{x}_{crit} = \mu_0 - t_\alpha s_{\bar{x}}$
Decision rule If $\bar{x} > \bar{x}_U$ or $\bar{x} < \bar{x}_L$, reject H_0 and accept H_1.
Decision rule If $\bar{x} > \bar{x}_{crit}$, reject H_0 and accept H_1.
Decision rule If $\bar{x} < \bar{x}_{crit}$, reject H_0 and accept H_1.

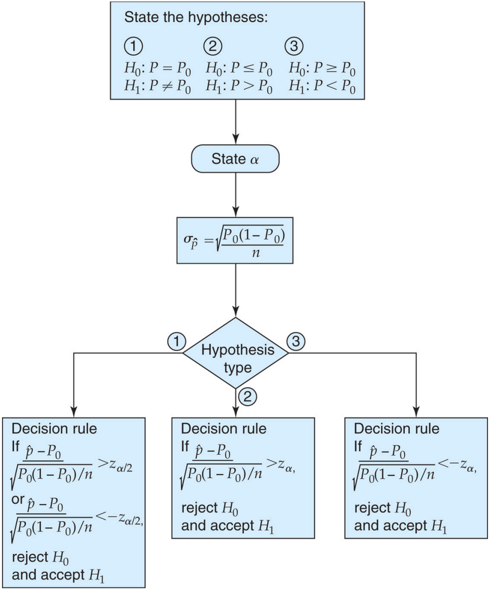
One Proportion, Large Samples

- Data: same as in the previous test, but x_i can only take 0 or 1 and follows the Bernoulli(p) distribution.
- The three pairs of hypotheses are the same as in the previous test, but here the population means are denoted as p.
- Test Statistic:

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 (1 - p_0) / n}},$$

which follows the $N(0,1)$ distribution under H_0 in large sample [$np_0 (1 - p_0) > 5$ with p_0 being the proportion under H_0], where $\hat{p} = \bar{x}$ is the sample proportion.
- Recall that the variance of the Bernoulli(p) distribution is $p(1-p)$, so under H_0, the variance of x_i is known. This is like testing one normal mean with known population variance.

- Decision Rule: reject H_0 if $z > z_\alpha$ in (i), if $z < -z_\alpha$ in (ii), and $|z| > z_{\alpha/2}$ in (iii).
- The p-value formulae are the same as in testing one normal mean with known population variance.


State the hypotheses:
① $H_0: P = P_0$ $H_1: P \neq P_0$
② $H_0: P \leq P_0$ $H_1: P > P_0$
③ $H_0: P \geq P_0$ $H_1: P < P_0$
State α
$\sigma_{\hat{p}} = \sqrt{\frac{P_0(1-P_0)}{n}}$
Hypothesis type
①
②
③
Decision rule
If $\frac{\hat{p}-P_0}{\sqrt{P_0(1-P_0)/n}} > z_{\alpha/2}$
or $\frac{\hat{p}-P_0}{\sqrt{P_0(1-P_0)/n}} < -z_{\alpha/2}$,
reject H_0
and accept H_1
Decision rule
If $\frac{\hat{p}-P_0}{\sqrt{P_0(1-P_0)/n}} > z_{\alpha}$,
reject H_0
and accept H_1
Decision rule
If $\frac{\hat{p}-P_0}{\sqrt{P_0(1-P_0)/n}} < -z_{\alpha}$,
reject H_0
and accept H_1

Example 9.5: Supermarket Shoppers Price Knowledge

- A supermarket wants to know whether shoppers are sensitive to the prices of goods. Among a random sample of 802 shoppers, 378 can state the correct price of an item immediately after putting it into their cart. Test at the 7% level the null that at least one-half of all shoppers can test the correct price.
- Solution: Our hypotheses are $H_0 : p \geq 0.5$ vs. $H_1 : p < 0.5$. The decision rule is

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 \left(1 - p_0\right) / n}} < -z_\alpha.$$

In this example, $\hat{p} = 378/802 = 0.471, p_0 = 0.5$ and $n = 802$, so

$$z = \frac{0.471 - 0.5}{\sqrt{0.5 \left(1 - 0.5\right) / 802}} = -1.64 < -1.474 = -z_{0.07},$$

and we reject the null. Or the p-value is $P\left(Z < z\right) = 0.051 < 0.07$.

Assessing the Power of a Test

- In testing one normal mean with known population variance, $H_0 : \mu = \mu_0$ vs. $H_1 : \mu > \mu_0$.
- Fix $\mu^* > \mu_0$,

$$\begin{aligned}\beta(\mu^*) &= P\left(\bar{x} < \bar{x}_c \middle| \mu^*\right) \\ &= P\left(\frac{\bar{x}-\mu^*}{\sigma/\sqrt{n}} < \frac{\bar{x}_c-\mu^*}{\sigma/\sqrt{n}} \middle| \mu^*\right) \\ &= P\left(Z < \frac{\bar{x}_c-\mu^*}{\sigma/\sqrt{n}}\right) \\ &= \Phi\left(\frac{\bar{x}_c-\mu^*}{\sigma/\sqrt{n}}\right).\end{aligned}$$

- $\pi(\mu^*) = 1-\beta(\mu^*) = P\left(\bar{x} > \bar{x}_c \middle| \mu^*\right) = 1-\Phi\left(\frac{\bar{x}_c-\mu^*}{\sigma/\sqrt{n}}\right) = \Phi\left(\frac{\mu^*-\bar{x}_c}{\sigma/\sqrt{n}}\right)$.

A Numerical Example

- Suppose $n = 16$, $\sigma = 0.1$, $\mu_0 = 5$ and $\alpha = 0.05$ (so $z_\alpha = 1.645$).
- Now, $\bar{x}_c = \mu_0 + z_\alpha \sigma / \sqrt{n} = 5 + 1.645 \times \left(0.1/\sqrt{16}\right) = 5.041$, so

$$\beta(\mu^*) = \Phi\left(\frac{5.041 - \mu^*}{0.1/\sqrt{16}}\right) \text{ and } \pi(\mu^*) = \Phi\left(\frac{\mu^* - 5.041}{0.1/\sqrt{16}}\right).$$

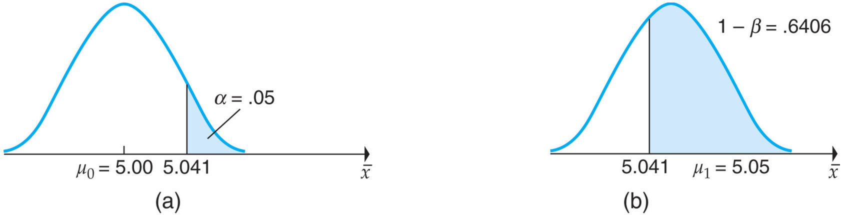


Figure: The Determination of $\pi(5.05)$

- Refer also to the figure on the trade-off between two types of errors.

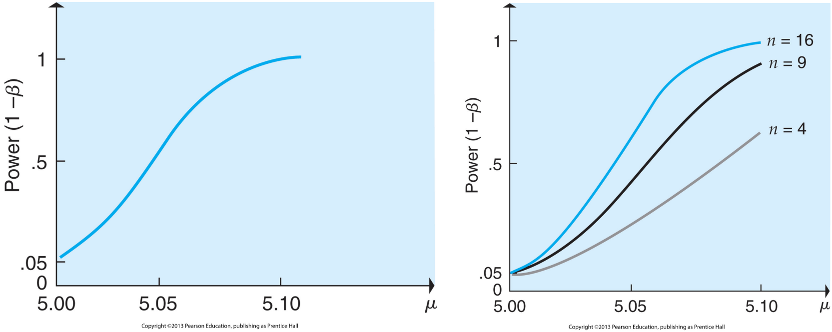


Figure: Power Functions for Test of $H_0 : \mu = 5$ vs. $H_1 : \mu > 5$ $(\alpha = 0.05, \sigma = 0.1, n = 16, 9, 4)$

- $\pi(\mu^*) = \Phi\left(\frac{\mu^* - \bar{x}_c}{\sigma/\sqrt{n}}\right) = \Phi\left(\frac{\mu^* - \mu_0 - z_\alpha \sigma/\sqrt{n}}{\sigma/\sqrt{n}}\right) = \Phi\left(\sqrt{n}\frac{\mu^* - \mu_0}{\sigma} - z_\alpha\right)$ is increasing in μ^*, n and α, and decreasing in σ^2, and $\pi(\bar{x}_c) = 0.5$. [why?]

Another Numerical Example

- Suppose we are interested in $H_0 : p = p_0 = 0.5$ vs. $H_1 : p \neq 0.5$, where p is, say, the proportion of forecasts made by a group of financial analysts that exceeded the actual level of earnings.
- The decision rule is to reject H_0 if $\left|\frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}\right| > z_{\alpha/2}$, where $\hat{p} = 382/600 = .637$ and $n = 600$.
- For $p_1 \neq p_0$,

$$\begin{aligned}
\beta(p_1) &= P\left(\left|\frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}\right| \leq z_{\alpha/2} \middle| p_1\right) \\
&= P\left(\left|\frac{\hat{p}-p_1+p_1-p_0}{\sqrt{p_1(1-p_1)/n}}\right| \leq z_{\alpha/2}\frac{\sqrt{p_0(1-p_0)/n}}{\sqrt{p_1(1-p_1)/n}} \middle| p_1\right) \\
&= P\left(\left|Z + \frac{p_1-p_0}{\sqrt{p_1(1-p_1)/n}}\right| \leq z_{\alpha/2}\sqrt{\frac{p_0(1-p_0)}{p_1(1-p_1)}}\right) \\
&= \Phi\left(z_{\alpha/2}\sqrt{\frac{p_0(1-p_0)}{p_1(1-p_1)}} - \frac{p_1-p_0}{\sqrt{p_1(1-p_1)/n}}\right) \\
&\quad -\Phi\left(-z_{\alpha/2}\sqrt{\frac{p_0(1-p_0)}{p_1(1-p_1)}} - \frac{p_1-p_0}{\sqrt{p_1(1-p_1)/n}}\right).
\end{aligned}$$

continue

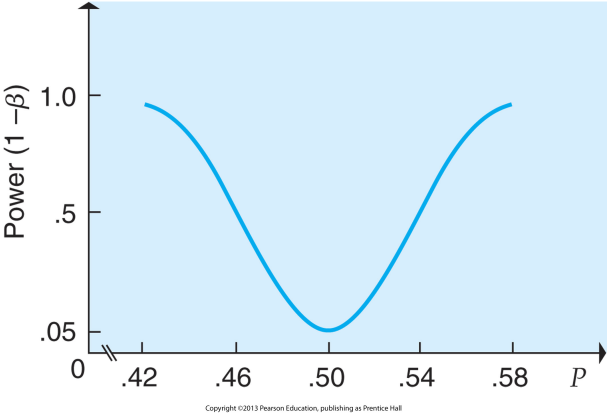


Figure: Power Functions for Test of $H_0 : P = .5$ vs. $H_1 : P \neq .5$ ($\alpha = 0.05, n = 600$)

- $\pi(p_1) = 1 - \beta(p_1)$ is increasing in $|p_1 - p_0|$.

One Normal Variance

- Data: same as in the one normal mean test.
- $H_0 : \sigma^2 = \sigma_0^2$ vs. (i) $H_1 : \sigma^2 > \sigma_0^2$, (ii) $\sigma^2 < \sigma_0^2$, and (iii) $\sigma^2 \neq \sigma_0^2$
  - Such hypotheses are useful in quality control.
- Test Statistic:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2},$$

  which follows the χ^2_{n-1} distribution under H_0.
- Decision Rule: reject H_0 if $\chi^2 > \chi^2_{n-1,\alpha}$ in (i), if $\chi^2 < \chi^2_{n-1,1-\alpha}$ in (ii), and $\chi^2 > \chi^2_{n-1,\alpha/2}$ or $\chi^2 < \chi^2_{n-1,1-\alpha/2}$ in (iii). [figure here]
  - The chi-square distribution tests are more sensitive to the normality assumption than the standard normal distribution tests.
- The p-value is $P\left(\chi^2_{n-1} > \chi^2\right)$ in (i), $P\left(\chi^2_{n-1} < \chi^2\right)$ in (ii), and in (iii)

$$2 \times \min\left\{P\left(\chi^2_{n-1} > \chi^2\right), P\left(\chi^2_{n-1} < \chi^2\right)\right\}$$

  [why? make sure the p-value approach is equivalent to the Neyman-Pearson approach].

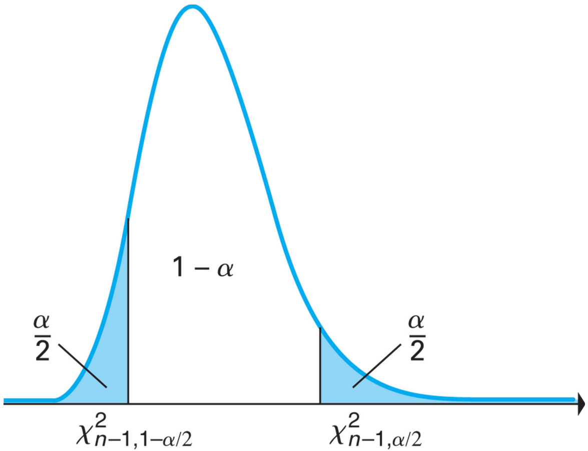


Figure: Chi-Square Distribution with $n-1$ Degrees of Freedom and Its **Upper** $\alpha/2$ and $1-\alpha/2$ Quantiles: the chi-square distribution is not symmetric, so there is no direct relation between $\chi^2_{n-1,\alpha/2}$ and $\chi^2_{n-1,1-\alpha/2}$ and we cannot use $|\chi^2|$ to describe the two-sided test

Hypothesis Testing: Two Populations

Matched Pair: Two Means

- Matched pair is a kind of dependent samples; apart from the factor under study, the pairs should resemble one another as closely as possible, such as twins.
  - Dependent samples can also be two measurements taken on the same person or object, e.g., a measurement is taken before an event and one after the event (e.g., the treatment on a patient), namely, repeated measurements.
- Data: $\{(x_i, y_i)\}_{i=1}^n$, where $x_i - y_i \sim N\left(\mu_x - \mu_y, \sigma_d^2\right)$ but x_i and y_i need not be normally distributed, and μ_x, μ_y and σ_d^2 are unknown.[4]
- (i) $H_0 : \mu_x - \mu_y = 0$ or $H_0 : \mu_x - \mu_y \leq 0$ vs. $H_1 : \mu_x - \mu_y > 0$
- (ii) $H_0 : \mu_x - \mu_y = 0$ or $H_0 : \mu_x - \mu_y \geq 0$ vs. $H_1 : \mu_x - \mu_y < 0$
- (ii) $H_0 : \mu_x - \mu_y = 0$ vs. $H_1 : \mu_x - \mu_y \neq 0$
- This is like testing one normal mean with unknown population variance.
  - x_i, μ, μ_0 and σ^2 there are like $x_i - y_i$, $\mu_x - \mu_y$, 0 and σ_d^2 here.

[4] Because x_i and y_i are not independent, σ_d^2 need not be $\sigma_x^2 + \sigma_y^2$.

continue

- Test Statistic:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}},$$

  which follows the t_{n-1} distribution under H_0, where $\bar{d} = \bar{x} - \bar{y}$, and s_d is the sample standard deviation of $\{(x_i - y_i)\}_{i=1}^n$.
- Decision Rule: reject H_0 if $t > t_{n-1,\alpha}$ in (i), if $t < -t_{n-1,\alpha}$ in (ii), and $|t| > t_{n-1,\alpha/2}$ in (iii).
- The p-value is $P(T > t)$ in (i), $P(T < t)$ in (ii), and $P(|T| > |t|)$ in (iii), where $T \sim t_{n-1}$.
- (*) Recall that the power of the t-test is inversely affected by σ_d^2, so a smaller σ_d^2 is favorable to the detection of the difference in μ_x and μ_y. Since

$$\sigma_d^2 = Var(x - y) = \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy},$$

  a positive σ_{xy} (as in our treatment example in the last slide) is helpful to our purpose. Intuitively, taking differences eliminates random fluctuations that are present in both the x- and y-components and do not interest us; after eliminating this variation, it is easier to discover a possible difference caused by the treatment.

Example 10.1: Analysis of Alternative Turkey-Feeding Programs

- Suppose we want to know whether a new feeding process can increase the mean weight of turkeys at the level 2.5% by using a random set of 25 matched turkey chicks hatched from the the same hen.
- Solution: Our hypotheses are $H_0 : \mu_x - \mu_y \leq 0$ vs. $H_1 : \mu_x - \mu_y > 0$. In this example, $\bar{d} = 1.489$, $s_d^2 = 3.226^2 + 2.057^2 - 2 \times 0.823 \times 3.226 \times 2.057 = 3.716$, so

$$t = \frac{1.489}{\sqrt{3.716}/\sqrt{25}} = \frac{1.489}{0.385} = 3.86 > 2.064 = t_{24,0.025},$$

and we reject the null and conclude that the new feeding program indeed increases the weight of turkey.

Independent Samples: Two Normal Means, Known Variances

- Data: $\{x_i\}_{i=1}^{n_x} \cup \{y_j\}_{j=1}^{n_y}$, where $x_i \sim N\left(\mu_x, \sigma_x^2\right)$, $y_j \sim N\left(\mu_y, \sigma_y^2\right)$, and x_i and y_j are independent for any i and j.
- The three pairs of hypotheses are the same as in the previous test.
- Test Statistic:

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}},$$

  which follows the $N(0,1)$ distribution under H_0 because $E[\bar{x} - \bar{y}] = \mu_x - \mu_y = 0$, $Var(\bar{x} - \bar{y}) = Var(\bar{x}) + Var(\bar{y}) = \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}$, and $\bar{x} - \bar{y}$ is normally distributed.
- Decision Rule: reject H_0 if $z > z_\alpha$ in (i), if $z < -z_\alpha$ in (ii), and $|z| > z_{\alpha/2}$ in (iii).
- The p-value is $P(Z > z)$ in (i), $P(Z < z)$ in (ii), and $P(|Z| > |z|)$ in (iii), where $Z \sim N(0,1)$.

Independent Samples: Two Normal Means, Unknown Equal Variances

- Data and hypotheses are the same as in the previous test, but $\sigma_x^2 = \sigma_y^2$ now.
- Test Statistic:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}},$$

  which follows the t_{n-2} distribution under H_0, where $n = n_x + n_y$, and the pooled sample variance

$$s_p^2 = \frac{(n_x - 1) s_x^2 + (n_y - 1) s_y^2}{n_x + n_y - 2} = \frac{\sum_{i=1}^{n_x} (x_i - \bar{x})^2 + \sum_{j=1}^{n_y} (y_j - \bar{y})^2}{n_x + n_y - 2}$$

  is the weighted average of s_x^2 and s_y^2.[5]
- Decision Rule: reject H_0 if $t > t_{n-2,\alpha}$ in (i), if $t < -t_{n-2,\alpha}$ in (ii), and $|t| > t_{n-2,\alpha/2}$ in (iii).
- The p-value $P(T > t)$ in (i), $P(T < t)$ in (ii), and $P(|T| > |t|)$ in (iii), where $T \sim t_{n-2}$.

[5] Why s_p^2 is better than $\frac{s_x^2 + s_y^2}{2}$?

Example 10.4: Example 10.1 continued

- The setup and data are the same as in Example 10.1 but we assume the two samples are independent now.
- In the current notation, $n_x = n_y = n_0 = 25$, and $n = 2n_0 = 50$.
- $\bar{x} - \bar{y}$ is still 1.489, but

$$s_d^2 = 3.226^2 + 2.057^2 = 14.638 > 3.716$$

and

$$\frac{s_d}{\sqrt{n_0}} = \frac{\sqrt{14.638}}{\sqrt{25}} = 0.765 > 0.385$$

or

$$s_p^2 = \frac{24 \times 3.226^2 + 24 \times 2.057^2}{48} = 7.319$$

and

$$\frac{s_d}{\sqrt{n_0}} = \sqrt{\frac{7.319}{25} + \frac{7.319}{25}} = 0.765 > 0.385.$$

continue

- The test statistic
$$t = \frac{1.489}{0.765} = 1.946 < 2.01 = t_{48,0.025},$$
so we cannot reject the null, a different conclusion from that in Example 10.1, where note that the df is 48 now.
- (*) $r_{xy} = 0.823 > \frac{2}{\sqrt{25}}$, indicating that x_i and y_i are not independent, so the t test here is not suitable. If x_i and y_i are indeed independent, the t test here is preferable because it has a larger df and thus a higher power (e.g., in this example, $t_{24,0.025} > t_{48,0.025}$).

Indpt Samples: Two Normal Means, Unknown Unequal Variances

- This is the famous Behrens-Fisher Problem.
- Data and hypotheses are the same as in the previous test.
- Test Statistic:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}},$$

  whose null distribution is very complicated, but Welch (1938) and Satterthwaite (1946) suggested to use the t_v distribution to approximate it, where

$$v = \frac{\left[\left(\frac{s_x^2}{n_x}\right) + \left(\frac{s_y^2}{n_y}\right)\right]^2}{\left(\frac{s_x^2}{n_x}\right)^2 / (n_x - 1) + \left(\frac{s_y^2}{n_y}\right)^2 / (n_y - 1)}$$

  is random and need not be an integer. [6]
- Decision Rule: reject H_0 if $t > t_{v,\alpha}$ in (i), if $t < -t_{v,\alpha}$ in (ii), and $|t| > t_{v,\alpha/2}$ in (iii).
- The p-value is $P(T > t)$ in (i), $P(T < t)$ in (ii), and $P(|T| > |t|)$ in (iii), where $T \sim t_v$.

[6] v reduces to $(n_0 - 1)\left(1 + \frac{2}{s_x^2/s_y^2 + s_y^2/s_x^2}\right)$ when $n_x = n_y = n_0$. If $s_x^2 \approx s_y^2$, then v further reduces to $2(n_0 - 1)$, which is the same as in the unknown equal population variances case.

Example 10.4 continued

- The setup and data are the same as in Example 10.4, but we assume the two unknown population variances are unequal now.
- The test statistic takes the same value 1.946; this is because when $n_x = n_y = n_0$,

$$\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y} = \frac{2}{n_0} \frac{(n_0 - 1) s_x^2 + (n_0 - 1) s_y^2}{n_0 + n_0 - 2} = \frac{s_x^2 + s_y^2}{n_0}.$$

- The only difference is that the df is 40 rather than 48 now.
- The p-value is equal to

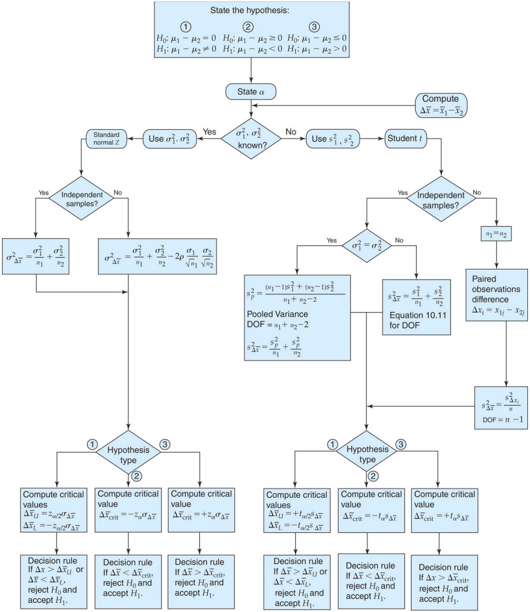
$$P(t_{40} > 1.946) = 0.02935 > 0.025,$$

so we cannot reject null, the same conclusion as in Example 10.4.
- For comparison, the p-value in Example 10.4 is

$$P(t_{48} > 1.946) = 0.02876 > 0.025.$$

(*) More Comments

- When $\sigma_x^2 \neq \sigma_y^2$, but we assume they are equal and apply the two-sample t-test with equal variances, then the true size may be different from the nominal size.
- Anyway, as $n_x = n_y \to \infty$, the true size of $t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_p^2}{n_x}+\frac{s_p^2}{n_y}}}$ converges to the nominal size for any pair of $\left(\sigma_x^2, \sigma_y^2\right)$.
- This leads to the advice to choose samples of equal size whenever possible.
- This is also wise when $\sigma_x^2 = \sigma_y^2$, because the power of the corresponding t-test is maximal when $n_x = n_y$ (for fixed $n_x + n_y$) as $\frac{1}{n_x} + \frac{1}{n_y}$ achieves the minimum.
- When $n_x, n_y \to \infty$ (need not be equal), by the central limit theorem, $t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}}}$ converges in distribution to $N(0,1)$.


State the hypothesis:
① $H_0: \mu_1 - \mu_2 = 0$ $H_1: \mu_1 - \mu_2 \neq 0$
② $H_0: \mu_1 - \mu_2 \geq 0$ $H_1: \mu_1 - \mu_2 < 0$
③ $H_0: \mu_1 - \mu_2 \leq 0$ $H_1: \mu_1 - \mu_2 > 0$
State α
Compute $\Delta\bar{x} = \bar{x}_1 - \bar{x}_2$
σ_1^2, σ_2^2 known?
Yes
No
Use σ_1^2, σ_2^2
Standard normal Z
Use s_1^2, s_2^2
Student t
Independent samples?
Yes
No
$\sigma^2_{\Delta\bar{x}} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$
$\sigma^2_{\Delta\bar{x}} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} - 2\rho \frac{\sigma_1}{\sqrt{n_1}} \frac{\sigma_2}{\sqrt{n_2}}$
Independent samples?
Yes
No
$\sigma_1^2 = \sigma_2^2$
Yes
No
$n_1 = n_2$
$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$
Pooled Variance
DOF = $n_1 + n_2 - 2$
$s^2_{\Delta\bar{x}} = \frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}$
$s^2_{\Delta\bar{x}} = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}$
Equation 10.11 for DOF
Paired observations difference
$\Delta x_i = x_{1j} - x_{2j}$
$s^2_{\Delta\bar{x}} = \frac{s^2_{\Delta x_i}}{n}$
DOF = $n - 1$
Hypothesis type
① ② ③
Compute critical values
$\Delta\bar{x}_U = z_{\alpha/2}\sigma_{\Delta\bar{x}}$
$\Delta\bar{x}_L = -z_{\alpha/2}\sigma_{\Delta\bar{x}}$
Compute critical value
$\Delta\bar{x}_{crit} = -z_\alpha\sigma_{\Delta\bar{x}}$
Compute critical value
$\Delta\bar{x}_{crit} = +z_\alpha\sigma_{\Delta\bar{x}}$
Decision rule
If $\Delta x > \Delta\bar{x}_U$ or $\Delta\bar{x} < \Delta\bar{x}_L$, reject H_0 and accept H_1.
Decision rule
If $\Delta\bar{x} < \Delta\bar{x}_{crit}$, reject H_0 and accept H_1.
Decision rule
If $\Delta\bar{x} > \Delta\bar{x}_{crit}$, reject H_0 and accept H_1.
Hypothesis type
① ② ③
Compute critical values
$\Delta\bar{x}_U = +t_{\alpha/2}s_{\Delta\bar{x}}$
$\Delta\bar{x}_L = -t_{\alpha/2}s_{\Delta\bar{x}}$
Compute critical value
$\Delta\bar{x}_{crit} = -t_\alpha s_{\Delta\bar{x}}$
Compute critical value
$\Delta\bar{x}_{crit} = +t_\alpha s_{\Delta\bar{x}}$
Decision rule
If $\Delta\bar{x} > \Delta\bar{x}_U$ or $\Delta\bar{x} < \Delta\bar{x}_L$, reject H_0 and accept H_1.
Decision rule
If $\Delta\bar{x} < \Delta\bar{x}_{crit}$, reject H_0 and accept H_1.
Decision rule
If $\Delta x > \Delta\bar{x}_{crit}$, reject H_0 and accept H_1.

Independent Samples: Two Proportions, Large Samples

- Data: same as in the previous test, but x_i and y_i can only take 0 or 1, so follows the Bernoulli(p_x) and Bernoulli(p_y) distributions.
- The three pairs of hypotheses are the same as in the previous test, but here the population means are denoted as p_x and p_y.
- Test Statistic:

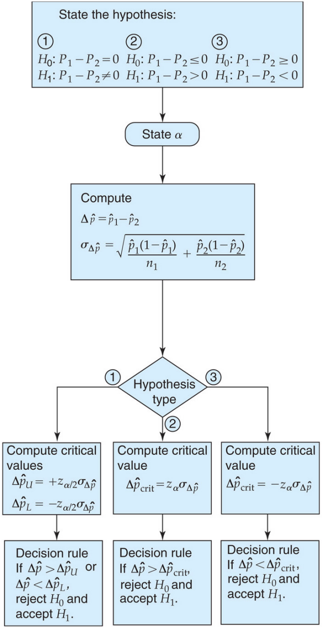
$$t = \frac{\hat{p}_x - \hat{p}_y}{\sqrt{\frac{\hat{p}_0(1-\hat{p}_0)}{n_x} + \frac{\hat{p}_0(1-\hat{p}_0)}{n_y}}},$$

which follows the $N(0,1)$ distribution under H_0 in large sample [$np_0(1-p_0) > 5$ with p_0 being the common proportion under H_0], where

$$\hat{p}_0 = \frac{n_x\hat{p}_x + n_y\hat{p}_y}{n_x + n_y} = \frac{\text{total number of successes in } \{x_i\}_{i=1}^{n_x} \cup \{y_j\}_{j=1}^{n_y}}{\text{total sample size of } \{x_i\}_{i=1}^{n_x} \cup \{y_j\}_{j=1}^{n_y}}.$$

- Recall that the variance of the Bernoulli(p) distribution is $p(1-p)$, so under H_0, the variances of x_i and y_i are also equal. This is like testing two normal means with unknown equal population variances.

- Decision Rule: reject H_0 if $t > z_\alpha$ in (i), if $t < -z_\alpha$ in (ii), and $|t| > z_{\alpha/2}$ in (iii).
- The p-value is $P(Z > t)$ in (i), $P(Z < t)$ in (ii), and $P(|Z| > |t|)$ in (iii), where $Z \sim N(0,1)$.


State the hypothesis:
① $H_0: P_1 - P_2 = 0$ $H_1: P_1 - P_2 \neq 0$
② $H_0: P_1 - P_2 \leq 0$ $H_1: P_1 - P_2 > 0$
③ $H_0: P_1 - P_2 \geq 0$ $H_1: P_1 - P_2 < 0$
State α
Compute
$\Delta\hat{p} = \hat{p}_1 - \hat{p}_2$
$\sigma_{\Delta\hat{p}} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$
Hypothesis type
①
②
③
Compute critical values
$\Delta\hat{p}_U = +z_{\alpha/2}\sigma_{\Delta\hat{p}}$
$\Delta\hat{p}_L = -z_{\alpha/2}\sigma_{\Delta\hat{p}}$
Compute critical value
$\Delta\hat{p}_{crit} = z_{\alpha}\sigma_{\Delta\hat{p}}$
Compute critical value
$\Delta\hat{p}_{crit} = -z_{\alpha}\sigma_{\Delta\hat{p}}$
Decision rule
If $\Delta\hat{p} > \Delta\hat{p}_U$ or $\Delta\hat{p} < \Delta\hat{p}_L$, reject H_0 and accept H_1.
Decision rule
If $\Delta\hat{p} > \Delta\hat{p}_{crit}$, reject H_0 and accept H_1.
Decision rule
If $\Delta\hat{p} < \Delta\hat{p}_{crit}$, reject H_0 and accept H_1.

Example 10.5: Change in Customer Recognition of New Products After an Advertising Campaign

- Before the advertising campaign, 50 of 270 random residents heard of the new product; after the campaign, 81 of 203 new random samples heard of the new product. Do these results indicate that customer recognition increased after the campaign at the 5% level?
- Solution: Our hypotheses are $H_0 : p_x - p_y \geq 0$ vs. $H_1 : p_x - p_y < 0$. In this example,

$$\begin{aligned}\hat{p}_x &= 50/270 = 0.185, \hat{p}_y = 81/203 = 0.399,\\ \hat{p}_0 &= \frac{n_x\hat{p}_x + n_y\hat{p}_y}{n_x + n_y} = \frac{50+81}{270+203} = 0.277,\end{aligned}$$

so

$$t = \frac{\hat{p}_x - \hat{p}_y}{\sqrt{\frac{\hat{p}_0(1-\hat{p}_0)}{n_x} + \frac{\hat{p}_0(1-\hat{p}_0)}{n_y}}} = \frac{0.185 - 0.399}{\sqrt{\frac{0.277(1-0.277)}{270} + \frac{0.277(1-0.277)}{203}}} = -5.15,$$

which is smaller than $-z_{0.05} = -1.645$, and we reject the null and conclude that the advertising campaign is effective.

Independent Samples: Two Normal Variances

- In testing two normal means with unknown equal population variances, we assume $\sigma_x^2 = \sigma_y^2$.
  - This hypothesis is also of interest in quality-control studies.
- Data: same as in testing two normal means, where $s_x^2 \geq s_y^2$.
- (i) $H_0 : \sigma_x^2 = \sigma_y^2$ or $H_0 : \sigma_x^2 \leq \sigma_y^2$ vs. $H_1 : \sigma_x^2 > \sigma_y^2$
  - Why don't we consider $H_0 : \sigma_x^2 = \sigma_y^2$ or $H_0 : \sigma_x^2 \geq \sigma_y^2$ vs. $H_1 : \sigma_x^2 < \sigma_y^2$?
- (ii) $H_0 : \sigma_x^2 = \sigma_y^2$ vs. $H_1 : \sigma_x^2 \neq \sigma_y^2$
- Test Statistic:

$$f = \frac{s_x^2}{s_y^2},$$

  which follows the F_{n_x-1,n_y-1} distribution [see the next slide] under H_0, where the larger sample variance is put in the numerator and the smaller in the denominator so that only the upper cutoff points of the F distribution are used [see below].
- Decision Rule: reject H_0 if $f > F_{n_x-1,n_y-1,\alpha}$ in (i), and if $f > F_{n_x-1,n_y-1,\alpha/2}$ in (ii).
- The p-value is $P\left(F > f\right)$ in (i), and $2 \cdot P\left(F > f\right)$ in (ii), where $F \sim F_{n_x-1,n_y-1}$.

F Distribution

- This distribution is named after R.A. Fisher.
- If X_1 follows a χ^2 distribution with d_1 degrees of freedom,

$$X_1 \sim \chi^2_{d_1},$$

and X_2 follows a χ^2 distribution with d_2 degrees of freedom,

$$X_2 \sim \chi^2_{d_2},$$

independent of X_1, then

$$\frac{X_1/d_1}{X_2/d_2} = \frac{\text{chi-square variable}/df}{\text{independent chi-square variable}/df} \sim F_{d_1,d_2},$$

an F-distribution with degrees of freedom d_1 and d_2.

- $f = \frac{s_x^2}{s_y^2} \overset{H_0}{=} \frac{s_x^2/\sigma_x^2}{s_y^2/\sigma_y^2} = \frac{\left[(n_x-1)s_x^2/\sigma_x^2\right]/(n_x-1)}{\left[(n_y-1)s_y^2/\sigma_y^2\right]/(n_y-1)} \sim \frac{\chi^2_{n_x-1}/(n_x-1)}{\chi^2_{n_y-1}/(n_y-1)} = F_{n_x-1,n_y-1}$.

- As in the t-distribution, $X_2/d_2 \to 1$ as $d_2 \to \infty$. So

$$F_{d_1,d_2} \to \chi^2_{d_1}/d_1$$

as $d_2 \to \infty$, i.e., $F_{d_1,d_2,\alpha} \approx \chi^2_{d_1,\alpha}/d_1$.

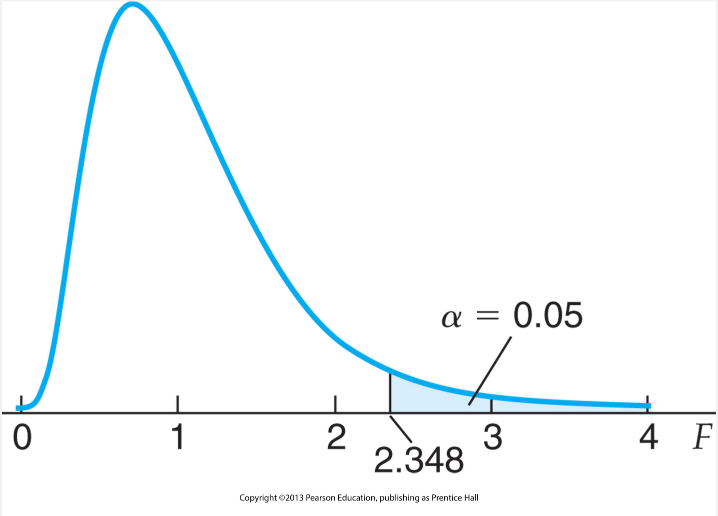


Figure: The Density of $F_{10,20}$

Example 10.6: Study of Maturity Variances

- We want to know whether the variance of the maturities of AAA-rated industrial bonds is different from that of CCC-rated ones at the 2% level.
- Solution: Our hypotheses are $H_0 : \sigma_x^2 = \sigma_y^2$ vs. $H_1 : \sigma_x^2 \neq \sigma_y^2$. In this example,

$$\begin{aligned} n_x &= 17, n_y = 11, \\ s_x^2 &= 123.35 \text{ and } s_y^2 = 8.02, \end{aligned}$$

so

$$f = \frac{s_x^2}{s_y^2} = \frac{123.35}{8.02} = 15.380 > 4.520 = F_{16,10,0.01},$$

or the p-value is $P\left(F_{16,10} > 15.380\right) = 0.00 < 0.01$. As a result, we reject the null and conclude that there is strong evidence that variances in maturities are different for these two types of bonds.

Some Comments on Hypothesis Testing

- A test with low power can result from:
  - Small sample size
  - Large variances in the underlying populations
  - Poor measurement procedures
- If sample sizes are large it is possible to find significant differences that are not practically important.
- Researchers should select the appropriate level of significance before computing *p*-values.