

**Applied Regression Modeling:
A Business Approach
Chapter 1: Foundations
Sections 1.1–1.4**

by Iain Pardoe

Identifying and summarizing data

1.1 Identifying and summarizing data

Identifying and summarizing data

Stem-and-leaf plot
Histogram for home prices example
Sample statistics
Sample standardized Z-values

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

- Overall task: analyze data to inform a (business) decision.
- Assume data relevant to the problem has been collected.
- Intermediate task: identify and summarize the data.
- Example: we've moved to a new city and wish to buy a home.
- Data: Y = selling price (in \$ thousands) for $n = 30$ *randomly sampled* single-family homes (**HOMES1**):

155.5	195.0	197.0	207.0	214.9	230.0
239.5	242.0	252.5	255.0	259.9	259.9
269.9	270.0	274.9	283.0	285.0	285.0
299.0	299.9	319.0	319.9	324.5	330.0
336.0	339.0	340.0	355.0	359.9	359.9

Stem-and-leaf plot

1.1 Identifying and summarizing data

Identifying and summarizing data

Stem-and-leaf plot

Histogram for home prices example

Sample statistics

Sample standardized Z-values

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

- Home prices example:

1 | 6

2 | 0011344

2 | 5666777899

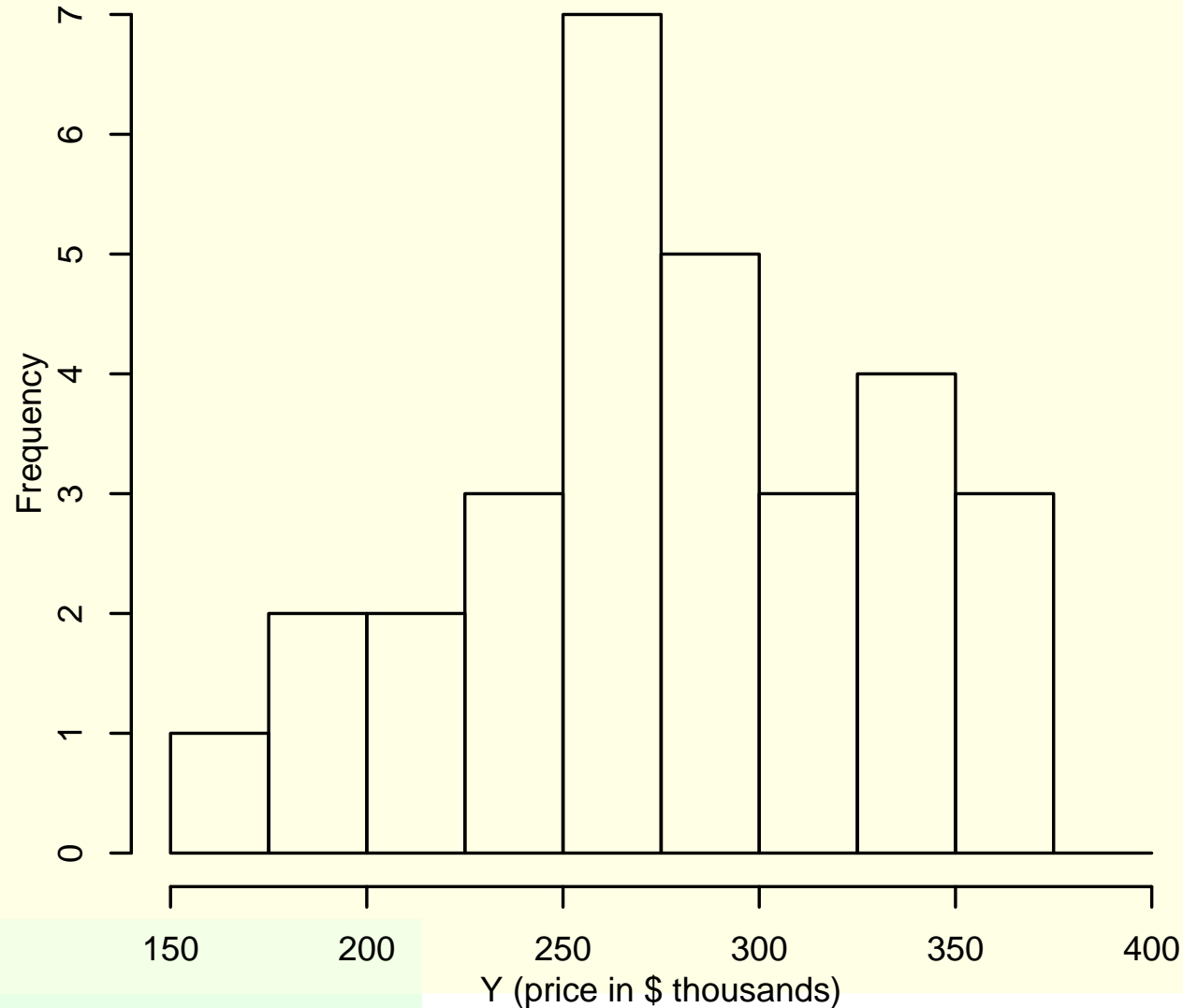
3 | 002223444

3 | 666

- Consider lowest home price represented by “1” in the stem and “6” in the leaf.
- This represents a number between 155 and 164.9 (thousand dollars).
- In particular, it is the lowest price of \$155,500.
- What does this graph tell you about home prices in this market?

Histogram for home prices example

Compare stem-and-leaf plot with a histogram:



1.1 Identifying and summarizing data

Identifying and summarizing data

Stem-and-leaf plot
Histogram for home prices example

Sample statistics

Sample standardized Z-values

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

- Sample mean, m_Y , measures “central tendency” of Y -values.
- Median also measures central tendency, but less sensitive to very small/large values.
- Sample standard deviation, s_Y , measures spread/variation.
- Minimum and maximum.
- Percentiles, e.g., 25th percentile: 25% of Y -values are smaller and 75% of Y -values are larger.
- Question: what’s another name for the 50th percentile?

Sample standardized Z -values

- Standardizing calibrates a list of numbers (Y) to a common scale.
- Subtract the mean and divide by the standard deviation:

$$Z = \frac{Y - m_Y}{s_Y}.$$

1.1 Identifying and summarizing data

Identifying and summarizing data

Stem-and-leaf plot
Histogram for home prices example

Sample statistics

Sample standardized Z -values

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Sample standardized Z -values

1.1 Identifying and summarizing data

Identifying and summarizing data

Stem-and-leaf plot
Histogram for home prices example

Sample statistics

Sample standardized Z -values

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

- Standardizing calibrates a list of numbers (Y) to a common scale.
- Subtract the mean and divide by the standard deviation:

$$Z = \frac{Y - m_Y}{s_Y}.$$

- Sample mean of Z -values?

Sample standardized Z -values

1.1 Identifying and summarizing data

Identifying and summarizing data

Stem-and-leaf plot
Histogram for home prices example

Sample statistics

Sample standardized Z -values

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

- Standardizing calibrates a list of numbers (Y) to a common scale.
- Subtract the mean and divide by the standard deviation:

$$Z = \frac{Y - m_Y}{s_Y}.$$

- Sample mean of Z -values? 0

Sample standardized Z -values

1.1 Identifying and summarizing data

Identifying and summarizing data

Stem-and-leaf plot
Histogram for home prices example

Sample statistics

Sample standardized Z -values

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

- Standardizing calibrates a list of numbers (Y) to a common scale.
- Subtract the mean and divide by the standard deviation:

$$Z = \frac{Y - m_Y}{s_Y}.$$

- Sample mean of Z -values? 0
- Sample standard deviation of Z -values?

Sample standardized Z -values

1.1 Identifying and summarizing data

Identifying and summarizing data

Stem-and-leaf plot
Histogram for home prices example

Sample statistics

Sample standardized Z -values

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

- Standardizing calibrates a list of numbers (Y) to a common scale.
- Subtract the mean and divide by the standard deviation:

$$Z = \frac{Y - m_Y}{s_Y}.$$

- Sample mean of Z -values? 0
- Sample standard deviation of Z -values? 1

Sample standardized Z -values

1.1 Identifying and summarizing data

Identifying and summarizing data

Stem-and-leaf plot
Histogram for home prices example

Sample statistics

Sample standardized Z -values

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

- Standardizing calibrates a list of numbers (Y) to a common scale.
- Subtract the mean and divide by the standard deviation:

$$Z = \frac{Y - m_Y}{s_Y}.$$

- Sample mean of Z -values? 0
- Sample standard deviation of Z -values? 1
- Exercise: use statistical software to create graphs, find summary statistics, and calculate standardized values for home prices example.

Population distributions

1.1 Identifying and summarizing data

1.2 Population distributions

Population distributions

Normal histogram for 1000 simulated home prices

Standard normal density curve

Critical values for standard normal distribution

Assessing normality
QQ-plot for home prices example

1.3 Selecting individuals at random—probability

1.4 Random sampling

- Population: entire collection of objects of interest.
- Sample: (random) subset of population.
- Statistical thinking: draw inferences about population by using sample data.
- Model: mathematical abstraction of the real world used to make statistical inferences.
- Assumptions:
 - model provides a reasonable fit to sample data,
 - sample is representative of population.
- Normal distribution: simple, effective model (“bell-curve”).

Normal histogram for 1000 simulated home prices

1.1 Identifying and summarizing data

1.2 Population distributions

Population distributions

Normal histogram for 1000 simulated home prices

Standard normal density curve

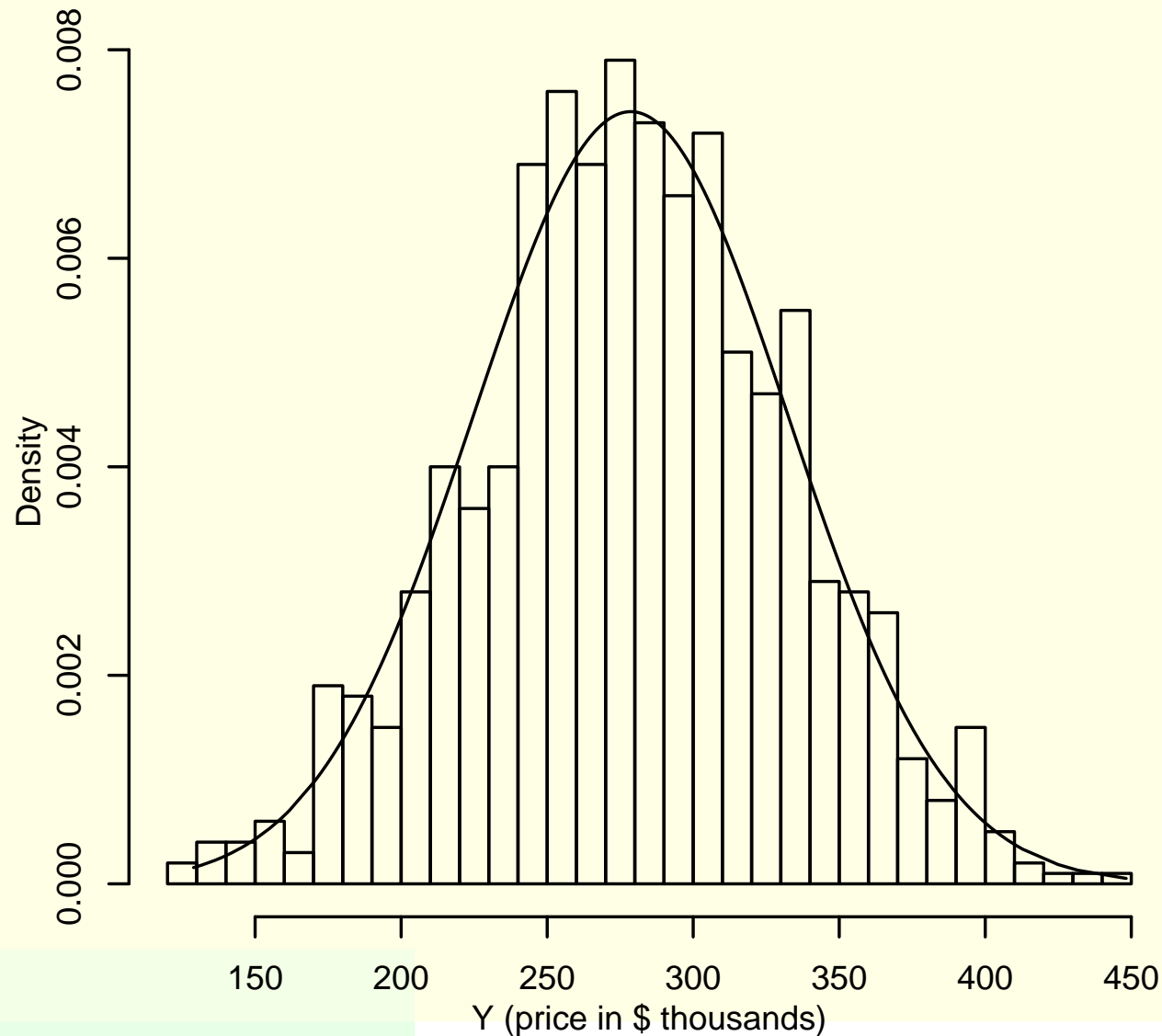
Critical values for standard normal distribution

Assessing normality
QQ-plot for home prices example

1.3 Selecting individuals at random—probability

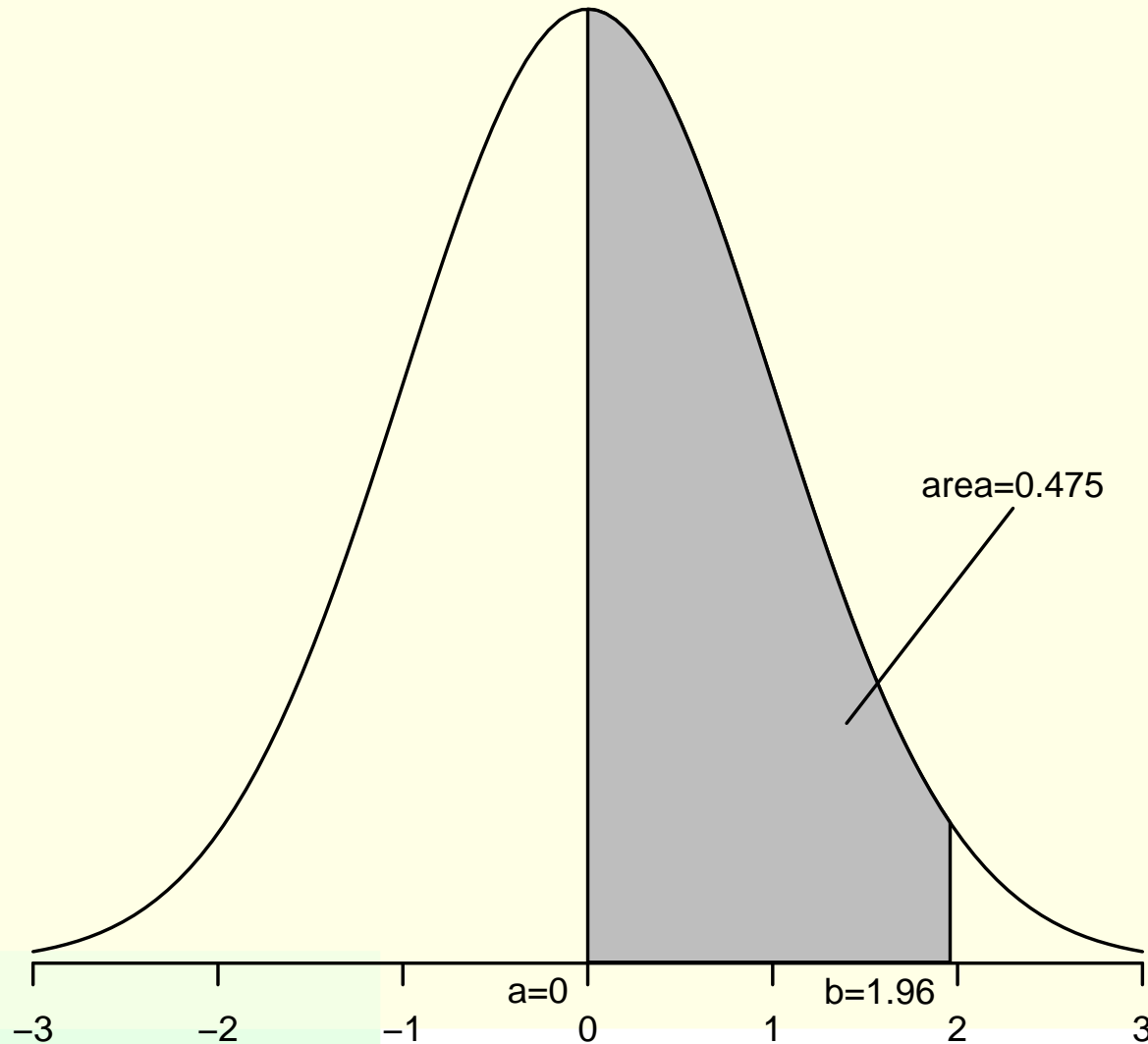
1.4 Random sampling

What happens to histogram as sample size increases?



Standard normal density curve

Shaded area = $\Pr(\text{standard normal is between } a \text{ and } b)$:



1.1 Identifying and summarizing data

1.2 Population distributions

Population distributions
Normal histogram for 1000 simulated home prices

Standard normal density curve

Critical values for standard normal distribution

Assessing normality
QQ-plot for home prices example

1.3 Selecting individuals at random—probability

1.4 Random sampling

Critical values for standard normal distribution

upper-tail area	0.1	0.05	0.025	0.01	0.005	0.001
horizontal axis value	1.282	1.645	1.960	2.326	2.576	3.090
two-tail area	0.2	0.1	0.05	0.02	0.01	0.002

- Horizontal axis values are called *critical values*.
- Tail areas (under the density curve) represent probabilities.
- Example: $\Pr(Z > 1.960) = 0.025$
and $\Pr(0 < Z < 1.960) = 1 - 0.5 - 0.025 = 0.475$.
- Exercises:
 - $\Pr(Z > 1.645) = ?$
 - $\Pr(Z < -2.326 \text{ or } > 2.326) = ?$
 - $\Pr(Z < ?) = 0.90$
(i.e., what is the 90th percentile?).

1.1 Identifying and summarizing data

1.2 Population distributions

Population distributions
Normal histogram for 1000 simulated home prices
Standard normal density curve

Critical values for standard normal distribution

Assessing normality
QQ-plot for home prices example

1.3 Selecting individuals at random—probability

1.4 Random sampling

1.1 Identifying and summarizing data

1.2 Population distributions

Population distributions
Normal histogram for 1000 simulated home prices
Standard normal density curve
Critical values for standard normal distribution

Assessing normality

QQ-plot for home prices example

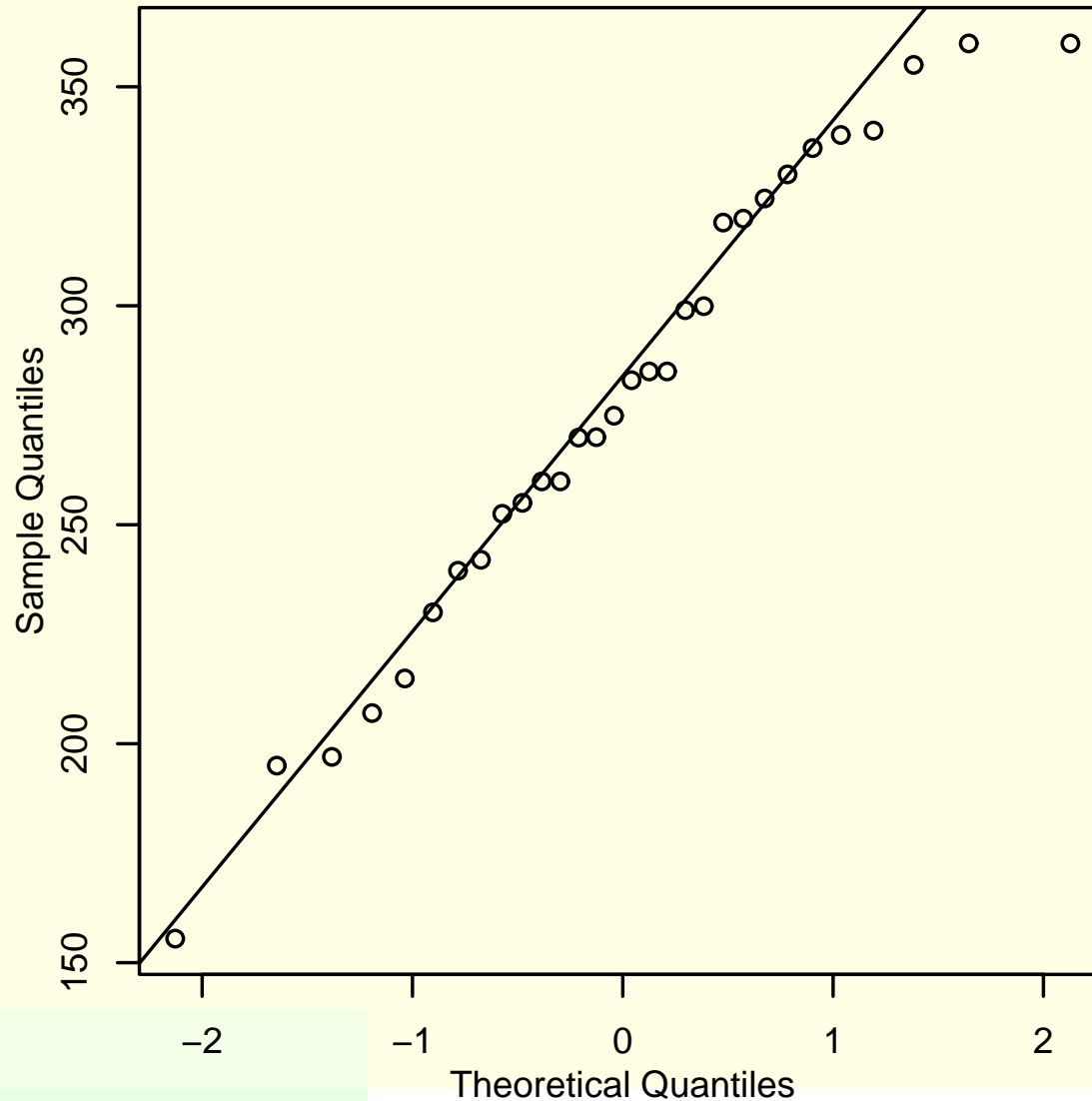
1.3 Selecting individuals at random—probability

1.4 Random sampling

- Previous slide showed how to make probability calculations for a standard normal distribution (mean 0, standard deviation 1).
- Section 1.3 shows similar calculations for a normal distribution with any mean and standard deviation.
- Such calculations are useful if our variable of interest (e.g., home price) has a normal distribution.
- How can we tell if a particular variable has a normal distribution?
 - Draw a histogram: is it *approximately* symmetric and bell-shaped? (see histogram for home prices example)
 - Draw a QQ-plot: do the points lie *reasonably* close to the line? (see next slide)

QQ-plot for home prices example

Do the points lie *reasonably* close to the line?



Normal probability and percentile calculations

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

Normal probability and percentile calculations

Finding probabilities

Finding percentiles

1.4 Random sampling

- Connection between normal distribution with any mean, $E(Y)$, and standard deviation, $SD(Y)$, and standard normal distribution:
 - Suppose $Y \sim \text{Normal}(E(Y), SD(Y)^2)$.
 - Then $Z = \frac{Y - E(Y)}{SD(Y)} \sim \text{Normal}(0, 1^2)$.
- Idea for finding probabilities: standardize Y into Z -units, then do probability calculation on Z (example next slide).
- Can also go other way to find percentiles: do probability calculation on Z , then unstandardize Z into Y -units (example subsequent slide).

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability
Normal probability and percentile calculations

Finding probabilities

Finding percentiles

1.4 Random sampling

- Assume home prices $Y \sim \text{Normal}(280, 50^2)$.
- Then $Z = \frac{Y-280}{50} \sim \text{Normal}(0, 1^2)$.
- What is the probability a home price is greater than \$360,000?

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability
Normal probability and percentile calculations

Finding probabilities

Finding percentiles

1.4 Random sampling

- Assume home prices $Y \sim \text{Normal}(280, 50^2)$.
- Then $Z = \frac{Y-280}{50} \sim \text{Normal}(0, 1^2)$.
- What is the probability a home price is greater than \$360,000?
- $\Pr(Y > 360)$

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability
Normal probability and percentile calculations

Finding probabilities

Finding percentiles

1.4 Random sampling

- Assume home prices $Y \sim \text{Normal}(280, 50^2)$.
- Then $Z = \frac{Y-280}{50} \sim \text{Normal}(0, 1^2)$.
- What is the probability a home price is greater than \$360,000?
- $$\Pr(Y > 360) = \Pr\left(\frac{Y - 280}{50} > \frac{360 - 280}{50}\right)$$

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability
Normal probability and percentile calculations

Finding probabilities

Finding percentiles

1.4 Random sampling

- Assume home prices $Y \sim \text{Normal}(280, 50^2)$.
- Then $Z = \frac{Y-280}{50} \sim \text{Normal}(0, 1^2)$.
- What is the probability a home price is greater than \$360,000?
- $$\begin{aligned}\Pr(Y > 360) &= \Pr\left(\frac{Y - 280}{50} > \frac{360 - 280}{50}\right) \\ &= \Pr(Z > 1.60)\end{aligned}$$

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability
Normal probability and percentile calculations

Finding probabilities

Finding percentiles

1.4 Random sampling

- Assume home prices $Y \sim \text{Normal}(280, 50^2)$.
- Then $Z = \frac{Y-280}{50} \sim \text{Normal}(0, 1^2)$.
- What is the probability a home price is greater than \$360,000?
- $$\begin{aligned}\Pr(Y > 360) &= \Pr\left(\frac{Y - 280}{50} > \frac{360 - 280}{50}\right) \\ &= \Pr(Z > 1.60) \\ &\approx 0.05.\end{aligned}$$

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability
Normal probability and percentile calculations

Finding probabilities
Finding percentiles

1.4 Random sampling

- Assume home prices $Y \sim \text{Normal}(280, 50^2)$.
- Then $Z = \frac{Y-280}{50} \sim \text{Normal}(0, 1^2)$.
- What is the probability a home price is greater than \$360,000?
- $$\begin{aligned}\Pr(Y > 360) &= \Pr\left(\frac{Y - 280}{50} > \frac{360 - 280}{50}\right) \\ &= \Pr(Z > 1.60) \\ &\approx 0.05.\end{aligned}$$
- What is the probability a home price is less than \$165,000?

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

Normal probability and percentile calculations

Finding probabilities

Finding percentiles

1.4 Random sampling

- Assume home prices $Y \sim \text{Normal}(280, 50^2)$.
- Then $Z = \frac{Y-280}{50} \sim \text{Normal}(0, 1^2)$.
- What is the 95th percentile of Y ?

Finding percentiles

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

Normal probability and percentile calculations

Finding probabilities

Finding percentiles

1.4 Random sampling

- Assume home prices $Y \sim \text{Normal}(280, 50^2)$.
- Then $Z = \frac{Y-280}{50} \sim \text{Normal}(0, 1^2)$.
- What is the 95th percentile of Y ?
- $\Pr(Z > 1.645) = 0.05$

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

Normal probability and percentile calculations

Finding probabilities

Finding percentiles

1.4 Random sampling

- Assume home prices $Y \sim \text{Normal}(280, 50^2)$.

- Then $Z = \frac{Y-280}{50} \sim \text{Normal}(0, 1^2)$.

- What is the 95th percentile of Y ?

- $\Pr(Z > 1.645) = 0.05$

$$\Pr\left(\frac{Y - 280}{50} > 1.645\right) = 0.05$$

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

Normal probability and percentile calculations

Finding probabilities

Finding percentiles

1.4 Random sampling

- Assume home prices $Y \sim \text{Normal}(280, 50^2)$.

- Then $Z = \frac{Y-280}{50} \sim \text{Normal}(0, 1^2)$.

- What is the 95th percentile of Y ?

- $$\Pr(Z > 1.645) = 0.05$$

$$\Pr\left(\frac{Y - 280}{50} > 1.645\right) = 0.05$$

$$\Pr(Y > 1.645(50) + 280) = 0.05$$

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

Normal probability and percentile calculations

Finding probabilities

Finding percentiles

1.4 Random sampling

- Assume home prices $Y \sim \text{Normal}(280, 50^2)$.

- Then $Z = \frac{Y-280}{50} \sim \text{Normal}(0, 1^2)$.

- What is the 95th percentile of Y ?

- $$\Pr(Z > 1.645) = 0.05$$

$$\Pr\left(\frac{Y - 280}{50} > 1.645\right) = 0.05$$

$$\Pr(Y > 1.645(50) + 280) = 0.05$$

$$\Pr(Y > 362) = 0.05.$$

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

Normal probability and percentile calculations

Finding probabilities

Finding percentiles

1.4 Random sampling

- Assume home prices $Y \sim \text{Normal}(280, 50^2)$.

- Then $Z = \frac{Y-280}{50} \sim \text{Normal}(0, 1^2)$.

- What is the 95th percentile of Y ?

- $$\Pr(Z > 1.645) = 0.05$$

$$\Pr\left(\frac{Y - 280}{50} > 1.645\right) = 0.05$$

$$\Pr(Y > 1.645(50) + 280) = 0.05$$

$$\Pr(Y > 362) = 0.05.$$

- What is the 90th percentile of Y ?

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling

Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action

Student's t-distribution
Critical values for t-distributions

Central limit theorem—t version

- Population parameters: numerical summary measures of the population, e.g.:
 - mean, $E(Y)$, and standard deviation, $SD(Y)$.
- Sample statistics: analogous sample measures, e.g.:
 - mean, m_Y , and standard deviation, s_Y .
- Statistical inference: use sample statistics to infer about (likely values of) population parameters.
- Example: the sample mean is an estimate of the population mean.
- Question: how far off might the estimate be?

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling

Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action

Student's t-distribution
Critical values for t-distributions

Central limit theorem—t version

- Population parameters: numerical summary measures of the population, e.g.:
 - mean, $E(Y)$, and standard deviation, $SD(Y)$.
- Sample statistics: analogous sample measures, e.g.:
 - mean, m_Y , and standard deviation, s_Y .
- Statistical inference: use sample statistics to infer about (likely values of) population parameters.
- Example: the sample mean is an estimate of the population mean.
- Question: how far off might the estimate be?
 - Could be a long way off if Y is very variable and/or sample size is small.

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling

Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action

Student's t-distribution
Critical values for t-distributions

Central limit theorem—t version

- Population parameters: numerical summary measures of the population, e.g.:
 - mean, $E(Y)$, and standard deviation, $SD(Y)$.
- Sample statistics: analogous sample measures, e.g.:
 - mean, m_Y , and standard deviation, s_Y .
- Statistical inference: use sample statistics to infer about (likely values of) population parameters.
- Example: the sample mean is an estimate of the population mean.
- Question: how far off might the estimate be?
 - Could be a long way off if Y is very variable and/or sample size is small.
- Quantify uncertainty using *sampling distributions*.

Central limit theorem—normal version

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling

Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action

Student's t-distribution

Critical values for t-distributions

Central limit theorem—t version

- Randomly sample Y_1, Y_2, \dots, Y_n from a population with mean, $E(Y)$, and standard deviation, $SD(Y)$.
- CLT: $m_Y \sim \text{Normal}(E(Y), SD(Y)^2/n)$,
so $Z = \frac{m_Y - E(Y)}{SD(Y)/\sqrt{n}} \sim \text{Normal}(0, 1^2)$.
- Assume home prices Y_1, Y_2, \dots, Y_{30} have $E(Y) = 280$ and $SD(Y) = 50$.
- What is the 95th percentile of m_Y ?

Central limit theorem—normal version

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling

Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action

Student's t-distribution

Critical values for t-distributions

Central limit theorem—t version

- Randomly sample Y_1, Y_2, \dots, Y_n from a population with mean, $E(Y)$, and standard deviation, $SD(Y)$.
- CLT: $m_Y \sim \text{Normal}(E(Y), SD(Y)^2/n)$,
so $Z = \frac{m_Y - E(Y)}{SD(Y)/\sqrt{n}} \sim \text{Normal}(0, 1^2)$.

- Assume home prices Y_1, Y_2, \dots, Y_{30} have $E(Y) = 280$ and $SD(Y) = 50$.
- What is the 95th percentile of m_Y ?

- $$\Pr(Z > 1.645) = 0.05$$

Central limit theorem—normal version

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling

Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action

Student's t-distribution

Critical values for t-distributions

Central limit theorem—t version

- Randomly sample Y_1, Y_2, \dots, Y_n from a population with mean, $E(Y)$, and standard deviation, $SD(Y)$.
- CLT: $m_Y \sim \text{Normal}(E(Y), SD(Y)^2/n)$,
so $Z = \frac{m_Y - E(Y)}{SD(Y)/\sqrt{n}} \sim \text{Normal}(0, 1^2)$.

- Assume home prices Y_1, Y_2, \dots, Y_{30} have $E(Y) = 280$ and $SD(Y) = 50$.
- What is the 95th percentile of m_Y ?

- $$\Pr(Z > 1.645) = 0.05$$
$$\Pr\left(\frac{m_Y - 280}{50/\sqrt{30}} > 1.645\right) = 0.05$$

Central limit theorem—normal version

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling

Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action

Student's t-distribution

Critical values for t-distributions

Central limit theorem—t version

- Randomly sample Y_1, Y_2, \dots, Y_n from a population with mean, $E(Y)$, and standard deviation, $SD(Y)$.
- CLT: $m_Y \sim \text{Normal}(E(Y), SD(Y)^2/n)$,
so $Z = \frac{m_Y - E(Y)}{SD(Y)/\sqrt{n}} \sim \text{Normal}(0, 1^2)$.

- Assume home prices Y_1, Y_2, \dots, Y_{30} have $E(Y) = 280$ and $SD(Y) = 50$.
- What is the 95th percentile of m_Y ?

- $$\Pr(Z > 1.645) = 0.05$$

$$\Pr\left(\frac{m_Y - 280}{50/\sqrt{30}} > 1.645\right) = 0.05$$

$$\Pr(m_Y > 1.645(50/\sqrt{30}) + 280) = 0.05$$

Central limit theorem—normal version

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling

Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action

Student's t-distribution

Critical values for t-distributions

Central limit theorem—t version

- Randomly sample Y_1, Y_2, \dots, Y_n from a population with mean, $E(Y)$, and standard deviation, $SD(Y)$.
- CLT: $m_Y \sim \text{Normal}(E(Y), SD(Y)^2/n)$,
so $Z = \frac{m_Y - E(Y)}{SD(Y)/\sqrt{n}} \sim \text{Normal}(0, 1^2)$.

- Assume home prices Y_1, Y_2, \dots, Y_{30} have $E(Y) = 280$ and $SD(Y) = 50$.
- What is the 95th percentile of m_Y ?

- $$\Pr(Z > 1.645) = 0.05$$

$$\Pr\left(\frac{m_Y - 280}{50/\sqrt{30}} > 1.645\right) = 0.05$$

$$\Pr(m_Y > 1.645(50/\sqrt{30}) + 280) = 0.05$$

$$\Pr(m_Y > 295) = 0.05.$$

Central limit theorem—normal version

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling

Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action

Student's t-distribution

Critical values for t-distributions

Central limit theorem—t version

- Randomly sample Y_1, Y_2, \dots, Y_n from a population with mean, $E(Y)$, and standard deviation, $SD(Y)$.
- CLT: $m_Y \sim \text{Normal}(E(Y), SD(Y)^2/n)$,
so $Z = \frac{m_Y - E(Y)}{SD(Y)/\sqrt{n}} \sim \text{Normal}(0, 1^2)$.

- Assume home prices Y_1, Y_2, \dots, Y_{30} have $E(Y) = 280$ and $SD(Y) = 50$.
- What is the 95th percentile of m_Y ?

- $$\Pr(Z > 1.645) = 0.05$$

$$\Pr\left(\frac{m_Y - 280}{50/\sqrt{30}} > 1.645\right) = 0.05$$

$$\Pr(m_Y > 1.645(50/\sqrt{30}) + 280) = 0.05$$

$$\Pr(m_Y > 295) = 0.05.$$

- What is the 90th percentile of m_Y ?

Finding sampling distribution probabilities

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling
Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action
Student's t-distribution
Critical values for t-distributions
Central limit theorem—t version

- Assume home prices Y_1, Y_2, \dots, Y_{30} have $E(Y) = 280$ and $SD(Y) = 50$.
- What is the probability the sample mean is greater than 295?

Finding sampling distribution probabilities

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling
Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action
Student's t-distribution
Critical values for t-distributions
Central limit theorem—t version

- Assume home prices Y_1, Y_2, \dots, Y_{30} have $E(Y) = 280$ and $SD(Y) = 50$.
- What is the probability the sample mean is greater than 295?
- CLT: $Z = \frac{m_Y - 280}{50/\sqrt{30}} \sim \text{Normal}(0, 1^2)$.

Finding sampling distribution probabilities

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling
Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action
Student's t-distribution
Critical values for t-distributions
Central limit theorem—t version

- Assume home prices Y_1, Y_2, \dots, Y_{30} have $E(Y) = 280$ and $SD(Y) = 50$.
- What is the probability the sample mean is greater than 295?
- CLT: $Z = \frac{m_Y - 280}{50/\sqrt{30}} \sim \text{Normal}(0, 1^2)$.
- $\Pr(m_Y > 295)$

Finding sampling distribution probabilities

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling
Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action
Student's t-distribution
Critical values for t-distributions
Central limit theorem—t version

- Assume home prices Y_1, Y_2, \dots, Y_{30} have $E(Y) = 280$ and $SD(Y) = 50$.
- What is the probability the sample mean is greater than 295?
- CLT: $Z = \frac{m_Y - 280}{50/\sqrt{30}} \sim \text{Normal}(0, 1^2)$.
- $\Pr(m_Y > 295) = \Pr\left(\frac{m_Y - 280}{50/\sqrt{30}} > \frac{295 - 280}{50/\sqrt{30}}\right)$

Finding sampling distribution probabilities

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling
Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action
Student's t-distribution
Critical values for t-distributions
Central limit theorem—t version

- Assume home prices Y_1, Y_2, \dots, Y_{30} have $E(Y) = 280$ and $SD(Y) = 50$.
- What is the probability the sample mean is greater than 295?
- CLT: $Z = \frac{m_Y - 280}{50/\sqrt{30}} \sim \text{Normal}(0, 1^2)$.
- $$\begin{aligned} \Pr(m_Y > 295) &= \Pr\left(\frac{m_Y - 280}{50/\sqrt{30}} > \frac{295 - 280}{50/\sqrt{30}}\right) \\ &= \Pr(Z > 1.643) \end{aligned}$$

Finding sampling distribution probabilities

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling
Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action
Student's t-distribution
Critical values for t-distributions
Central limit theorem—t version

- Assume home prices Y_1, Y_2, \dots, Y_{30} have $E(Y) = 280$ and $SD(Y) = 50$.
- What is the probability the sample mean is greater than 295?
- CLT: $Z = \frac{m_Y - 280}{50/\sqrt{30}} \sim \text{Normal}(0, 1^2)$.
- $$\begin{aligned} \Pr(m_Y > 295) &= \Pr\left(\frac{m_Y - 280}{50/\sqrt{30}} > \frac{295 - 280}{50/\sqrt{30}}\right) \\ &= \Pr(Z > 1.643) \\ &\approx 0.05. \end{aligned}$$

Finding sampling distribution probabilities

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling
Central limit theorem—normal version

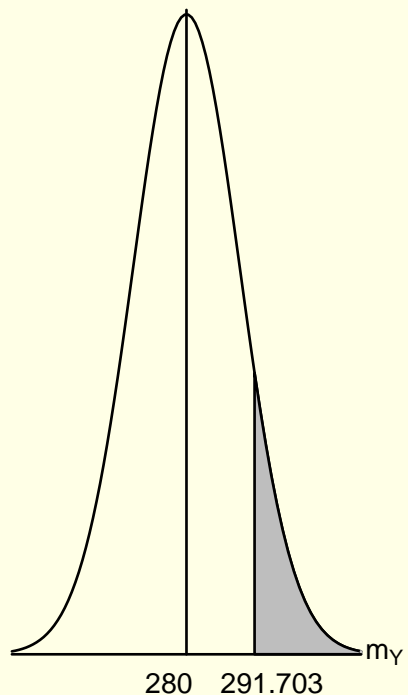
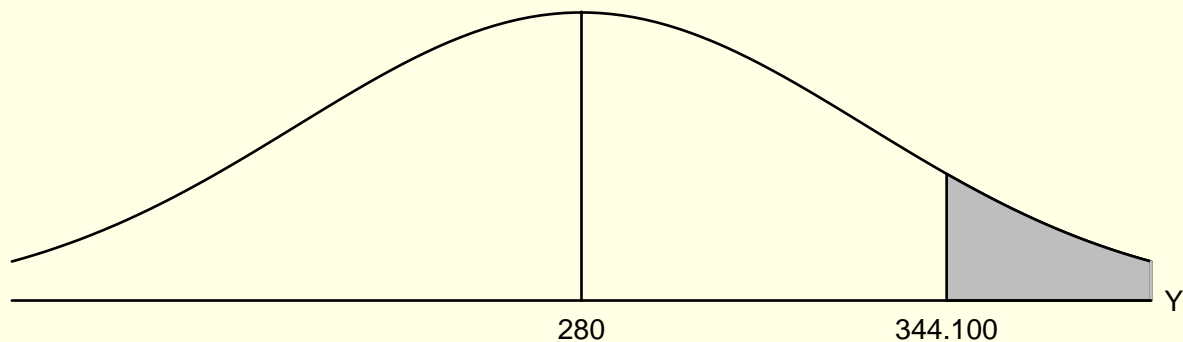
Finding sampling distribution probabilities

The central limit theorem in action
Student's t-distribution
Critical values for t-distributions
Central limit theorem—t version

- Assume home prices Y_1, Y_2, \dots, Y_{30} have $E(Y) = 280$ and $SD(Y) = 50$.
- What is the probability the sample mean is greater than 295?
- CLT: $Z = \frac{m_Y - 280}{50/\sqrt{30}} \sim \text{Normal}(0, 1^2)$.
- $$\begin{aligned} \Pr(m_Y > 295) &= \Pr\left(\frac{m_Y - 280}{50/\sqrt{30}} > \frac{295 - 280}{50/\sqrt{30}}\right) \\ &= \Pr(Z > 1.643) \\ &\approx 0.05. \end{aligned}$$
- What is the probability the sample mean is greater than 292?

The central limit theorem in action

Top: Y population distn. Bottom: m_Y sampling distn.



1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling

Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action

Student's t-distribution

Critical values for t-distributions

Central limit theorem—t version

Student's t-distribution

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling

Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action

Student's t-distribution

Critical values for t-distributions

Central limit theorem—t version

- Drawback to CLT: need to know population standard deviation, $SD(Y)$, to use it.
- Since we rarely know $SD(Y)$, what would be a good estimate to use instead?

Student's t-distribution

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling

Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action

Student's t-distribution

Critical values for t-distributions

Central limit theorem—t version

- Drawback to CLT: need to know population standard deviation, $SD(Y)$, to use it.
- Since we rarely know $SD(Y)$, what would be a good estimate to use instead? The sample s.d., s_Y .

Student's t-distribution

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling

Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action

Student's t-distribution

Critical values for t-distributions

Central limit theorem—t version

- Drawback to CLT: need to know population standard deviation, $SD(Y)$, to use it.
- Since we rarely know $SD(Y)$, what would be a good estimate to use instead? The sample s.d., s_Y .
- Replacing $SD(Y)$ with s_Y requires use of a t-distribution rather than the normal:
 - t-distribution is like normal but more spread out (fatter tails) to reflect additional uncertainty;
 - additional uncertainty is due to using s_Y instead of assuming we know $SD(Y)$;
 - s_Y is a better estimate of $SD(Y)$ for large n ;
 - t-distribution accounts for this using *degrees of freedom* ($df = n - 1$ in this case);
 - as df becomes large, t-distribution looks more and more like normal.

Critical values for t-distributions

upper-tail area	0.1	0.05	0.025	0.01	0.005	0.001
df = 3	1.638	2.353	3.182	4.541	5.841	10.215
df = 15	1.341	1.753	2.131	2.602	2.947	3.733
df = 29	1.311	1.699	2.045	2.462	2.756	3.396
df = 60	1.296	1.671	2.000	2.390	2.660	3.232
df = ∞ (normal)	1.282	1.645	1.960	2.326	2.576	3.090
two-tail area	0.2	0.1	0.05	0.02	0.01	0.002

- Horizontal axis values are called *critical values*.
- Tail areas (under the density curve) represent probabilities.
- Example: $\Pr(t_{29} > 1.699) = 0.05$.
- Note that critical values get closer to those for the normal as df gets larger.

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling
Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action
Student's t-distribution

Critical values for t-distributions

Central limit theorem—t version

Central limit theorem—t version

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling

Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action

Student's t-distribution

Critical values for t-distributions

Central limit theorem—t version

- Randomly sample Y_1, Y_2, \dots, Y_n from a population with mean, $E(Y)$.
- CLT: t-statistic = $\frac{m_Y - E(Y)}{s_Y / \sqrt{n}} \sim t_{n-1}$
(t-distribution with $n - 1$ df).
- Assume home prices Y_1, \dots, Y_{30} have $E(Y) = 280$.
- Sample standard deviation, s_Y , is 53.8656.
- What is the 95th percentile of m_Y ?

Central limit theorem—t version

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling

Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action

Student's t-distribution

Critical values for t-distributions

Central limit theorem—t version

- Randomly sample Y_1, Y_2, \dots, Y_n from a population with mean, $E(Y)$.

- CLT: t-statistic = $\frac{m_Y - E(Y)}{s_Y / \sqrt{n}} \sim t_{n-1}$
(t-distribution with $n - 1$ df).

- Assume home prices Y_1, \dots, Y_{30} have $E(Y) = 280$.

- Sample standard deviation, s_Y , is 53.8656.

- What is the 95th percentile of m_Y ?

- $$\Pr(t_{29} > 1.699) = 0.05$$

Central limit theorem—t version

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling

Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action

Student's t-distribution

Critical values for t-distributions

Central limit theorem—t version

- Randomly sample Y_1, Y_2, \dots, Y_n from a population with mean, $E(Y)$.
- CLT: t-statistic = $\frac{m_Y - E(Y)}{s_Y / \sqrt{n}} \sim t_{n-1}$ (t-distribution with $n - 1$ df).
- Assume home prices Y_1, \dots, Y_{30} have $E(Y) = 280$.
- Sample standard deviation, s_Y , is 53.8656.
- What is the 95th percentile of m_Y ?

$$\Pr(t_{29} > 1.699) = 0.05$$
$$\Pr\left(\frac{m_Y - 280}{53.8656 / \sqrt{30}} > 1.699\right) = 0.05$$

Central limit theorem—t version

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling

Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action

Student's t-distribution

Critical values for t-distributions

Central limit theorem—t version

- Randomly sample Y_1, Y_2, \dots, Y_n from a population with mean, $E(Y)$.

- CLT: t-statistic = $\frac{m_Y - E(Y)}{s_Y / \sqrt{n}} \sim t_{n-1}$
(t-distribution with $n - 1$ df).

- Assume home prices Y_1, \dots, Y_{30} have $E(Y) = 280$.

- Sample standard deviation, s_Y , is 53.8656.

- What is the 95th percentile of m_Y ?

- $$\Pr(t_{29} > 1.699) = 0.05$$

$$\Pr\left(\frac{m_Y - 280}{53.8656 / \sqrt{30}} > 1.699\right) = 0.05$$

$$\Pr(m_Y > 1.699(53.8656 / \sqrt{30}) + 280) = 0.05$$

Central limit theorem—t version

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling

Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action

Student's t-distribution

Critical values for t-distributions

Central limit theorem—t version

- Randomly sample Y_1, Y_2, \dots, Y_n from a population with mean, $E(Y)$.

- CLT: t-statistic = $\frac{m_Y - E(Y)}{s_Y / \sqrt{n}} \sim t_{n-1}$
(t-distribution with $n - 1$ df).

- Assume home prices Y_1, \dots, Y_{30} have $E(Y) = 280$.

- Sample standard deviation, s_Y , is 53.8656.

- What is the 95th percentile of m_Y ?

- $$\Pr(t_{29} > 1.699) = 0.05$$

$$\Pr\left(\frac{m_Y - 280}{53.8656 / \sqrt{30}} > 1.699\right) = 0.05$$

$$\Pr(m_Y > 1.699(53.8656 / \sqrt{30}) + 280) = 0.05$$

$$\Pr(m_Y > 297) = 0.05.$$

Central limit theorem—t version

1.1 Identifying and summarizing data

1.2 Population distributions

1.3 Selecting individuals at random—probability

1.4 Random sampling

Random sampling

Central limit theorem—normal version

Finding sampling distribution probabilities

The central limit theorem in action

Student's t-distribution

Critical values for t-distributions

Central limit theorem—t version

Central limit theorem—t version

Central limit theorem—t version

- Randomly sample Y_1, Y_2, \dots, Y_n from a population with mean, $E(Y)$.

- CLT: t-statistic = $\frac{m_Y - E(Y)}{s_Y / \sqrt{n}} \sim t_{n-1}$
(t-distribution with $n - 1$ df).

- Assume home prices Y_1, \dots, Y_{30} have $E(Y) = 280$.

- Sample standard deviation, s_Y , is 53.8656.

- What is the 95th percentile of m_Y ?

- $$\Pr(t_{29} > 1.699) = 0.05$$

$$\Pr\left(\frac{m_Y - 280}{53.8656 / \sqrt{30}} > 1.699\right) = 0.05$$

$$\Pr(m_Y > 1.699(53.8656 / \sqrt{30}) + 280) = 0.05$$

$$\Pr(m_Y > 297) = 0.05.$$

- What is the 90th percentile of m_Y ?