# HYPOTHESIS TESTING

# HYPOTHESIS TESTING

Null and Alternative Hypotheses

Test Statistic

*P*-Value

Significance Level

One-Sample *z* Test

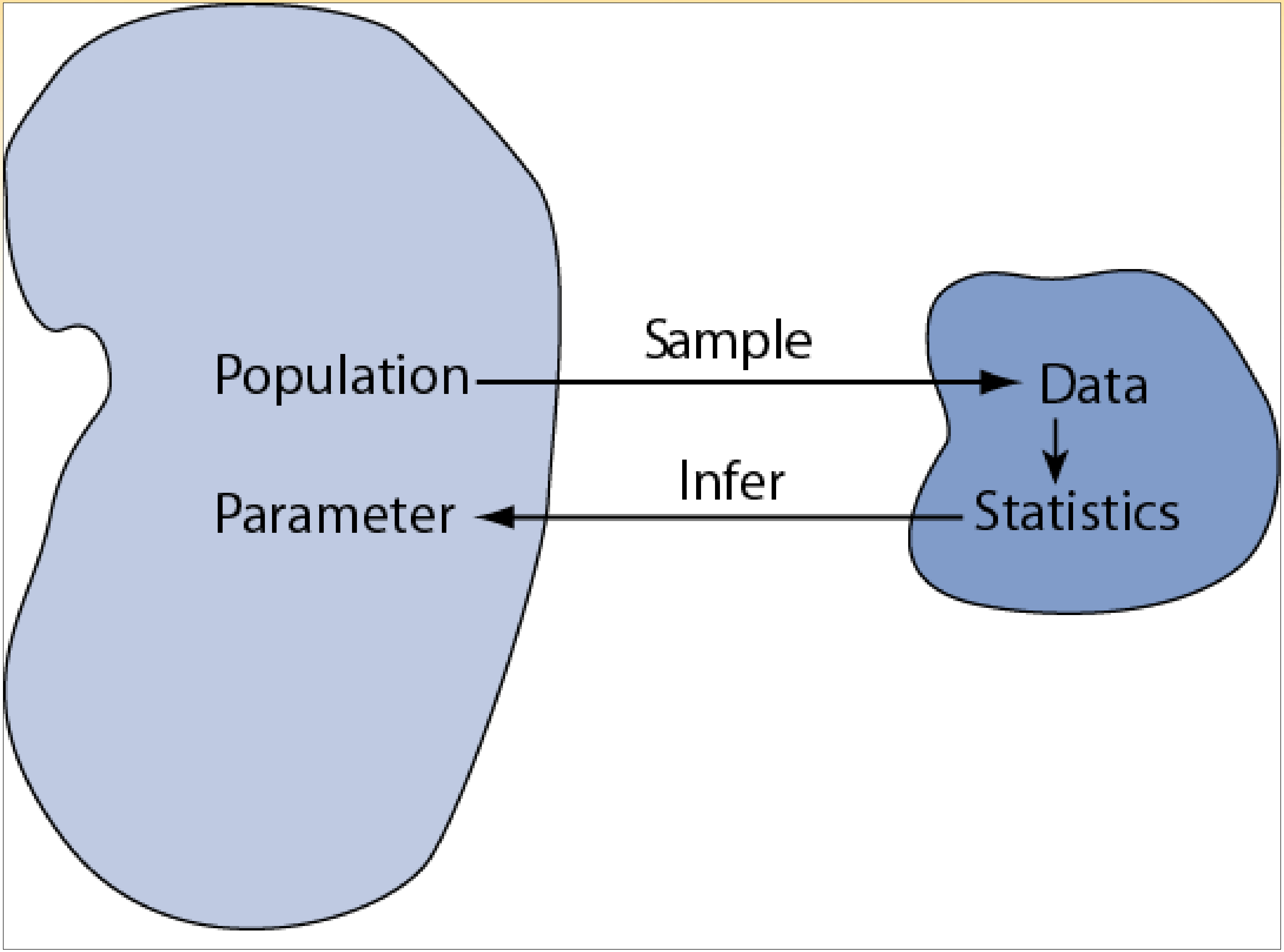
Power and Sample Size

# Terms Introduce in Prior Chapter

- **Population** $\equiv$ all possible values
- **Sample** $\equiv$ a portion of the population
- **Statistical inference** $\equiv$ generalizing from a sample to a population with calculated degree of certainty
- Two forms of statistical inference
  - **Hypothesis testing**
  - **Estimation**
- **Parameter** $\equiv$ a characteristic of population, e.g., population mean µ
- **Statistic** $\equiv$ calculated from data in the sample, e.g., sample mean ($\bar{x}$)

# Distinctions Between Parameters and Statistics (Chapter 8 review)

| | Parameters | Statistics |
|---|---|---|
| Source | Population | Sample |
| Notation | Greek (e.g., μ) | Roman (e.g., *xbar*) |
| Vary | No | Yes |
| Calculated | No | Yes |

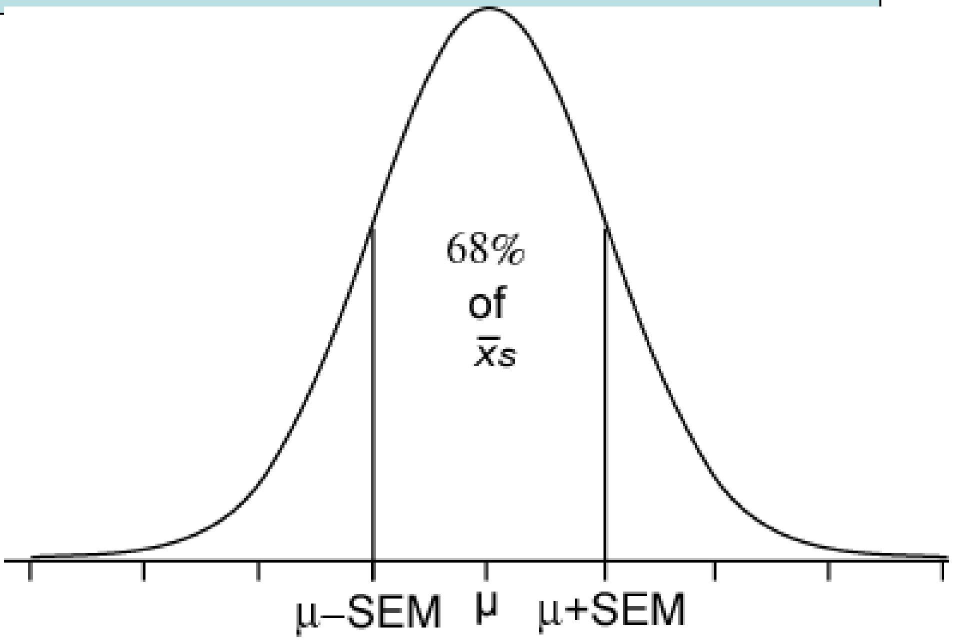

Sample
Population
Data
Infer
Parameter
Statistics

# Sampling Distributions of a Mean (Introduced in Ch 8)

The **sampling distributions of a mean (SDM)** describes the behavior of a sampling mean

$$\bar{x} \sim N\left(\mu, SE_{\bar{x}}\right)$$

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



# Hypothesis Testing

- Is also called *significance testing*
- Tests a claim about a parameter using evidence (data in a sample
- The technique is introduced by considering a one-sample z test
- The procedure is broken into four steps
- *Each* element of the procedure must be understood

# Hypothesis Testing Steps

A. Null and alternative hypotheses

B. Test statistic

C. P-value and interpretation

D. Significance level (optional)

# Null and Alternative Hypotheses

- Convert the research question to null and alternative hypotheses
- The **null hypothesis ($H_0$)** is a claim of "no difference in the population"
- The **alternative hypothesis ($H_a$)** claims "$H_0$ is false"
- Collect data and seek evidence against $H_0$ as a way of bolstering $H_a$ (deduction)

# Illustrative Example: "Body Weight"

- **The problem:** In the 1970s, 20–29 year old men in the U.S. had a mean μ body weight of 170 pounds. Standard deviation σ was 40 pounds. We test whether mean body weight in the population now differs.
- **Null hypothesis** $H_{0:}$ $\mu = 170$ ("no difference")
- The **alternative hypothesis** can be either
  $H_{a:}$ $\mu > 170$ (**one-sided test**) or
  $H_{a:}$ $\mu \neq 170$ (**two-sided test**)

# Test Statistic

This is an example of a one-sample test of a mean when σ is known. Use this statistic to test the problem:

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

$$\text{where } \mu_0 \equiv \text{population mean assuming } H_0 \text{ is true}$$

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

# Illustrative Example: *z* statistic

- For the illustrative example, $\mu_0 = 170$
- We know $\sigma = 40$
- Take an SRS of $n = 64$. Therefore

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

- If we found a sample mean of 173, then

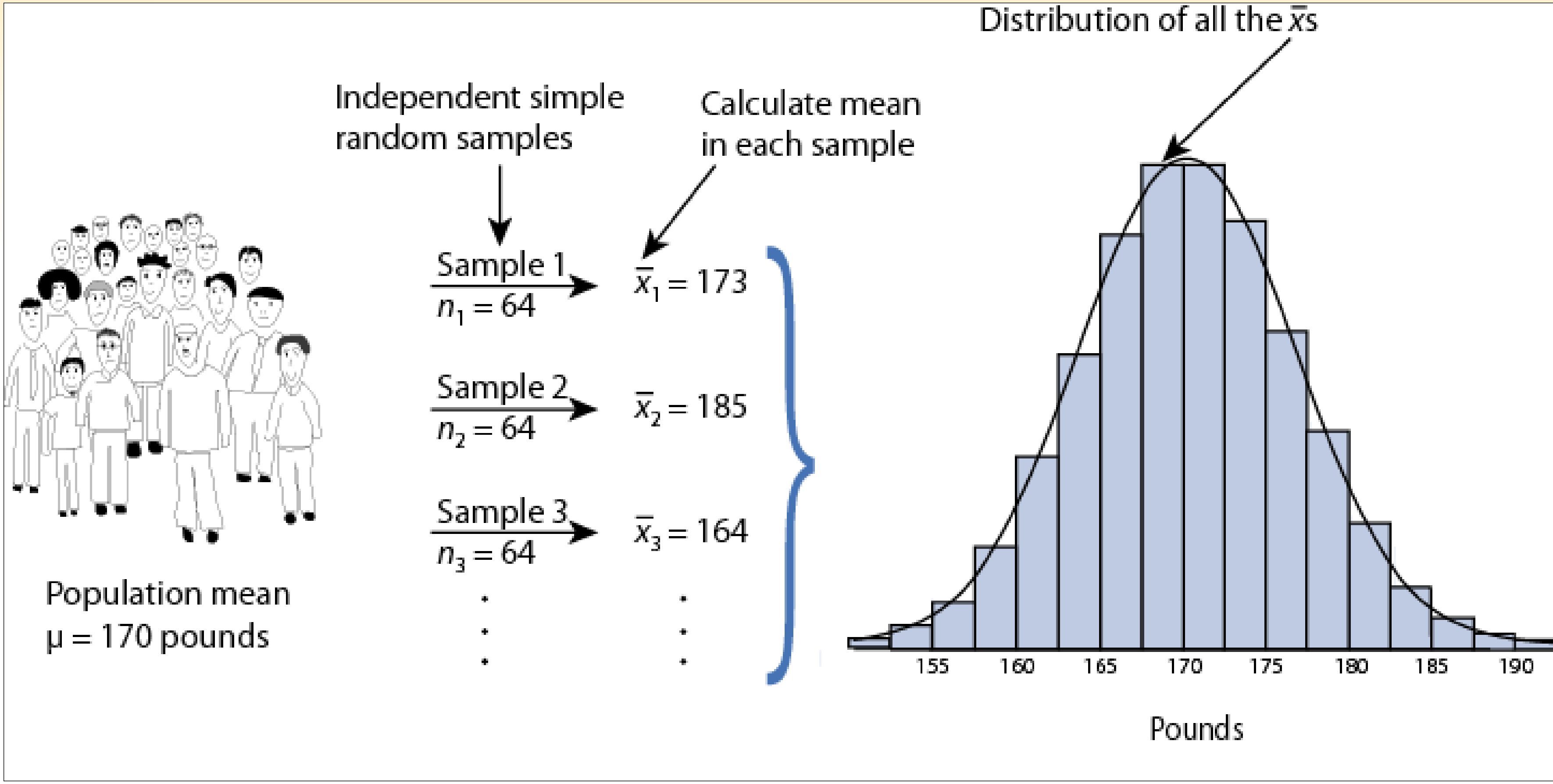
$$z_{\text{stat}} = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}} = \frac{173 - 170}{5} = 0.60$$

# Illustrative Example: *z* statistic

If we found a sample mean of 185, then

$$z_{\text{stat}} = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}} = \frac{185 - 170}{5} = 3.00$$

# Reasoning Behinμ$z_{stat}$




Sampling distribution of xbar under $H_0$: μ = 170 for $n$ = 64 ⇒

$$\bar{x} \sim N(170,5)$$

# §9.3 *P*-value

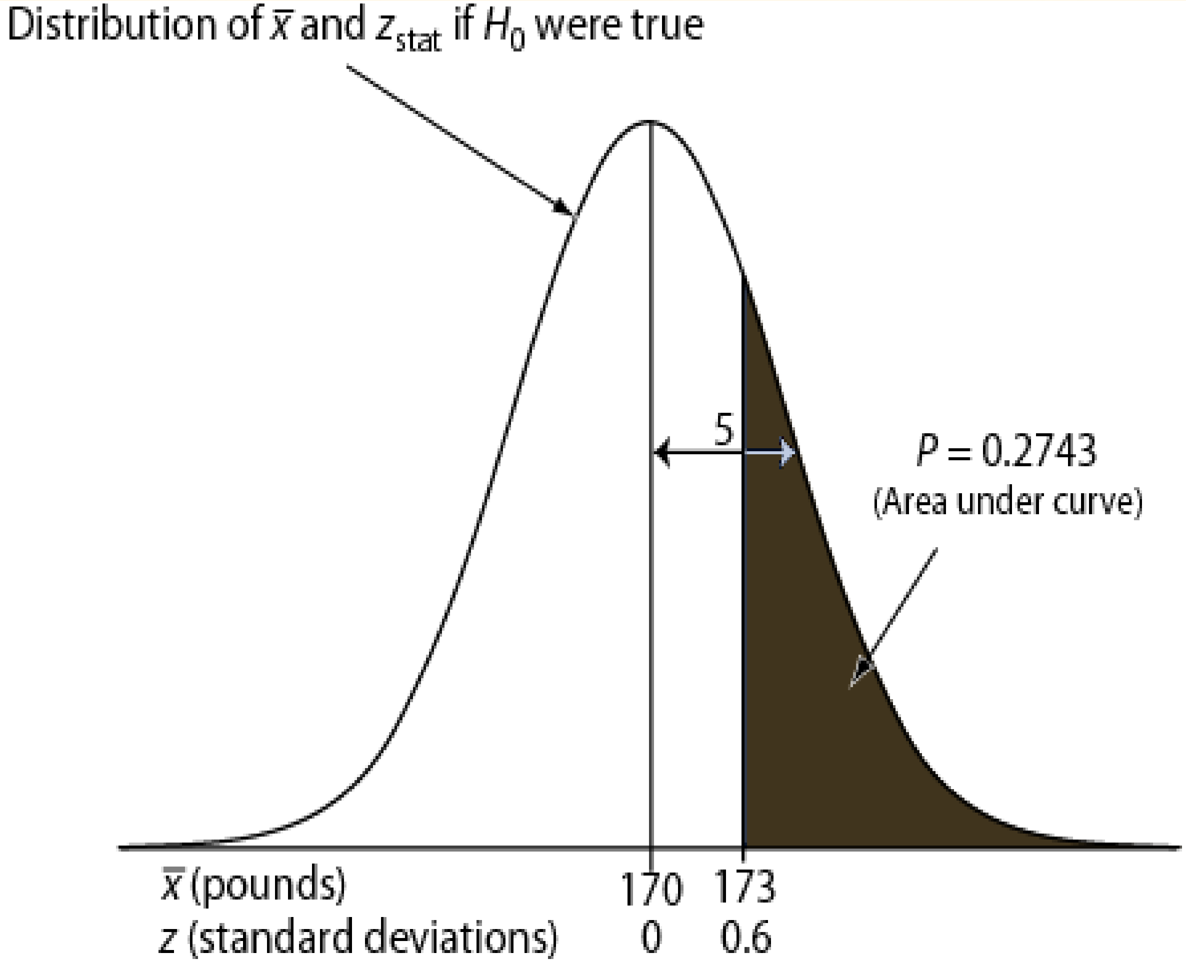
- The *P*-value answer the question: What is the probability of the observed test statistic or one more extreme **when $H_0$ *is true*?**
- This corresponds to the AUC in the tail of the Standard Normal distribution beyond the $z_{stat.}$
- Convert $z$ statistics to *P*-value :

  For $H_a$: $\mu > \mu_0 \Rightarrow P = \Pr(Z > z_{stat})$ = right-tail beyond $z_{stat}$

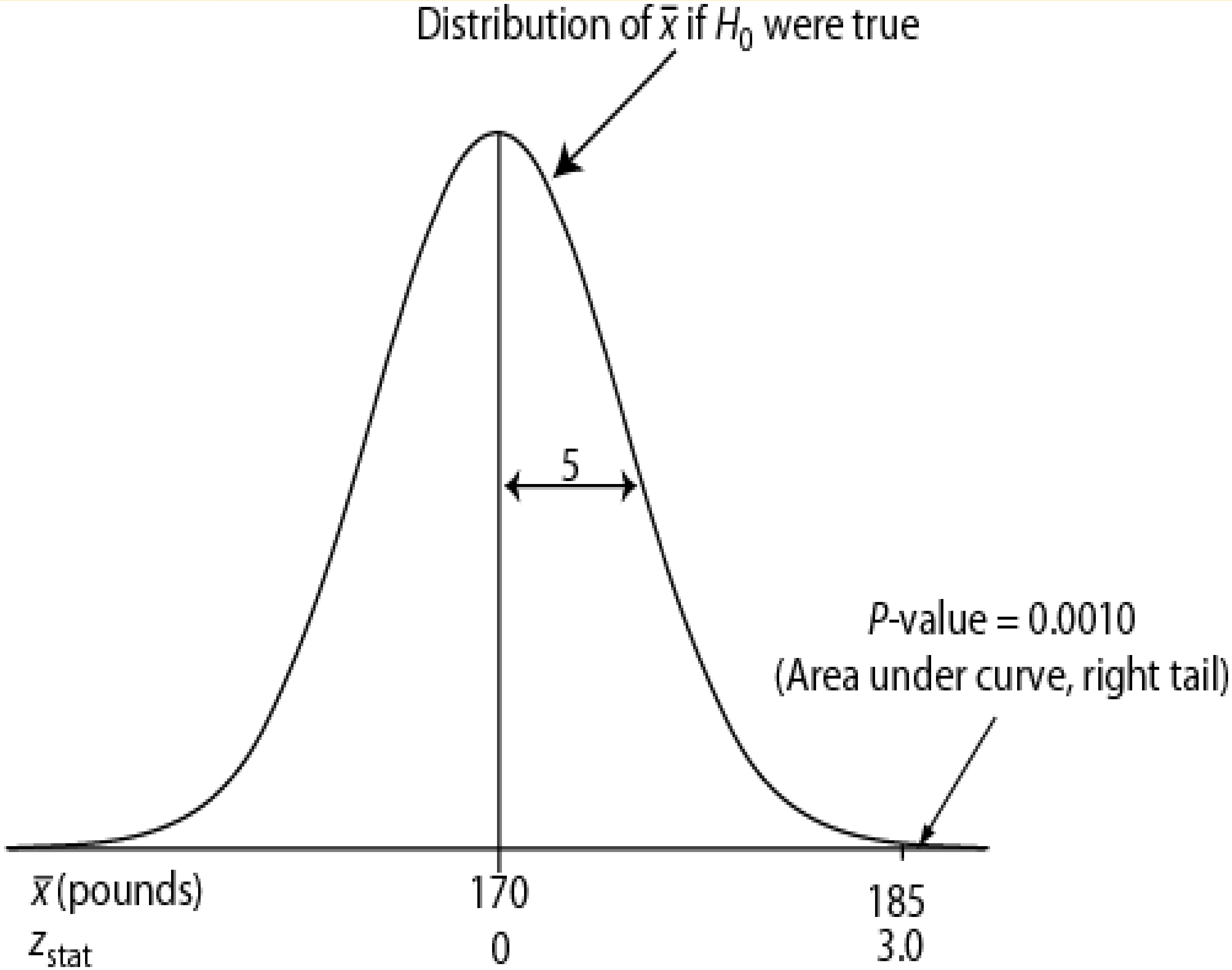
  For $H_a$: $\mu < \mu_0 \Rightarrow P = \Pr(Z < z_{stat})$ = left tail beyond $z_{stat}$

  For $H_a$: $\mu \neq \mu_0 \Rightarrow P = 2 \times$ one-tailed *P*-value
- Use Table B or software to find these probabilities (next two slides).

# One-sided *P*-value for $z_{stat}$ of 0.6

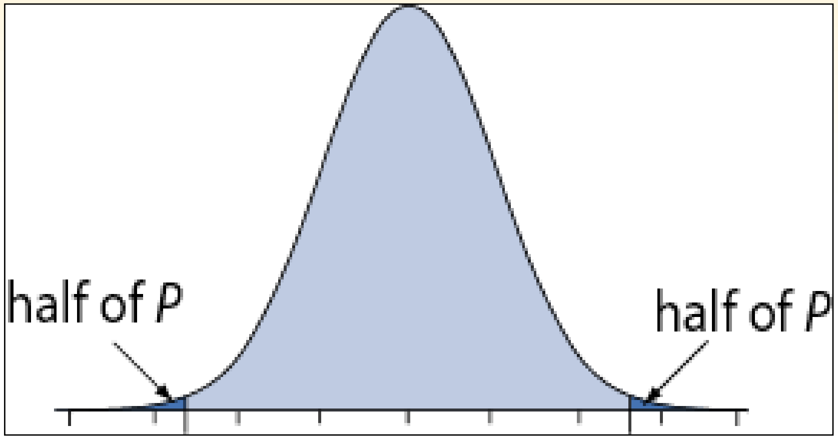


# One-sided *P*-value for $z_{stat}$ of 3.0



# Two-Sided $P$-Value

- One-sided $H_a \Rightarrow$ AUC in tail beyond $z_{stat}$
- Two-sided $H_a \Rightarrow$ consider potential deviations in both directions $\Rightarrow$ double the one-sided $P$-value




Examples: If one-sided $P = 0.0010$, then two-sided $P = 2 \times 0.0010 = 0.0020$. If one-sided $P = 0.2743$, then two-sided $P = 2 \times 0.2743 = 0.5486$.

# Interpretation

- *P*-value answer the question: What is the probability of the observed test statistic … **when $H_0$ *is true*?**
- Thus, smaller and smaller *P*-values provide stronger and stronger evidence against $H_0$
- Small *P*-value $\Rightarrow$ strong evidence

# Interpretation

**Conventions***

$P > 0.10 \Rightarrow$ non-significant evidence against $H_0$

$0.05 < P \leq 0.10 \Rightarrow$ marginally significant evidence

$0.01 < P \leq 0.05 \Rightarrow$ significant evidence against $H_0$

$P \leq 0.01 \Rightarrow$ highly significant evidence against $H_0$

**Examples**

$P = .27 \Rightarrow$ non-significant evidence against $H_0$

$P = .01 \Rightarrow$ highly significant evidence against $H_0$

*** It is *unwise* to draw firm borders for "significance"**

# α-Level (Used in some situations)

- Let $\alpha \equiv$ probability of erroneously rejecting $H_0$
- Set α threshold (e.g., let $\alpha = .10$, .05, *or whatever*)
- Reject $H_0$ when $P \le \alpha$
- Retain $H_0$ when $P > \alpha$
- Example: Set $\alpha = .10$. Find $P = 0.27 \Rightarrow$ retain $H_0$
- Example: Set $\alpha = .01$. Find $P = .001 \Rightarrow$ reject $H_0$

# (Summary) One-Sample *z* Test

A. Hypothesis statements

$H_0$: $\mu = \mu_0$ vs.

$H_a$: $\mu \neq \mu_0$ (two-sided) or

$H_a$: $\mu < \mu_0$ (left-sided) or

$H_a$: $\mu > \mu_0$ (right-sided)

B. Test statistic

$$z_{\text{stat}} = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}} \text{ where } SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

C. P-value: convert $z_{\text{stat}}$ to P value

D. Significance statement (usually not necessary)

# Conditions for z test

- σ known (not from data)
- Population approximately Normal or large sample (central limit theorem)
- SRS (or facsimile)
- Data valid

# The Lake Wobegon Example

## “where all the children are above average”

- Let X represent Weschler Adult Intelligence scores (WAIS)
- Typically, $X \sim N(100, 15)$
- Take SRS of $n = 9$ from Lake Wobegon population
- Data ⇒ {116, 128, 125, 119, 89, 99, 105, 116, 118}
- Calculate: x-bar = 112.8
- Does sample mean provide strong evidence that population mean $\mu > 100$?

# Example: "Lake Wobegon"

**A. Hypotheses:**

$H_0$: μ = 100 versus

$H_a$: μ > 100 (one-sided)

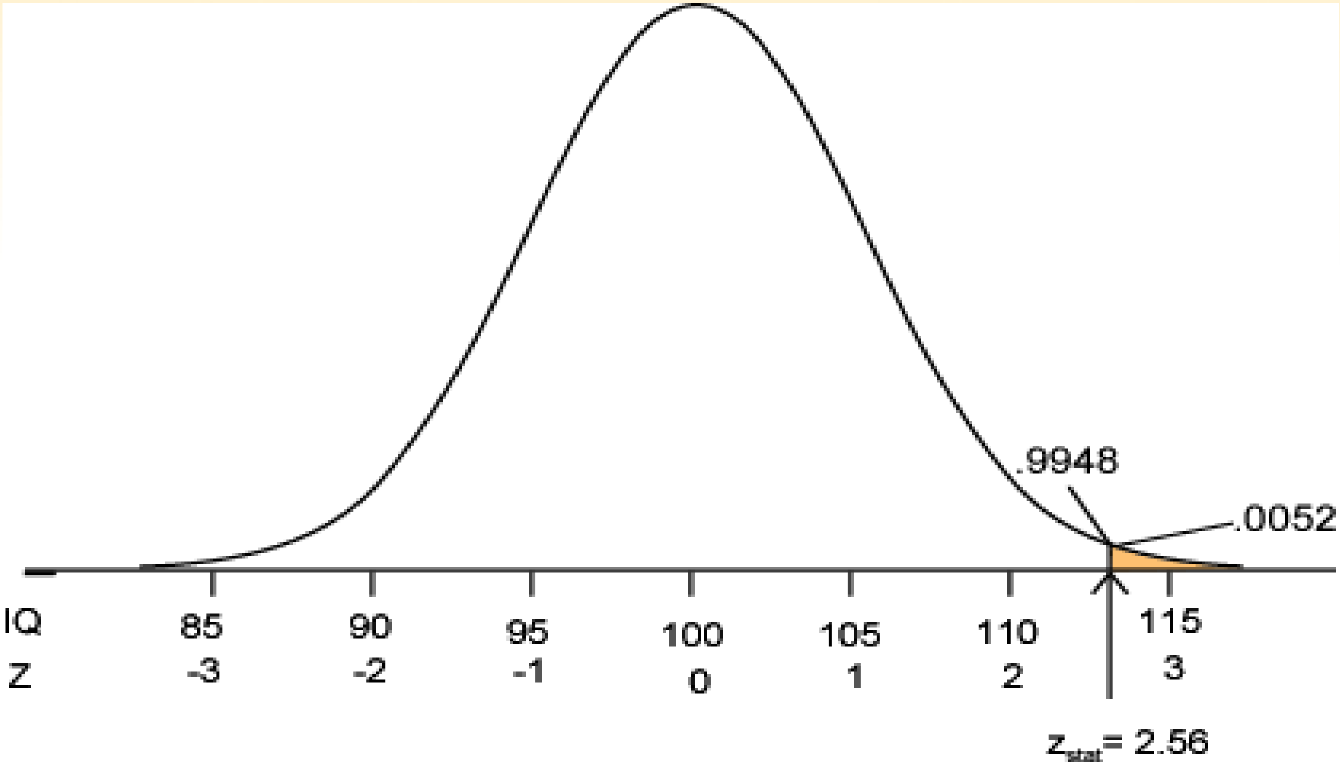
$H_a$: μ ≠ 100 (two-sided)

**B. Test statistic:**

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{9}} = 5$$

$$z_{\text{stat}} = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}} = \frac{112.8 - 100}{5} = 2.56$$

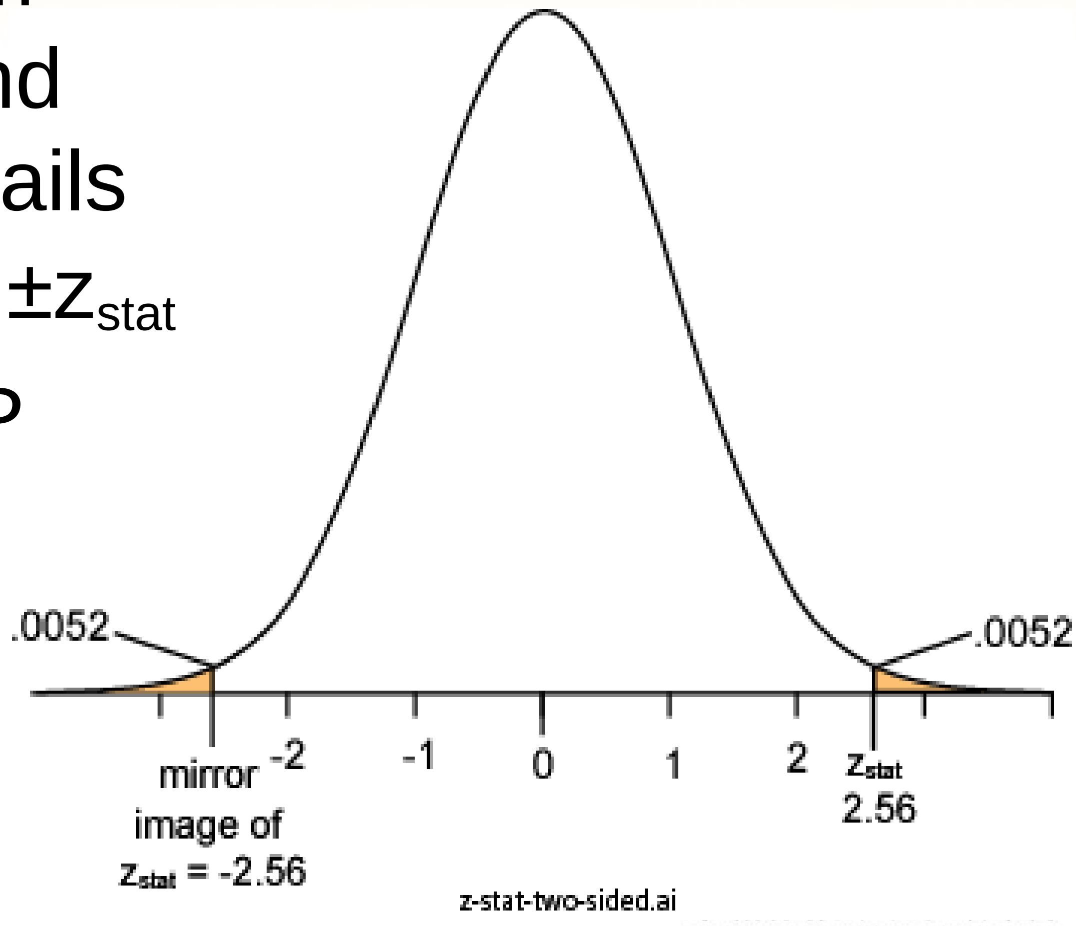
**C. *P*-value:** $P = \Pr(Z \geq 2.56) = 0.0052$




$P$ =.0052 ⇒ it is unlikely the sample came from this null distribution ⇒ strong evidence against $H_0$

# Two-Sided *P*-value: Lake Wobegon

- $H_a$: $\mu \neq 100$
- Considers random deviations "up" and "down" from $\mu_0 \Rightarrow$ tails above and below $\pm z_{stat}$
- Thus, two-sided *P* $= 2 \times 0.0052$ $= 0.0104$



# 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

- $\beta \equiv$ 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

- $\Phi(z)$ represent the cumulative probability of Standard Normal $Z$
- $\mu_0$ represent the population mean under the null hypothesis
- $\mu_a$ represents the population mean under the alternative hypothesis

# Calculating Power: Example

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

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

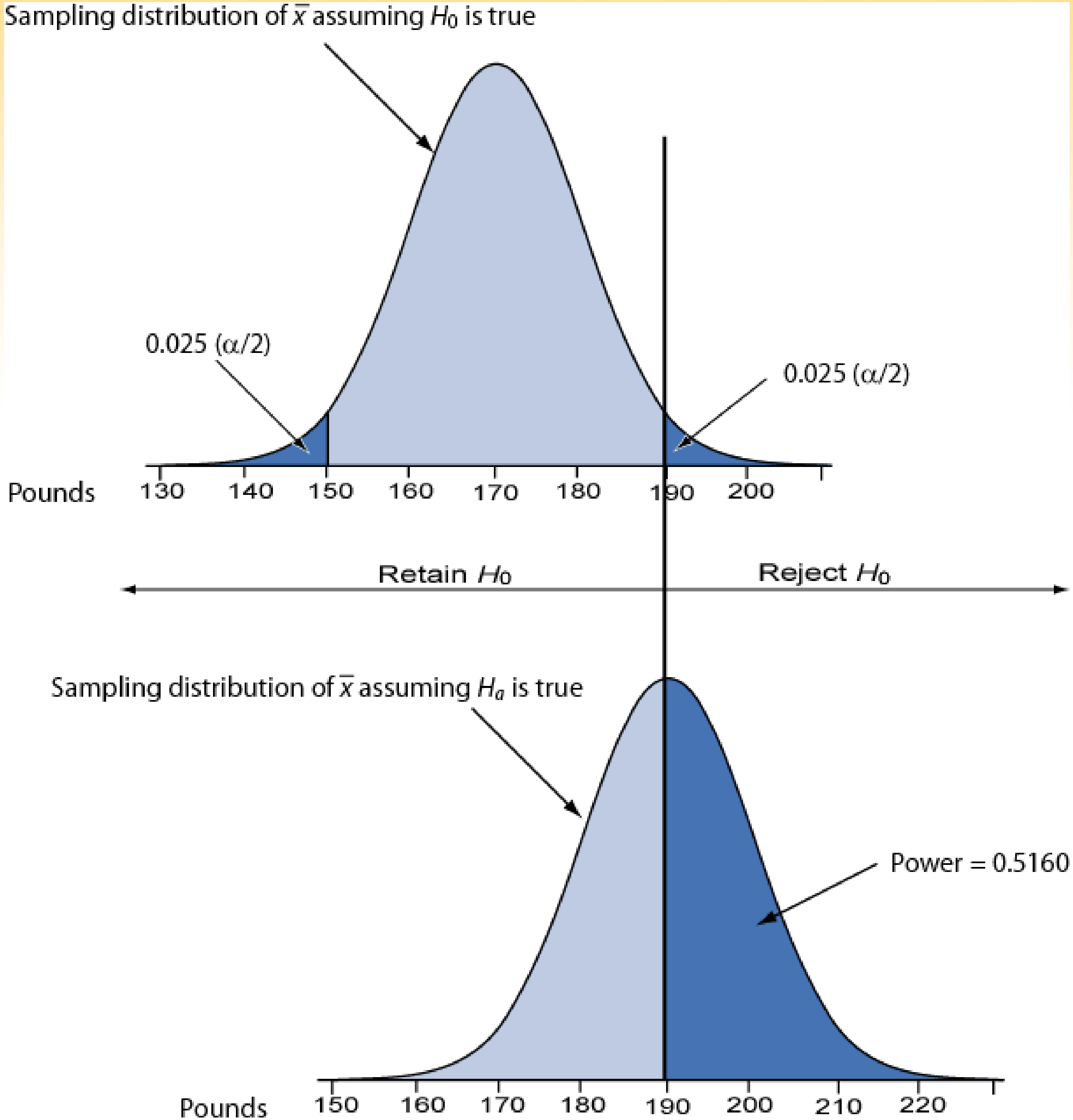
$$= \Phi\left( -1.96 + \frac{|170 - 190|\sqrt{16}}{40} \right)$$

$$= \Phi(0.04)$$

$$= 0.5160$$

# Reasoning Behind Power

- Competing sampling distributions
  Top curve (next page) assumes $H_0$ is true
  Bottom curve assumes $H_a$ is true
  α is set to 0.05 (two-sided)
- We will reject $H_0$ when a sample mean exceeds 189.6 (right tail, top curve)
- The probability of getting a value greater than 189.6 on the bottom curve is 0.5160, corresponding to the power of the test


Sampling distribution of $\overline{x}$ assuming $H_0$ is true
0.025 ($\alpha$/2)
0.025 ($\alpha$/2)
Pounds
130
140
150
160
170
180
190
200
Retain $H_0$
Reject $H_0$
Sampling distribution of $\overline{x}$ assuming $H_a$ is true
Power = 0.5160
Pounds
150
160
170
180
190
200
210
220

# 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 for a one-sample *z* test with 90% power and $\alpha = 0.05$ (two-tailed) when $\sigma = 40$? Let $H_0$: $\mu = 170$ and $H_a$: $\mu = 190$ (thus, $\Delta = \mu_0 - \mu_a = 170 - 190 = -20$)

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

Round up to 42 to ensure adequate power.

Illustration: conditions for 90% power.