STOCKHOLM UNIVERSITY
Department of Statistics

Emma Pettersson
2017-11-24

Solutions to Plenary Exercises: Plenary Exercise 5

Basic Statistics for Economists, 15 ECTS, STE101

EXERCISE 1

Let X_k be the weight of package k. The weights of different packages are iid, where $X_k \sim N(\mu_X, \sigma_X^2)$.

μ_X is unknown, but the variance $\sigma_X^2 = 0.1^2$ is known.

Use that $\bar{X} \sim N\left(\mu_X, \frac{\sigma_X^2}{n}\right)$.

We want to test on the $\alpha = 5\%$ significance level:

$$H_0: \mu = 2$$
$$H_A: \mu \neq 2$$

The test variable:

$$Z = \frac{\bar{X} - \mu_0}{\sigma / n} \sim N(0,1)$$

We reject the null hypothesis if the observed value is more extreme than the critical value, $|z_{obs}| > z_{\alpha/2}$.

The critical value:

$$z_{\alpha/2} = z_{0.025} = \text{[table 2]} = 1.96$$

The observed value:

$$|z_{obs}| = \left|\frac{\bar{x} - \mu_0}{\sigma / n}\right| = \left|\frac{1.95 - 2}{0.1/\sqrt{12}}\right| = \left|-\frac{\sqrt{12}}{2}\right| = \sqrt{3} = 1.732$$

Decision:

$$1.732 < 1.96$$
$$|z_{obs}| < z_{\alpha/2}$$

We cannot reject the null hypothesis on the 5% significance level, we cannot reject the hypothesis that the true mean weight is 2 kg.

EXERCISE 2

X_k is a random variable for item k such that if $X_k = 0$ the item works fine, and if $X_k = 1$ the item is defective. The probability that $X_k = 1$ is P, and the probability that $X_k = 0$ is $1 - P$. Assume that all items are independent of one another.

Let Y be the number of defective items in a sample of $n = 160$ items. Y then has a binomial distribution with parameters n and p. $Y \sim Bin(n, P)$.

A point estimate for P is $\hat{p} = \frac{Y}{n} = \sum X_k/n = \bar{X}$. And according to the CLT, we can approximate the distribution of $\hat{p}$ with a normal distribution when n is large.

Hypotheses: (We use a one-sided test because the goal was probably to decreases the error rate)

$$H_0 : P = 0.08$$

$$H_A : P < 0.08$$

$$\alpha = 0.05$$

We reject the null hypothesis if the observed value of the test statistic is more extreme than the critical value. In other words, we reject H_0 if $z_{obs} < -z_\alpha$. This is the same as saying we reject H_0 when $|z_{obs}| > z_\alpha$.

The point estimate:

$$\hat{p} = \frac{Y}{n} = \frac{10}{160} = 0.0625$$

The test statistic:

$$Z = \frac{\hat{p} - P_0}{\sqrt{\frac{P_0(1 - P_0)}{n}}} \sim N(0,1)$$

The observed value:

$$z_{obs} = \frac{0.0625 - 0.08}{\sqrt{\frac{0.08(0.92)}{160}}} = \frac{0.0625 - 0.08}{\sqrt{0.00046}} = -0.8159$$

The critical value:

$$-z_\alpha = -z_{0.05} = -1.6449$$

Conclusion:

$$z_{obs} > -z_\alpha$$

$$-0.8159 > -1.6449$$

We cannot reject the null hypothesis on the 5% significance level, and we cannot say that the new process had a significant effect on decreasing the error rate.

EXERCISE 3

Let X_k be the life span of tool k. We assume that all tools are iid, and that $X_k \sim N(\mu_X, \sigma_X^2)$.

If the life span of a tool is normally distributed, then the average life span of several tools will also be normally distributed. $\bar{X} \sim N(\mu_X, \sigma_X^2/n)$. However, note that both μ_X and σ_X^2 are unknown, and that our sample size $n = 6$ is very small.

Hypotheses:

$$H_0: \mu = 3000$$

$$H_A: \mu < 3000$$

$$\alpha = 0.05$$

We reject the null hypothesis if the observed value of the test statistic is more extreme than the critical value. In other words, we reject H_0 if $t_{obs} < -t_{n-1;\alpha}$, this is the same as saying that we will reject H_0 if $|t_{obs}| > t_{n-1;\alpha}$.

The test statistic:

$$t_{n-1} = \frac{\bar{X} - \mu_0}{s_x/\sqrt{n}} \sim t(n-1)$$

The critical value:

$$t_{n-1;\alpha} = t_{5;0.05} = 2.015$$

The observed value:

$$\bar{x} = \sum \frac{x_k}{n} = \frac{2970 + \cdots + 2925}{6} = \frac{17760}{6} = 2960$$

$$s_x^2 = \frac{\sum(x_k - \bar{x})^2}{n-1} = \frac{(2970 - 2960)^2 + \cdots + (2925 - 2960)^2}{5}$$
$$= \frac{10^2 + 60^2 + 45^2 + 60^2 + 20^2 + 35^2}{5} = \frac{10950}{5} = 2190$$

$$t_{obs} = \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}} = \frac{2960 - 3000}{\sqrt{2910/6}} = -2.094$$

Conclusion:

$$-2.094 < -2.015$$
$$t_{obs} < -t_{5;0.05}$$

We reject the null hypothesis at the 5% significance level, the average life span of the tool is significantly less than 3000.

EXERCISE 4

Let P be the proportion of individuals in the market that know about Rokak's cookies.

A point estimate for P, from a iid sample with n subjects is: $\hat{p} = \frac{Y}{n}$. Where Y is the number of people who know about the product.

If n is large, according to the CLT $\hat{p}$ with be approximately normally distributed, $\hat{p} \sim N(nP, nP(1-P))$.

a. Hypotheses:

$$H_0: P = 0.4$$
$$H_A: P < 0.4$$

The test statistic:

The sample size of $n = 500$ is very large (rule of thumb: $nP_0(1-P_0) = 120 > 5$), and we can make the following normal approximation:

$$Z = \frac{\hat{p} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} \sim N(0,1)$$

The point estimate:

$$\hat{p} = \frac{Y}{n} = \frac{180}{500} = 0.36$$

The observed value:

$$z_{obs} = \frac{0.36 - 0.4}{\sqrt{\frac{(0.4)(0.6)}{500}}} = -\frac{0.04}{\sqrt{0.00048}} = -1.8257$$

The critical value:

A significance level is not given in the question, so let's find the critical values when using significance levels of 5% and 2.5%.

$$z_{0.05} = 1.6449$$

$$z_{0.025} = 1.96$$

We reject the null hypothesis if $|z_{obs}| > z_\alpha$.

$1.8257 > 1.6449 \rightarrow |z_{obs}| > z_{0.05}$. We reject the null hypothesis on the 5% significance level.

$1.8257 < 1.96 \rightarrow |z_{obs}| < z_{0.025}$. We cannot reject the null hypothesis on the 2.5% significance level.

Our choice to reject the null hypothesis depends on which significance level we choose.

We can also calculate the p-value:

$$P(Z < -1.83) = P(Z > 1.83) = 1 - P(Z < 1.83) = 1 - 0.96638 = \mathbf{0.03362}$$

Given that the null hypothesis is true, the probability that we reject H_0 is around 3.36%. In general, when p-values are low, we reject the null hypothesis, and when p-values are high we do not reject.

b. Sample 1: $\hat{p}_1 = \frac{180}{500} = 0.36$
   Sample 2: $\hat{p}_2 = \frac{195}{500} = 0.39$

   All the assumptions made for sample 1 also hold for sample 2. In addition to this, the samples are independent of each other.

   Hypotheses:

$$H_0: P_2 = P_1 \quad \Leftrightarrow \quad H_0: P_2 - P_1 = 0$$

$$H_A: P_2 > P_1 \quad \Leftrightarrow \quad H_0: P_2 - P_1 > 0$$

   Test statistic:

$$Z = \frac{\hat{p}_2 - \hat{p}_1 - 0}{\sqrt{\hat{p}_0(1-\hat{p}_0)\left(\frac{1}{n_2} + \frac{1}{n_1}\right)}} \sim N(0,1) \quad where \quad \hat{p}_0 = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

   The estimate for $\hat{p}_0$, is a weighted average of the estimates from both samples, we weigh them together to form one estimate, that is used in the denominator to estimate the variance.

   Observed value:

$$\hat{p}_0 = \frac{500(0.36) + 500(0.39)}{500 + 500} = 0.375$$

$$z_{obs} = \frac{0.39 - 0.36}{\sqrt{0.375(0.625)\left(\frac{2}{500}\right)}} = 0.9797959 \approx \mathbf{0.98}$$

   Critical value (same as in a.) $z_\alpha = z_{0.05} = 1.6449$

   Since our observed value is less than the critical value ($z_{obs} < z_\alpha$; $0.98 < 1.6449$), we cannot reject the null hypothesis at the 5% significance level.