Example:

Fifth grade students from two neighboring counties took a placement exam.

- Group 1, from County 1, consisted of 57 students. The sample mean score for these students was 77.2 and the true variance is known to be 15.3.

- Group 2, from County 2, consisted of 63 students and had a sample mean score of 75.3 and the true variance is known to be 19.7.

Example:

From previous years of data, it is believed that the scores for both counties are normally distributed.

Derive a test to determine whether or not the two population means are the same.

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

- Suppose that $X_{1,1}, X_{1,2}, \ldots, X_{1,n_1}$ is a random sample of size n_1 from the normal distribution with mean μ_1 and variance σ_1^2.

- Suppose that $X_{2,1}, X_{2,2}, \ldots, X_{2,n_2}$ is a random sample of size n_2 from the normal distribution with mean μ_2 and variance σ_2^2.

- Suppose that σ_1^2 and σ_2^2 are known and that the samples are independent.

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

$$\Downarrow$$

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 \neq 0$$

Think of this as

$\theta = 0$ **versus** $\theta \neq 0$

for

$$\theta = \mu_1 - \mu_2$$

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

Step One:

Choose an estimator for $\theta = \mu_1 - \mu_2$.

$$\widehat{\theta} = \overline{X}_1 - \overline{X}_2$$

Step Two:

Give the "form" of the test.

Reject H_0, in favor of H_1 if either

$$\widehat{\theta} > c \quad \text{or} \quad \widehat{\theta} < -c$$

for some c to be determined.

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

Step Three:

Find c using α.

Will be working with the random variable

$$\overline{X}_1 - \overline{X}_2$$

We need to know its distribution…

Distribution of the statistic:

- $\overline{X}_1 - \overline{X}_2$ is normally distributed

Mean:

$$E\left[\overline{X}_1 - \overline{X}_2\right] = E[\overline{X}_1] - E[\overline{X}_2]$$

$$= \mu_1 - \mu_2$$

Distribution of the statistic:

- $\overline{X}_1 - \overline{X}_2$ is normally distributed

Variance:

$$\mathrm{Var}\left[\overline{X}_1 - \overline{X}_2\right] = \mathrm{Var}[\overline{X}_1 + (-1)\overline{X}_2]$$

$$\overset{\text{indep}}{=} \mathrm{Var}[\overline{X}_1] + \mathrm{Var}[(-1)\overline{X}_2]$$

$$= \mathrm{Var}[\overline{X}_1] + (-1)^2\mathrm{Var}[\overline{X}_2]$$

$$= \mathrm{Var}[\overline{X}_1] + \mathrm{Var}[\overline{X}_2]$$

Distribution of the statistic:

- $\overline{X}_1 - \overline{X}_2$ is normally distributed

Variance:

$$\text{Var}\left[\overline{X}_1 - \overline{X}_2\right] \overset{\text{indep}}{=} \text{Var}[\overline{X}_1] + \text{Var}[\overline{X}_2]$$

$$= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

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

Step Three:

Find c using α.

$\overline{X}_1 - \overline{X}_2$ **is normally distributed**

$$\overline{X}_1 - \overline{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_1}\right)$$

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

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

Step Three:

$$\alpha = P(\text{ Type I Error })$$

$$= P(\text{ Reject } H_0 \; ; \theta = 0)$$

$$= P(\overline{X}_1 - \overline{X}_2 > c \text{ or } \overline{X}_1 - \overline{X}_2 < -c ; \theta = 0)$$

$$= 1 - P(-c \leq \overline{X}_1 - \overline{X}_2 \leq c \; ; \underbrace{\theta = 0}_{\mu_1 - \mu_2 = 0})$$

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

Step Three:

$$= 1 - P(-c \leq \overline{X}_1 - \overline{X}_2 \leq c \; ; \theta = 0)$$

- **Subtract** $\mu_1 - \mu_2$ **(which is 0)**
- **Divide by**

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

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

Step Three:

$$\alpha = 1 - P\left(\frac{-c}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq Z \leq \frac{c}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right)$$

$$1 - \alpha = P\left(\frac{-c}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq Z \leq \frac{c}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right)$$

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

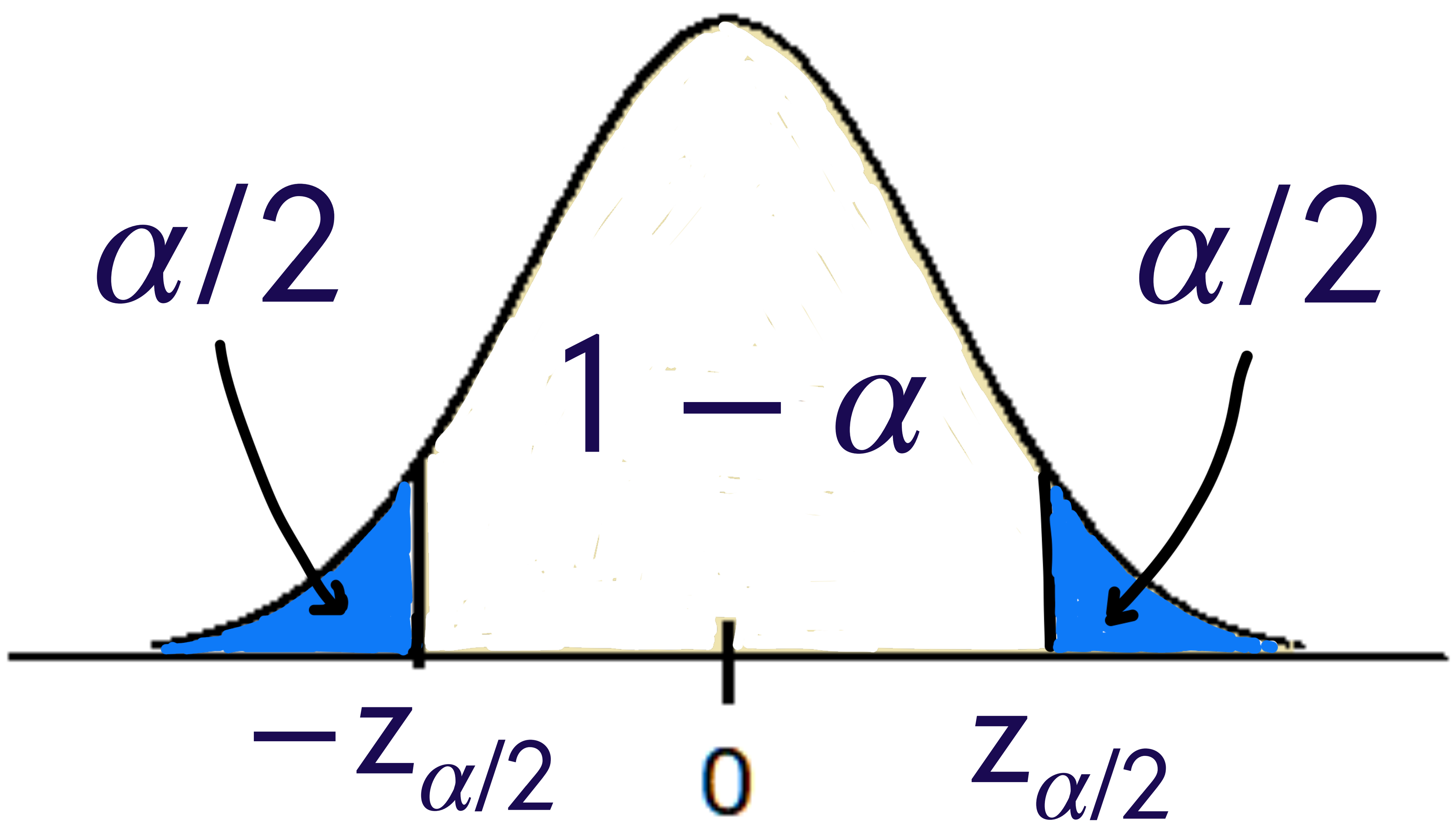
Step Three:

$$\alpha = 1 - P\left(\frac{-c}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq Z \leq \frac{c}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right)$$

$$1 - \alpha = P\left(\frac{-c}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq Z \leq \frac{c}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right)$$

Step Three:

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



$$\Rightarrow \quad \frac{c}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = z_{\alpha/2}$$

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

Step Four:

Conclusion:

Reject H_0**, in favor of** H_1**, if**

$$\overline{X}_1 - \overline{X}_2 > z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

or

$$\overline{X}_1 - \overline{X}_2 < - z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Example:

Fifth grade students from two neighboring counties took a placement exam.

Group 1, from County 1, consisted of 57 students. The sample mean score for these students was 77.2. Group 2, from County 2, consisted of 63 students and had a sample mean score of 75.3.

From previous years of data, it is believed that the scores for both counties are normally distributed, and that the variances of scores from Counties A and B, respectively, are 15.3 and 19.7.

Example:

Fifth grade students from two neighboring counties took a placement exam.

Derive a test to figure out whether or not the two population means are the same.

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

Example:

$$n_1 = 57 \qquad n_2 = 63$$

$$\bar{x}_1 = 77.2 \qquad \bar{x}_2 = 75.3$$

$$\sigma_1^2 = 15.3 \qquad \sigma_2^2 = 19.7$$

Suppose that $\alpha = 0.05$**.**

$$z_{\alpha/2} = z_{0.025} = 1.96$$

$$z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 1.49$$

Example:

$$z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} = 1.49$$

$$\bar{x}_1 - \bar{x}_2 = 77.2 - 75.3 = 1.9$$

So,

$$\bar{x}_1 - \bar{x}_2 > z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

and we reject H_0. The data suggests that the true mean scores for the counties are different!