MATH 205: Statistical methods

November 10th, 2021

Lecture 19: Comparing the mean of two populations

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.

Hypothesis testing: an analogy

In a criminal trial, there are to contradictory assertions

- the accused individual is innocent
- the accused individual is guilty

$\rightarrow$ the claim of innocence is the favored or protected 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

z-test

- Given a random sample of size n from a distribution with mean μ (unknown) and either
  - the distribution is normal and population standard deviation σ is known, or
  - sample size $n > 40$
- Assume that the data is collected with the measured sample mean $\bar{x}$ and we want to test

$$H_0 : \mu = \mu_0$$
$$H_a : \mu < \mu_0$$

z-test: normal distribution with known σ

- If the null hypothesis $\mu = \mu_0$ is true, then $X_i \sim N(\mu_0, \sigma^2)$
- A sample would be more contradictory to the null hypothesis than the current sample we have if

$$\bar{X} \leq \bar{x} \quad \text{or} \quad \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \leq \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

- Thus, the p-value in this case is

$$P\left[Z \leq \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right] = \Phi\left(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right)$$

P-values for z-tests

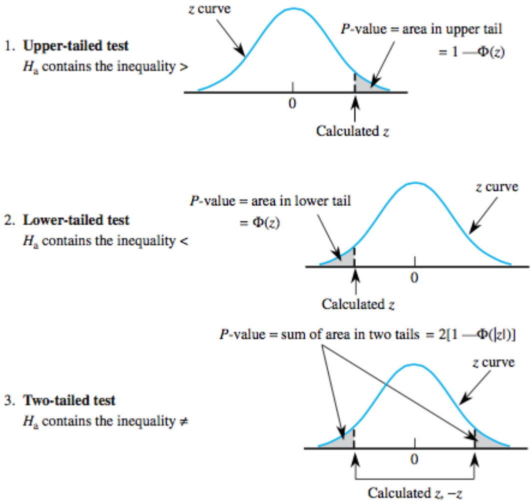

Figure 9.7 Determination of the P-value for a z test

Practice problem

Problem

The target thickness for silicon wafers used in a certain type of integrated circuit is 245 μ*m. A sample of 50 wafers is obtained and the thickness of each one is determined, resulting in a sample mean thickness of 246.18* μ*m and a sample standard deviation of 3.60* μ*m.*
At signififcant level $\alpha = 0.01$*, does this data suggest that true average wafer thickness is something other than the target value?*

$\Phi(z)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997

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.

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.

Does this data suggest that true average wafer thickness is larger than the target value? Carry out a test of significance at level .01 and provide the corresponding P-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

Problem 1

A company that makes cola drinks states that the mean caffeine content per one 12-ounce bottle of cola is 40 milligrams. You work as a quality control manager and are asked to test this claim. During your tests, you find that a random sample of 30 bottles of cola (12-ounce) has a mean caffeine content of 39.2 milligrams. From a previous study, you know that the standard deviation of the population is $\sigma = 7.5$ milligrams. We assume that the caffeine content is normally distributed.

(a) (20 points) At α= 1% level of significant, can you reject the company's claim? What is the P-value associated with the test?

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

Assume further that the distributions of X and Y are normal and σ_1, σ_2 are known:

Problem

(a) *What is the distribution of*

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

(b) *Compute*

$$P\left[-1.96 \leq \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \leq 1.96\right]$$

(c) *Construct a 95% CI for $\mu_1 - \mu_2$ (in terms of $\bar{x}$, $\bar{y}$, m, n, σ_1, σ_2).*

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

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

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. ■