Handout for Lecture 14

Hypothesis Testing & p-values

ECON 340: Economic Research Methods — Instructor: Div Bhagia

Hypothesis Testing

1. Set up null hypothesis and alternative hypothesis

$$\text{Null Hypothesis:} \quad H_0 : \mu = \mu_0$$
$$\text{Alternative Hypothesis:} \quad H_1 : \mu \neq \mu_0$$

2. Construct test statistic Z if true population variance is known, else use T-statistic.

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \quad \text{and} \quad t_0 = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$$

3. Under the null if $\bar{X} \sim N(\mu_0, \sigma^2/n)$, then $Z \sim N(0,1)$ and $T \sim t_{n-1}$. In case of known population variance, reject the null if $|z_0| > z_{\alpha/2}$. In the case of unknown population variance, reject the null if $|t_0| > t_{n-1,\alpha/2}$.

Note: When $n \geq 100$ you can reject the null if $|t_0| > z_{\alpha/2}$ as in large sample t distribution looks like the standard normal.

p-Value:

p-Value is the probability of obtaining an outcome even more surprising under the null hypothesis than the one you got.

- Known variance: $p = 2Pr(Z > |z_0|)$
- Unknown variance, $n < 100$: $p = 2Pr(T > |t_0|)$
- Unknown variance, $n \geq 100$: $p = 2Pr(T > |t_0|) = 2Pr(Z > |t_0|)$

Question 1: A car manufacturer wants to estimate the mean CO2 emissions of a new model of car. A sample of 196 cars is randomly selected and their CO2 emissions are measured. The sample mean and standard deviation are 120 g/km and 20 g/km, respectively. The car manufacturer had initially claimed that the average CO2 emissions of this model would be 115 g/km. Test the manufacturer's claim at a 5% level of significance.

Answer: Null and alternative hypothesis:

$$H_0 : \mu = 115 \qquad H_1 : \mu \neq 115$$

Note that here, $\bar{x} = 120$, $S = 20$, and $n = 196$. We can calculate the t-statistic as follows:

$$t_0 = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} = \frac{120 - 115}{20/\sqrt{196}} = 3.5$$

Under the null $T \sim t_{195}$. However, since the sample size is large enough, we can just use the normal distribution to find the critical value. Critical value: $z_{0.025} = 1.96$ leaves 2.5% area in the upper tail. Since $|t_0| = 3.5 > 1.96$ we will reject the null at 5% level of significance.

Question 2: Find the p-value associated with your test statistic in the previous question.

Answer: Here the p-value is given by:

$$p = 2Pr(T > |t_0|) = 2Pr(Z > |t_0|) = 2Pr(Z > 3.5) = 2 \times 0.002 = 0.004$$

So only 0.4% outcomes for the sample mean would be more surprising than 120 g/km that we found if the true population mean was indeed 115 g/km.