Two Sample z-test Examples

January 31, 2022

1. A car manufacturer aims to improve the quality of the products by reducing the defects and also increase the customer satisfaction. Therefore, he monitors the efficiency of two assembly lines in the shop floor. In line A there are 18 defects reported out of 200 samples. While the line B shows 25 defects out of 600 cars. At α 5%, is the differences between two assembly procedures are significant?

Solution: State: $\hat{p}_c = \dfrac{18+25}{200+600}$, p_c is used as we are assuming $p_1 = p_2 = p_c$ under the null. $\alpha = 0.05$.

$$H_0 : p_1 - p_2 = 0$$
$$H_a : p_1 - p_2 \neq 0$$

Plan: We are running a two sample z-test for $p_1 - p_2$. The difference in efficiency proportions of the two assemble lines. Let's check conditions to ensure this type of inference is appropriate:

Random Sampling: We will assume our samples from both assemble lines are random and representative of their respective populations.
Independence: The samples of $n_1 = 200$ and $n_2 = 600$ cars are clearly both less than their respective populations (all cars manufactured in both lines).
Normality: $n_1(\hat{p}_c)$, $n_1(1-\hat{p}_c)$,$n_2(\hat{p}_c)$, $n_2(1-\hat{p}_c)$ are all greater than 10 which means we may assume $\hat{p}$ is normally distributed with a mean of 0, and a standard error of $\sqrt{\hat{p}_c(1-\hat{p}_c)\left(\frac{1}{200}+\frac{1}{600}\right)}$. (Note actually values of $\hat{p}_c$ and calculations should be shown.)

Do: Test statistic:

$$\begin{aligned} z &= \frac{\text{statistic} - \text{parameter}}{\text{standard error of statistic}} \\ &= \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}_c(1-\hat{p}_c)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \\ &= \frac{\left(\frac{18}{200} - \frac{25}{600}\right) - 0}{\sqrt{\frac{43}{800}\left(\frac{757}{800}\right)\left(\frac{1}{200}+\frac{1}{600}\right)}} \\ &= 2.624824 \end{aligned}$$

p-value:

$$\begin{aligned} p\text{-value} &= 2P(Z > 2.624824) \\ &= 0.008669377 \end{aligned}$$

Conclude: Our p-value is less than our significance level of 0.05 so we reject the null hypothesis in favour of the alternative. Assuming that there is no difference in the proportion of defects between the two car lines, there is a roughly 1% chance of observing a sample that is just as extreme or more extreme than our own. This is strong enough evidence to reject the null hypothesis and suggest that the proportions differ between the two assembly lines.

2. Researchers want to test the side effects of a new COVID vaccine. In clinical trail 62 out of 300 individuals taking X1 vaccine report side effects. While 48 individuals out of 300 taking X2 vaccine report side effects. At 95% confidence level, is the X1 vaccine working any differently than the X2?

Solution: State: $\hat{p}_c = \dfrac{62+48}{300+300}$, $\hat{p}_c$ is used as we are assuming $p_1 = p_2 = p_c$ under the null. $\alpha = 0.05$.

$$H_0 : p_1 - p_2 = 0$$
$$H_a : p_1 - p_2 \neq 0$$

Plan: We are running a two sample z-test for $p_1 - p_2$. The difference in proportion of people who experience side effects between the two vaccines. Let's check conditions to ensure this type of inference is appropriate:

Random Sampling: We will assume our clinical trials are random and representative of their respective populations (all people who take the respective vaccines).
Independence: The samples of $n_1 = 300$ and $n_2 = 300$ subjects are clearly both less than their respective populations.
Normality: $n_1(\hat{p}_c)$, $n_1(1-\hat{p}_c)$,$n_2(\hat{p}_c)$, $n_2(1-\hat{p}_c)$ are all greater than 10 which means we may assume $\hat{p}$ is normally distributed with a mean of 0, and a standard error of $\sqrt{\hat{p}_c(1-\hat{p}_c)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$. (Note actually values of $\hat{p}_c$ and calculations should be shown.)
Do: Test statistic:

$$\begin{aligned} z &= \frac{\text{statistic} - \text{parameter}}{\text{standard error of statistic}} \\ &= \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}_c(1-\hat{p}_c)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \\ &= \frac{\left(\frac{62}{300} - \frac{48}{300}\right) - 0}{\sqrt{\frac{110}{600}\left(\frac{490}{600}\right)\left(\frac{1}{300}+\frac{1}{300}\right)}} \\ &= 1.48 \end{aligned}$$

p-value:

$$\begin{aligned} p\text{-value} &= 2P(Z > 2.624824) \\ &= 0.1388732 \end{aligned}$$

Conclude: Our p-value is greater than our significance level of 0.05 so we fail to reject the null hypothesis. Assuming that there is no difference in the proportion of people who experience side effects between the two vaccines there is a roughly 14% chance of observing a clinical trial as extreme or more extreme than our own. This is not strong enough evidence to reject H_0.