STA238 Tutorial 4

Luis Ledesma

2023-02-15

1 Announcements

- You can upload your work on Crowdmark from the end of the tutorial session to 5pm Friday of that week.
- All questions must be solved using RStudio.

2 Recall: Last tutorial

Last tutorial: based on a random sample, we obtained some point estimates and computed some probability estimates assuming that the data was obtained from the normal distribution. We also demonstrated through simulation, that the sum of independent normal random variables is also a normal random variable.

Main takeaways:

1. The sample mean $\hat{\mu}$ will be the point estimate of the population mean μ, and the sample variance s^2 will be the point estimate of the population variance σ^2.
2. The standard error and the sample variance will be different concepts (albeit related to each other).
3. Assuming that X follows a certain probability distribution, the number x' such that $P(X < x') = p$ is known as the p-th quantile (in a way, it is an inverse function).
4. The sum of independent and identically distributed normal random variables will be a random variable itself.

3 Tutorial activity

In broad strokes, we want to:

1. Compare the means between two populations from independent random samples.
2. Conducting the respective two-sided and one-sided hypothesis tests.
3. Understanding the assumptions that must hold in order to carry out the hypothesis tests and carry out statistical inference.

3.1 Computing confidence interval of the difference between means

Given both independent samples in the problem, the confidence interval at α significance for the difference between the means $\mu_1 - \mu_2$ will be:

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Where s_1 and s_2 are the sample means from the first and second samples, respectively.

Question: Why can we use the quantiles from the normal distribution, rather than the ones from the t-distribution?

Intuitively speaking, the difference between the sample means will be $\bar{x}_1 - \bar{x}_2 = 35$. We may expect the hypothesis test of $H_0 : \mu_1 - \mu_2 = 0$ against $H_0 : \mu_1 - \mu_2 \neq 0$ to yield a significant result (obviously, this would depend on the sample sizes).

After simplifying the above formula of the 95% confidence interval with the given values:

$$35 \pm 1.96 \sqrt{\frac{150^2}{400} + \frac{200^2}{400}} = 35 \pm 24.5 = (10.5, 59.5)$$

3.2 Hypothesis tests: two-sided tests

Suppose we want to test $H_0 : \mu_1 - \mu_2 = 0$ against $H_A : \mu_1 - \mu_2 \neq 0$. The test statistic will be given by:

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0, 1)$$

Using the given values, the test statistic will be $z_c = \frac{35}{12.5} = 2.8$.

Now, what would be the rejection region? Since we are doing a two-sided hypothesis test as $H_A : \mu_1 - \mu_2 \neq 0$, it would be $|z_c| \geq z_{\frac{\alpha}{2}} = 1.96$. Thus, we have enough evidence to reject the null hypothesis at 0.05 significance.

Alternatively, the p-value would be $P(|X| > 2.8) = 0.0052$, which is lower than 0.05. Thus, we have enough evidence to reject $H_0 : \mu_1 - \mu_2 = 0$ at this significance level, and conclude that $\mu_1 - \mu_2 \neq 0$ at this significance level.

3.3 Hypothesis tests: one-sided tests

What if we now want to test $H_0 : \mu_1 - \mu_2 = 0$ against $H_A : \mu_1 - \mu_2 > 0$?

The test statistic would be the same, however the rejection region would be different. In this case, it would now be $z_c \geq z_\alpha = 1.64$, so we would still have enough evidence to reject that $\mu_1 - \mu_2 = 0$, and conclude that $\mu_1 - \mu_2 > 0$ at this significance level.

The p-value would be $P(X > 2.8) = 0.0026$, which is lower than 0.05, so we would be able to reject H_0 at this significance level.

3.4 Hypothesis test: Other hypotheses

Suppose that we now want to test $H_0 : \mu_1 - \mu_2 = 25$ against $H_A : \mu_1 - \mu_2 \neq 25$. The test statistic would be:

$$\frac{(\bar{x}_1 - \bar{x}_2) - 25}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0, 1)$$

After plugging in our values, $z_c = \frac{10}{2.5} = 0.8$. This value would be outside the rejection region of the two-sided test, which is $|z_c| \geq z_{\frac{\alpha}{2}} = 1.96$, so we do not have enough evidence to reject the null hypothesis that $H_0 : \mu_1 - \mu_2 = 25$.

The two-sided p-value would be $P(|X| > 0.8) = 0.42$, which is higher than 0.05, so we do not have enough evidence to reject H_0 at this significance level.

Question: Carry out the one-sided hypothesis test $H_0 : \mu_1 - \mu_2 = 25$ against $H_A : \mu_1 - \mu_2 > 25$. What would be the p-value?

3.5 Assumptions of hypothesis tests

These hypothesis tests rely on the fact that sample 1 and sample 2 are independent, and that the sample sizes are high enough that the test statistic follows a normal distribution by the Central Limit Theorem.