Independent Two-sample z-test

1. Assumptions

- **Experimental Design:** The sample forms two independent treatment groups.
- **Null Hypothesis:** The population means of the two treatment groups are not significantly different from each other.
- **Population Distribution:** Arbitrary distribution within each treatment group.
- **Sample Size:** Size of each treatment group is equal to or greater than 30.

2. Inputs for independent two-sample z-test

- Sample sizes of treatment groups: n_1 and n_2
- Sample means of treatment groups: $\bar{x}_1$ and $\bar{x}_2$
- Standard deviations of treatment groups: s_1 and s_2
- Standard errors of treatment group means:

$$\mathrm{SE}_1 = \frac{s_1}{\sqrt{n_1}} \quad \text{and} \quad \mathrm{SE}_2 = \frac{s_2}{\sqrt{n_2}}$$

- Standard error of the differences: $s_{\text{diff}} = \sqrt{\mathrm{SE}_1^2 + \mathrm{SE}_2^2}$
- Null hypothesis: $\mu_1 = \mu_2$
- The level of the test: α

3. The Five Steps for Performing the Test of Hypothesis

1. State null and alternative hypotheses:

$$H_0 : \mu_1 = \mu_2, \quad H_1 : \mu_1 \neq \mu_2$$

2. Compute test statistic:

$$z = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)}{\mathrm{SE}_{\text{diff}}},$$

   assuming the null hypothesis.

3. Compute $100(1-\alpha)\%$ confidence interval I for z.

4. If $z \in I$, accept H_0; if $z \notin I$, reject H_1 and accept H_0.

5. Compute p-value.

4. Discussion

Since we are assuming that $n \geq 30$, the sample standard deviations s_1 and s_1 are close approximations to the population standard deviations σ_1 and σ_2, so we will assume that the population standard deviations are known and equal to the respective sample standard deviations. Furthermore

$$\mathrm{E}(\bar{x}_2 - \bar{x}_1) = \mathrm{E}(\bar{x}_2) - \mathrm{E}(\bar{x}_1) = \mu_2 - \mu_1.$$

Also, if the two treatment groups are independent,

$$\mathrm{Var}(\bar{x}_2 - \bar{x}_1) = 1^2 \cdot \mathrm{Var}(\bar{x}_2) + (-1)^2\,\mathrm{Var}(\bar{x}_1) = \frac{\sigma_2^2}{n_1} + \frac{\sigma_1^2}{n_2},$$

the standard deviation of $\bar{x}_2 - \bar{x}_1$ is

$$\mathrm{SE}_{\text{diff}} = \sqrt{\mathrm{Var}(\bar{x}_2 - \bar{x}_1)} = \sqrt{\frac{\sigma_2^2}{n_1} + \frac{\sigma_1^2}{n_2}}.$$

Finally, because the expected value and variance of

$$z = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)}{\mathrm{SE}_{\text{diff}}^2}$$

are $\mu_2 - \mu_1$ and $\mathrm{SE}_{\text{diff}}^2$, respectively, $\mathrm{E}(z) = 0$ and $\sigma_z = 1$. By the central limit theorem, $z \sim N(0,1)$, so we can use the standard normal table to find confidence intervals and p-values for z.

4. A Sample Problem

Freedman, Pisani, and Purves, p. 510: Freshman at public universities work 12.2 hours per week for pay on the average, with a standard deviation of 10.5. At private universities, the average for freshman is 9.2 hours, with a standard deviation of 9.9 hours. The sample size for each is 1,000. Is the difference between the averages real or is it just chance variation. Perform a level 0.05 independent two-sample test to find out.

Solution: We have these inputs:

$$
\begin{aligned}
n_1 &= 1,000 \qquad n_2 = 1,000 \qquad \bar{x}_1 = 12.2 \qquad \bar{x}_2 = 9.2 \\
s_1 &= 10.2 \qquad s_2 = 9.9 \qquad \alpha = 0.05 \\
\mathrm{SE}_1 &= \frac{\sigma_1}{\sqrt{n_1}} = \frac{10.5}{\sqrt{1,000}} = 0.332 \qquad \mathrm{SE}_2 = \frac{\sigma_2}{\sqrt{n_1}} = \frac{9.9}{\sqrt{1,000}} = 0.332 \\
\mathrm{SE}_{\text{diff}} &= \sqrt{\mathrm{SE}_1^2 + \mathrm{SE}_2^2} = \sqrt{0.332^2 + 0.313^2} = 0.463
\end{aligned}
$$

The five steps of the z-test:

1. State the null and alternative hypotheses:

$$H_0 : \mu_1 = \mu_2, \qquad H_1 : \mu_1 \neq \mu_2$$

2. Compute the test statistic:

$$z = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)}{\mathrm{SE}_{\text{diff}}} = \frac{(10.2 - 12.2)}{0.463} = -4.320$$

3. Compute a $100(1-\alpha)\%$ confidence interval $I : [-1.96, 1.96]$, using the standard normal table.

4. Determine whether to accept or reject H_0: $-4.320 \notin [-1.96, 1.96]$, so reject H_0.

5. Compute the p-value: if u is standard normal,

$$P(u \leq -z) = P(u \leq -4.32) = 0.00000780.$$

By the symmetry of the normal curve,

$$P(z \leq u) = P(4.32 \leq u) = 0.00000780.$$

Thus $p = 0.00000780 + 0.00000780 = 0.00001460$.