Example:

A random sample of 500 people in a certain county which is about to have a national election were asked whether they preferred “Candidate A” or “Candidate B”.

From this sample, 320 people responded that they preferred Candidate A.

A random sample of 400 people in a second county which is about to have a national election were asked whether they preferred “Candidate A” or “Candidate B”.

Example:

From this second county sample, 268 people responded that they preferred Candidate A.

$$\hat{p}_1 = \frac{320}{500} = 0.64$$

$$\hat{p}_2 = \frac{268}{400} = 0.67$$

Test

$$H_0 : p_1 = p_2 \qquad H_1 : p_1 \neq p_2$$

Change to:

$$H_0 : p_1 - p_2 = 0$$

$$H_1 : p_1 - p_2 \neq 0$$

Estimate $p_1 - p_2$ with $\hat{p}_1 - \hat{p}_2$.

For large enough samples,

$$\hat{p}_1 \overset{\text{approx}}{\sim} N\left(p_1, \frac{p_1(1-p_1)}{n_1}\right)$$

and

$$\hat{p}_2 \overset{\text{approx}}{\sim} N\left(p_2, \frac{p_2(1-p_2)}{n_1}\right)$$

$$\hat{p}_1 - \hat{p}_2 \sim N(?, ?)$$

$$E[\hat{p}_1 - \hat{p}_2] = E[\hat{p}_1] - E[\hat{p}_2] = p_1 - p_2$$

$$\mathrm{Var}[\hat{p}_1 - \hat{p}_2] \overset{\text{indep}}{=} \mathrm{Var}[\hat{p}_1] + \mathrm{Var}[\hat{p}_2]$$

$$= \frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}$$

$= 0$ under H_0

$$\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}} \sim N(0, 1)$$

$= ?$ under H_0

Use estimators for p_1 and p_2 assuming they are the same.

- Call the common value p.
- Estimate by putting both groups together.

In the example with

$$\hat{p}_1 = \frac{320}{500} = 0.64 \qquad \hat{p}_2 = \frac{268}{400} = 0.67$$

we have

$$\hat{p} = \frac{320 + 268}{500 + 400} = \frac{588}{900} = \frac{49}{75}$$

$$\approx 0.6533$$

Use

approx

$$Z := \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}} \sim N(0, 1)$$

$$= \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Two-tailed test with z-critical values…

Example:

320 people out of a random sample of size 500 from County 1 prefer Candidate A.

268 people out of a random sample of size 400 from County prefer Candidate A.

$$\hat{p}_1 = \frac{320}{500} = 0.64 \qquad \hat{p}_2 = \frac{268}{400} = 0.67$$

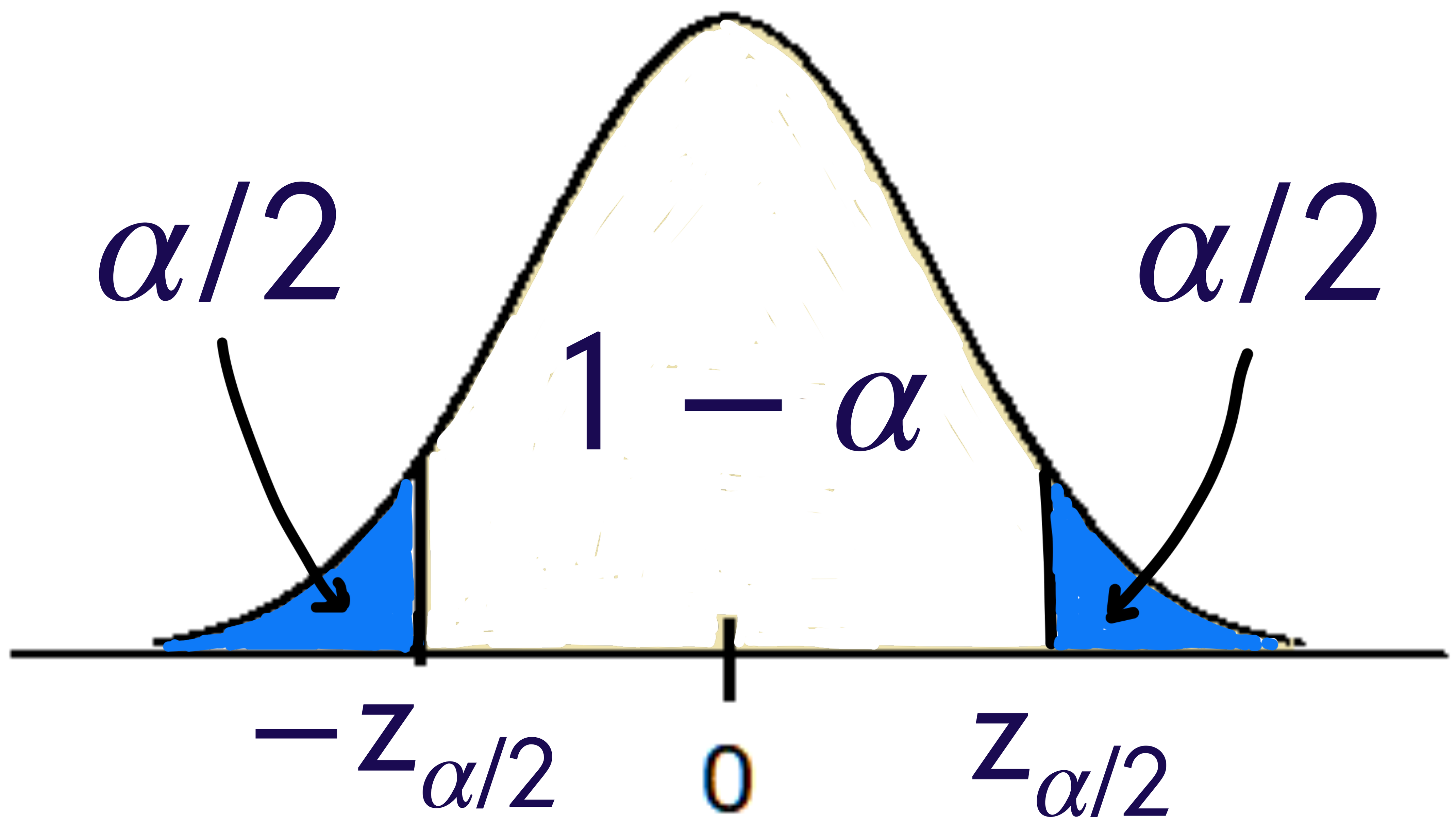
Use $\alpha = 0.05$.

Example:

$$\hat{p} = \frac{320 + 268}{500 + 400} = \frac{588}{900} = \frac{49}{75}$$

$$Z = \frac{0.64 - 0.67 - 0}{\sqrt{0.6533(1 - 0.6533)\left(\frac{1}{500} + \frac{1}{400}\right)}}$$

$$\approx -0.9397$$



$$z_{0.025} = 1.96$$

$$\text{qnorm}(1 - 0.05/2)$$

Z=-0.9397 does not fall in the rejection region!

Example:

The data suggests that the true proportion of people who like Candidate A is not different for the 2 counties.

```
rop.test(x = c(320, 268), n = c(500, 400))

     2-sample test for equality of proportions with

ta:  c(320, 268) out of c(500, 400)
squared = 0.75556, df = 1, p-value = 0.3847
ternative hypothesis: two.sided
 percent confidence interval:
.094648  0.034648
mple estimates:
op 1 prop 2
.64   0.67
```