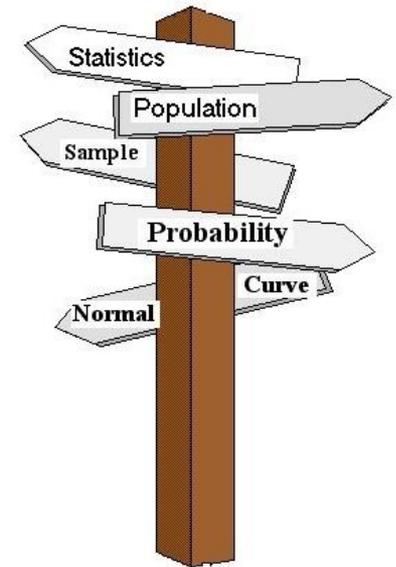


Hypothesis Testing

Hypothesis Testing

- Objective: To test a *claim* about a *population parameter*
- Hypothesis testing steps
 - A. Hypothesis statements
 - B. Test statistic
 - C. P-value and interpretation
 - D. Significance level (optional)



Step A: Hypotheses

- Convert research question to null and alternative hypotheses
- The **null hypothesis (H_0)** is a claim of “no difference”
- The **alternative hypothesis (H_a)** says “ H_0 is false”
- The hypotheses address the population parameter (μ), NOT the sample statistic (x-bar)

Step A: Hypotheses

- **Research question:** Is mean body weight of a particular population of men higher than expected?
- **Expected norm:** Prior research (*before collecting data*) has established that the population should have **mean $\mu = 170$** pounds with standard deviation $\sigma = 40$ pounds.
- **Beware 🖐️: Hypotheses are always based on research questions and expected norms, NOT on data!**

Null hypothesis $H_0: \mu = 170$

Alternative hypothesis :

$H_a: \mu > 170$ (one-sided) *OR*

$H_a: \mu \neq 170$ (two-sided)



Step B: Test Statistic

For one sample test of μ when σ is known, use this test statistic:

$$Z_{\text{stat}} = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}}$$

where

$\bar{x} \equiv$ sample mean

$\mu_0 \equiv$ population mean assuming H_0 is true

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Step B: Test Statistic

- For our example, $\mu_0 = 170$ and $\sigma = 40$
- Take an SRS of $n = 64$
- Calculate a sample mean (\bar{x}) of 173

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{40}{\sqrt{64}} = 5$$

$$z_{\text{stat}} = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}} = \frac{173 - 170}{5} = 0.60$$

Step C: *P*-Value

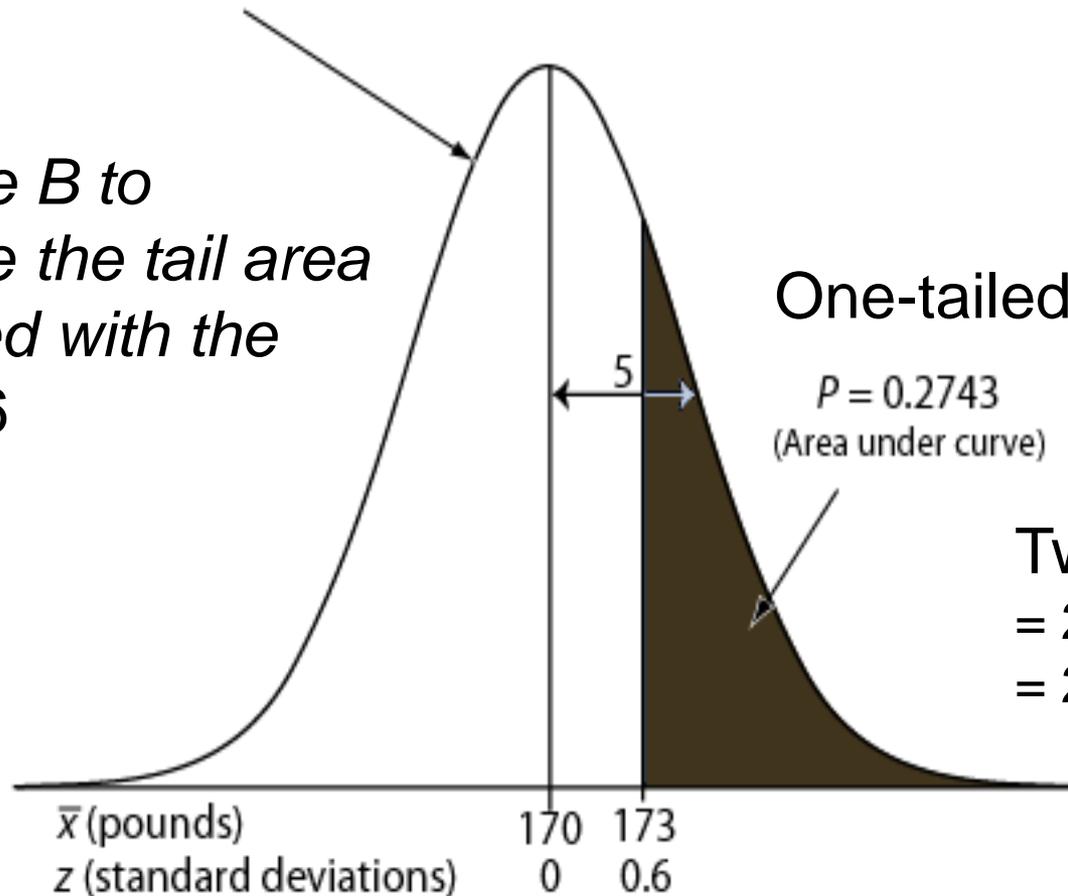
Convert *z* statistics to a *P*-value:

- For $H_a: \mu > \mu_0$
 $P\text{-value} = \Pr(Z > z_{\text{stat}}) = \text{right-tail beyond } z_{\text{stat}}$
- For $H_a: \mu < \mu_0$
 $P\text{-value} = \Pr(Z < z_{\text{stat}}) = \text{left tail beyond } z_{\text{stat}}$
- For $H_a: \mu \neq \mu_0$
 $P\text{-value} = 2 \times \text{one-tailed } P\text{-value}$

Step C: P -value (example)

Distribution of \bar{x} and z_{stat} if H_0 were true

Use Table B to determine the tail area associated with the z_{stat} of 0.6



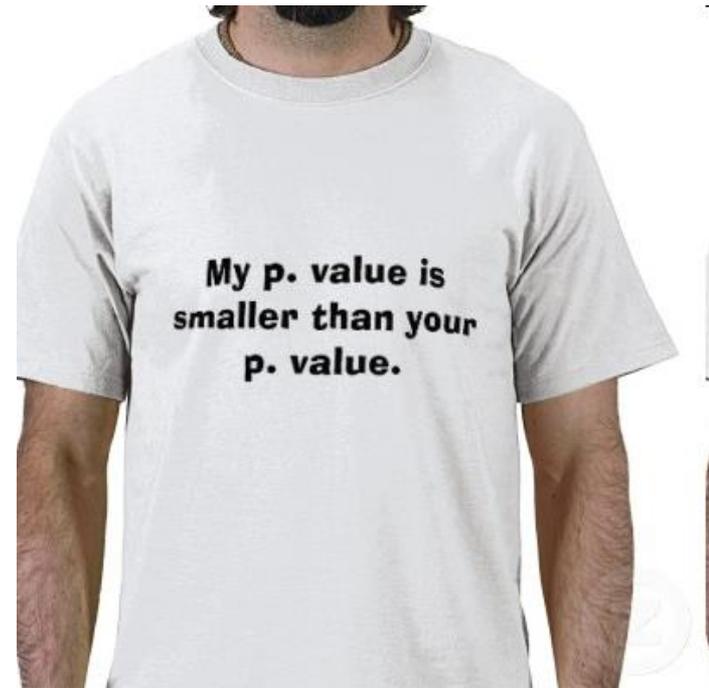
One-tailed $P = .2743$

$P = 0.2743$
(Area under curve)

Two-tailed P
 $= 2 \times \text{one-tailed } P$
 $= 2 \times .2743 = .5486$

Step C: P -values

- P -value answer the question: What is the probability of the observed test statistic ... **when H_0 is true?**
- Smaller and smaller P -values provide stronger and stronger evidence against H_0



Step C: P-values

Conventions*

$P > 0.10 \Rightarrow$ poor evidence against H_0

$0.05 < P \leq 0.10 \Rightarrow$ marginally evidence against H_0

$0.01 < P < 0.05 \Rightarrow$ good evidence against H_0

$P \leq 0.01 \Rightarrow$ very good evidence against H_0

Examples

$P = .27 \Rightarrow$ poor evidence against H_0

$P = .01 \Rightarrow$ very good evidence against H_0

*** It is *unwise* to draw firm borders for “significance”**

Summary

z TEST FOR A POPULATION MEAN

Draw an SRS of size n from a Normal population that has unknown mean μ and known standard deviation σ . To test the null hypothesis that μ has a specified value,

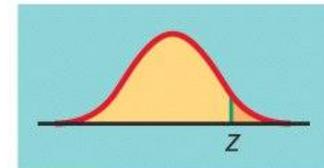
$$H_0: \mu = \mu_0$$

calculate the **one-sample z statistic**

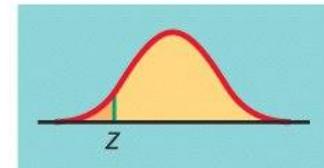
$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

In terms of a variable Z having the standard Normal distribution, the P -value for a test of H_0 against

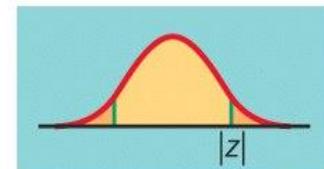
$$H_a: \mu > \mu_0 \text{ is } P(Z \geq z)$$



$$H_a: \mu < \mu_0 \text{ is } P(Z \leq z)$$



$$H_a: \mu \neq \mu_0 \text{ is } 2P(Z \geq |z|)$$



Step D (optional) Significance Level

- Let $\alpha \equiv$ threshold for “significance”
- If $P\text{-value} \leq \alpha \Rightarrow$ evidence *is significant*
- If $P\text{-value} > \alpha \Rightarrow$ evidence *not significant*

Example:

If $\alpha = 0.01$ and $P\text{-value} = 0.27 \Rightarrow$ evidence *not significant*

If $\alpha = 0.01$ and $P\text{-value} = 0.0027 \Rightarrow$ evidence *is significant*

Power and Sample Size

Two types of decision errors:

Type I error = erroneous rejection of true H_0

Type II error = erroneous retention of false H_0

Decision	Truth	
	H_0 true	H_0 false
Retain H_0	Correct retention	Type II error
Reject H_0	Type I error	Correct rejection

$\alpha \equiv$ Probability of a Type I error

$\beta \equiv$ Probability of a Type II error

Power

Power is defined as the probability that a statistical test will reject the null hypothesis or the ability of a statistical test to detect an effect

In order to reject the null hypothesis (which states that there is no relationship between the variables of interest), **power should be at least .80**

Power

A **power analysis** provides information for determining the **minimum number of subjects** you need to collect in order to make your study worthwhile

Power

You should always perform a power analysis before you begin your data collection, ideally when you are designing your study

This is referred to as an a **priori power analysis** which is done before you conduct your study

Power

A **post-hoc power analysis** is done after you have completed your research.

When doing a post hoc power analysis you need to know the alpha, the power you would like to achieve (.e.g., .80) and the effect size (small, medium or large)

Power

You should also have a sense of how much power is required to detect **an effect (difference)** when conducting a power analysis.

This is expressed as $\text{power} = 1 - \beta$,

where β is the probability of a Type II error

Power

The Power of a test in Inferential Statistics is its **Probability of correctly accepting (Failing to Reject) the Null Hypothesis** when, in reality, the Null is true

Power is good. Higher Power gives you a more accurate test.

Power

- Power is the opposite of Beta
- Beta (β) is the Probability of making a Beta (False Negative) Error. Power is the Probability of not making a Beta Error.
- Power = $1 - \beta$

Power

	Beta Error (False Negative)
	<p>Smoking doesn't cause cancer.</p> 
What it means	<p>The error of <u>concluding that there is nothing</u> – no difference, or no change, or no effect – <u>when, in reality, there is</u></p>
Null Hypothesis	<p>is Accepted (Fail to Reject) when it is false</p>
Probability of making this Error	β
Probability of <u>not</u> making this Error	$\text{Power} = 1 - \beta$

Power

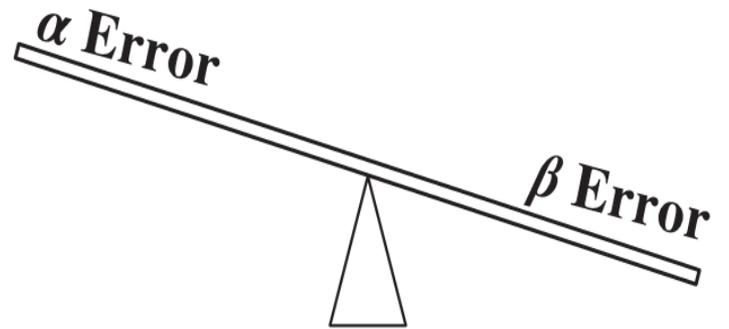
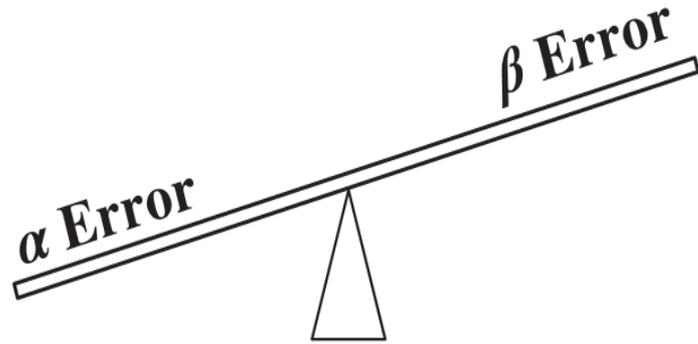
p is the Probability of an Alpha Error	β is the Probability of a Beta Error
α is the maximum tolerable Probability for an Alpha Error	
$1 - \alpha$ is called the Confidence Level	$1 - \beta$ is called the Power of the test

Power

- Power is affected by three factors –
Directly by Significance Level (α) and Sample Size (n),

and **Inversely** by the Effect Size (ES)

Power

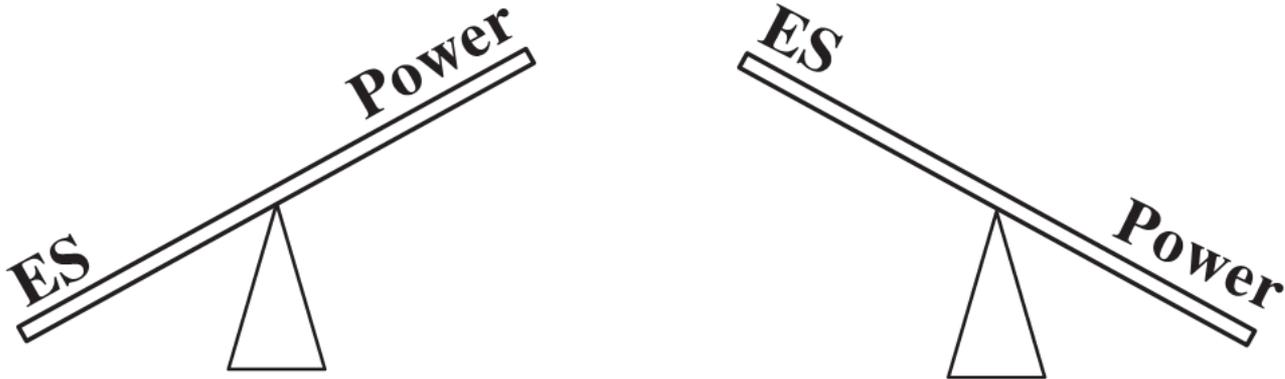


Power

- If we require a low Probability of an Alpha Error (at left), we select a low value for Alpha (α). This causes β to increase. If β increases, then Power (which is $1 - \beta$) decreases.
- So, decreasing Alpha causes Power to decrease (if Effect Size and Sample Size remain the same). And, increasing Alpha causes Power to increase

Power

- Power and Effect Size, on the other hand, have an inverse relationship.



Power

- If we want to be able to detect a small difference, change or effect (small Effect Size), then we'll need a test with more Power than we would need for larger Effect Sizes.
- This can be achieved by increasing the Sample Size.

Power

- Samples Size (n) affects Power directly. All other things being equal, increasing the Sample Size increases the Power of the test. Reducing the Sample Size reduces the Power.
- Therefore, Power is a function of Alpha, Effect Size, and Sample Size.

Power = f (Alpha, Effect Size, Sample Size)

Power

- α is the probability of alpha error, used for analyzing the test results
- β is the probability of beta error, used to determine the minimum Sample Size needed to detect a difference, change, or effect of a given Effect Size

Power

- $\beta \equiv$ probability of a Type II error
 $\beta = \Pr(\text{retain } H_0 \mid H_0 \text{ false})$
(the “|” is read as “given”)
- $1 - \beta =$ “Power” \equiv probability of avoiding a Type II error
 $1 - \beta = \Pr(\text{reject } H_0 \mid H_0 \text{ false})$

Power of a z test

$$1 - \beta = \Phi \left(-z_{1-\frac{\alpha}{2}} + \frac{|\mu_0 - \mu_a| \sqrt{n}}{\sigma} \right)$$

where

- $\Phi(z) \equiv$ cumulative probability of Standard Normal value z
- $\mu_0 \equiv$ population mean under H_0
- $\mu_a \equiv$ population mean under H_a

Calculating Power: Example

A study of $n = 16$ retains $H_0: \mu = 170$ at $\alpha = 0.05$ (two-sided); σ is 40. What was the power of test to identify a population mean of 190?

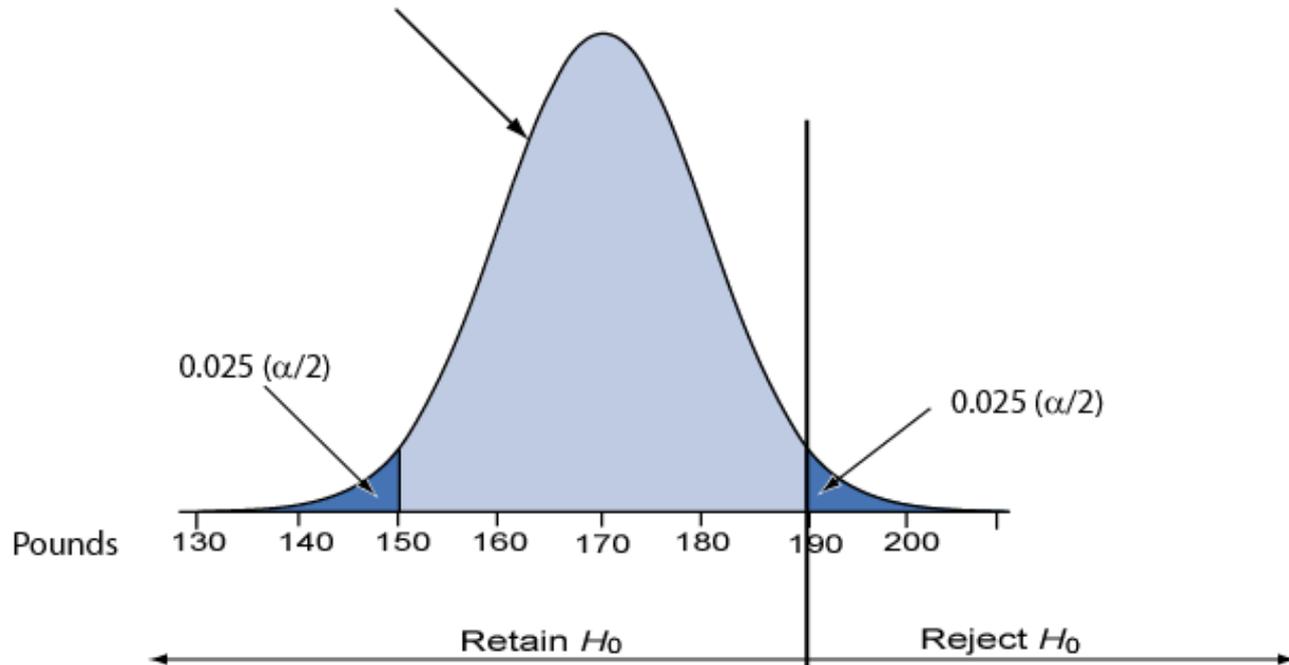
$$\begin{aligned}1 - \beta &= \Phi\left(-z_{1-\frac{\alpha}{2}} + \frac{|\mu_0 - \mu_a| \sqrt{n}}{\sigma}\right) \\&= \Phi\left(-1.96 + \frac{|170 - 190| \sqrt{16}}{40}\right) \\&= \Phi(0.04)\end{aligned}$$

\Rightarrow look up cumulative probability on Table B $\Rightarrow = 0.5160$

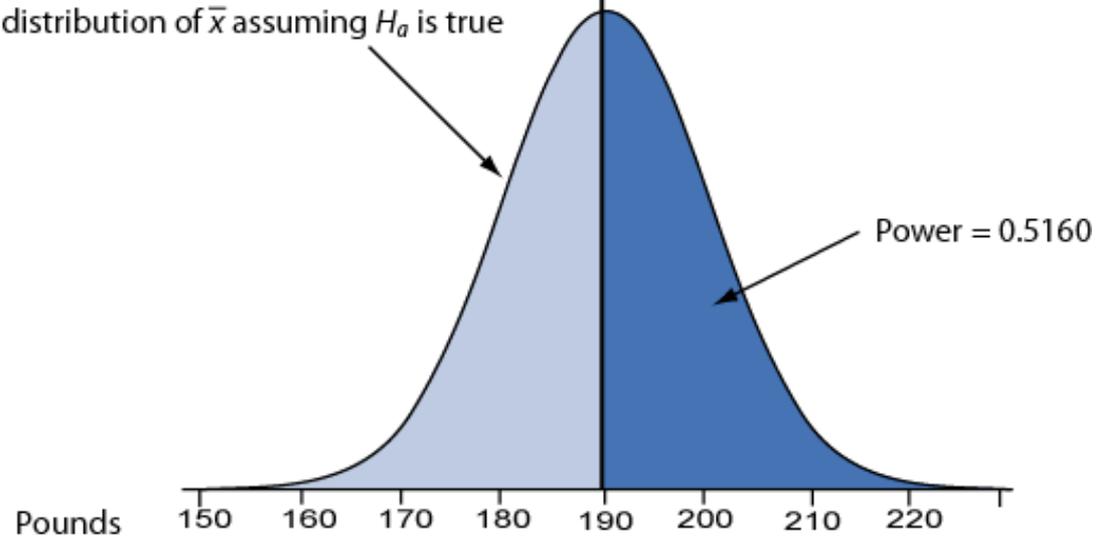
Reasoning of Power Calculation

- Competing “theories”
 - Top curve (next page) assumes H_0 is true
 - Bottom curve assumes H_a is true
 - α set to 0.05 (two-sided)
- Reject H_0 when sample mean exceeds 189.6 (right tail, top curve)
- Probability of a value greater than 189.6 on the bottom curve is 0.5160, corresponding to the power of the test

Sampling distribution of \bar{x} assuming H_0 is true



Sampling distribution of \bar{x} assuming H_a is true



Sample Size Requirements

Sample size for one-sample z test:

$$n = \frac{\sigma^2 \left(z_{1-\beta} + z_{1-\frac{\alpha}{2}} \right)^2}{\Delta^2}$$

where

$1 - \beta \equiv$ desired power

$\alpha \equiv$ desired significance level (two-sided)

$\sigma \equiv$ population standard deviation

$\Delta = \mu_0 - \mu_a \equiv$ the **difference worth detecting**

Example: Sample Size Requirement

How large a sample is needed to test $H_0: \mu = 170$ versus $H_a: \mu = 190$ with 90% power and $\alpha = 0.05$ (two-tailed) when $\sigma = 40$?

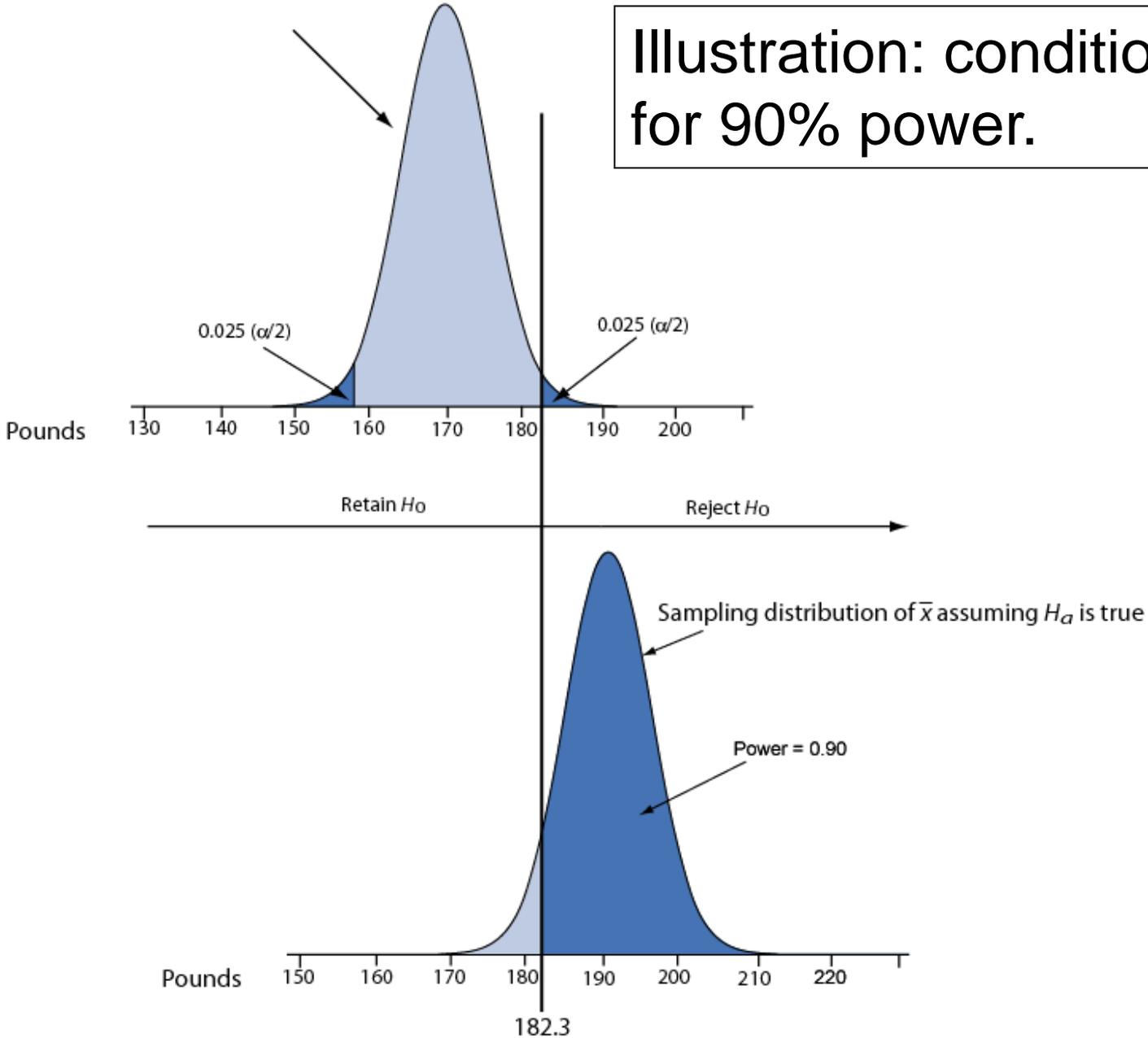
Note: $\Delta = \mu_0 - \mu_a = 170 - 190 = -20$

$$n = \frac{\sigma^2 \left(z_{1-\beta} + z_{1-\frac{\alpha}{2}} \right)^2}{\Delta^2} = \frac{40^2 (1.28 + 1.96)^2}{-20^2} = 41.99$$

Round up to 42 to ensure adequate power.

Sampling distribution of \bar{x} assuming H_0 is true

Illustration: conditions for 90% power.



Sample Size Determination

- The sample size calculated from a sample size formula is known as the minimum number of sample units **needed to detect a significant effect** of the test statistic result
- In other words, to prevent committing type II error

Sample size for Proportions

A Proportion is a percentage expressed as a decimal. So 50% is 0.50 and 100% is 1.0.

Statistical formulas usually use the Proportion format.

Examples of Proportions of Count Data

Count	Sample Size	Proportion
66 people said they'd vote for Candidate A	120 people were surveyed	$66/120 = 0.55$
8 people preferred strawberry ice cream	20 people in a focus group	$8/20 = 0.40$
6 defective items	production run of 1000	$6/1000 = 0.006$

Sample size for Proportions

- The symbol for a Proportion is p . That is also the symbol for Probability.
- If the Proportion of people favoring Candidate A is 0.55 then the Probability of any one person favoring Candidate A is 0.55.

Sample size for Proportions

- The formula for calculating n , the Sample Size, includes four symbols, α , p , MOE, and z .

Alpha, α

- Alpha is called the Level of Significance, and it is involved with the concept of Alpha Error.
- Can be thought of as the “Lack-of-Confidence” Level

Sample size for Proportions

- In statistics, we don't have a symbol for the Confidence Level, but we do for its opposite. The symbol is the Greek letter Alpha (α).

Level of Confidence = $1 - \alpha$, so $\alpha = 1 - \text{Level of Confidence}$

The person performing the statistical analysis selects a value for Alpha *prior to collecting data for the Sample.*

If we want to be 95% Confident that our Sample is representative of the overall Population or Process, that means that we are willing to accept the fact that 5% of the time we will collect a Sample which is not representative.

In that case, we would select $\alpha = 5\%$. (That is the value most commonly selected.)

Sample size for Proportions

You may be wondering why you have to settle for 5% or some other number. Why can't you have 0%?

Statistically, the only way you can get $\alpha = 0\%$ is if you poll 100% of the Population or Process.

MOE is the Margin of Error, sometimes denoted by “E,” instead of MOE

The actual Population or Process Proportion will very likely not be the exact number which we calculate for p from Sample data. But, we would like it limited to a narrow range. How narrow is specified by the Margin of Error, MOE.

So, if we say “plus or minus 3%,” then MOE = 3%

Sample size for Proportions

We specify the values of Alpha and Margin of Error prior to collecting the Sample data and beginning the analysis

Sample size for Proportions

When there **is** an estimate, \hat{p} , for the Population/Process Proportion, the formula for the minimum Sample Size is

$$n = (\hat{p})(1 - \hat{p})(z_{\alpha/2})^2 / (\text{MOE})^2$$

Sample size for Proportions

When there **is not** an estimate, or if you want to take the most conservative approach, set $\hat{p} = 0.5$ and the formula becomes

$$n = (0.25)(z_{\alpha/2})^2 / (\text{MOE})^2$$

Sample size for Proportions

- This formula assumes you don't know the Population Size (N). If you do know N , divide the n above by $1 + n/N$

Sample size for Proportions

\hat{p} is an **estimate** (or a default) for the actual Proportion of the Population or Process

The **estimate** must be from the same Population or Process. (It can be another survey, an earlier sample)

Sample size for Proportions

With an estimate, there is always a chance for error. Some things may have been different in the other survey, or some Factors may have changed since the earlier production run.

Sample size for Proportions

- The most conservative approach would be to use the default value of \hat{p} which gives the maximum value for the product $(\hat{p})(1 - \hat{p})$. That is, $\hat{p} = 0.5$ (50%).
- As demonstrated in the following table, the closer p is to 0.5, the larger the value of the product of \hat{p} multiplied by $1 - \hat{p}$, and thus, the larger the value of n

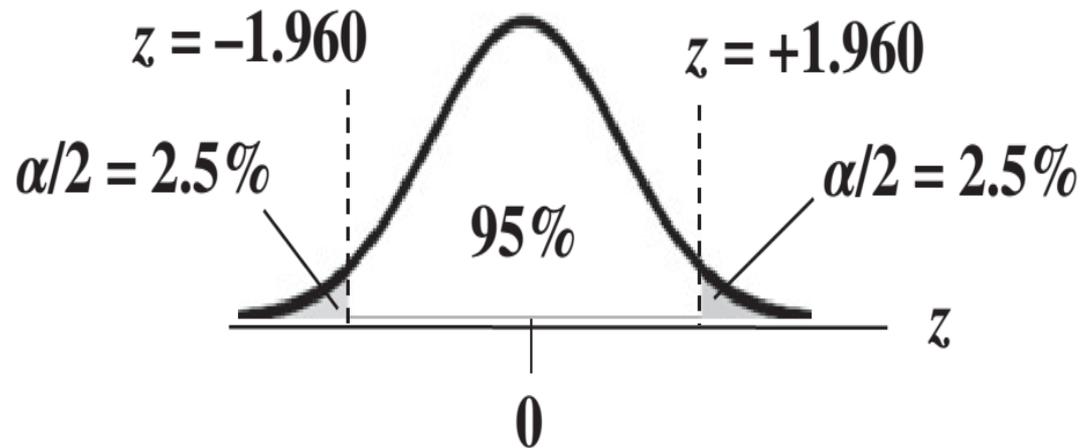
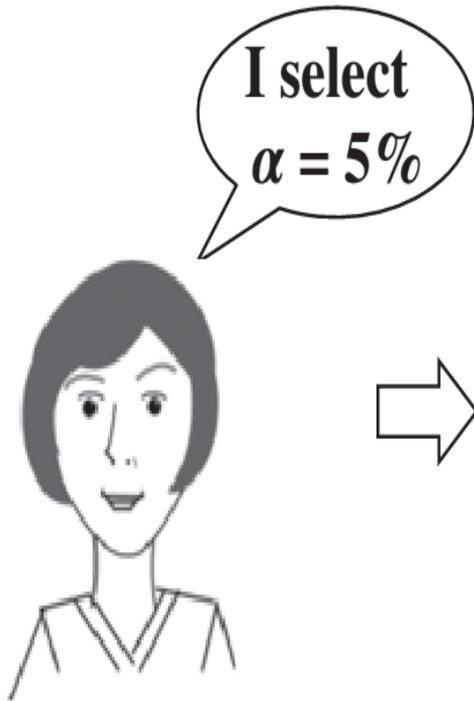
Sample size for Proportions

\hat{p}	$1 - \hat{p}$	$\hat{p} (1 - \hat{p})$
0.10	0.90	0.0900
0.25	0.75	0.1875
0.40	0.60	0.2400
0.50	0.50	0.2500
0.90	0.10	0.0900

Sample size for Proportions

**Note: There are websites
which will do all these
calculations for you**

$z_{\alpha/2}$ is the value of z for a given value of $\alpha/2$



Sample size for Proportions

So if we select $\alpha = 5\%$, we get $z = 1.960$. If we specify that we want the Margin of Error to be 3%, we can use the default formula to calculate a Sample Size:

Sample size for Proportions

$$\begin{aligned}n &= (0.25)(z_{\alpha/2})^2 / (\text{MOE})^2 \\ &= (0.25)(1.960)^2 / (0.03)^2 \\ &= 1067.11\end{aligned}$$

Round up (always) to **1068**

So, we will need to poll at least 1068 people to be 95% Confident with a 3% Margin of Error

Sample size for Proportions

What if we can't afford the time or money to collect data of the calculated Sample Size?

We can sacrifice some accuracy in either the Level of Confidence or the Margin of Error or both

Sample size for Proportions

- Let's say we can only afford to poll 625 people. What can we do? With a little algebra, we see that we can plug in a value for n and then calculate either MOE or $z\alpha/2$.
- Here's how we calculate MOE, given $n = 625$, $\hat{p} = 0.5$, and $z\alpha/2$:

Sample size for Proportions

$$n = (0.25)(z_{\alpha/2})^2 / \text{MOE}^2$$

$$625 = (0.25)(1.96)^2 / \text{MOE}^2$$

$$\text{MOE}^2 = (0.25)(1.96)^2 / 625$$

$$\text{MOE} = (0.5)(1.96) / 25 = 0.0392$$

So, if we reduce the minimum Sample Size from 1068 to 625, that increases the Margin of Error from 3% to about 4%

Sample size for Proportions

- For a given Sample Size, Alpha and Margin of Error affect each other inversely
- If we select a lower value of Alpha (which means a higher Level of Confidence), the Margin of Error increases.
- If we select a higher value for Alpha, the Margin of Error decreases

Sample size for Proportions

- The only way to reduce both is to increase the Sample Size



Sample size for Proportions

- The following things increase the minimum Sample Size:
 - **Higher Level of Confidence** (i.e., smaller value of Alpha) selected
 - **Smaller Margin of Error** specified
 - **Estimated Proportion** closer to 0.5

Sample size for quantitative data

The minimum Sample Size calculations can be done using

- α , the selected Level of Significance
- σ , Standard Deviation of the Population or Process (or an estimate of it)
- MOE, the desired Margin of Error