Lecture 16 - Decisions and Power

Sta102 / BME 102

March 7th, 2016

Colin Rundel & Mine Çetinkaya-Rundel

Decisions and Decision Errors

Decision errors for HTs

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

Decision errors for HTs

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true		
	H_A true		

Decision errors for HTs

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	
	H_A true		

Decision errors for HTs

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	
	H_A true		✓

Decision errors for HTs

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	*Type 1 Error*
	H_A true		✓

- A *Type 1 Error* is rejecting the null hypothesis when H_0 is true.

Decision errors for HTs

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	*Type 1 Error*
	H_A true	*Type 2 Error*	✓

- A *Type 1 Error* is rejecting the null hypothesis when H_0 is true.
- A *Type 2 Error* is failing to reject the null hypothesis when H_A is true.

Type 1 error rate

As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* (α) of 0.05.

Type 1 error rate

As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* (α) of 0.05.

This means that, for those cases where H_0 is actually true, we will incorrectly reject it at most 5% of the time.

Type 1 error rate

As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* (α) of 0.05.

This means that, for those cases where H_0 is actually true, we will incorrectly reject it at most 5% of the time.

$$P(\text{Type 1 error}) = P(\text{Rejecting } H_0 \mid H_0 \text{ is true}) = \alpha$$

Type 1 error rate

As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* (α) of 0.05.

This means that, for those cases where H_0 is actually true, we will incorrectly reject it at most 5% of the time.

$$P(\text{Type 1 error}) = P(\text{Rejecting } H_0 \mid H_0 \text{ is true}) = \alpha$$

This is why we prefer small values of α – decreasing α decreases our Type 1 error rate.

Filling in the table...

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true		
	H_A true		

Filling in the table...

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true		α
	H_A true		

Filling in the table...

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true		α
	H_A true	β	

Filling in the table...

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	$1 - \alpha$	α
	H_A true	β	

Filling in the table...

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	$1-\alpha$	α
	H_A true	β	$1-\beta$

Filling in the table...

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	$1-\alpha$	α
	H_A true	β	$1-\beta$

Type 1 error rate $= \alpha = P(\text{Rejecting } H_0 \mid H_0 \text{ is true})$

Type 2 error rate $= \beta = P(\text{Failing to reject } H_0 \mid H_A \text{ is true})$

Power $= 1 - \beta = P(\text{Rejecting } H_0 \mid H_A \text{ is true})$

Type 2 error rate

The type 2 error is defined as

$$\beta = P(\text{Failing to reject } H_0 \mid H_A \text{ is true}) = \;?$$

Type 2 error rate

The type 2 error is defined as

$$\beta = P(\text{Failing to reject } H_0 \mid H_A \text{ is true}) = \, ?$$

It is not immediately obvious how to calculate this probability (or its complement) but we can come up with some basic rules:

- If the true population average is very close to the null hypothesis value (δ likely to be small), it will be difficult to detect the difference (and reject H_0).
- If the true population average is very different from the null hypothesis value (δ likely to be large), it will be easy to detect the difference.

Type 2 error rate - intuition

Intuitively, β depends on

- δ (effect size)
- α (significance level)
- n (sample size)

Type 2 error rate - intuition

Intuitively, β depends on

- δ (effect size)
- α (significance level)
- n (sample size)

to increase power / decrease β:

- increase n,
- increase δ, and/or
- increase α

Power

Example - Blood Pressure

Blood pressure oscillates with the beating of the heart, and the systolic pressure is defined as the peak pressure when a person is at rest. The average systolic blood pressure for people in the U.S. is about 130 mmHg with a standard deviation of about 25 mmHg.

We are interested in finding out if the average blood pressure of employees at a certain company is *greater* than the national average, so we collect a random sample of 100 employees and measure their systolic blood pressure. What are the hypotheses?

Example - Blood Pressure

Blood pressure oscillates with the beating of the heart, and the systolic pressure is defined as the peak pressure when a person is at rest. The average systolic blood pressure for people in the U.S. is about 130 mmHg with a standard deviation of about 25 mmHg.

We are interested in finding out if the average blood pressure of employees at a certain company is *greater* than the national average, so we collect a random sample of 100 employees and measure their systolic blood pressure. What are the hypotheses?

$$H_0 : \mu = 130$$
$$H_A : \mu > 130$$

Example - Blood Pressure

Blood pressure oscillates with the beating of the heart, and the systolic pressure is defined as the peak pressure when a person is at rest. The average systolic blood pressure for people in the U.S. is about 130 mmHg with a standard deviation of about 25 mmHg.

We are interested in finding out if the average blood pressure of employees at a certain company is *greater* than the national average, so we collect a random sample of 100 employees and measure their systolic blood pressure. What are the hypotheses?

$$H_0 : \mu = 130$$
$$H_A : \mu > 130$$

We'll start with a very specific question – "What is the power of this hypothesis test to correctly detect an *increase* of 2 mmHg in average blood pressure?"

Calculating power

The preceeding question can be rephrased as – How likely is it that this test will reject H_0 when the true average systolic blood pressure for employees at this company is 132 mmHg?

$$\text{Power} = P(\text{Rejecting } H_0 \mid \mu = 132 \ (H_A))$$

Calculating power

The preceeding question can be rephrased as – How likely is it that this test will reject H_0 when the true average systolic blood pressure for employees at this company is 132 mmHg?

$$\text{Power} = P(\text{Rejecting } H_0 \mid \mu = 132\ (H_A))$$

Let's break this down into two simpler problems:

Calculating power

The preceeding question can be rephrased as – How likely is it that this test will reject H_0 when the true average systolic blood pressure for employees at this company is 132 mmHg?

$$\text{Power} = P(\text{Rejecting } H_0 \mid \mu = 132 \ (H_A))$$

Let's break this down into two simpler problems:

1. Problem 1: Which values of $\bar{x}$ represent sufficient evidence to reject H_0?

Calculating power

The preceeding question can be rephrased as – How likely is it that this test will reject H_0 when the true average systolic blood pressure for employees at this company is 132 mmHg?

$$\text{Power} = P(\text{Rejecting } H_0 \mid \mu = 132 \ (H_A))$$

Let's break this down into two simpler problems:

1. Problem 1: Which values of $\bar{x}$ represent sufficient evidence to reject H_0?
2. Problem 2: What is the probability that we would reject H_0 if $\bar{x}$ had come from a distribution with $\mu = 132$, i.e. what is the probability that we can obtain such an $\bar{x}$ from this distribution?

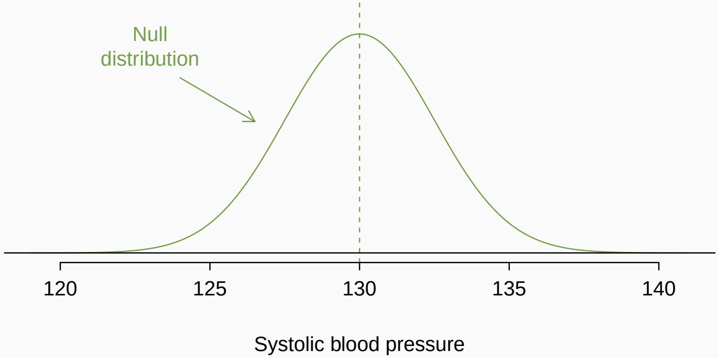
Problem 1

Which values of $\bar{x}$ represent sufficient evidence to reject H_0?
(Remember $H_0 : \mu = 130$, $H_A : \mu > 130$)

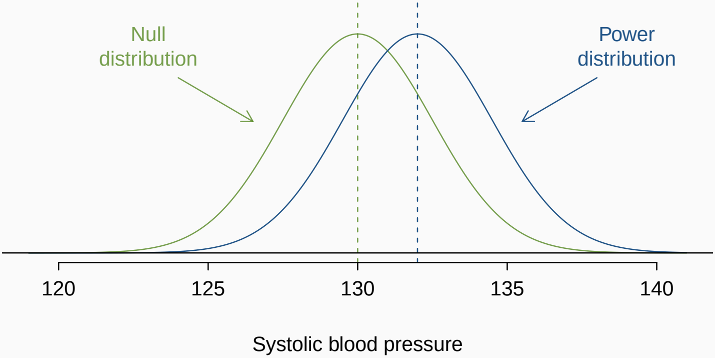
Problem 2

What is the probability that we would reject H_0 if $\bar{x}$ came from a distribution where $\mu = 132$.

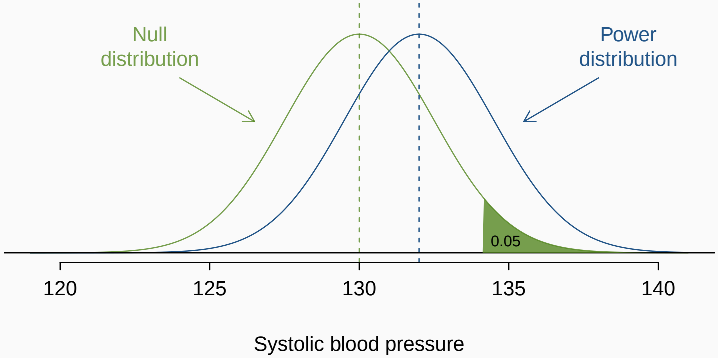
Putting it all together



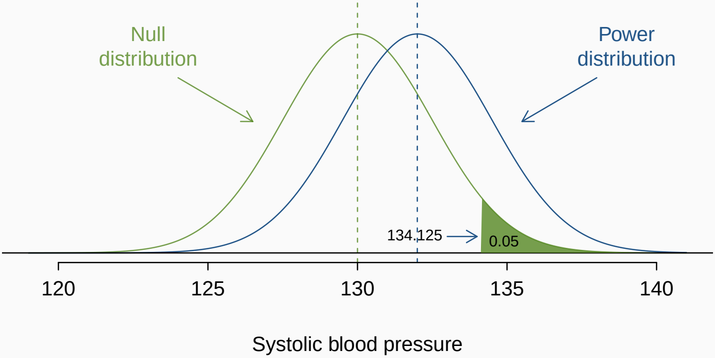
Putting it all together



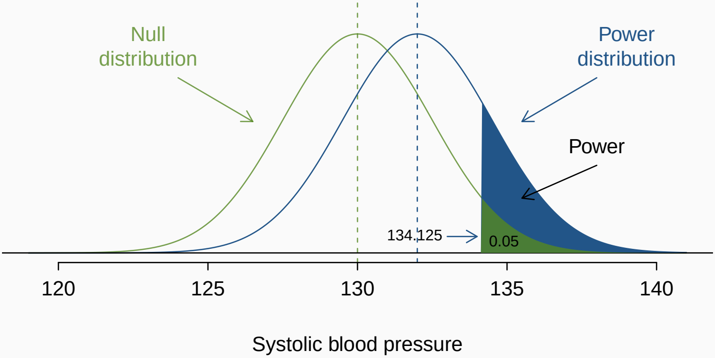
Putting it all together



Putting it all together



Putting it all together



Recap - Calculating Power

- *Step 0:* Pick a meaningful effect size δ and a significance level α
- *Step 1:* Calculate the range of values for the point estimate beyond which you would reject H_0 at the chosen α level.
- *Step 2:* Calculate the probability of observing a value from preceding step if the sample was derived from a population where $\mu = \mu_{H_0} + \delta$

Example - Power for a two sided hypothesis test

Going back to the blood pressure example, what would the power be to detect a 4 mmHg increase in average blood pressure for the hypothesis that the population average is different from 130 mmHg at a 95% significance level for a sample of 625 patients?

Example - Power for a two sided hypothesis test

Going back to the blood pressure example, what would the power be to detect a 4 mmHg increase in average blood pressure for the hypothesis that the population average is different from 130 mmHg at a 95% significance level for a sample of 625 patients?

Step 0:

$H_0 : \mu = 130, \quad H_A : \mu \neq 130, \quad \alpha = 0.05, \quad n = 625, \quad \sigma = 25, \quad \delta = 4, \quad 1 - \beta = ?$

Step 1:

$$P(T > t \text{ or } T < -t) < 0.05 \quad \Rightarrow \quad t > 1.96$$
$$\bar{x} > 130 + 1.96 \frac{25}{\sqrt{625}} \text{ or } \bar{x} < 130 - 1.96 \frac{25}{\sqrt{625}}$$
$$\bar{x} > 131.96 \text{ or } \bar{x} < 128.04$$

Step 2: Assume $\mu = \mu_{H_0} + \delta = 134$

$$\begin{aligned} P(\bar{x} > 131.96 \text{ or } \bar{x} < 128.04) &= P(T > [131.96 - 134]/1) + P(T < [128.04 - 134]/1) \\ &= P(T > -2.04) + P(T < -5.96) \\ &= 0.979 + 0 = 0.979 \end{aligned}$$

Example - Using power to determine sample size

Going back to the blood pressure example, how large a sample would you need if you wanted 90% power to detect a 4 mmHg increase in average blood pressure for the hypothesis that the population average is different from 130 mmHg at a 95% significance level?

Example - Using power to determine sample size

Going back to the blood pressure example, how large a sample would you need if you wanted 90% power to detect a 4 mmHg increase in average blood pressure for the hypothesis that the population average is different from 130 mmHg at a 95% significance level?

Step 0:

$$H_0 : \mu = 130, \quad H_A : \mu \neq 130, \quad \alpha = 0.05, \quad \beta = 0.10, \quad \sigma = 25, \quad \delta = 4, \quad n = ?$$

Step 1:

$$P(T > t \text{ or } T < -t) < 0.05 \quad \Rightarrow \quad t > 1.96$$
$$\bar{x} > 130 + 1.96\frac{25}{\sqrt{n}} \text{ or } \bar{x} < 130 - 1.96\frac{25}{\sqrt{n}}$$

Step 2: Assume $\mu = \mu_{H_0} + \delta = 134$

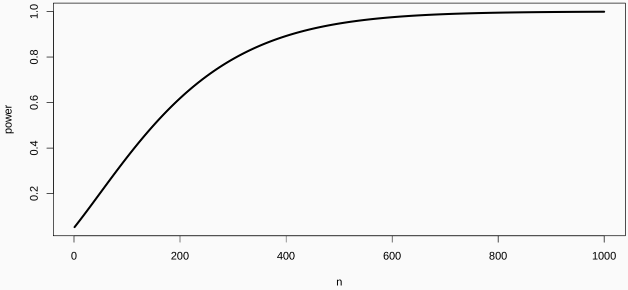
$$P\left(\bar{x} > 130 + 1.96\frac{25}{\sqrt{n}} \text{ or } \bar{x} < 130 - 1.96\frac{25}{\sqrt{n}}\right) = 0.9$$
$$P\left(T > 1.96 - 4\frac{\sqrt{n}}{25} \text{ or } T < -1.96 - 4\frac{\sqrt{n}}{25}\right) = 0.9$$

Example - Using power to determine sample size (cont.)

So we are left with an equation we cannot solve directly, how do we evaluate it?

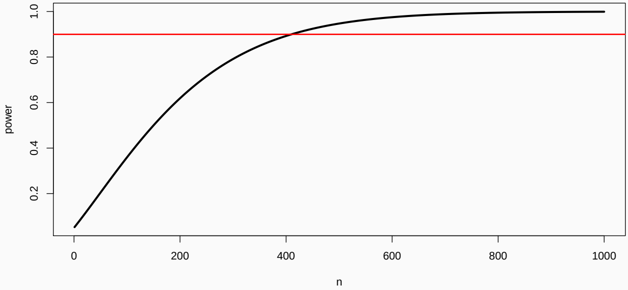
Example - Using power to determine sample size (cont.)

So we are left with an equation we cannot solve directly, how do we evaluate it?



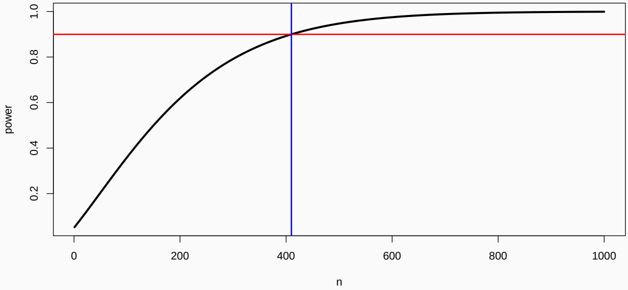
Example - Using power to determine sample size (cont.)

So we are left with an equation we cannot solve directly, how do we evaluate it?



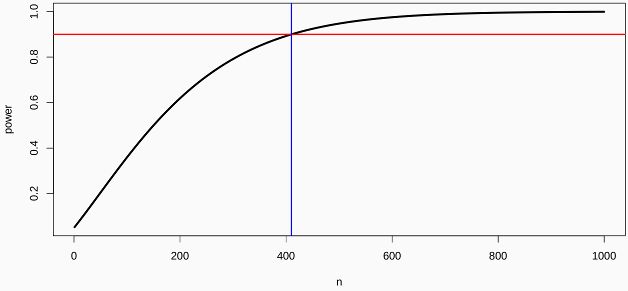
Example - Using power to determine sample size (cont.)

So we are left with an equation we cannot solve directly, how do we evaluate it?



Example - Using power to determine sample size (cont.)

So we are left with an equation we cannot solve directly, how do we evaluate it?



For $n = 410$ the power $= 0.8996$, therefore we need 411 subjects in our sample to achieve the desired level of power for the given circumstance.

Back to the GSS

Yesterday we discussed using the GSS for assessing what percentage of Americans would an answer about experiment design correctly. We performed a hypothesis test to examine whether more than 80% of Americans have a good intuition about experimental design.

Back to the GSS

Yesterday we discussed using the GSS for assessing what percentage of Americans would an answer about experiment design correctly. We performed a hypothesis test to examine whether more than 80% of Americans have a good intuition about experimental design.

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

If we are planning to sample 670 individuals and ask them the same question, what is the power of our study to detect an effect size of 0.05?

Calculating power - Step 0 and 1

What is the power of our hypotheses and data to detect a difference of 0.05 in p?

Calculating power - Step 0 and 1

What is the power of our hypotheses and data to detect a difference of 0.05 in p?

Step 0: What do we know?

$H_0 : p = 0.8,\ \ H_A : p > 0.8,\ \ \alpha = 0.05,\ \ n = 670,\ \ SE = 0.0154,\ \ \delta = 0.05,\ \ 1-\beta = ?$

Calculating power - Step 0 and 1

What is the power of our hypotheses and data to detect a difference of 0.05 in p?

Step 0: What do we know?

$H_0 : p = 0.8, \;\; H_A : p > 0.8, \;\; \alpha = 0.05, \;\; n = 670, \;\; SE = 0.0154, \;\; \delta = 0.05, \;\; 1 - \beta = ?$

Step 1: What values of $\hat{p}$ would let us reject H_0?

Calculating power - Step 0 and 1

What is the power of our hypotheses and data to detect a difference of 0.05 in p?

Step 0: What do we know?

$H_0 : p = 0.8, \;\; H_A : p > 0.8, \;\; \alpha = 0.05, \;\; n = 670, \;\; SE = 0.0154, \;\; \delta = 0.05, \;\; 1 - \beta = ?$

Step 1: What values of $\hat{p}$ would let us reject H_0?

$$P(Z > z) < 0.05 \quad \Rightarrow \quad z > 1.645$$

Calculating power - Step 0 and 1

What is the power of our hypotheses and data to detect a difference of 0.05 in p?

Step 0: What do we know?

$H_0 : p = 0.8,\ \ H_A : p > 0.8,\ \ \alpha = 0.05,\ \ n = 670,\ \ SE = 0.0154,\ \ \delta = 0.05,\ \ 1 - \beta = ?$

Step 1: What values of $\hat{p}$ would let us reject H_0?

$$P(Z > z) < 0.05 \quad \Rightarrow \quad z > 1.645$$

$$P\left(\frac{\hat{p} - 0.8}{0.0154} > 1.645\right) = 0.05$$

Calculating power - Step 0 and 1

What is the power of our hypotheses and data to detect a difference of 0.05 in p?

Step 0: What do we know?

$H_0 : p = 0.8,\ \ H_A : p > 0.8,\ \ \alpha = 0.05,\ \ n = 670,\ \ SE = 0.0154,\ \ \delta = 0.05,\ \ 1-\beta = ?$

Step 1: What values of $\hat{p}$ would let us reject H_0?

$$P(Z > z) < 0.05 \quad \Rightarrow \quad z > 1.645$$

$$P\left(\frac{\hat{p} - 0.8}{0.0154} > 1.645\right) = 0.05$$

$$\hat{p} > 0.8 + 0.0154 \times 1.645$$

Calculating power - Step 0 and 1

What is the power of our hypotheses and data to detect a difference of 0.05 in p?

Step 0: What do we know?

$H_0 : p = 0.8,\ \ H_A : p > 0.8,\ \ \alpha = 0.05,\ \ n = 670,\ \ SE = 0.0154,\ \ \delta = 0.05,\ \ 1-\beta = ?$

Step 1: What values of $\hat{p}$ would let us reject H_0?

$$P(Z > z) < 0.05 \quad \Rightarrow \quad z > 1.645$$

$$P\left(\frac{\hat{p} - 0.8}{0.0154} > 1.645\right) = 0.05$$

$$\hat{p} > 0.8 + 0.0154 \times 1.645$$

$$\hat{p} > 0.825$$

Calculating power - Step 0 and 1

What is the power of our hypotheses and data to detect a difference of 0.05 in p?

Step 0: What do we know?

$H_0: p = 0.8,\ \ H_A: p > 0.8,\ \ \alpha = 0.05,\ \ n = 670,\ \ SE = 0.0154,\ \ \delta = 0.05,\ \ 1-\beta = ?$

Step 1: What values of $\hat{p}$ would let us reject H_0?

$$P(Z > z) < 0.05 \quad \Rightarrow \quad z > 1.645$$

$$P\left(\frac{\hat{p} - 0.8}{0.0154} > 1.645\right) = 0.05$$

$$\hat{p} > 0.8 + 0.0154 \times 1.645$$

$$\hat{p} > 0.825$$

Calculating power - Step 2

What is the power of our hypotheses and data to detect a difference of 0.05 in p?

Calculating power - Step 2

What is the power of our hypotheses and data to detect a difference of 0.05 in p?

Step 0: What do we know?

$H_0 : p = 0.8,\ \ H_A : p > 0.8,\ \ \alpha = 0.05,\ \ n = 670,\ \ SE = 0.0154,\ \ \delta = 0.05,\ \ 1-\beta = ?$

Calculating power - Step 2

What is the power of our hypotheses and data to detect a difference of 0.05 in p?

Step 0: What do we know?

$H_0 : p = 0.8,\ \ H_A : p > 0.8,\ \ \alpha = 0.05,\ \ n = 670,\ \ SE = 0.0154,\ \ \delta = 0.05,\ \ 1-\beta = ?$

Step 1: What values of $\hat{p}$ would let us reject H_0?

$$\hat{p} > 0.825$$

Calculating power - Step 2

What is the power of our hypotheses and data to detect a difference of 0.05 in p?

Step 0: What do we know?

$H_0: p = 0.8,\ \ H_A: p > 0.8,\ \ \alpha = 0.05,\ \ n = 670,\ \ SE = 0.0154,\ \ \delta = 0.05,\ \ 1-\beta = ?$

Step 1: What values of $\hat{p}$ would let us reject H_0?

$$\hat{p} > 0.825$$

Step 2: Assume $p = 0.8 + \delta = 0.85$, what is the probability we reject H_0? Since p changed, so does $SE = \sqrt{0.85(1-0.85)/670} = 0.0138$.

$$\begin{aligned} &P(\hat{p} > 0.825 | p = 0.85) \\ &= P\left(Z > \frac{0.825 - 0.85}{0.0138}\right) \\ &= P(Z > -1.811) \\ &= 0.965 \end{aligned}$$