Two sample Z test for difference means

Function of the test:

The function of the test is to determine whether two samples drawn from the two independent populations have same mean or not, provided that population standard deviations are known.

Test Assumptions

1. Sampling method for each sample is simple random sampling
2. Samples are independent
3. Two populations are normally distributed if samples are of sizes less than 30 but do not require this condition if sample are larger than 30 (The sampling distribution of difference of two means is normally distribution, generally it is true by virtue of central limit theorem, if sample of at least 30 observations taken)
4. Population standard deviations are known

Hypothesis to test

Let μ_1 and μ_2 be the population mean of population I and population II respectively from which samples are drawn independently.

Two-tailed test	**Left-tailed test**	**Right-tailed test**
H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$	H_0: $\mu_1 \geq \mu_2$ H_1: $\mu_1 < \mu_2$	H_0: $\mu_1 \leq \mu_2$ H_1: $\mu_1 > \mu_2$

Test Statistic

The appropriate test statistic is given by,

$$Z = \frac{\text{Sample difference of means - Hypothesized difference of means}}{\text{Standard error of difference of means}}$$

$$= \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

In practice, the two-sample *z*-test is not used often, because the two population standard deviations σ_1 and σ_2 are usually unknown. Instead, sample standard deviations and the *t*-distribution are used. However, if samples are of large sizes, even though population standard deviations are not known, we can still use Z distribution replacing population standard deviations by respective sample standard deviations. The test statistic is given by,

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Decision Rule

We adopt following decision rule:

Hypothesis	Decision Rule
Case I (Two-sided test) **H_0: $\mu_1 = \mu_2$ vs $\mu_1 \neq \mu_2$**	Reject H_0 if cal Z $\leq$ - Z_α Or cal Z $\geq$ + Z_α Alternatively, reject H_0 if |cal Z| $\geq Z_{\alpha/2}$
Case II (Left sided test) **H_0: $\mu_1 \geq \mu_2$ vs $\mu_1 < \mu_2$**	Reject H_0 if cal Z $\leq$ - Z_α Alternatively, reject H_0 if |cal Z| $\geq Z_\alpha$
Case III (Right sided test) **H_0: $\mu_1 \leq \mu_2$ vs $\mu_1 > \mu_2$**	Reject H_0 if cal Z $\geq$ + Z_α Alternatively, reject H_0 if |cal Z| $\geq Z_\alpha$

Problem

To compare the starting salaries of college graduates majoring in engineering and computer science a random sample of 50 recent college graduates in each major were selected and the following information obtained.

Major	Mean ($)	SD($)
Engineering	56,202	2225
Computer Science	50,657	2375

Do the data provide sufficient evidence to indicate a **difference** in average starting annual salaries for college graduates who majored in engineering and computer science? Test using α = 0.05.

Solution:

Let X1 be the annual salary of students with Engineering major and X2 be the annual salary of students with Computer Science major

Sample mean ($\overline{X}_1$) = 56202 and Sample mean ($\overline{X}_2$) = 50657

Sample SD (S1) = 2225 and Sample SD (S2) = 2375

Sample size (n1) = 50 and Sample size (n2) = 50

Step 1: Setting up Null and Alternative hypothesis

Let μ_1 and μ_2 be the population mean of population I and population II respectively from which samples are drawn independently.

Null Hypothesis

H_0: $\mu_1 = \mu_2$ (There is no significant difference in the annual salary of Engineering major group and Computer science major group)

Alternative Hypothesis

H_1: $\mu_1 \neq \mu_2$ (There is significant difference in the annual salary of Engineering major group and Computer science major group)

Step 2: Choice of α for the test

The level of significance for the test (α) = 5 % = 0.05

Step 3: Test statistic

The appropriate test statistic for this test is given by

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Although, population standard deviations are not known, we can still use Z distribution as each sample size exceed 30

The test statistic has SND with mean = 0 and SD = 1

Step 4: Critical value of Z

The test is two-sides and level of significance (α) = 5 %. So there are two rejection regions

The critical value of Z from Z table is given by,

Z = 1.96

AR: -196 < Cal Z < + 1.96

RR: Either Z ≥ + 1.96 (URR) Or Z ≤ - 1.96 (LRR)

Decision Rule is that accept H0 if cal Z falls in the acceptance region otherwise reject it.

Step 5: Calculated Z

$$Z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{56202 - 50657}{\sqrt{\frac{2225^2}{50} + \frac{2375^2}{50}}} = +\ 12.047$$

Hence, cal Z = + 12.047

Step 6: Statistical Decision

Since, calculated Z = + 12.047 falls in the upper rejection region (Z > + 1.96) i.e. $|12.047| = 12.047 >> Critical Z = 1.96$, we strongly reject null hypothesis H0 at % % level of significance and in favor of alternative hypothesis.

Step 7 : Conclusions

There is significant difference in the mean annual salary between Engineering major group and Computer science major group. Since the average annual salary of Engineering is 56,202 which is greater that of computer science major (50,657) we can say graduates with engineering majors are getting significantly higher salary than graduates with computer science.