- 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 $X_{2,1}, X_{2,2}, \ldots, X_{2,n}$ 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.

Find a $100(1-\alpha)\%$ confidence interval for the difference $\mu_1 - \mu_2$.

Step One:

An estimator: $\overline{X}_1 - \overline{X}_2$

Step Two:

Distribution of the estimator:

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

Step Two:

Distribution of the estimator:

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

Mean:

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

$$= \mu_1 - \mu_2$$

Step Two:

Distribution of the estimator:

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

Variance:

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

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

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

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

Step Two:

Distribution of the estimator:

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

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

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

Step Three:

Critical values:

$$-z_{\alpha/2} < \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < z_{\alpha/2}$$

Step Four:

Solve for $\mu_1 - \mu_2$ "in the middle".

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

$$< \mu_1 - \mu_2$$

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

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

Example:

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

Group 1, from County A, consisted of 57 students. The sample mean score for these students was 77.2. Group 2, from County B, 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.

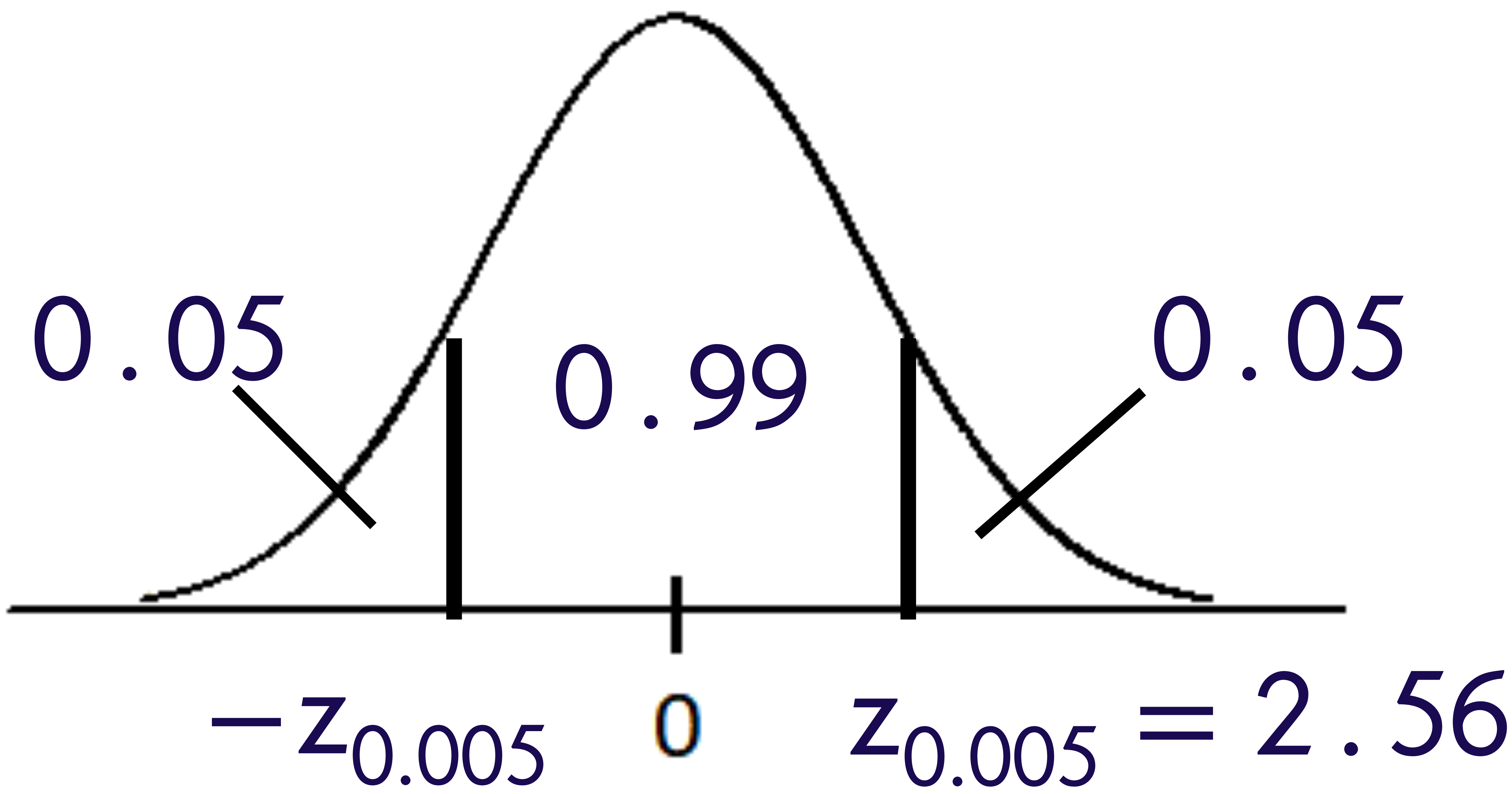
Example:

Find and interpret a 99% confidence interval for $\mu_1 - \mu_2$, the difference in the means for the counties.

$$n_1 = 57,\ \bar{x}_1 = 77.2,\ \sigma_1^2 = 15.3$$

$$n_2 = 63,\ \bar{x}_2 = 75.3,\ \sigma_2^2 = 19.7$$

Z-curve: (standard normal curve)



qnorm(0.995)

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

gives $(-0.0515,\ 3.8515)$

$(-0.0515, \ 3.8515)$

This is an interval of “plausible values” for the difference $\mu_1 - \mu_2$.

Since it contains the value 0, it is plausible that $\mu_1 - \mu_2 = 0$.

i.e. It is plausible that $\mu_1 = \mu_2$.

$$(-0.0515,\ 3.8515)$$

It also seems like we can remove this possibility if we change the percentage for our confidence interval.

DATA SNOOPING!

For large samples $(n_1 > 30\ n_2 > 30)$

$$\boxed{\overline{X}_1 - \overline{X}_2 \pm z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

is an approximate $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$.