Example:

A random sample of 500 people in a certain country 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.

Let p be the true proportion of the people in the country who prefer Candidate A.

Test the hypotheses

$$H_0 : p \leq 0.65$$ versus

$$H_1 : p > 0.65$$

Use level of significance 0.10.

We have an estimate

$$\hat{p} = \frac{320}{500} = \frac{16}{25}$$

The Model:

Take a random sample of size n.

Record $X_1, X_2, \ldots, X_n$ where

$$X_i = \begin{cases} 1 & \text{person i likes Candidate A} \\ 0 & \text{person i likes Candidate B} \end{cases}$$

Then $X_1, X_2, \ldots, X_n$ is a random sample from the Bernoulli distribution with parameter p.

The Model:

Note that, with these 1's and 0's,

$$\hat{p} = \frac{\textbf{\# in the sample who like A}}{\textbf{\# in the sample}}$$

$$= \frac{\sum_{i=1}^{n} X_i}{n} = \overline{X}$$

By the Central Limit Theorem, $\hat{p} = \overline{X}$ has, for large samples, an approximately normal distribution.

The Model: $\hat{p} = \overline{X}$

$$E[\hat{p}] = E[X_1] = p$$

$$Var[\hat{p}] = \frac{Var[X_1]}{n} = \frac{p(1-p)}{n}$$

So, $$\hat{p} \overset{\text{approx}}{\sim} N\left(p, \frac{p(1-p)}{n}\right)$$

The Model: $\hat{p} = \overline{X}$

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

In particular,

$$\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

behaves roughly like a N(0,1) as n gets large

The Model: $\hat{p} = \overline{X}$

What does "large" mean?

"n>30" is a rule of thumb to apply to all distributions, but we can (and should!) do better with specific distributions.

- $\hat{p}$ lives between 0 and 1.
- the normal distribution lives between $-\infty$ and ∞

- $\hat{p}$ lives between 0 and 1.
- The normal distribution lives between $-\infty$ and ∞.
- However, 99.7% of the area under a N(0,1) curve lies between -3 and 3,

i.e. “99.7% of the probability for a normal distribution is within 3 standard deviations of it’s mean

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

$$\Rightarrow \quad \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Go forward using normality if the interval

$$\left(\hat{p} - 3\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + 3\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

is completely contained within [0,1].

Step One:

$$H_0 : p \leq 0.65$$
$$H_1 : p > 0.65$$

Choose a statistic.

$\hat{p}$ = sample proportion for Candidate A

Step Two:

Form of the test.

Reject H_0, in favor of H_1, if $\hat{p} > c$.

Step Three:

$H_0: p \leq 0.65$

$H_1: p > 0.65$

Use α to find c.

Assume normality of $\hat{p}$?

- It is a sample mean and $n > 30$.
- The interval

$$\left(\hat{p} - 3\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + 3\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

is $(0.5756, 0.7044)$ ✓

$H_0 : p \leq 0.65$

$H_1 : p > 0.65$

Step Three:

Use α to find c.

$$\alpha = \max_{p \in H_0} P(\text{Type I Error})$$

$$= \max_{p \leq 0.65} P(\text{Reject } H_0 ; p)$$

$$= \max_{p \leq 0.65} P(\hat{p} > c ; p)$$

$$H_0 : p \leq 0.65$$
$$H_1 : p > 0.65$$

Step Three:

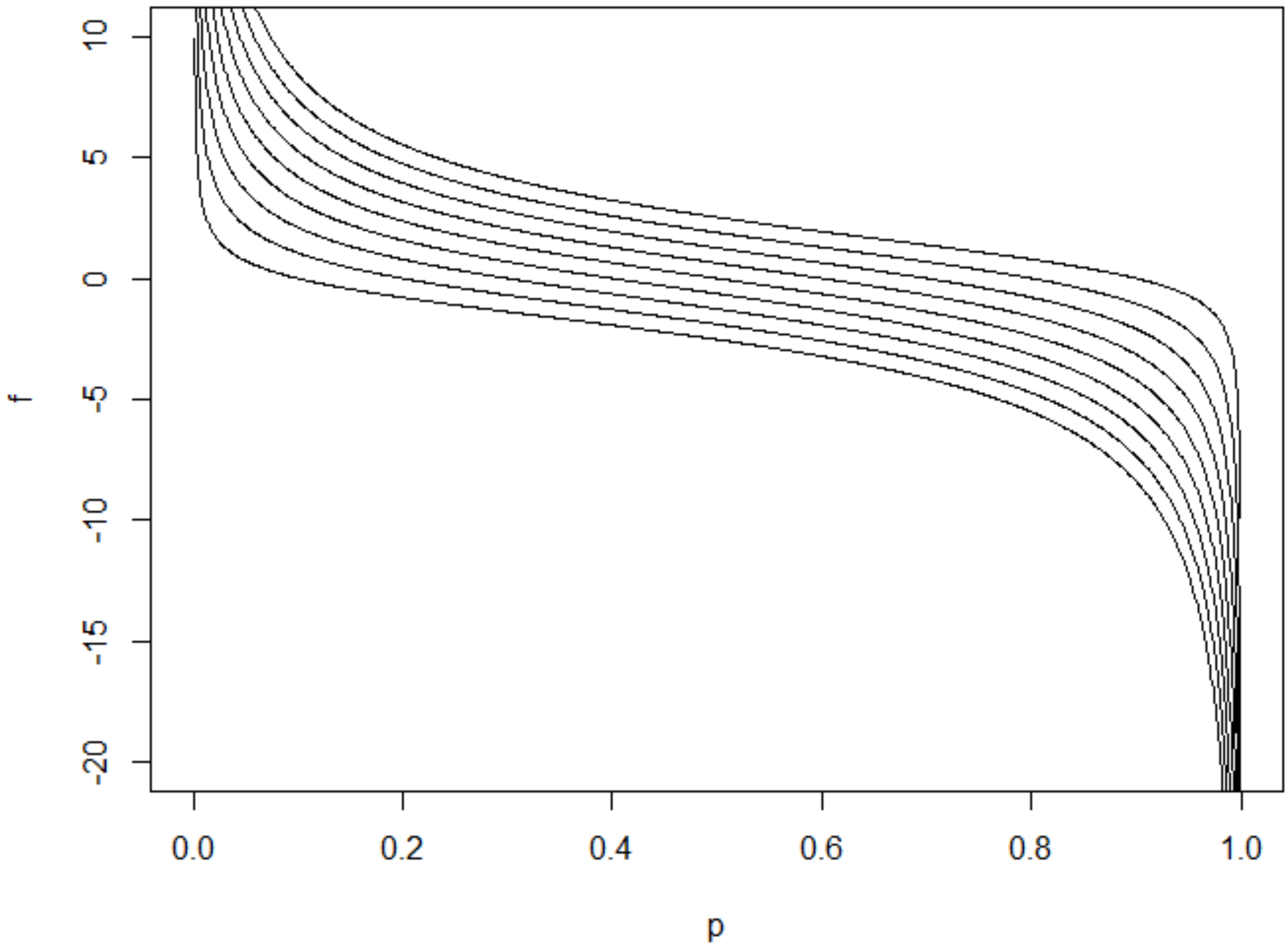
Use α to find c.

$$\alpha = \max_{p \leq 0.65} P\left(\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} > \frac{c - p}{\sqrt{\frac{p(1-p)}{n}}} ; p \right)$$

$$\approx \max_{p \leq 0.65} P\left(Z > \frac{c - p}{\sqrt{\frac{p(1-p)}{n}}} \right)$$

$$0 < c < 1$$

$$\frac{c - p}{\sqrt{\frac{p(1-p)}{n}}}$$



$$0.10 = \max_{p \le 0.65} P\left(Z > \frac{c-p}{\sqrt{\frac{p(1-p)}{n}}}\right)$$

$$= P\left(Z > \frac{c-0.65}{\sqrt{\frac{0.65(1-0.65)}{n}}}\right)$$

$$\Rightarrow \quad \frac{c-0.65}{\sqrt{\frac{0.65(1-0.65)}{n}}} = z_{0.10}$$

Reject H_0 if

$$\hat{p} > 0.65 + z_{0.10}\sqrt{\frac{0.65(1-0.65)}{n}}$$

Back to the example:

Let p be the true proportion of the people in the country who prefer Candidate A.

$$n = 500 \qquad \hat{p} = \frac{16}{25} = 0.64$$

$$\alpha = 0.10 \qquad z_{0.10} = 1.28$$

$$0.65 + z_{0.10}\sqrt{\frac{0.65(1-0.65)}{n}} = 0.6773$$

Reject H_0 if

$$\underset{\uparrow\; 0.64}{\hat{p}} > \underbrace{0.65 + z_{0.10}\sqrt{\frac{0.65(1-0.65)}{n}}}_{0.6773}$$

We fail to reject H_0, in favor of H_1.

The data do not suggest that the true proportion of people who like Candidate A is greater than 0.65.