MATH 205: Statistical methods

November 15th, 2021

Lecture 20: Comparing the mean of two populations (cont.)

Chapter 7: Significance of evidence

Hypothesis testing

In a hypothesis-testing problem, there are two contradictory hypotheses under consideration

- The null hypothesis, denoted by H_0, is the claim that is initially assumed to be true
- The alternative hypothesis, denoted by H_a, is the assertion that is contradictory to H_0.
- The null hypothesis will be rejected in favor of the alternative hypothesis only if sample evidence suggests that H_0 is false.
- If the sample does not strongly contradict H_0, we will continue to believe in the probability of the null hypothesis.

Test about a population mean

- Null hypothesis

$$H_0 : \mu = \mu_0$$

- The alternative hypothesis will be either:
  - $H_a : \mu > \mu_0$
  - $H_a : \mu < \mu_0$
  - $H_a : \mu \neq \mu_0$

Note: μ_0 here denotes a constant, and μ denotes the population mean (unknown)

Statistical "Proof by contradiction"

Ideas:

- Assume that a hypothesis (the null hypothesis) is true
- We ask ourselves, what is the probability that we'll see a dataset as contradictory as (or more contradictory than) the current one?
- That probability is referred to as the p-value (also called **observed significance level**) of the test
- If the p-value is less than a predetermined threshold (called **significant level**, often denoted by α), then we reject the null hypothesis

Note: "contradictory" is a relative concept and is reflected through the alternative hypothesis

P-values for z-tests

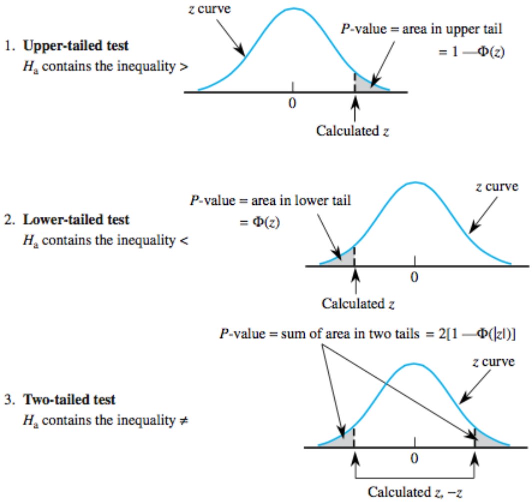

Figure 9.7 Determination of the P-value for a z test

Practice problem

Problem

The target thickness for silicon wafers used in a certain type of integrated circuit is 245 μm. A sample of 50 wafers is obtained and the thickness of each one is determined, resulting in a sample mean thickness of 246.18 μm and a sample standard deviation of 3.60 μm.

At significant level $\alpha = 0.01$, does this data suggest that true average wafer thickness is something other than the target value?

P-values for z-tests

1. Parameter of interest: μ = true average wafer thickness
2. Null hypothesis: H_0: $\mu = 245$
3. Alternative hypothesis: H_a: $\mu \neq 245$
4. Formula for test statistic value: $z = \dfrac{\bar{x} - 245}{s/\sqrt{n}}$
5. Calculation of test statistic value: $z = \dfrac{246.18 - 245}{3.60/\sqrt{50}} = 2.32$
6. Determination of P-value: Because the test is two-tailed,

$$P\text{-value} = 2[1 - \Phi(2.32)] = .0204$$

7. Conclusion: Using a significance level of .01, H_0 would not be rejected since .0204 > .01. At this significance level, there is insufficient evidence to conclude that true average thickness differs from the target value.

Interpreting P-values

A P-value:

- is not the probability that H_0 is true
- is not the probability of rejecting H_0
- is the probability, calculated assuming that H_0 is true, of obtaining a test statistic value at least as contradictory to the null hypothesis as the value that actually resulted

Two-sample inference: example

Example

Let μ_1 and μ_2 denote true average decrease in cholesterol for two drugs. From two independent samples $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$, we want to test:

$$H_0 : \mu_1 = \mu_2$$
$$H_a : \mu_1 \neq \mu_2$$

Settings

Assumption

1. *$X_1, X_2, \ldots, X_m$ is a random sample from a population with mean μ_1 and variance σ_1^2.*
2. *$Y_1, Y_2, \ldots, Y_n$ is a random sample from a population with mean μ_2 and variance σ_2^2.*
3. *The X and Y samples are independent of each other.*

Analysis

Problem

Assume that

- $X_1, X_2, \ldots, X_m$ *is a random sample from a population with mean* μ_1 *and variance* σ_1^2.
- $Y_1, Y_2, \ldots, Y_n$ *is a random sample from a population with mean* μ_2 *and variance* σ_2^2.
- *The* X *and* Y *samples are independent of each other.*

Compute (in terms of $\mu_1, \mu_2, \sigma_1, \sigma_2, m, n$*)*

(a) $E[\bar{X} - \bar{Y}]$

(b) $Var[\bar{X} - \bar{Y}]$ *and* $\sigma_{\bar{X}-\bar{Y}}$

Properties of $\bar{X} - \bar{Y}$

Proposition

The expected value of $\overline{X} - \overline{Y}$ is $\mu_1 - \mu_2$, so $\overline{X} - \overline{Y}$ is an unbiased estimator of $\mu_1 - \mu_2$. The standard deviation of $\overline{X} - \overline{Y}$ is

$$\sigma_{\overline{X}-\overline{Y}} = \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$$

Confidence intervals

When both population distributions are normal, standardizing $\overline{X} - \overline{Y}$ gives a random variable Z with a standard normal distribution. Since the area under the z curve between $-z_{\alpha/2}$ and $z_{\alpha/2}$ is $1 - \alpha$, it follows that

$$P\left(-z_{\alpha/2} < \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} < z_{\alpha/2}\right) = 1 - \alpha$$

Manipulation of the inequalities inside the parentheses to isolate $\mu_1 - \mu_2$ yields the equivalent probability statement

$$P\left(\overline{X} - \overline{Y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} < \mu_1 - \mu_2 < \overline{X} - \overline{Y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}\right) = 1 - \alpha$$

Testing the difference between two population means

- Setting: independent normal random samples $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ with known values of σ_1 and σ_2. Constant Δ_0.
- Null hypothesis:

$$H_0 : \mu_1 - \mu_2 = \Delta_0$$

- Alternative hypothesis:
  (a) $H_a : \mu_1 - \mu_2 > \Delta_0$
  (b) $H_a : \mu_1 - \mu_2 < \Delta_0$
  (c) $H_a : \mu_1 - \mu_2 \neq \Delta_0$
- When $\Delta = 0$, the test (c) becomes

$$H_0 : \mu_1 = \mu_2$$

$$H_a : \mu_1 \neq \mu_2$$

Testing the difference between two population means

Assume that we want to test the null hypothesis $H_0 : \mu_1 - \mu_2 = \Delta_0$ against each of the following alternative hypothesis

(a) $H_a : \mu_1 - \mu_2 > \Delta_0$

(b) $H_a : \mu_1 - \mu_2 < \Delta_0$

(c) $H_a : \mu_1 - \mu_2 \neq \Delta_0$

We use the test statistic:

$$z = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}.$$

and derive the p-value in the same way as the one-sample tests.

Practice problem

Each student in a class of 21 responded to a questionnaire that requested their GPA and the number of hours each week that they studied. For those who studied less than 10 h/week the GPAs were

$$2.80, 3.40, 4.00, 3.60, 2.00, 3.00, 3.47, 2.80, 2.60, 2.00$$

and for those who studied at least 10 h/week the GPAs were

$$3.00, 3.00, 2.20, 2.40, 4.00, 2.96, 3.41, 3.27, 3.80, 3.10, 2.50$$

Assume that the distribution of GPA for each group is normal and both distributions have standard deviation $\sigma_1 = \sigma_2 = 0.6$. Treating the two samples as random, is there evidence that true average GPA differs for the two study times? Carry out a test of significance at level .05.

Solution

1. The parameter of interest is $\mu_1 - \mu_2$, the difference between true mean GPA for the < 10 (conceptual) population and true mean GPA for the $\geq$10 population.
2. The null hypothesis is H_0: $\mu_1 - \mu_2 = 0$.
3. The alternative hypothesis is H_a: $\mu_1 - \mu_2 \neq 0$; if H_a is true then μ_1 and μ_2 are different. Although it would seem unlikely that $\mu_1 - \mu_2 > 0$ (those with low study hours have higher mean GPA) we will allow it as a possibility and do a two-tailed test.
4. With $\Delta_0 = 0$, the test statistic value is

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$$

5. The inequality in H_a implies that the test is two-tailed. For $\alpha = .05$, $\alpha/2 = .025$ and $z_{\alpha/2} = z_{.025} = 1.96$. H_0 will be rejected if $z \geq 1.96$ or $z \leq -1.96$.

Solution

6. Substituting $m = 10$, $\bar{x} = 2.97$, $\sigma_1^2 = .36$, $n = 11$, $\bar{y} = 3.06$, and $\sigma_2^2 = .36$ into the formula for z yields

$$z = \frac{2.97 - 3.06}{\sqrt{\frac{.36}{10} + \frac{.36}{11}}} = \frac{-.09}{.262} = -.34$$

That is, the value of $\bar{x} - \bar{y}$ is only one-third of a standard deviation below what would be expected when H_0 is true.

7. Because the value of z is not even close to the rejection region, there is no reason to reject the null hypothesis. This test shows no evidence of any relationship between study hours and GPA. ■

Large-sample tests/confidence intervals

- Central Limit Theorem: $\bar{X}$ and $\bar{Y}$ are approximately normal when $n > 30 \rightarrow$ so is $\bar{X} - \bar{Y}$. Thus

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$$

  is approximately standard normal
- When n is sufficiently large $S_1 \approx \sigma_1$ and $S_2 \approx \sigma_2$
- Conclusion:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}}$$

  is approximately standard normal when n is sufficiently large

If $m, n > 40$, we can ignore the normal assumption and replace σ by S

Large-sample tests

Proposition

Use of the test statistic value

$$z = \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$$

along with the previously stated upper-, lower-, and two-tailed rejection regions based on z critical values gives large-sample tests whose significance levels are approximately α. These tests are usually appropriate if both $m > 40$ and $n > 40$. A P-value is computed exactly as it was for our earlier z tests.

Large-sample CIs

Proposition

Provided that m and n are both large, a CI for $\mu_1 - \mu_2$ with a confidence level of approximately $100(1-\alpha)\%$ is

$$\bar{x} - \bar{y} \ \pm \ z_{\alpha/2}\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$$

where $-$gives the lower limit and $+$ the upper limit of the interval. An upper or lower confidence bound can also be calculated by retaining the appropriate sign and replacing $z_{\alpha/2}$ by z_α.

Example

Example

Let μ_1 and μ_2 denote true average tread lives for two competing brands of size P205/65R15 radial tires.

(a) Test

$$H_0 : \mu_1 = \mu_2$$
$$H_a : \mu_1 \neq \mu_2$$

at level 0.05 using the following data: $m = 45$, $\bar{x} = 42,500$, $s_1 = 2200$, $n = 45$, $\bar{y} = 40,400$, and $s_2 = 1900$.

(b) Construct a 95% CI for $\mu_1 - \mu_2$.

Example

The article "Gender Differences in Individuals with Comorbid Alcohol Dependence and Post-Traumatic Stress Disorder" (Amer. J. Addiction, 2003: 412–423) reported the accompanying data on total score on the Obsessive-Compulsive Drinking Scale (OCSD).

Gender	Sample Size	Sample Mean	Sample SD
Male	44	19.93	7.74
Female	40	16.26	7.58

Formulate hypotheses and carry out an appropriate analysis. Does your conclusion depend on whether a significance level of .05 or .01 was employed?

Example

Research has shown that good hip range of motion and strength in throwing athletes results in improved performance and decreased body stress. The article "Functional Hip Characteristics of Baseball Pitchers and Position Players" (Am. J. Sport. Med., 2010: 383–388) reported on a study involving samples of 40 professional pitchers and 40 professional position players.

For the pitchers, the sample mean trail leg total arc of motion (degrees) was 75.6 with a sample standard deviation of 5.9, whereas the sample mean and sample standard deviation for position players were 79.6 and 7.6, respectively.
Assuming normality, test appropriate hypotheses to decide whether true average range of motion for the pitchers is less than that for the position players (as hypothesized by the investigators).

Example

A letter in the Journal of the American Medical Association (May 19, 1978) reports that of 215 male physicians who were Harvard graduates and died between November 1974 and October 1977, the 125 in full-time practice lived an average of 48.9 years beyond graduation, whereas the 90 with academic affiliations lived an average of 43.2 years beyond graduation.

Does the data suggest that the mean lifetime after graduation for doctors in full-time practice exceeds the mean lifetime for those who have an academic affiliation