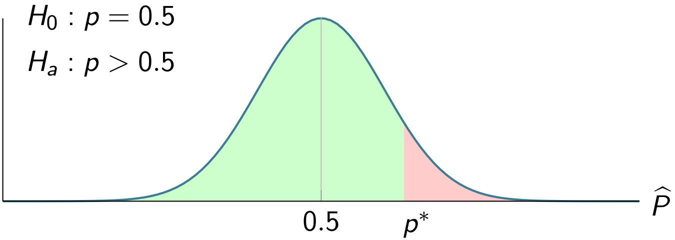
Hypotheses and sample sizes

Chapter 23

Ričardas Zitikis

School of Mathematical and Statistical Sciences
Western University, Ontario



Sample size n, and sample proportion $\widehat{p}$

Significance level α, say $\alpha = 0.05$

Problem: find p^* such that

- retain H_0 if $\widehat{p} < p^*$
- reject H_0 if $\widehat{p} > p^*$

Answer: p^* is the number such that

$$\Pr\left(\widehat{P} > p^*\right) = 0.05$$

Reducing to the standard normal

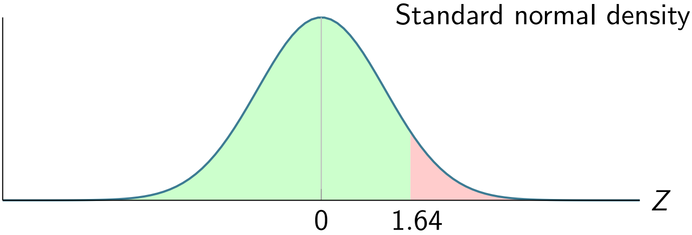
$$\begin{aligned}
\Pr\left(\widehat{P} > p^*\right) &= \Pr\left(\frac{\widehat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} > \frac{p^* - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right) \quad \text{with} \quad p_0 = 0.5 \\
&\approx \Pr\left(Z > \frac{p^* - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right) \qquad \text{(approximately normal)} \\
&= \Pr\left(Z > \frac{p^* - 0.5}{0.5/\sqrt{n}}\right)
\end{aligned}$$

Our task is to find p^* such that

$$\Pr\left(Z > \frac{p^* - 0.5}{0.5/\sqrt{n}}\right) = 0.05$$

In the case of the standard normal, we have (Table B)

$$\Pr(Z > 1.64) = 0.05$$



Hence

$$\frac{p^* - 0.5}{0.5/\sqrt{n}} = 1.64$$

which is equivalent to

$$p^* = 0.5 + 1.64\frac{0.5}{\sqrt{n}}$$

Decision rule

We now have the decision rule

- retain H_0 if $\widehat{p} < 0.5 + 1.64 \frac{0.5}{\sqrt{n}} = 0.5 + \frac{0.82}{\sqrt{n}}$
- reject H_0 if $\widehat{p} > 0.5 + 1.64 \frac{0.5}{\sqrt{n}} = 0.5 + \frac{0.82}{\sqrt{n}}$

whose construction has been based on the assumption that the null H_0 is true.

Alternatives in the H_0 rejection region

Given the decision rule

- retain H_0 if $\widehat{p} < 0.5 + 1.64\frac{0.5}{\sqrt{n}} = 0.5 + \frac{0.82}{\sqrt{n}}$
- reject H_0 if $\widehat{p} > 0.5 + 1.64\frac{0.5}{\sqrt{n}} = 0.5 + \frac{0.82}{\sqrt{n}}$

what is the **sample size** that (rightly) puts the alternatives

- $H_a : p = 0.7$
- $H_a : p = 0.502$

into the H_0 rejection region?

- The alternative value $p = 0.7$ is in the H_0 rejection region when

$$0.7 > 0.5 + \frac{0.82}{\sqrt{n}}$$

which is equivalent to

$$n \geq 16.81.$$

- The alternative value $p = 0.502$ is in the H_0 rejection region when

$$0.502 > 0.5 + \frac{0.82}{\sqrt{n}}$$

which is equivalent to

$$n \geq 168,100$$

Hence, we need $10,000$ times more observations than previously

The probability of (rightly) accepting H_a

Given the decision rule

- retain H_0 if $\widehat{p} < 0.5 + 1.64\frac{0.5}{\sqrt{n}} = 0.5 + \frac{0.82}{\sqrt{n}}$
- reject H_0 if $\widehat{p} > 0.5 + 1.64\frac{0.5}{\sqrt{n}} = 0.5 + \frac{0.82}{\sqrt{n}}$

what is the **probability** that we shall (rightly) accept the alternatives

- $H_a : p = 0.7$
- $H_a : p = 0.502$

when they are actually true?

Let's slightly generalize the problem: given the decision rule

- retain H_0 if $\widehat{p} < 0.5 + 1.64\frac{0.5}{\sqrt{n}}$
- reject H_0 if $\widehat{p} > 0.5 + 1.64\frac{0.5}{\sqrt{n}}$

what is the probability that we shall accept the alternative

- $H_a : p = p_a$

when it is actually true? The probability is equal to

$$\Pr\left(\widehat{P} > 0.5 + 1.64\frac{0.5}{\sqrt{n}}\right) = \Pr\left(\frac{\widehat{P} - p_a}{\sqrt{\frac{p_a(1-p_a)}{n}}} > \frac{0.5 - p_a + 1.64\frac{0.5}{\sqrt{n}}}{\sqrt{\frac{p_a(1-p_a)}{n}}}\right)$$

$$\approx \Pr\left(Z > \frac{\sqrt{n}(0.5 - p_a) + 0.82}{\sqrt{p_a(1 - p_a)}}\right)$$

When $p_a = 0.7$, we have

$$\begin{aligned}\Pr\left(\widehat{P} > 0.5 + 1.64\frac{0.5}{\sqrt{n}}\right) &\approx \Pr\left(Z > \frac{\sqrt{n}(0.5 - p_a) + 0.82}{\sqrt{p_a(1-p_a)}}\right)\\ &= \Pr\left(Z > -\sqrt{n}\,0.436 + 1.789\right)\end{aligned}$$

When $p_a = 0.502$, we have

$$\begin{aligned}\Pr\left(\widehat{P} > 0.5 + 1.64\frac{0.5}{\sqrt{n}}\right) &\approx \Pr\left(Z > \frac{\sqrt{n}(0.5 - p_a) + 0.82}{\sqrt{p_a(1-p_a)}}\right)\\ &= \Pr\left(Z > -\sqrt{n}\,0.004 + 1.640\right)\end{aligned}$$

For example, when $n = 16$, the first probability is around 0.5, whereas the second probability is just around 0.05

The required sample size for a given probability

Question: What is the sample size that makes the probability of accepting the alternative when it is true to be at least 0.95?

We accept the alternative $H_a : p = p_a$ with the probability

$$\Pr\left(Z > \frac{\sqrt{n}(0.5 - p_a) + 0.82}{\sqrt{p_a(1 - p_a)}}\right) \geq 0.95$$

This condition is equivalent to

$$\Pr\left(Z \leq \frac{\sqrt{n}(0.5 - p_a) + 0.82}{\sqrt{p_a(1 - p_a)}}\right) \leq 0.05$$

which is equivalent to

$$\frac{\sqrt{n}(0.5 - p_a) + 0.82}{\sqrt{p_a(1 - p_a)}} \leq -1.64$$

The condition

$$\frac{\sqrt{n}(0.5 - p_a) + 0.82}{\sqrt{p_a(1 - p_a)}} \leq -1.64$$

is equivalent to

$$\sqrt{n}(0.5 - p_a) + 0.82 \leq -1.64\sqrt{p_a(1 - p_a)}$$

which is equivalent to

$$\sqrt{n} \geq \frac{0.82 + 1.64\sqrt{p_a(1 - p_a)}}{p_a - 0.5}$$

where $\leq$ has turned into $\geq$ because $0.5 - p_a < 0$ (i.e. $p_a > 0.5$)

Let's now see what this means when $p_a = 0.7$ and $p_a = 0.502$

When $p_a = 0.7$, the condition

$$\sqrt{n} \geq \frac{0.82 + 1.64\sqrt{p_a(1 - p_a)}}{p_a - 0.5}$$

is equivalent to

$$n \geq 61.74364$$

When $p_a = 0.502$, the same condition is equivalent to

$$n \geq 672,394.6$$

As we see, given our decision rule based on the null $H_0 : p = 0.5$, we rightly accept the alternative $H_a : p = 0.7$ when it is true with probability at least 0.95 only when we have at least $n = 62$ observations, and do the same with the alternative $H_a : p = 0.502$ only when we have at least $n = 672,395$ observations, which is nearly $11,000$ times more than in the previous case.