IFN501 - System Modeling and Simulation

Session 5: Introduction to Statistics (Part 2)

Daniel Febrian Sengkey

Department of Electrical Engineering
Faculty of Engineering
Universitas Sam Ratulangi

Outline

Outline

Outline

Outline

The Five-Number Summary

- As described in the previous course, there are 2 types of statistics.

The Five-Number Summary

- As described in the previous course, there are 2 types of statistics.
- Descriptive statistics is used to describe the data and giving insights about the data.

The Five-Number Summary

- As described in the previous course, there are 2 types of statistics.
- Descriptive statistics is used to describe the data and giving insights about the data.
- There are five numbers commonly used to summarize the data:

The Five-Number Summary

- As described in the previous course, there are 2 types of statistics.
- Descriptive statistics is used to describe the data and giving insights about the data.
- There are five numbers commonly used to summarize the data:
    - **Min**imum value

The Five-Number Summary

- As described in the previous course, there are 2 types of statistics.
- Descriptive statistics is used to describe the data and giving insights about the data.
- There are five numbers commonly used to summarize the data:
  - **Min**imum value
  - First quartile

The Five-Number Summary

- As described in the previous course, there are 2 types of statistics.
- Descriptive statistics is used to describe the data and giving insights about the data.
- There are five numbers commonly used to summarize the data:
  - **Min**imum value
  - First quartile
  - Median

The Five-Number Summary

- As described in the previous course, there are 2 types of statistics.
- Descriptive statistics is used to describe the data and giving insights about the data.
- There are five numbers commonly used to summarize the data:
  - **Min**imum value
  - First quartile
  - Median
  - Third Quartile

The Five-Number Summary

- As described in the previous course, there are 2 types of statistics.
- Descriptive statistics is used to describe the data and giving insights about the data.
- There are five numbers commonly used to summarize the data:
    - **Min**imum value
    - First quartile
    - Median
    - Third Quartile
    - **Max**imum value

The Five-Number Summary

Finding Min and Max in R

- We use our `IFC6503.A.2016` dataset as example.

- We use our `IFC6503.A.2016` dataset as example.
- To find the minimum value in the `Nilai_Akhir` data we use

The Five-Number Summary

Finding Min and Max in R

- We use our `IFC6503.A.2016` dataset as example.
- To find the minimum value in the `Nilai_Akhir` data we use

```
min(final.scores)

## [1] 0
```

The Five-Number Summary

Finding Min and Max in R

- We use our `IFC6503.A.2016` dataset as example.
- To find the minimum value in the `Nilai_Akhir` data we use

```
min(final.scores)

## [1] 0
```

- To find the maximum value in the `Nilai_Akhir` data we use

The Five-Number Summary

Finding Min and Max in R

- We use our `IFC6503.A.2016` dataset as example.
- To find the minimum value in the `Nilai_Akhir` data we use

```r
min(final.scores)

## [1] 0
```

- To find the maximum value in the `Nilai_Akhir` data we use

```r
max(final.scores)

## [1] 96.7
```

- To find the median in the `Nilai_Akhir` data we use

[1] Denoted by μ for population data; $\bar{x}$ for sample data.

The Five-Number Summary

Finding Median and Mean in R

- To find the median in the `Nilai_Akhir` data we use

```
median(final.scores)

## [1] 74.06
```

[1]Denoted by μ for population data; $\bar{x}$ for sample data.

The Five-Number Summary

Finding Median and Mean in R

- To find the median in the `Nilai_Akhir` data we use

```
median(final.scores)

## [1] 74.06
```

- To find the mean[1] of the `Nilai_Akhir` data we use

[1] Denoted by μ for population data; $\bar{x}$ for sample data.

The Five-Number Summary

Finding Median and Mean in R

- To find the median in the `Nilai_Akhir` data we use

```r
median(final.scores)

## [1] 74.06
```

- To find the mean[1] of the `Nilai_Akhir` data we use

```r
mean(final.scores)

## [1] 66.71
```

[1] Denoted by μ for population data; $\bar{x}$ for sample data.

The Five-Number Summary

Using `fivenum()` and `summary()` Function

- To check the five-number summary at once we can use

The Five-Number Summary

Using `fivenum()` and `summary()` Function

- To check the five-number summary at once we can use

```
fivenum(final.scores)

## [1]  0.00 56.55 74.06 85.85 96.70
```

The Five-Number Summary

Using `fivenum()` and `summary()` Function

- To check the five-number summary at once we can use

```r
fivenum(final.scores)
```

```
## [1]  0.00 56.55 74.06 85.85 96.70
```

- R has another function that includes mean in the report:

```r
summary(final.scores)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0    56.6    74.1    66.7    85.6    96.7
```

Outline

Outliers

- Outliers: value that are very small or very large relative to the majority of the values in a dataset [1].

Outliers

- Outliers: value that are very small or very large relative to the majority of the values in a dataset [1].
- Outliers also known as *Extreme Values*.

Outliers

- Outliers: value that are very small or very large relative to the majority of the values in a dataset [1].
- Outliers also known as *Extreme Values*.
- There are several method to identify outliers.

Outliers

- Outliers: value that are very small or very large relative to the majority of the values in a dataset [1].
- Outliers also known as *Extreme Values*.
- There are several method to identify outliers.
- In univariate methods, the outliers are the values outside $1.5 \times IQR$

Outliers

- Outliers: value that are very small or very large relative to the majority of the values in a dataset [1].
- Outliers also known as *Extreme Values*.
- There are several method to identify outliers.
- In univariate methods, the outliers are the values outside $1.5 \times IQR$
- In R, we can use the `boxplot.stats()` function to identify these values.

Outliers

Using The `boxplot.stats()` Function

```
boxplot.stats(final.scores)

## $stats
## [1] 18.10 56.55 74.06 85.85 96.70
##
## $n
## [1] 42
##
## $conf
## [1] 66.92 81.21
##
## $out
## [1] 0.00 0.00 5.16 5.00
```

Outliers

- The outliers are the values in the `$out` vector.

Outliers

- The outliers are the values in the `$out` vector.
- If we are only interested in the outliers, then we can use

Outliers

- The outliers are the values in the `$out` vector.
- If we are only interested in the outliers, then we can use

```
boxplot.stats(final.scores)$out

## [1] 0.00 0.00 5.16 5.00
```

Outline

Variance and Standard Deviation [1]

- Standard deviation[2] is the most-used measure of dispersion.

[2] Denoted by σ for population data; s for sample data.

Variance and Standard Deviation [1]

- Standard deviation[2] is the most-used measure of dispersion.
- Standard deviation is calculated by taking the positive square root of variance:

[2]Denoted by σ for population data; s for sample data.

Variance and Standard Deviation [1]

- Standard deviation[2] is the most-used measure of dispersion.
- Standard deviation is calculated by taking the positive square root of variance:
- The variance for population data is:

[2]Denoted by σ for population data; s for sample data.

Variance and Standard Deviation [1]

- Standard deviation[2] is the most-used measure of dispersion.
- Standard deviation is calculated by taking the positive square root of variance:
- The variance for population data is:

$$\sigma^2 = \frac{\Sigma(x-\mu)^2}{N} \tag{1}$$

[2]Denoted by σ for population data; s for sample data.

Variance and Standard Deviation [1]

- Standard deviation[2] is the most-used measure of dispersion.
- Standard deviation is calculated by taking the positive square root of variance:
- The variance for population data is:

$$\sigma^2 = \frac{\Sigma(x-\mu)^2}{N} \tag{1}$$

- and for sample data

[2]Denoted by σ for population data; s for sample data.

Variance and Standard Deviation [1]

- Standard deviation[2] is the most-used measure of dispersion.
- Standard deviation is calculated by taking the positive square root of variance:
- The variance for population data is:

$$\sigma^2 = \frac{\Sigma(x - \mu)^2}{N} \tag{1}$$

- and for sample data

$$s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1} \tag{2}$$

[2] Denoted by σ for population data; s for sample data.

- To calculate standard deviation in R we use:

- To calculate standard deviation in R we use:

```
sd(final.scores)

## [1] 26.88068439
```

- To calculate standard deviation in R we use:

```r
sd(final.scores)
```

```
## [1] 26.88068439
```

- And for variance

- To calculate standard deviation in R we use:

```
sd(final.scores)

## [1] 26.88068439
```

- And for variance

```
var(final.scores)

## [1] 722.5711934
```

Outline

Boxplot

- Boxplot (or the box-whisker plot) is a graph that shows the data summary based on the five-number statistics and the unusual observations [2].

Boxplot

- Boxplot (or the box-whisker plot) is a graph that shows the data summary based on the five-number statistics and the unusual observations [2].
- To create a boxplot in R, we use the `boxplot()` function.

Boxplot

- Boxplot (or the box-whisker plot) is a graph that shows the data summary based on the five-number statistics and the unusual observations [2].
- To create a boxplot in R, we use the `boxplot()` function.

```r
boxplot(final.scores)
```

Boxplot

- Boxplot (or the box-whisker plot) is a graph that shows the data summary based on the five-number statistics and the unusual observations [2].
- To create a boxplot in R, we use the `boxplot()` function.

```r
boxplot(final.scores)
```

- The resulting plot is shown at Figure 1.

Boxplot

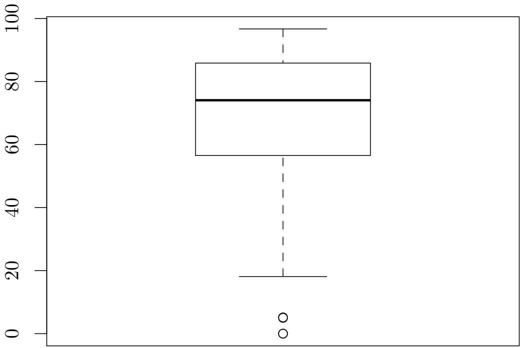

Figure 1 : Boxplot for the final scores data.

Boxplot and Stripchart

- We can have all the data points plotted by using `stripchart()`.

Boxplot and Stripchart

- We can have all the data points plotted by using `stripchart()`.
- Stripchart is a 1-D scatter plot.

Boxplot and Stripchart

- We can have all the data points plotted by using `stripchart()`.
- Stripchart is a 1-D scatter plot.
- In R, commonly we can put a plot above another plot by passing parameter `add=T` to the function.

Boxplot and Stripchart

- We can have all the data points plotted by using `stripchart()`.
- Stripchart is a 1-D scatter plot.
- In R, commonly we can put a plot above another plot by passing parameter `add=T` to the function.
- So, to have a stripchart over our plotted boxplot, the command is:

Boxplot and Stripchart

- We can have all the data points plotted by using `stripchart()`.
- Stripchart is a 1-D scatter plot.
- In R, commonly we can put a plot above another plot by passing parameter `add=T` to the function.
- So, to have a stripchart over our plotted boxplot, the command is:

```
stripchart(
        final.scores,
        vertical = T,
        col='orange',
        at=0.75,
        add=T
        )
```

Boxplot and Stripchart

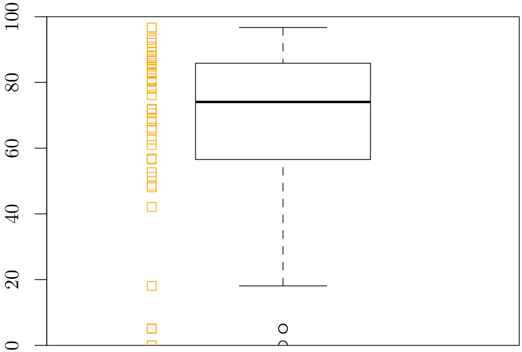

Figure 2 : Boxplot and stripchart for the final scores data.

Boxplot and Histogram

- The shape of a boxplot has the same trend with the histogram for the same data.

Boxplot and Histogram

- The shape of a boxplot has the same trend with the histogram for the same data.
- We can use the `layout()` function to have both boxplot and histogram in a single plot.

Boxplot and Histogram

- The shape of a boxplot has the same trend with the histogram for the same data.
- We can use the `layout()` function to have both boxplot and histogram in a single plot.

```
nf <- layout(mat = matrix(c(1, 2), 2, 1, byrow = T), height = c(1,
    2))
old.mar <- par(mar = c(4, 4, 0, 1) + 0.1)
boxplot(final.scores, horizontal = T, outline = T, frame = F,
    col = "green1")
hist(final.scores, col = "pink", main = NULL)
par(old.mar)
```

Boxplot and Histogram

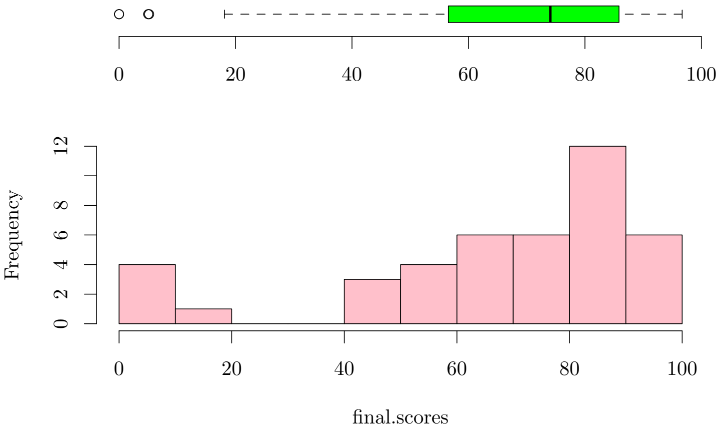

Figure 3 : Boxplot and histogram for the final scores data.

Boxplot, Histogram and Stripchart

And if we want the stripchart over the histogram:

```
nf <- layout(mat = matrix(c(1, 2), 2, 1, byrow = T), height = c(1,
    2))
old.mar <- par(mar = c(4, 4, 0, 1) + 0.1)
boxplot(final.scores, horizontal = T, outline = T, frame = F,
    col = "green1")
hist(final.scores, col = "pink", main = NULL, xlab = "Final Scores")
stripchart(final.scores, col = "purple", pch = 19, method = "stack",
    add = T)
par(old.mar)
```

Boxplot, Histogram and Stripchart

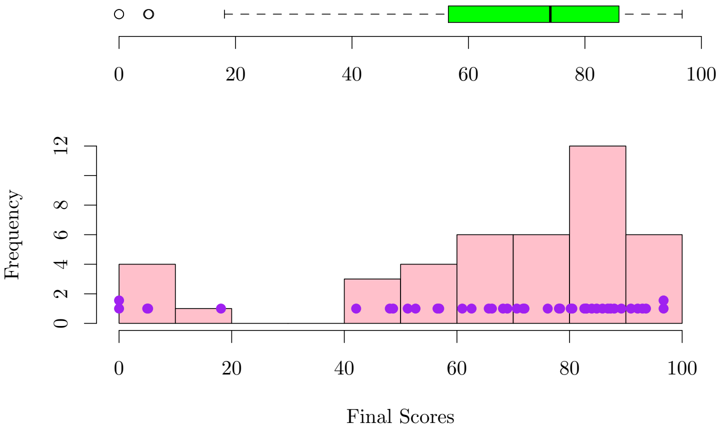

Figure 4 : Boxplot, histogram, and stripchart for the final scores data.

Outline

Exercise 3

1. Load the `IFC6510` dataset and:

Exercise 3

1. Load the `IFC6510` dataset and:
   - Explore the five-number summary from the `Nilai_Akhir` data.

Exercise 3

1. Load the `IFC6510` dataset and:
   - Explore the five-number summary from the `Nilai_Akhir` data.
   - Create boxplot from the `Nilai_Akhir` data.

Exercise 3

1. Load the `IFC6510` dataset and:
   - Explore the five-number summary from the `Nilai_Akhir` data.
   - Create boxplot from the `Nilai_Akhir` data.
   - Superimpose the boxplot with stripchart.

Exercise 3

1. Load the `IFC6510` dataset and:
   - Explore the five-number summary from the `Nilai_Akhir` data.
   - Create boxplot from the `Nilai_Akhir` data.
   - Superimpose the boxplot with stripchart.
   - Create the histogram, stacked with boxplot, superimposed with stripchart using the `Nilai_Akhir` data.

Exercise 3

1. Load the `IFC6510` dataset and:
   - Explore the five-number summary from the `Nilai_Akhir` data.
   - Create boxplot from the `Nilai_Akhir` data.
   - Superimpose the boxplot with stripchart.
   - Create the histogram, stacked with boxplot, superimposed with stripchart using the `Nilai_Akhir` data.
2. Learn how to put multiple graphs in a single plot.

Exercise 3

1. Load the `IFC6510` dataset and:
   - Explore the five-number summary from the `Nilai_Akhir` data.
   - Create boxplot from the `Nilai_Akhir` data.
   - Superimpose the boxplot with stripchart.
   - Create the histogram, stacked with boxplot, superimposed with stripchart using the `Nilai_Akhir` data.
2. Learn how to put multiple graphs in a single plot.
   - Create superimposed histogram by using `Nilai_Akhir` data from `IFC6510` and `IFC6503.A.2016` datasets.

Exercise 3

1. Load the `IFC6510` dataset and:
   - Explore the five-number summary from the `Nilai_Akhir` data.
   - Create boxplot from the `Nilai_Akhir` data.
   - Superimpose the boxplot with stripchart.
   - Create the histogram, stacked with boxplot, superimposed with stripchart using the `Nilai_Akhir` data.
2. Learn how to put multiple graphs in a single plot.
   - Create superimposed histogram by using `Nilai_Akhir` data from `IFC6510` and `IFC6503.A.2016` datasets.
   - Create side-by-side boxplots from both datasets.

Exercise 3

1. Load the `IFC6510` dataset and:
   - Explore the five-number summary from the `Nilai_Akhir` data.
   - Create boxplot from the `Nilai_Akhir` data.
   - Superimpose the boxplot with stripchart.
   - Create the histogram, stacked with boxplot, superimposed with stripchart using the `Nilai_Akhir` data.
2. Learn how to put multiple graphs in a single plot.
   - Create superimposed histogram by using `Nilai_Akhir` data from `IFC6510` and `IFC6503.A.2016` datasets.
   - Create side-by-side boxplots from both datasets.
   - Create the histogram, stacked with boxplot, superimposed with stripchart from both dataset in a single plot.

Exercise 3

1. Load the `IFC6510` dataset and:
   - Explore the five-number summary from the `Nilai_Akhir` data.
   - Create boxplot from the `Nilai_Akhir` data.
   - Superimpose the boxplot with stripchart.
   - Create the histogram, stacked with boxplot, superimposed with stripchart using the `Nilai_Akhir` data.
2. Learn how to put multiple graphs in a single plot.
   - Create superimposed histogram by using `Nilai_Akhir` data from `IFC6510` and `IFC6503.A.2016` datasets.
   - Create side-by-side boxplots from both datasets.
   - Create the histogram, stacked with boxplot, superimposed with stripchart from both dataset in a single plot.
   - Compare the five-number summary from both datasets.

Exercise 3

1. Load the `IFC6510` dataset and:
   - Explore the five-number summary from the `Nilai_Akhir` data.
   - Create boxplot from the `Nilai_Akhir` data.
   - Superimpose the boxplot with stripchart.
   - Create the histogram, stacked with boxplot, superimposed with stripchart using the `Nilai_Akhir` data.
2. Learn how to put multiple graphs in a single plot.
   - Create superimposed histogram by using `Nilai_Akhir` data from `IFC6510` and `IFC6503.A.2016` datasets.
   - Create side-by-side boxplots from both datasets.
   - Create the histogram, stacked with boxplot, superimposed with stripchart from both dataset in a single plot.
   - Compare the five-number summary from both datasets.
3. Load the `10000rep.RData`

Exercise 3

1. Load the `IFC6510` dataset and:
   - Explore the five-number summary from the `Nilai_Akhir` data.
   - Create boxplot from the `Nilai_Akhir` data.
   - Superimpose the boxplot with stripchart.
   - Create the histogram, stacked with boxplot, superimposed with stripchart using the `Nilai_Akhir` data.
2. Learn how to put multiple graphs in a single plot.
   - Create superimposed histogram by using `Nilai_Akhir` data from `IFC6510` and `IFC6503.A.2016` datasets.
   - Create side-by-side boxplots from both datasets.
   - Create the histogram, stacked with boxplot, superimposed with stripchart from both dataset in a single plot.
   - Compare the five-number summary from both datasets.
3. Load the `10000rep.RData`
   - This dataset was produced from an experiment in OMNeT++ with 10,000 repetitions for each method.

Exercise 3

1. Load the `IFC6510` dataset and:
   - Explore the five-number summary from the `Nilai_Akhir` data.
   - Create boxplot from the `Nilai_Akhir` data.
   - Superimpose the boxplot with stripchart.
   - Create the histogram, stacked with boxplot, superimposed with stripchart using the `Nilai_Akhir` data.
2. Learn how to put multiple graphs in a single plot.
   - Create superimposed histogram by using `Nilai_Akhir` data from `IFC6510` and `IFC6503.A.2016` datasets.
   - Create side-by-side boxplots from both datasets.
   - Create the histogram, stacked with boxplot, superimposed with stripchart from both dataset in a single plot.
   - Compare the five-number summary from both datasets.
3. Load the `10000rep.RData`
   - This dataset was produced from an experiment in OMNeT++ with 10,000 repetitions for each method.
   - There are 2 vectors, `hopCount.normal` and `hopCount.nsb`.

Exercise 3

1. Load the `IFC6510` dataset and:
   - Explore the five-number summary from the `Nilai_Akhir` data.
   - Create boxplot from the `Nilai_Akhir` data.
   - Superimpose the boxplot with stripchart.
   - Create the histogram, stacked with boxplot, superimposed with stripchart using the `Nilai_Akhir` data.
2. Learn how to put multiple graphs in a single plot.
   - Create superimposed histogram by using `Nilai_Akhir` data from `IFC6510` and `IFC6503.A.2016` datasets.
   - Create side-by-side boxplots from both datasets.
   - Create the histogram, stacked with boxplot, superimposed with stripchart from both dataset in a single plot.
   - Compare the five-number summary from both datasets.
3. Load the `10000rep.RData`
   - This dataset was produced from an experiment in OMNeT++ with 10,000 repetitions for each method.
   - There are 2 vectors, `hopCount.normal` and `hopCount.nsb`.
   - Repeat task number 2 with these data.

Outline

Outline

Probability [2]

- Probability is a foundation for statistics.

Probability [2]

- Probability is a foundation for statistics.
- Several cases of probability:

Probability [2]

- Probability is a foundation for statistics.
- Several cases of probability:
  1. A “die”, the singular of dice, is a cube with six faces numbered 1, 2, 3, 4, 5, and 6. **What is the chance of getting 1 when rolling a die?**

Probability [2]

- Probability is a foundation for statistics.
- Several cases of probability:
  1. A “die”, the singular of dice, is a cube with six faces numbered 1, 2, 3, 4, 5, and 6. **What is the chance of getting 1 when rolling a die?**
     If the die is fair, then the chance of a 1 is as good as the chance of any other number. Since there are six outcomes, the chance must be 1-in-6 or, equivalently, $1/6$.

Probability [2]

- Probability is a foundation for statistics.
- Several cases of probability:
  1. A “die”, the singular of dice, is a cube with six faces numbered 1, 2, 3, 4, 5, and 6. **What is the chance of getting 1 when rolling a die?**
     If the die is fair, then the chance of a 1 is as good as the chance of any other number. Since there are six outcomes, the chance must be 1-in-6 or, equivalently, $1/6$.
  2. **What is the chance of getting a 1 or 2 in the next roll?**

Probability [2]

- Probability is a foundation for statistics.
- Several cases of probability:
    1. A “die”, the singular of dice, is a cube with six faces numbered 1, 2, 3, 4, 5, and 6. **What is the chance of getting 1 when rolling a die?**
       If the die is fair, then the chance of a 1 is as good as the chance of any other number. Since there are six outcomes, the chance must be 1-in-6 or, equivalently, $1/6$.
    2. **What is the chance of getting a 1 or 2 in the next roll?**
       1 and 2 constitute two of the six equally likely possible outcomes, so the chance of getting one of these two outcomes must be $2/6 = 1/3$.

Probability [2]

- Probability is a foundation for statistics.
- Several cases of probability:
  1. A "die", the singular of dice, is a cube with six faces numbered 1, 2, 3, 4, 5, and 6. **What is the chance of getting 1 when rolling a die?**
     If the die is fair, then the chance of a 1 is as good as the chance of any other number. Since there are six outcomes, the chance must be 1-in-6 or, equivalently, $1/6$.
  2. **What is the chance of getting a 1 or 2 in the next roll?**
     1 and 2 constitute two of the six equally likely possible outcomes, so the chance of getting one of these two outcomes must be $2/6 = 1/3$.
  3. Consider rolling two dice. If $1/6^{th}$ of the time the first die is a 1 and $1/6^{th}$ of those times the second die is a 1, **what is the chance of getting two 1s?**

Probability [2]

- Probability is a foundation for statistics.
- Several cases of probability:
  1. A "die", the singular of dice, is a cube with six faces numbered 1, 2, 3, 4, 5, and 6. **What is the chance of getting 1 when rolling a die?**
     If the die is fair, then the chance of a 1 is as good as the chance of any other number. Since there are six outcomes, the chance must be 1-in-6 or, equivalently, $1/6$.
  2. **What is the chance of getting a 1 or 2 in the next roll?**
     1 and 2 constitute two of the six equally likely possible outcomes, so the chance of getting one of these two outcomes must be $2/6 = 1/3$.
  3. Consider rolling two dice. If $1/6^{th}$ of the time the first die is a 1 and $1/6^{th}$ of those times the second die is a 1, **what is the chance of getting two 1s?**
     If $16.\bar{6}\%$ of the time the first die is a 1 and $1/6^{th}$ of <u>those</u> times the second die is also a 1, then the chance that both dice are 1 is $(1/6) \times (1/6)$ or $1/36$.

Probability [2]

Probability

The **probability** of an outcome is the proportion of times the outcome would occur if we observed the random process an infinite number of times.

- We use probability to build tools to describe and understand apparent randomness.

Probability [2]

Probability

The **probability** of an outcome is the proportion of times the outcome would occur if we observed the random process an infinite number of times.

- We use probability to build tools to describe and understand apparent randomness.
- Rolling a die or flipping a coin is a seemingly random process and each gives rise to an outcome.

Probability [2]

Probability

The **probability** of an outcome is the proportion of times the outcome would occur if we observed the random process an infinite number of times.

- We use probability to build tools to describe and understand apparent randomness.
- Rolling a die or flipping a coin is a seemingly random process and each gives rise to an outcome.
- Probability is defined as a proportion, and it always takes values between 0 and 1 (inclusively). It may also be displayed as a percentage between 0% and 100%.

The Law of Large Numbers [2]

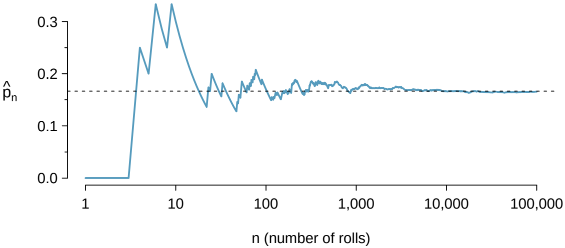

Figure 5 : The fraction of die rolls that are 1 at each stage in a simulation. The proportion tends to get closer to the probability $1/6 \approx 0.167$ as the number of rolls increases.

The Law of Large Numbers

As more observations are collected, the proportion $\hat{p}_n$ of occurrences with a particular outcome converges to the probability p of that outcome.

Probability [2]

- Above we write p as the probability of rolling a 1. We can also write this probability as

$$P(\text{rolling a } 1) \tag{3}$$

Probability [2]

- Above we write p as the probability of rolling a 1. We can also write this probability as

$$P(\text{rolling a } 1) \tag{3}$$

- As we become more comfortable with this notation, we will abbreviate it further.

Probability [2]

- Above we write p as the probability of rolling a 1. We can also write this probability as

$$P(\text{rolling a } 1) \tag{3}$$

- As we become more comfortable with this notation, we will abbreviate it further.
- For instance, if it is clear that the process is “rolling a die”, we could abbreviate $P(\text{rolling a } 1)$ as $P(1)$.

Outline

Disjoint or Mutually Exclusive Outcomes [2]

- Two outcomes are called **disjoint** or **mutually exclusive** if they cannot both happen.

Disjoint or Mutually Exclusive Outcomes [2]

- Two outcomes are called **disjoint** or **mutually exclusive** if they cannot both happen.
- For instance

Disjoint or Mutually Exclusive Outcomes [2]

- Two outcomes are called **disjoint** or **mutually exclusive** if they cannot both happen.
- For instance
  - we roll a die, the outcomes 1 and 2 are disjoint since they cannot both occur.

Disjoint or Mutually Exclusive Outcomes [2]

- Two outcomes are called **disjoint** or **mutually exclusive** if they cannot both happen.
- For instance
  - we roll a die, the outcomes 1 and 2 are disjoint since they cannot both occur.
  - On the other hand, the outcomes 1 and "rolling an odd number" are not disjoint since both occur if the outcome of the roll is a 1.

Disjoint or Mutually Exclusive Outcomes [2]

- Two outcomes are called **disjoint** or **mutually exclusive** if they cannot both happen.
- For instance
  - we roll a die, the outcomes 1 and 2 are disjoint since they cannot both occur.
  - On the other hand, the outcomes 1 and “rolling an odd number” are not disjoint since both occur if the outcome of the roll is a 1.
- The terms <u>disjoint</u> and <u>mutually exclusive</u> are equivalent and interchangeable.

Disjoint or Mutually Exclusive Outcomes [2]

Rolling a Die

- Calculating the probability of disjoint outcomes is easy.

Disjoint or Mutually Exclusive Outcomes [2]

Rolling a Die

- Calculating the probability of disjoint outcomes is easy.
- When rolling a die, the outcomes 1 and 2 are disjoint, and we compute the probability that one of these outcomes will occur by adding their separate probabilities:

Disjoint or Mutually Exclusive Outcomes [2]

Rolling a Die

- Calculating the probability of disjoint outcomes is easy.
- When rolling a die, the outcomes 1 and 2 are disjoint, and we compute the probability that one of these outcomes will occur by adding their separate probabilities:

$$P(1 \text{ or } 2) = P(1) + P(2) = 1/6 + 1/6 = 1/3 \tag{4}$$

Disjoint or Mutually Exclusive Outcomes [2]

Rolling a Die

- Calculating the probability of disjoint outcomes is easy.
- When rolling a die, the outcomes 1 and 2 are disjoint, and we compute the probability that one of these outcomes will occur by adding their separate probabilities:

$$P(1 \text{ or } 2) = P(1) + P(2) = 1/6 + 1/6 = 1/3 \tag{4}$$

- What about the probability of rolling a 1, 2, 3, 4, 5, or 6?

Disjoint or Mutually Exclusive Outcomes [2]

Rolling a Die

- Calculating the probability of disjoint outcomes is easy.
- When rolling a die, the outcomes 1 and 2 are disjoint, and we compute the probability that one of these outcomes will occur by adding their separate probabilities:

$$P(1 \text{ or } 2) = P(1) + P(2) = 1/6 + 1/6 = 1/3 \tag{4}$$

- What about the probability of rolling a 1, 2, 3, 4, 5, or 6? Here again, all of the outcomes are disjoint so we add the probabilities:

Disjoint or Mutually Exclusive Outcomes [2]

Rolling a Die

- Calculating the probability of disjoint outcomes is easy.
- When rolling a die, the outcomes 1 and 2 are disjoint, and we compute the probability that one of these outcomes will occur by adding their separate probabilities:

$$P(1 \text{ or } 2) = P(1) + P(2) = 1/6 + 1/6 = 1/3 \tag{4}$$

- What about the probability of rolling a 1, 2, 3, 4, 5, or 6? Here again, all of the outcomes are disjoint so we add the probabilities:

$$P(1 \text{ or } 2 \text{ or } 3 \text{ or } 4 \text{ or } 5 \text{ or } 6)$$

Disjoint or Mutually Exclusive Outcomes [2]

Rolling a Die

- Calculating the probability of disjoint outcomes is easy.
- When rolling a die, the outcomes 1 and 2 are disjoint, and we compute the probability that one of these outcomes will occur by adding their separate probabilities:

$$P(1 \text{ or } 2) = P(1) + P(2) = 1/6 + 1/6 = 1/3 \tag{4}$$

- What about the probability of rolling a 1, 2, 3, 4, 5, or 6? Here again, all of the outcomes are disjoint so we add the probabilities:

$$\begin{aligned} & P(1 \text{ or } 2 \text{ or } 3 \text{ or } 4 \text{ or } 5 \text{ or } 6) \\ & \quad = P(1) + P(2) + P(3) + P(4) + P(5) + P(6) \end{aligned}$$

Disjoint or Mutually Exclusive Outcomes [2]

Rolling a Die

- Calculating the probability of disjoint outcomes is easy.
- When rolling a die, the outcomes 1 and 2 are disjoint, and we compute the probability that one of these outcomes will occur by adding their separate probabilities:

$$P(1 \text{ or } 2) = P(1) + P(2) = 1/6 + 1/6 = 1/3 \tag{4}$$

- What about the probability of rolling a 1, 2, 3, 4, 5, or 6? Here again, all of the outcomes are disjoint so we add the probabilities:

$$\begin{aligned} & P(1 \text{ or } 2 \text{ or } 3 \text{ or } 4 \text{ or } 5 \text{ or } 6) \\ & \quad = P(1) + P(2) + P(3) + P(4) + P(5) + P(6) \\ & \quad = 1/6 + 1/6 + 1/6 + 1/6 + 1/6 + 1/6 \end{aligned}$$

Disjoint or Mutually Exclusive Outcomes [2]

Rolling a Die

- Calculating the probability of disjoint outcomes is easy.
- When rolling a die, the outcomes 1 and 2 are disjoint, and we compute the probability that one of these outcomes will occur by adding their separate probabilities:

$$P(1 \text{ or } 2) = P(1) + P(2) = 1/6 + 1/6 = 1/3 \tag{4}$$

- What about the probability of rolling a 1, 2, 3, 4, 5, or 6? Here again, all of the outcomes are disjoint so we add the probabilities:

$$\begin{aligned} &P(1 \text{ or } 2 \text{ or } 3 \text{ or } 4 \text{ or } 5 \text{ or } 6) \\ &\quad = P(1) + P(2) + P(3) + P(4) + P(5) + P(6) \\ &\quad = 1/6 + 1/6 + 1/6 + 1/6 + 1/6 + 1/6 \\ &\quad = 1 \end{aligned}$$

Disjoint or Mutually Exclusive Outcomes [2]

Rolling a Die

- Calculating the probability of disjoint outcomes is easy.
- When rolling a die, the outcomes 1 and 2 are disjoint, and we compute the probability that one of these outcomes will occur by adding their separate probabilities:

$$P(1 \text{ or } 2) = P(1) + P(2) = 1/6 + 1/6 = 1/3 \tag{4}$$

- What about the probability of rolling a 1, 2, 3, 4, 5, or 6? Here again, all of the outcomes are disjoint so we add the probabilities:

$$\begin{aligned} &P(1 \text{ or } 2 \text{ or } 3 \text{ or } 4 \text{ or } 5 \text{ or } 6) \\ &\quad = P(1) + P(2) + P(3) + P(4) + P(5) + P(6) \\ &\quad = 1/6 + 1/6 + 1/6 + 1/6 + 1/6 + 1/6 \\ &\quad = 1 \end{aligned}$$

- The **Addition Rule** guarantees the accuracy of this approach when the outcomes are disjoint.

Disjoint or Mutually Exclusive Outcomes [2]

Addition Rule

Addition Rule of Disjoint Outcomes

If A_1 and A_2 represent two disjoint outcomes, then the probability that one of them occurs is given by

$$P(A_1 \text{ or } A_2) = P(A_1) + P(A_2) \tag{5}$$

Disjoint or Mutually Exclusive Outcomes [2]

Addition Rule

Addition Rule of Disjoint Outcomes

If A_1 and A_2 represent two disjoint outcomes, then the probability that one of them occurs is given by

$$P(A_1 \text{ or } A_2) = P(A_1) + P(A_2) \tag{5}$$

If there are many disjoint outcomes A_1, ..., A_k, then the probability that one of these outcomes will occur is

$$P(A_1) + P(A_2) + \cdots + P(A_k) \tag{6}$$

Outline

Probabilities when Events are Not Disjoint [2]

Cards in a Deck

Table 1 : Representations of the 52 unique cards in a deck.

2♣	3♣	4♣	5♣	6♣	7♣	8♣	9♣	10♣	J♣	Q♣	K♣	A♣
2♢	3♢	4♢	5♢	6♢	7♢	8♢	9♢	10♢	J♢	Q♢	K♢	A♢
2♡	3♡	4♡	5♡	6♡	7♡	8♡	9♡	10♡	J♡	Q♡	K♡	A♡
2♠	3♠	4♠	5♠	6♠	7♠	8♠	9♠	10♠	J♠	Q♠	K♠	A♠

As exercise, consider

- What is the probability that a randomly selected card is a diamond?

Probabilities when Events are Not Disjoint [2]

Cards in a Deck

Table 1 : Representations of the 52 unique cards in a deck.

2♣	3♣	4♣	5♣	6♣	7♣	8♣	9♣	10♣	J♣	Q♣	K♣	A♣
2♢	3♢	4♢	5♢	6♢	7♢	8♢	9♢	10♢	J♢	Q♢	K♢	A♢
2♡	3♡	4♡	5♡	6♡	7♡	8♡	9♡	10♡	J♡	Q♡	K♡	A♡
2♠	3♠	4♠	5♠	6♠	7♠	8♠	9♠	10♠	J♠	Q♠	K♠	A♠

As exercise, consider

- What is the probability that a randomly selected card is a diamond?
- What is the probability that a randomly selected card is a face card?

Probabilities when Events are Not Disjoint [2]

Cards in a Deck - Venn Diagram

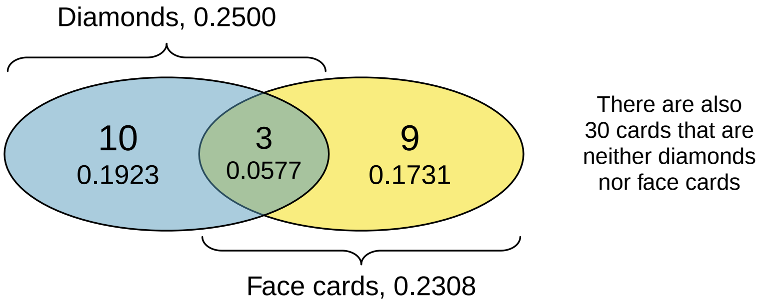

Figure 6 : A Venn diagram for diamonds and face cards.

Disjoint or Mutually Exclusive Outcomes [2]

- Let A represent the event that a randomly selected card is a diamond and B represent the event that it is a face card.

Disjoint or Mutually Exclusive Outcomes [2]

- Let A represent the event that a randomly selected card is a diamond and B represent the event that it is a face card.
- We could not use the addition rule for disjoint events since we there are several cards that fall in both categories[3].

Disjoint or Mutually Exclusive Outcomes [2]

- Let A represent the event that a randomly selected card is a diamond and B represent the event that it is a face card.
- We could not use the addition rule for disjoint events since we there are several cards that fall in both categories[3].

[3] There are three cards that are in both events were counted twice, once in each probability.

Disjoint or Mutually Exclusive Outcomes [2]

- Let A represent the event that a randomly selected card is a diamond and B represent the event that it is a face card.
- We could not use the addition rule for disjoint events since we there are several cards that fall in both categories[3].

$$P(A \text{ or } B) = P(\diamondsuit \text{ or face card})$$

[3] There are three cards that are in both events were counted twice, once in each probability.

Disjoint or Mutually Exclusive Outcomes [2]

- Let A represent the event that a randomly selected card is a diamond and B represent the event that it is a face card.
- We could not use the addition rule for disjoint events since we there are several cards that fall in both categories[3].

$$
\begin{aligned}
P(A \text{ or } B) &= P(\diamondsuit \text{ or face card}) \\
&= P(\diamondsuit) + P(\text{face card}) - P(\diamondsuit \text{ and face card}) \quad (7)
\end{aligned}
$$

[3]There are three cards that are in both events were counted twice, once in each probability.

Disjoint or Mutually Exclusive Outcomes [2]

- Let A represent the event that a randomly selected card is a diamond and B represent the event that it is a face card.
- We could not use the addition rule for disjoint events since we there are several cards that fall in both categories[3].

$$\begin{aligned} P(A \text{ or } B) &= P(\diamondsuit \text{ or face card}) \\ &= P(\diamondsuit) + P(\text{face card}) - P(\diamondsuit \text{ and face card}) \quad (7) \\ &= 13/52 + 12/52 - 3/52 \end{aligned}$$

[3] There are three cards that are in both events were counted twice, once in each probability.

Disjoint or Mutually Exclusive Outcomes [2]

- Let A represent the event that a randomly selected card is a diamond and B represent the event that it is a face card.
- We could not use the addition rule for disjoint events since we there are several cards that fall in both categories[3].

$$\begin{aligned} P(A \text{ or } B) &= P(\diamondsuit \text{ or face card}) \\ &= P(\diamondsuit) + P(\text{face card}) - P(\diamondsuit \text{ and face card}) \quad (7) \\ &= 13/52 + 12/52 - 3/52 \\ &= 22/52 = 11/26 \end{aligned}$$

[3] There are three cards that are in both events were counted twice, once in each probability.

General Addition Rule

If A and B are any two events, disjoint or not, then the probability that at least one of them will occur is

General Addition Rule

If A and B are any two events, disjoint or not, then the probability that at least one of them will occur is

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) \tag{8}$$

General Addition Rule

If A and B are any two events, disjoint or not, then the probability that at least one of them will occur is

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) \tag{8}$$

where $P(A \text{ and } B)$ is the probability that both events occur.

Outline

Probability Distributions [2]

Table 2 : Probability distribution for the sum of two dice.

Dice sum	2	3	4	5	6	7	8	9	10	11	12
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

- **Probability distribution** is a table of all disjoint outcomes and their associated probabilities.

Probability Distributions [2]

Table 2 : Probability distribution for the sum of two dice.

Dice sum	2	3	4	5	6	7	8	9	10	11	12
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

- **Probability distribution** is a table of all disjoint outcomes and their associated probabilities.
- Table 2 shows the probability distribution for the sum of two dice.

Probability Distributions [2]

Table 2 : Probability distribution for the sum of two dice.

Dice sum	2	3	4	5	6	7	8	9	10	11	12
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

- **Probability distribution** is a table of all disjoint outcomes and their associated probabilities.
- Table 2 shows the probability distribution for the sum of two dice.
- Rules for probability distributions:

Probability Distributions [2]

Table 2 : Probability distribution for the sum of two dice.

Dice sum	2	3	4	5	6	7	8	9	10	11	12
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

- **Probability distribution** is a table of all disjoint outcomes and their associated probabilities.
- Table 2 shows the probability distribution for the sum of two dice.
- Rules for probability distributions:
    1. The outcomes listed must be disjoint.

Probability Distributions [2]

Table 2 : Probability distribution for the sum of two dice.

Dice sum	2	3	4	5	6	7	8	9	10	11	12
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

- **Probability distribution** is a table of all disjoint outcomes and their associated probabilities.
- Table 2 shows the probability distribution for the sum of two dice.
- Rules for probability distributions:
  1. The outcomes listed must be disjoint.
  2. Each probability must be between 0 and 1.

Probability Distributions [2]

Table 2 : Probability distribution for the sum of two dice.

Dice sum	2	3	4	5	6	7	8	9	10	11	12
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

- **Probability distribution** is a table of all disjoint outcomes and their associated probabilities.
- Table 2 shows the probability distribution for the sum of two dice.
- Rules for probability distributions:
  1. The outcomes listed must be disjoint.
  2. Each probability must be between 0 and 1.
  3. The probabilities must total 1.

Table 3 : Proposed distributions of US household incomes.

Income range ($1000s)	0-25	25-50	50-100	100+
(a)	0.18	0.39	0.33	0.16
(b)	0.38	-0.27	0.52	0.37
(c)	0.28	0.27	0.29	0.16

- Table 3 suggests three distributions for household income in the United States.

Table 3 : Proposed distributions of US household incomes.

Income range ($1000s)	0-25	25-50	50-100	100+
(a)	0.18	0.39	0.33	0.16
(b)	0.38	-0.27	0.52	0.37
(c)	0.28	0.27	0.29	0.16

- Table 3 suggests three distributions for household income in the United States.
- Only one is correct. **Which one must it be?**

Table 3 : Proposed distributions of US household incomes.

Income range ($1000s)	0-25	25-50	50-100	100+
(a)	0.18	0.39	0.33	0.16
(b)	0.38	-0.27	0.52	0.37
(c)	0.28	0.27	0.29	0.16

- Table 3 suggests three distributions for household income in the United States.
- Only one is correct. **Which one must it be?**
- What is wrong with the other two?

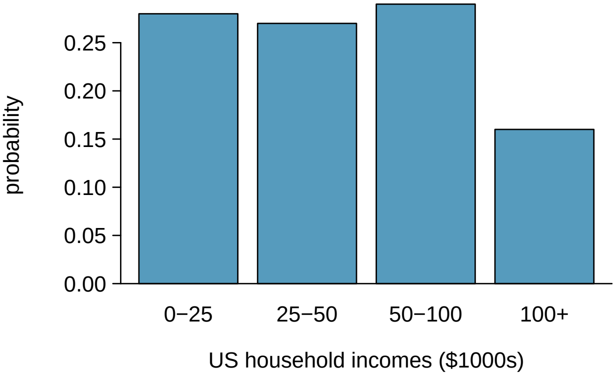

Figure 7 : The probability distribution of US household income.

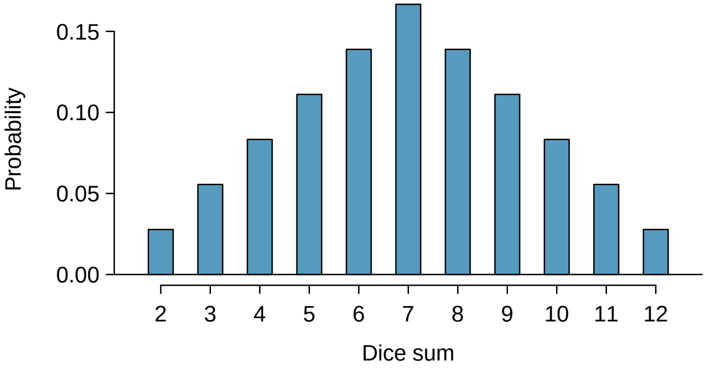

Figure 8 : The probability distribution of the sum of two dice (Table 2).

Outline

Random Variable [2]

Example

- **Case 1:**

Random Variable [2]

Example

- **Case 1:**
  - Two books are assigned for a statistics class: a textbook and its corresponding study guide.

Random Variable [2]

Example

- **Case 1:**
  - Two books are assigned for a statistics class: a textbook and its corresponding study guide.
  - The university bookstore determined 20% of enrolled students do not buy either book,

Random Variable [2]

Example

- **Case 1:**
  - Two books are assigned for a statistics class: a textbook and its corresponding study guide.
  - The university bookstore determined 20% of enrolled students do not buy either book, 55% buy the textbook only,

Random Variable [2]

Example

- **Case 1:**
  - Two books are assigned for a statistics class: a textbook and its corresponding study guide.
  - The university bookstore determined 20% of enrolled students do not buy either book, 55% buy the textbook only, and 25% buy both books,

Random Variable [2]

Example

- **Case 1:**
  - Two books are assigned for a statistics class: a textbook and its corresponding study guide.
  - The university bookstore determined 20% of enrolled students do not buy either book, 55% buy the textbook only, and 25% buy both books, and these percentages are relatively constant from one term to another.

Random Variable [2]

Example

- **Case 1:**
  - Two books are assigned for a statistics class: a textbook and its corresponding study guide.
  - The university bookstore determined 20% of enrolled students do not buy either book, 55% buy the textbook only, and 25% buy both books, and these percentages are relatively constant from one term to another.
  - If there are 100 students enrolled, how many books should the bookstore expect to sell to this class?

Random Variable [2]

Example

- **Case 1:**
  - Two books are assigned for a statistics class: a textbook and its corresponding study guide.
  - The university bookstore determined 20% of enrolled students do not buy either book, 55% buy the textbook only, and 25% buy both books, and these percentages are relatively constant from one term to another.
  - If there are 100 students enrolled, how many books should the bookstore expect to sell to this class?
- **Answer**

Random Variable [2]

Example

- **Case 1:**
  - Two books are assigned for a statistics class: a textbook and its corresponding study guide.
  - The university bookstore determined 20% of enrolled students do not buy either book, 55% buy the textbook only, and 25% buy both books, and these percentages are relatively constant from one term to another.
  - If there are 100 students enrolled, how many books should the bookstore expect to sell to this class?
- **Answer**
  - Around 20 students will not buy either book (0 books total),

Random Variable [2]

Example

- **Case 1:**
  - Two books are assigned for a statistics class: a textbook and its corresponding study guide.
  - The university bookstore determined 20% of enrolled students do not buy either book, 55% buy the textbook only, and 25% buy both books, and these percentages are relatively constant from one term to another.
  - If there are 100 students enrolled, how many books should the bookstore expect to sell to this class?
- **Answer**
  - Around 20 students will not buy either book (0 books total),
  - about 55 will buy one book (55 books total),

Random Variable [2]

Example

- **Case 1:**
  - Two books are assigned for a statistics class: a textbook and its corresponding study guide.
  - The university bookstore determined 20% of enrolled students do not buy either book, 55% buy the textbook only, and 25% buy both books, and these percentages are relatively constant from one term to another.
  - If there are 100 students enrolled, how many books should the bookstore expect to sell to this class?
- **Answer**
  - Around 20 students will not buy either book (0 books total),
  - about 55 will buy one book (55 books total),
  - and approximately 25 will buy two books (totaling 50 books for these 25 students).

Random Variable [2]

Example

- **Case 1:**
  - Two books are assigned for a statistics class: a textbook and its corresponding study guide.
  - The university bookstore determined 20% of enrolled students do not buy either book, 55% buy the textbook only, and 25% buy both books, and these percentages are relatively constant from one term to another.
  - If there are 100 students enrolled, how many books should the bookstore expect to sell to this class?
- **Answer**
  - Around 20 students will not buy either book (0 books total),
  - about 55 will buy one book (55 books total),
  - and approximately 25 will buy two books (totaling 50 books for these 25 students).
  - The bookstore should expect to sell about 105 books for this class.

Random Variable [2]

Example

- **Case 2:**

Random Variable [2]

Example

- **Case 2:**
  - The textbook costs \$137 and the study guide \$33.

Random Variable [2]

Example

- **Case 2:**
  - The textbook costs \$137 and the study guide \$33.
  - How much revenue should the bookstore expect from this class of 100 students?

Random Variable [2]

Example

- **Case 2:**
  - The textbook costs \$137 and the study guide \$33.
  - How much revenue should the bookstore expect from this class of 100 students?
- **Answer**

Random Variable [2]

Example

- **Case 2:**
  - The textbook costs \$137 and the study guide \$33.
  - How much revenue should the bookstore expect from this class of 100 students?
- **Answer**
  - About 55 students will just buy a textbook, providing revenue of

- **Case 2:**
  - The textbook costs \$137 and the study guide \$33.
  - How much revenue should the bookstore expect from this class of 100 students?
- **Answer**
  - About 55 students will just buy a textbook, providing revenue of

$$\$137 \times 55 = \$7,535$$

Random Variable [2]

Example

- **Case 2:**
  - The textbook costs \$137 and the study guide \$33.
  - How much revenue should the bookstore expect from this class of 100 students?
- **Answer**
  - About 55 students will just buy a textbook, providing revenue of

$$\$137 \times 55 = \$7,535$$

  - The roughly 25 students who buy both the textbook and the study guide would pay a total of

Random Variable [2]

Example

- **Case 2:**
  - The textbook costs \$137 and the study guide \$33.
  - How much revenue should the bookstore expect from this class of 100 students?
- **Answer**
  - About 55 students will just buy a textbook, providing revenue of

$$\$137 \times 55 = \$7,535$$

  - The roughly 25 students who buy both the textbook and the study guide would pay a total of

$$(\$137 + \$33) \times 25 = \$170 \times 25 = \$4,250$$

Random Variable [2]

Example

- **Case 2:**
  - The textbook costs \$137 and the study guide \$33.
  - How much revenue should the bookstore expect from this class of 100 students?
- **Answer**
  - About 55 students will just buy a textbook, providing revenue of

$$\$137 \times 55 = \$7,535$$

  - The roughly 25 students who buy both the textbook and the study guide would pay a total of

$$(\$137 + \$33) \times 25 = \$170 \times 25 = \$4,250$$

  - Thus, the bookstore should expect to generate about

- **Case 2:**
  - The textbook costs \$137 and the study guide \$33.
  - How much revenue should the bookstore expect from this class of 100 students?
- **Answer**
  - About 55 students will just buy a textbook, providing revenue of

    $$\$137 \times 55 = \$7,535$$

  - The roughly 25 students who buy both the textbook and the study guide would pay a total of

    $$(\$137 + \$33) \times 25 = \$170 \times 25 = \$4,250$$

  - Thus, the bookstore should expect to generate about

    $$\$7,535 + \$4,250 = \$11,785$$

    from these 100 students for this one class.

Random Variable [2]

Example

- **Case 2:**
  - The textbook costs $137 and the study guide $33.
  - How much revenue should the bookstore expect from this class of 100 students?
- **Answer**
  - About 55 students will just buy a textbook, providing revenue of

$$\$137 \times 55 = \$7,535$$

  - The roughly 25 students who buy both the textbook and the study guide would pay a total of

$$(\$137 + \$33) \times 25 = \$170 \times 25 = \$4,250$$

  - Thus, the bookstore should expect to generate about

$$\$7,535 + \$4,250 = \$11,785$$

    from these 100 students for this one class.
  - However, there might be some sampling variability so the actual amount may differ by a little bit.

Random Variable [2]

Example

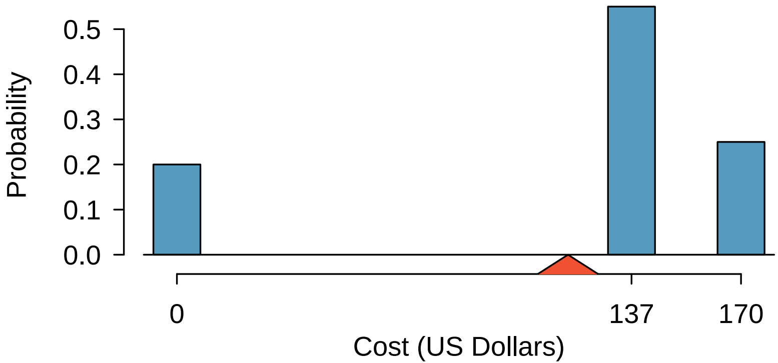

Figure 9 : Probability distribution for the bookstore's revenue from a single student. The distribution balances on a triangle representing the average revenue per student.

Expectation [2]

Random Variable

A random process or variable with a numerical outcome.

- We call a variable with a numerical outcome a random variable.

Expectation [2]

Random Variable

A random process or variable with a numerical outcome.

- We call a variable with a numerical outcome a random variable.
- Random variable is usually represented with a capital letter such as X, Y, or Z.

Expectation [2]

Random Variable

A random process or variable with a numerical outcome.

- We call a variable with a numerical outcome a random variable.
- Random variable is usually represented with a capital letter such as X, Y, or Z.
- The previous cases, amount of money a single student will spend on his/her statistics books is a random variable.

Expectation [2]

Table 4 : The probability distribution for the random variable X, representing the bookstore's revenue from a single student.

i	1	2	3	Total
x_i	\$0	\$137	\$170	–
$P(X = x_i)$	0.20	0.55	0.25	1.00

- Suppose we represent the amount of money a student will spend on his/her statistics book as X.

Expectation [2]

Table 4 : The probability distribution for the random variable X, representing the bookstore's revenue from a single student.

i	1	2	3	Total
x_i	\$0	\$137	\$170	–
$P(X = x_i)$	0.20	0.55	0.25	1.00

- Suppose we represent the amount of money a student will spend on his/her statistics book as X.
- The possible outcomes of X are labeled with a corresponding lower case letter x and subscripts.

Expectation [2]

Table 4 : The probability distribution for the random variable X, representing the bookstore's revenue from a single student.

i	1	2	3	Total
x_i	\$0	\$137	\$170	–
$P(X = x_i)$	0.20	0.55	0.25	1.00

- Suppose we represent the amount of money a student will spend on his/her statistics book as X.
- The possible outcomes of X are labeled with a corresponding lower case letter x and subscripts.
- For example, we write $x_1 = \$0$, $x_2 = \$137$, and $x_3 = \$170$, which occur with probabilities 0.20, 0.55, and 0.25.

Expectation [2]

Table 4 : The probability distribution for the random variable X, representing the bookstore's revenue from a single student.

i	1	2	3	Total
x_i	\$0	\$137	\$170	–
$P(X = x_i)$	0.20	0.55	0.25	1.00

- Suppose we represent the amount of money a student will spend on his/her statistics book as X.
- The possible outcomes of X are labeled with a corresponding lower case letter x and subscripts.
- For example, we write $x_1 = \$0$, $x_2 = \$137$, and $x_3 = \$170$, which occur with probabilities 0.20, 0.55, and 0.25.
- The distribution of X is summarized in Figure 9 and Table **??**.

Expectation [2]

Expected value of a Discrete Random Variable

If X takes outcomes x_1, ..., x_k with probabilities $P(X = x_1)$, ..., $P(X = x_k)$, the expected value of X is the sum of each outcome multiplied by its corresponding probability:

Expectation [2]

Expected value of a Discrete Random Variable

If X takes outcomes x_1, ..., x_k with probabilities $P(X = x_1)$, ..., $P(X = x_k)$, the expected value of X is the sum of each outcome multiplied by its corresponding probability:

$$\begin{aligned} E(X) &= x_1 \times P(X = x_1) + \cdots + x_k \times P(X = x_k) \\ &= \sum_{i=1}^{k} x_i P(X = x_i) \end{aligned} \tag{9}$$

Expectation [2]

Expected value of a Discrete Random Variable

If X takes outcomes x_1, ..., x_k with probabilities $P(X = x_1)$, ..., $P(X = x_k)$, the expected value of X is the sum of each outcome multiplied by its corresponding probability:

$$\begin{aligned} E(X) &= x_1 \times P(X = x_1) + \cdots + x_k \times P(X = x_k) \\ &= \sum_{i=1}^{k} x_i P(X = x_i) \end{aligned} \tag{9}$$

The Greek letter μ may be used in place of the notation $E(X)$.

Expectation [2]

- The expected total revenue is \$ 11,785 and there are 100 students.

Expectation [2]

- The expected total revenue is \$ 11,785 and there are 100 students. Therefore the expected revenue per student is

Expectation [2]

- The expected total revenue is \$ 11,785 and there are 100 students. Therefore the expected revenue per student is

$$\$11,785/100 = \$117.85$$

Expectation [2]

- The expected total revenue is $ 11,785 and there are 100 students. Therefore the expected revenue per student is

$$\$11,785/100 = \$117.85$$

- The average outcome of X as \$117.85 is called as the **expected value** of X.

Expectation [2]

- The expected total revenue is \$ 11,785 and there are 100 students. Therefore the expected revenue per student is

$$\$11,785/100 = \$117.85$$

- The average outcome of X as \$117.85 is called as the **expected value** of X.
- For our case, the expected value of a random variable is computed by:

Expectation [2]

- The expected total revenue is \$ 11,785 and there are 100 students. Therefore the expected revenue per student is

$$\$11,785/100 = \$117.85$$

- The average outcome of X as \$117.85 is called as the **expected value** of X.
- For our case, the expected value of a random variable is computed by:

$$E(X) = 0 \times P(X = 0) + 137 \times P(X = 137) + 170 \times P(X = 170)$$

Expectation [2]

- The expected total revenue is \$ 11,785 and there are 100 students. Therefore the expected revenue per student is

$$\$11,785/100 = \$117.85$$

- The average outcome of X as \$117.85 is called as the **expected value** of X.
- For our case, the expected value of a random variable is computed by:

$$\begin{aligned} E(X) &= 0 \times P(X = 0) + 137 \times P(X = 137) + 170 \times P(X = 170) \\ &= 0 \times 0.20 + 137 \times 0.55 + 170 \times 0.25 \end{aligned}$$

Expectation [2]

- The expected total revenue is \$ 11,785 and there are 100 students. Therefore the expected revenue per student is

$$\$11,785/100 = \$117.85$$

- The average outcome of X as \$117.85 is called as the **expected value** of X.
- For our case, the expected value of a random variable is computed by:

$$\begin{aligned} E(X) &= 0 \times P(X = 0) + 137 \times P(X = 137) + 170 \times P(X = 170) \\ &= 0 \times 0.20 + 137 \times 0.55 + 170 \times 0.25 \\ &= 117.85 \end{aligned}$$

Outline

Exercise 4

- Load the `30rep.Rdata`, `100rep.Rdata`, `1000rep.Rdata`, `10000rep.Rdata`, datasets.

Exercise 4

- Load the `30rep.Rdata`, `100rep.Rdata`, `1000rep.Rdata`, `10000rep.Rdata`, datasets.
- Create histogram for each dataset.

Exercise 4

- Load the `30rep.Rdata`, `100rep.Rdata`, `1000rep.Rdata`, `10000rep.Rdata`, datasets.
- Create histogram for each dataset.
- Compare the shapes of the histograms.

Exercise 4

- Load the `30rep.Rdata`, `100rep.Rdata`, `1000rep.Rdata`, `10000rep.Rdata`, datasets.
- Create histogram for each dataset.
- Compare the shapes of the histograms.
- Since the number 30, 100, 1000, and 10,000 represent the number of repetitions for each method, then what is the relation between the number of experiments with the frequency distribution?

Outline

Outline

Normal Distribution [2]

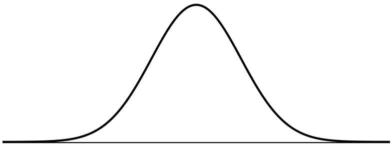

Figure 10 : A normal curve.

- Normal distribution is the most common distribution we see in practice[4].

Normal Distribution [2]

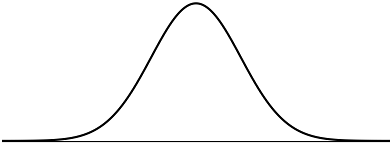

Figure 10 : A normal curve.

- Normal distribution is the most common distribution we see in practice[4].

[4] Also known as Gaussian distribution.

Normal Distribution [2]

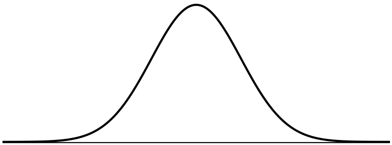

Figure 10 : A normal curve.

- Normal distribution is the most common distribution we see in practice[4].
- If plotted, the distribution will form a symmetric curve known as the normal curve, and sometimes called as bell curve as shown in Figure 10.

[4] Also known as Gaussian distribution.

Normal Distribution [2]

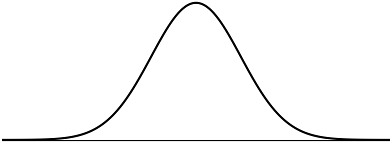

Figure 10 : A normal curve.

- Normal distribution is the most common distribution we see in practice[4].
- If plotted, the distribution will form a symmetric curve known as the normal curve, and sometimes called as bell curve as shown in Figure 10.
- The fact is, many variables are normal, but none are exactly normal.

[4] Also known as Gaussian distribution.

Outline

Normal Distribution Model [2]

- Normal distribution always describes a symmetric, unimodal, bell-shaped curve.

Normal Distribution Model [2]

- Normal distribution always describes a symmetric, unimodal, bell-shaped curve.
- However, these curves can look different depending on the details of the model.

Normal Distribution Model [2]

- Normal distribution always describes a symmetric, unimodal, bell-shaped curve.
- However, these curves can look different depending on the details of the model.
- Specifically, the normal distribution model can be adjusted using two parameters: mean and standard deviation.

Normal Distribution Model [2]

- Normal distribution always describes a symmetric, unimodal, bell-shaped curve.
- However, these curves can look different depending on the details of the model.
- Specifically, the normal distribution model can be adjusted using two parameters: mean and standard deviation.
- As you can probably guess, changing the mean shifts the bell curve to the left or right, while changing the standard deviation stretches or constricts the curve.

Normal Distribution Model [2]

- Normal distribution always describes a symmetric, unimodal, bell-shaped curve.
- However, these curves can look different depending on the details of the model.
- Specifically, the normal distribution model can be adjusted using two parameters: mean and standard deviation.
- As you can probably guess, changing the mean shifts the bell curve to the left or right, while changing the standard deviation stretches or constricts the curve.
- Figure 11 shows the normal distribution with mean 0 and standard deviation 1 in the left panel and the normal distributions with mean 19 and standard deviation 4 in the right panel.

Normal Distribution Model [2]

- Normal distribution always describes a symmetric, unimodal, bell-shaped curve.
- However, these curves can look different depending on the details of the model.
- Specifically, the normal distribution model can be adjusted using two parameters: mean and standard deviation.
- As you can probably guess, changing the mean shifts the bell curve to the left or right, while changing the standard deviation stretches or constricts the curve.
- Figure 11 shows the normal distribution with mean 0 and standard deviation 1 in the left panel and the normal distributions with mean 19 and standard deviation 4 in the right panel.
- Figure 12 shows these distributions on the same axis.

Normal Distribution Model [2]

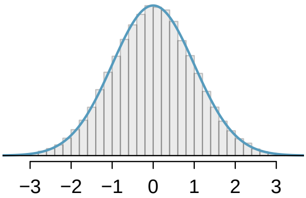

Figure 11 : Both curves represent the normal distribution, however, they differ in their center and spread. The normal distribution with mean 0 and standard deviation 1 is called the **standard normal distribution**.

Normal Distribution Model [2]

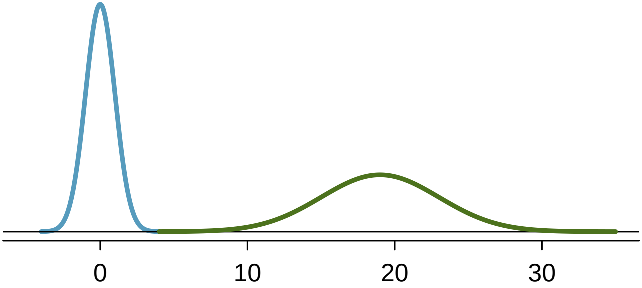

Figure 12 : The normal models shown in Figure 11 but plotted together and on the same scale.

Normal Distribution Model [2]

- If a normal distribution has mean μ and standard deviation σ, we may write the distribution as

Normal Distribution Model [2]

- If a normal distribution has mean μ and standard deviation σ, we may write the distribution as

$$N(\mu, \sigma)$$

Normal Distribution Model [2]

- If a normal distribution has mean μ and standard deviation σ, we may write the distribution as

$$N(\mu, \sigma)$$

- The two distributions in Figure 12 can be written as

Normal Distribution Model [2]

- If a normal distribution has mean μ and standard deviation σ, we may write the distribution as

$$N(\mu, \sigma)$$

- The two distributions in Figure 12 can be written as

$$N(\mu = 0, \sigma = 1) \quad \text{and} \quad N(\mu = 19, \sigma = 4)$$

Normal Distribution Model [2]

- If a normal distribution has mean μ and standard deviation σ, we may write the distribution as

$$N(\mu, \sigma)$$

- The two distributions in Figure 12 can be written as

$$N(\mu = 0, \sigma = 1) \quad \text{and} \quad N(\mu = 19, \sigma = 4)$$

- Because the mean and standard deviation describe a normal distribution exactly, they are called the distribution's parameters.

Outline

Standardizing with Z-scores [2]

Table 5 : Mean and standard deviation for the SAT and ACT. Please refer to [2] for the source of the data.

	SAT	ACT
Mean	1500	21
SD	300	5

- Table 5 shows the mean and standard deviation for total scores on the SAT and ACT.

Standardizing with Z-scores [2]

Table 5 : Mean and standard deviation for the SAT and ACT. Please refer to [2] for the source of the data.

	SAT	ACT
Mean	1500	21
SD	300	5

- Table 5 shows the mean and standard deviation for total scores on the SAT and ACT.
- The distribution of SAT and ACT scores are both nearly normal.

Standardizing with Z-scores [2]

Table 5 : Mean and standard deviation for the SAT and ACT. Please refer to [2] for the source of the data.

	SAT	ACT
Mean	1500	21
SD	300	5

- Table 5 shows the mean and standard deviation for total scores on the SAT and ACT.
- The distribution of SAT and ACT scores are both nearly normal.
- Suppose Ann scored 1800 on her SAT and Tom scored 24 on his ACT. Who performed better?

Standardizing with Z-scores [2]

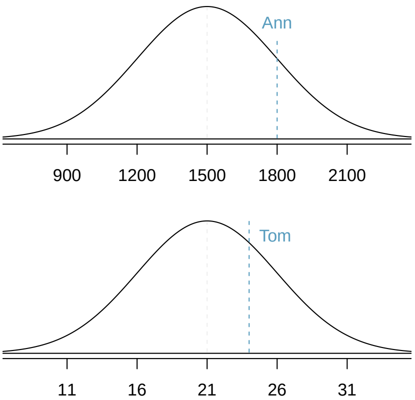

Figure 13 : Ann's and Tom's scores shown with the distributions of SAT and ACT scores.

Standardizing with Z-scores [2]

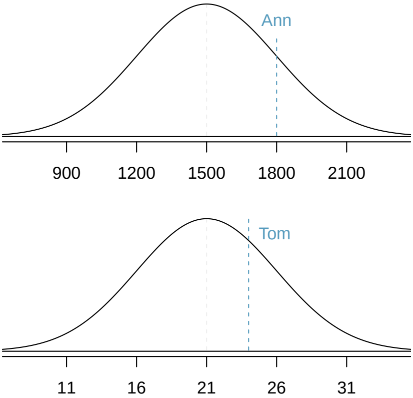

Figure 13 : Ann's and Tom's scores shown with the distributions of SAT and ACT scores.

- We use the standard deviation as a guide.

Standardizing with Z-scores [2]

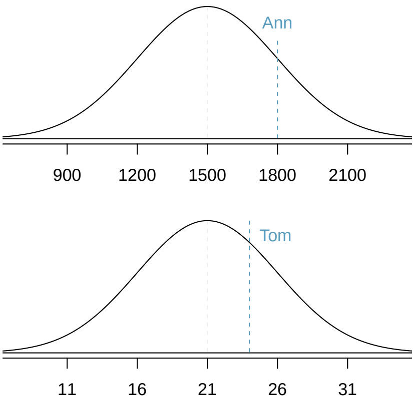

Figure 13 : Ann's and Tom's scores shown with the distributions of SAT and ACT scores.

- We use the standard deviation as a guide.
- Ann is 1 standard deviation above average on the SAT: $1500 + 300 = 1800$.

Standardizing with Z-scores [2]

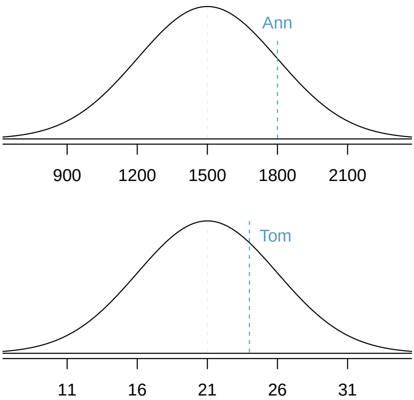

Figure 13 : Ann's and Tom's scores shown with the distributions of SAT and ACT scores.

- We use the standard deviation as a guide.
- Ann is 1 standard deviation above average on the SAT: $1500 + 300 = 1800$.
- Tom is 0.6 standard deviations above the mean on the ACT: $21 + 0.6 \times 5 = 24$.

Standardizing with Z-scores [2]

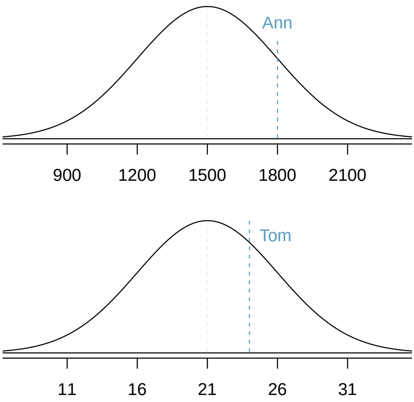

Figure 13 : Ann's and Tom's scores shown with the distributions of SAT and ACT scores.

- We use the standard deviation as a guide.
- Ann is 1 standard deviation above average on the SAT: $1500 + 300 = 1800$.
- Tom is 0.6 standard deviations above the mean on the ACT: $21 + 0.6 \times 5 = 24$.
- In Figure 13, we can see that Ann tends to do better with respect to everyone else than Tom did, so her score was better.

Standardizing with Z-scores [2]

- The previous case used a standardization technique called a Z-score.

Standardizing with Z-scores [2]

- The previous case used a standardization technique called a Z-score.
- Z-score is a method most commonly employed for nearly normal observations but that may be used with any distribution.

Standardizing with Z-scores [2]

- The previous case used a standardization technique called a Z-score.
- Z-score is a method most commonly employed for nearly normal observations but that may be used with any distribution.
- Z-score of an observation is defined as the number of standard deviations it falls above or below the mean.

Standardizing with Z-scores [2]

- The previous case used a standardization technique called a Z-score.
- Z-score is a method most commonly employed for nearly normal observations but that may be used with any distribution.
- Z-score of an observation is defined as the number of standard deviations it falls above or below the mean.
- If the observation is one standard deviation above the mean, its Z-score is 1.

Standardizing with Z-scores [2]

- The previous case used a standardization technique called a Z-score.
- Z-score is a method most commonly employed for nearly normal observations but that may be used with any distribution.
- Z-score of an observation is defined as the number of standard deviations it falls above or below the mean.
- If the observation is one standard deviation above the mean, its Z-score is 1.
- If it is 1.5 standard deviations below the mean, then its Z-score is -1.5.

Standardizing with Z-scores [2]

- If x is an observation from a distribution $N(\mu, \sigma)$, we define the Z-score mathematically as

Standardizing with Z-scores [2]

- If x is an observation from a distribution $N(\mu, \sigma)$, we define the Z-score mathematically as

$$Z = \frac{x - \mu}{\sigma} \tag{10}$$

Standardizing with Z-scores [2]

- If x is an observation from a distribution $N(\mu, \sigma)$, we define the Z-score mathematically as

$$Z = \frac{x - \mu}{\sigma} \tag{10}$$

- Using $\mu_{SAT} = 1500$, $\sigma_{SAT} = 300$, and $x_{Ann} = 1800$, we find Ann's Z-score:

$$Z_{Ann} =$$

Standardizing with Z-scores [2]

- If x is an observation from a distribution $N(\mu, \sigma)$, we define the Z-score mathematically as

$$Z = \frac{x - \mu}{\sigma} \tag{10}$$

- Using $\mu_{SAT} = 1500$, $\sigma_{SAT} = 300$, and $x_{Ann} = 1800$, we find Ann's Z-score:

$$Z_{Ann} = \frac{x_{Ann} - \mu_{SAT}}{\sigma_{SAT}} =$$

Standardizing with Z-scores [2]

- If x is an observation from a distribution $N(\mu, \sigma)$, we define the Z-score mathematically as

$$Z = \frac{x - \mu}{\sigma} \tag{10}$$

- Using $\mu_{SAT} = 1500$, $\sigma_{SAT} = 300$, and $x_{Ann} = 1800$, we find Ann's Z-score:

$$Z_{Ann} = \frac{x_{Ann} - \mu_{SAT}}{\sigma_{SAT}} = \frac{1800 - 1500}{300} =$$

Standardizing with Z-scores [2]

- If x is an observation from a distribution $N(\mu, \sigma)$, we define the Z-score mathematically as

$$Z = \frac{x - \mu}{\sigma} \tag{10}$$

- Using $\mu_{SAT} = 1500$, $\sigma_{SAT} = 300$, and $x_{Ann} = 1800$, we find Ann's Z-score:

$$Z_{Ann} = \frac{x_{Ann} - \mu_{SAT}}{\sigma_{SAT}} = \frac{1800 - 1500}{300} = 1$$

Standardizing with Z-scores [2]

- If x is an observation from a distribution $N(\mu, \sigma)$, we define the Z-score mathematically as

$$Z = \frac{x - \mu}{\sigma} \tag{10}$$

- Using $\mu_{SAT} = 1500$, $\sigma_{SAT} = 300$, and $x_{Ann} = 1800$, we find Ann's Z-score:

$$Z_{Ann} = \frac{x_{Ann} - \mu_{SAT}}{\sigma_{SAT}} = \frac{1800 - 1500}{300} = 1$$

- and for Tom's ACT score:

$$Z_{Tom} =$$

Standardizing with Z-scores [2]

- If x is an observation from a distribution $N(\mu, \sigma)$, we define the Z-score mathematically as

$$Z = \frac{x - \mu}{\sigma} \tag{10}$$

- Using $\mu_{SAT} = 1500$, $\sigma_{SAT} = 300$, and $x_{Ann} = 1800$, we find Ann's Z-score:

$$Z_{Ann} = \frac{x_{Ann} - \mu_{SAT}}{\sigma_{SAT}} = \frac{1800 - 1500}{300} = 1$$

- and for Tom's ACT score:

$$Z_{Tom} = \frac{x_{Tom} - \mu_{ACT}}{\sigma_{ACT}} =$$

Standardizing with Z-scores [2]

- If x is an observation from a distribution $N(\mu, \sigma)$, we define the Z-score mathematically as

$$Z = \frac{x - \mu}{\sigma} \tag{10}$$

- Using $\mu_{SAT} = 1500$, $\sigma_{SAT} = 300$, and $x_{Ann} = 1800$, we find Ann's Z-score:

$$Z_{Ann} = \frac{x_{Ann} - \mu_{SAT}}{\sigma_{SAT}} = \frac{1800 - 1500}{300} = 1$$

- and for Tom's ACT score:

$$Z_{Tom} = \frac{x_{Tom} - \mu_{ACT}}{\sigma_{ACT}} = \frac{24 - 21}{5} =$$

Standardizing with Z-scores [2]

- If x is an observation from a distribution $N(\mu, \sigma)$, we define the Z-score mathematically as

$$Z = \frac{x - \mu}{\sigma} \tag{10}$$

- Using $\mu_{SAT} = 1500$, $\sigma_{SAT} = 300$, and $x_{Ann} = 1800$, we find Ann's Z-score:

$$Z_{Ann} = \frac{x_{Ann} - \mu_{SAT}}{\sigma_{SAT}} = \frac{1800 - 1500}{300} = 1$$

- and for Tom's ACT score:

$$Z_{Tom} = \frac{x_{Tom} - \mu_{ACT}}{\sigma_{ACT}} = \frac{24 - 21}{5} = 0.6$$

Outline

Normal Probability Table [2]

- A **normal probability table**, which lists Z-scores and corresponding percentiles, can be used to identify a percentile based on the Z-score (and vice versa). Statistical software can also be used.

Normal Probability Table [2]

- A **normal probability table**, which lists Z-scores and corresponding percentiles, can be used to identify a percentile based on the Z-score (and vice versa). Statistical software can also be used.
- We can use the normal model to find percentiles.

Normal Probability Table [2]

- A **normal probability table**, which lists Z-scores and corresponding percentiles, can be used to identify a percentile based on the Z-score (and vice versa). Statistical software can also be used.
- We can use the normal model to find percentiles.
- We use this table to identify the percentile corresponding to any particular Z-score.

Normal Probability Table [2]

- A **normal probability table**, which lists Z-scores and corresponding percentiles, can be used to identify a percentile based on the Z-score (and vice versa). Statistical software can also be used.
- We can use the normal model to find percentiles.
- We use this table to identify the percentile corresponding to any particular Z-score.
- For instance:

Normal Probability Table [2]

- A **normal probability table**, which lists Z-scores and corresponding percentiles, can be used to identify a percentile based on the Z-score (and vice versa). Statistical software can also be used.
- We can use the normal model to find percentiles.
- We use this table to identify the percentile corresponding to any particular Z-score.
- For instance:
  - The percentile of $Z = 0.43$ is shown in row 0.4 and column 0.03 in Table 6: 0.6664, or the 66.64^{th} percentile.

Normal Probability Table [2]

- A **normal probability table**, which lists Z-scores and corresponding percentiles, can be used to identify a percentile based on the Z-score (and vice versa). Statistical software can also be used.
- We can use the normal model to find percentiles.
- We use this table to identify the percentile corresponding to any particular Z-score.
- For instance:
  - The percentile of $Z = 0.43$ is shown in row 0.4 and column 0.03 in Table 6: 0.6664, or the 66.64^{th} percentile.
  - Generally,

Normal Probability Table [2]

- A **normal probability table**, which lists Z-scores and corresponding percentiles, can be used to identify a percentile based on the Z-score (and vice versa). Statistical software can also be used.
- We can use the normal model to find percentiles.
- We use this table to identify the percentile corresponding to any particular Z-score.
- For instance:
  - The percentile of $Z = 0.43$ is shown in row 0.4 and column 0.03 in Table 6: 0.6664, or the 66.64^{th} percentile.
  - Generally, we round Z to two decimals,

Normal Probability Table [2]

- A **normal probability table**, which lists Z-scores and corresponding percentiles, can be used to identify a percentile based on the Z-score (and vice versa). Statistical software can also be used.
- We can use the normal model to find percentiles.
- We use this table to identify the percentile corresponding to any particular Z-score.
- For instance:
  - The percentile of $Z = 0.43$ is shown in row 0.4 and column 0.03 in Table 6: 0.6664, or the 66.64^{th} percentile.
  - Generally, we round Z to two decimals, identify the proper row in the normal probability table up through the first decimal,

Normal Probability Table [2]

- A **normal probability table**, which lists Z-scores and corresponding percentiles, can be used to identify a percentile based on the Z-score (and vice versa). Statistical software can also be used.
- We can use the normal model to find percentiles.
- We use this table to identify the percentile corresponding to any particular Z-score.
- For instance:
  - The percentile of $Z = 0.43$ is shown in row 0.4 and column 0.03 in Table 6: 0.6664, or the 66.64^{th} percentile.
  - Generally, we round Z to two decimals, identify the proper row in the normal probability table up through the first decimal, and then determine the column representing the second decimal value.

Normal Probability Table [2]

- A **normal probability table**, which lists Z-scores and corresponding percentiles, can be used to identify a percentile based on the Z-score (and vice versa). Statistical software can also be used.
- We can use the normal model to find percentiles.
- We use this table to identify the percentile corresponding to any particular Z-score.
- For instance:
  - The percentile of $Z = 0.43$ is shown in row 0.4 and column 0.03 in Table 6: 0.6664, or the 66.64^{th} percentile.
  - Generally, we round Z to two decimals, identify the proper row in the normal probability table up through the first decimal, and then determine the column representing the second decimal value.
  - The intersection of this row and column is the percentile of the observation.

Normal Probability Table [2]

Table 6 : A section of the normal probability table. The percentile for a normal random variable with $Z = 0.43$ has been highlighted, and the percentile closest to 0.8000 has also been **highlighted**.

	Second decimal place of Z									
Z	0.00	0.01	0.02	0.03	**0.04**	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	**0.7995**	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Normal Probability Table [2]

- Recall the SAT and ACT example:

Normal Probability Table [2]

- Recall the SAT and ACT example:
  - Ann earned a score of 1800 on her SAT with a corresponding $Z = 1$.

Normal Probability Table [2]

- Recall the SAT and ACT example:
  - Ann earned a score of 1800 on her SAT with a corresponding $Z = 1$.
  - She would like to know what percentile she falls in among all SAT test-takers.

Normal Probability Table [2]

- Recall the SAT and ACT example:
  - Ann earned a score of 1800 on her SAT with a corresponding $Z = 1$.
  - She would like to know what percentile she falls in among all SAT test-takers.
- Recall the SAT and ACT example:

Normal Probability Table [2]

- Recall the SAT and ACT example:
  - Ann earned a score of 1800 on her SAT with a corresponding $Z = 1$.
  - She would like to know what percentile she falls in among all SAT test-takers.
- Recall the SAT and ACT example:
  - Ann's **percentile** is the percentage of people who earned a lower SAT score than Ann[5].

Normal Probability Table [2]

- Recall the SAT and ACT example:
  - Ann earned a score of 1800 on her SAT with a corresponding $Z = 1$.
  - She would like to know what percentile she falls in among all SAT test-takers.
- Recall the SAT and ACT example:
  - Ann's **percentile** is the percentage of people who earned a lower SAT score than Ann[5].

[5] We shade the area representing those individuals in Figure 14

Normal Probability Table [2]

- Recall the SAT and ACT example:
  - Ann earned a score of 1800 on her SAT with a corresponding $Z = 1$.
  - She would like to know what percentile she falls in among all SAT test-takers.
- Recall the SAT and ACT example:
  - Ann's **percentile** is the percentage of people who earned a lower SAT score than Ann[5].
  - The total area under the normal curve is always equal to 1, and the proportion of people who scored below Ann on the SAT is equal to the area shaded in Figure 14: 0.8413.

[5] We shade the area representing those individuals in Figure 14

Normal Probability Table [2]

- Recall the SAT and ACT example:
  - Ann earned a score of 1800 on her SAT with a corresponding $Z = 1$.
  - She would like to know what percentile she falls in among all SAT test-takers.
- Recall the SAT and ACT example:
  - Ann's **percentile** is the percentage of people who earned a lower SAT score than Ann[5].
  - The total area under the normal curve is always equal to 1, and the proportion of people who scored below Ann on the SAT is equal to the area shaded in Figure 14: 0.8413.
  - In other words, Ann is in the 84^{th} percentile of SAT takers.

[5] We shade the area representing those individuals in Figure 14

Normal Probability Table [2]

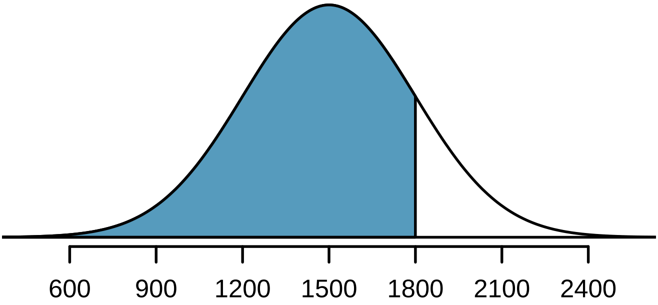

Figure 14 : The normal model for SAT scores, shading the area of those individuals who scored below Ann.

Normal Probability Table [2]

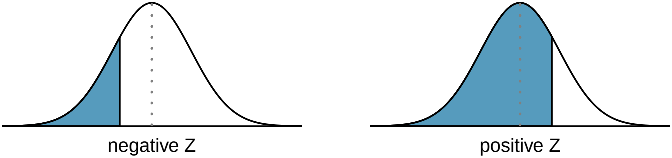

Figure 15 : The area to the left of Z represents the percentile of the observation.

68-95-99.7 Rule [2]

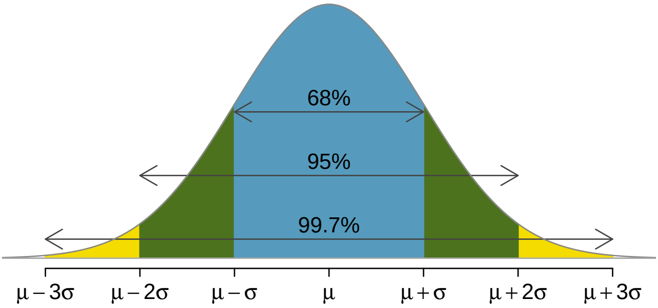

Figure 16 : Probabilities for falling within 1, 2, and 3 standard deviations of the mean in a normal distribution.

Outline

Outline

Bernoulli Distribution [2]

The Milgram's Experiment

Stanley Milgram began a series of experiments in 1963 to estimate what proportion of people would willingly obey an authority and give severe shocks to a stranger. Milgram found that about 65% of people would obey the authority and give such shocks. Over the years, additional research suggested this number is approximately consistent across communities and time.[6]

- Each person in Milgram's experiment can be thought of as a **trial**.

[6] Find further information on Milgram's experiment at `http:`

The Milgram's Experiment

Stanley Milgram began a series of experiments in 1963 to estimate what proportion of people would willingly obey an authority and give severe shocks to a stranger. Milgram found that about 65% of people would obey the authority and give such shocks. Over the years, additional research suggested this number is approximately consistent across communities and time.[6]

- Each person in Milgram's experiment can be thought of as a **trial**.
- We label a person a **success** if she refuses to administer the worst shock and **failure** if she administers the worst shock.

[6] Find further information on Milgram's experiment at `http:`

Bernoulli Distribution [2]

The Milgram's Experiment

Stanley Milgram began a series of experiments in 1963 to estimate what proportion of people would willingly obey an authority and give severe shocks to a stranger. Milgram found that about 65% of people would obey the authority and give such shocks. Over the years, additional research suggested this number is approximately consistent across communities and time.[6]

- Each person in Milgram's experiment can be thought of as a **trial**.
- We label a person a **success** if she refuses to administer the worst shock and **failure** if she administers the worst shock.
- Because only 35% of individuals refused to administer the most severe shock, we denote the **probability of a success** with $p = 0.35$.

[6] Find further information on Milgram's experiment at `http:`

Bernoulli Distribution [2]

The Milgram's Experiment

Stanley Milgram began a series of experiments in 1963 to estimate what proportion of people would willingly obey an authority and give severe shocks to a stranger. Milgram found that about 65% of people would obey the authority and give such shocks. Over the years, additional research suggested this number is approximately consistent across communities and time.[6]

- Each person in Milgram's experiment can be thought of as a **trial**.
- We label a person a **success** if she refuses to administer the worst shock and **failure** if she administers the worst shock.
- Because only 35% of individuals refused to administer the most severe shock, we denote the **probability of a success** with $p = 0.35$.
- The probability of a failure is sometimes denoted with $q = 1 - p$.

---

[6] Find further information on Milgram's experiment at `http:`

Bernoulli Distribution [2]

The Milgram's Experiment

Stanley Milgram began a series of experiments in 1963 to estimate what proportion of people would willingly obey an authority and give severe shocks to a stranger. Milgram found that about 65% of people would obey the authority and give such shocks. Over the years, additional research suggested this number is approximately consistent across communities and time.[6]

- Each person in Milgram's experiment can be thought of as a **trial**.
- We label a person a **success** if she refuses to administer the worst shock and **failure** if she administers the worst shock.
- Because only 35% of individuals refused to administer the most severe shock, we denote the **probability of a success** with $p = 0.35$.
- The probability of a failure is sometimes denoted with $q = 1 - p$.
- When an individual trial only has two possible outcomes, it is called a **Bernoulli random variable**.

[6] Find further information on Milgram's experiment at `http:`

Bernoulli Distribution [2]

Bernoulli Random Variable

A Bernoulli random variable has exactly two possible outcomes. We typically label one of these outcomes a “success” and the other outcome a “failure”. We may also denote a success by `1` and a failure by `0`.

- Bernoulli random variables are often denoted as `1` for a success and `0` for a failure.

Bernoulli Distribution [2]

Bernoulli Random Variable

A Bernoulli random variable has exactly two possible outcomes. We typically label one of these outcomes a “success” and the other outcome a “failure”. We may also denote a success by `1` and a failure by `0`.

- Bernoulli random variables are often denoted as `1` for a success and `0` for a failure.
- In addition to being convenient in entering data, it is also mathematically handy.

Bernoulli Distribution [2]

Bernoulli Random Variable

A Bernoulli random variable has exactly two possible outcomes. We typically label one of these outcomes a “success” and the other outcome a “failure”. We may also denote a success by `1` and a failure by `0`.

- Bernoulli random variables are often denoted as `1` for a success and `0` for a failure.
- In addition to being convenient in entering data, it is also mathematically handy.
- Suppose we observe ten trials:

Bernoulli Distribution [2]

Bernoulli Random Variable

A Bernoulli random variable has exactly two possible outcomes. We typically label one of these outcomes a "success" and the other outcome a "failure". We may also denote a success by 1 and a failure by 0.

- Bernoulli random variables are often denoted as 1 for a success and 0 for a failure.
- In addition to being convenient in entering data, it is also mathematically handy.
- Suppose we observe ten trials:

$$0111101100$$

Bernoulli Distribution [2]

Bernoulli Random Variable

A Bernoulli random variable has exactly two possible outcomes. We typically label one of these outcomes a “success” and the other outcome a “failure”. We may also denote a success by `1` and a failure by `0`.

- Bernoulli random variables are often denoted as `1` for a success and `0` for a failure.
- In addition to being convenient in entering data, it is also mathematically handy.
- Suppose we observe ten trials:

  `0111101100`

- Then the **sample proportion**, $\hat{p}$, is the sample mean of these observations:

Bernoulli Distribution [2]

Bernoulli Random Variable

A Bernoulli random variable has exactly two possible outcomes. We typically label one of these outcomes a "success" and the other outcome a "failure". We may also denote a success by `1` and a failure by `0`.

- Bernoulli random variables are often denoted as `1` for a success and `0` for a failure.
- In addition to being convenient in entering data, it is also mathematically handy.
- Suppose we observe ten trials:

  `0111101100`

- Then the **sample proportion**, $\hat{p}$, is the sample mean of these observations:

$$\hat{p} =$$

Bernoulli Distribution [2]

Bernoulli Random Variable

A Bernoulli random variable has exactly two possible outcomes. We typically label one of these outcomes a "success" and the other outcome a "failure". We may also denote a success by 1 and a failure by `0`.

- Bernoulli random variables are often denoted as 1 for a success and `0` for a failure.
- In addition to being convenient in entering data, it is also mathematically handy.
- Suppose we observe ten trials:

$$0111101100$$

- Then the **sample proportion**, $\hat{p}$, is the sample mean of these observations:

$$\hat{p} = \frac{\text{\# of successes}}{\text{\# of trials}} =$$

Bernoulli Distribution [2]

Bernoulli Random Variable

A Bernoulli random variable has exactly two possible outcomes. We typically label one of these outcomes a "success" and the other outcome a "failure". We may also denote a success by 1 and a failure by 0.

- Bernoulli random variables are often denoted as 1 for a success and 0 for a failure.
- In addition to being convenient in entering data, it is also mathematically handy.
- Suppose we observe ten trials:

$$0111101100$$

- Then the **sample proportion**, $\hat{p}$, is the sample mean of these observations:

$$\hat{p} = \frac{\text{\# of successes}}{\text{\# of trials}} = \frac{0+1+1+1+1+0+1+1+0+0}{10} =$$

Bernoulli Distribution [2]

Bernoulli Random Variable

A Bernoulli random variable has exactly two possible outcomes. We typically label one of these outcomes a "success" and the other outcome a "failure". We may also denote a success by 1 and a failure by 0.

- Bernoulli random variables are often denoted as 1 for a success and 0 for a failure.
- In addition to being convenient in entering data, it is also mathematically handy.
- Suppose we observe ten trials:

  0111101100

- Then the **sample proportion**, $\hat{p}$, is the sample mean of these observations:

$$\hat{p} = \frac{\text{\# of successes}}{\text{\# of trials}} = \frac{0+1+1+1+1+0+1+1+0+0}{10} = 0.6$$

Bernoulli Distribution [2]

Bernoulli Random Variable

If X is a random variable that takes value 1 with probability of success p and 0 with probability $1 - p$, then X is a Bernoulli random variable with mean and standard deviation

$$\mu = p \qquad \sigma = \sqrt{p(1-p)} \tag{11}$$

Outline

Geometric Distribution [2]

Dr. Smith's Experiment

- Dr. Smith wants to repeat Milgram's experiments but she only wants to sample people until she finds someone who will not inflict the worst shock.[7]

Geometric Distribution [2]

Dr. Smith's Experiment

- Dr. Smith wants to repeat Milgram's experiments but she only wants to sample people until she finds someone who will not inflict the worst shock.[7]

[7] This is hypothetical since, in reality, this sort of study probably would not be permitted any longer under current ethical standards.

Geometric Distribution [2]

Dr. Smith's Experiment

- Dr. Smith wants to repeat Milgram's experiments but she only wants to sample people until she finds someone who will not inflict the worst shock.[7]
- If the probability a person will not give the most severe shock is still 0.35 and the subjects are independent, what are the chances that she will stop the study after the first person? The second person? The third?

[7] This is hypothetical since, in reality, this sort of study probably would not be permitted any longer under current ethical standards.

Geometric Distribution [2]

Dr. Smith's Experiment

- Dr. Smith wants to repeat Milgram's experiments but she only wants to sample people until she finds someone who will not inflict the worst shock.[7]
- If the probability a person will not give the most severe shock is still 0.35 and the subjects are independent, what are the chances that she will stop the study after the first person? The second person? The third?
- What about if it takes her $n-1$ individuals who will administer the worst shock before finding her first success, i.e. the first success is on the n^{th} person? (If the first success is the fifth person, then we say $n=5$.)

[7] This is hypothetical since, in reality, this sort of study probably would not be permitted any longer under current ethical standards.

Geometric Distribution [2]

Dr. Smith's Experiment

- The probability of stopping after the first person is just the chance the first person will not administer the worst shock: $1 - 0.65 = 0.35$.

Geometric Distribution [2]

Dr. Smith's Experiment

- The probability of stopping after the first person is just the chance the first person will not administer the worst shock: $1 - 0.65 = 0.35$.
- The probability it will be the second person is

Geometric Distribution [2]

Dr. Smith's Experiment

- The probability of stopping after the first person is just the chance the first person will not administer the worst shock: $1 - 0.65 = 0.35$.
- The probability it will be the second person is

$$P(\text{second person is the first to not administer the worst shock})$$
$$= P(\text{the first will, the second won't}) = (0.65)(0.35) = 0.228$$

Geometric Distribution [2]

Dr. Smith's Experiment

- The probability of stopping after the first person is just the chance the first person will not administer the worst shock: $1 - 0.65 = 0.35$.
- The probability it will be the second person is

$$
\begin{aligned}
&P(\text{second person is the first to not administer the worst shock}) \\
&\quad = P(\text{the first will, the second won't}) = (0.65)(0.35) = 0.228
\end{aligned}
$$

- Likewise, the probability it will be the third person is $(0.65)(0.65)(0.35) = 0.148$.

Geometric Distribution [2]

Dr. Smith's Experiment

- The probability of stopping after the first person is just the chance the first person will not administer the worst shock: $1 - 0.65 = 0.35$.
- The probability it will be the second person is

$$P(\text{second person is the first to not administer the worst shock}) \\ = P(\text{the first will, the second won't}) = (0.65)(0.35) = 0.228$$

- Likewise, the probability it will be the third person is $(0.65)(0.65)(0.35) = 0.148$.
- If the first success is on the n^{th} person, then there are $n - 1$ failures and finally 1 success, which corresponds to the probability $(0.65)^{n-1}(0.35)$.

Geometric Distribution [2]

Dr. Smith's Experiment

- The probability of stopping after the first person is just the chance the first person will not administer the worst shock: $1 - 0.65 = 0.35$.
- The probability it will be the second person is

$$
\begin{aligned}
&P(\text{second person is the first to not administer the worst shock}) \\
&\quad = P(\text{the first will, the second won't}) = (0.65)(0.35) = 0.228
\end{aligned}
$$

- Likewise, the probability it will be the third person is $(0.65)(0.65)(0.35) = 0.148$.
- If the first success is on the n^{th} person, then there are $n - 1$ failures and finally 1 success, which corresponds to the probability $(0.65)^{n-1}(0.35)$.
- This is the same as $(1 - 0.35)^{n-1}(0.35)$.

Geometric Distribution [2]

Dr. Smith's Experiment

- Previous example illustrates what is called the geometric distribution, which describes the waiting time until a success for **independent and identically distributed (iid)** Bernoulli random variables.

Geometric Distribution [2]

Dr. Smith's Experiment

- Previous example illustrates what is called the geometric distribution, which describes the waiting time until a success for **independent and identically distributed (iid)** Bernoulli random variables.
- In this case, the independence aspect just means the individuals in the example don't affect each other, and identical means they each have the same probability of success.

- Previous example illustrates what is called the geometric distribution, which describes the waiting time until a success for **independent and identically distributed (iid)** Bernoulli random variables.
- In this case, the independence aspect just means the individuals in the example don't affect each other, and identical means they each have the same probability of success.
- The geometric distribution from the previous example is shown in Figure 17.

- Previous example illustrates what is called the geometric distribution, which describes the waiting time until a success for **independent and identically distributed (iid)** Bernoulli random variables.
- In this case, the independence aspect just means the individuals in the example don't affect each other, and identical means they each have the same probability of success.
- The geometric distribution from the previous example is shown in Figure 17.
- In general, the probabilities for a geometric distribution decrease **exponentially** fast.

Geometric Distribution [2]

Dr. Smith's Experiment

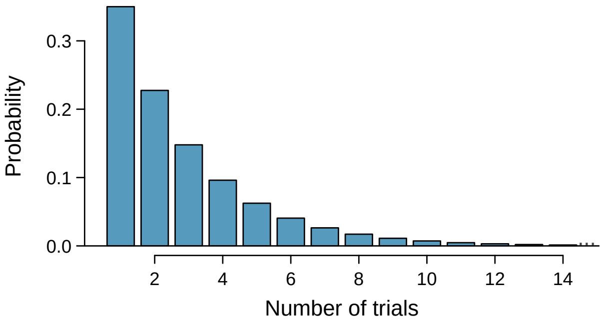

Figure 17 : The geometric distribution when the probability of success is $p = 0.35$.

Geometric Distribution [2]

Dr. Smith's Experiment

Geometric Distribution

If the probability of a success in one trial is p and the probability of a failure is $1-p$, then the probability of finding the first success in the n^{th} trial is given by

Geometric Distribution

If the probability of a success in one trial is p and the probability of a failure is $1 - p$, then the probability of finding the first success in the n^{th} trial is given by

$$(1 - p)^{n-1}p \tag{12}$$

Geometric Distribution

If the probability of a success in one trial is p and the probability of a failure is $1 - p$, then the probability of finding the first success in the n^{th} trial is given by

$$(1-p)^{n-1}p \tag{12}$$

The mean (i.e. expected value), variance, and standard deviation of this wait time are given by

Geometric Distribution [2]

Dr. Smith's Experiment

Geometric Distribution

If the probability of a success in one trial is p and the probability of a failure is $1 - p$, then the probability of finding the first success in the n^{th} trial is given by

$$(1 - p)^{n-1} p \tag{12}$$

The mean (i.e. expected value), variance, and standard deviation of this wait time are given by

$$\mu = \frac{1}{p} \qquad \sigma^2 = \frac{1-p}{p^2} \qquad \sigma = \sqrt{\frac{1-p}{p^2}} \tag{13}$$

Outline

Binomial Distribution [2]

Binomial Distribution

Suppose the probability of a single trial being a success is p. Then the probability of observing exactly k successes in n independent trials is given by

Binomial Distribution [2]

Binomial Distribution

Suppose the probability of a single trial being a success is p. Then the probability of observing exactly k successes in n independent trials is given by

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \tag{14}$$

Binomial Distribution [2]

Binomial Distribution

Suppose the probability of a single trial being a success is p. Then the probability of observing exactly k successes in n independent trials is given by

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \tag{14}$$

Additionally, the mean, variance, and standard deviation of the number of observed successes are

Binomial Distribution [2]

Binomial Distribution

Suppose the probability of a single trial being a success is p. Then the probability of observing exactly k successes in n independent trials is given by

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \tag{14}$$

Additionally, the mean, variance, and standard deviation of the number of observed successes are

$$\mu = np \qquad \sigma^2 = np(1-p) \qquad \sigma = \sqrt{np(1-p)} \tag{15}$$

Binomial Distribution [2]

To check if a distribution is binomial or not, check for the following conditions:

1. The trials are independent.

Binomial Distribution [2]

To check if a distribution is binomial or not, check for the following conditions:

1. The trials are independent.
2. The number of trials, n, is fixed.

Binomial Distribution [2]

To check if a distribution is binomial or not, check for the following conditions:

1. The trials are independent.
2. The number of trials, n, is fixed.
3. Each trial outcome can be classified as a success or failure.

Binomial Distribution [2]

To check if a distribution is binomial or not, check for the following conditions:

1. The trials are independent.
2. The number of trials, n, is fixed.
3. Each trial outcome can be classified as a success or failure.
4. The probability of a success, p, is the same for each trial.

Outline

Random Number Generator

- Random number generators is an essential part in computer simulation.

Random Number Generator

- Random number generators is an essential part in computer simulation.
- By using several randomly generated numbers, we can expect different results from our experiments.

Random Number Generator

- Random number generators is an essential part in computer simulation.
- By using several randomly generated numbers, we can expect different results from our experiments.
- The results then can be examined, compared, and analyzed to see the differences between researched methods.

Random Number Generator

- Random number generators is an essential part in computer simulation.
- By using several randomly generated numbers, we can expect different results from our experiments.
- The results then can be examined, compared, and analyzed to see the differences between researched methods.
- (Pseudo) Random numbers are generated by computer algorithm (e.g Mersenne-Twister).

Random Number Generator

- Random number generators is an essential part in computer simulation.
- By using several randomly generated numbers, we can expect different results from our experiments.
- The results then can be examined, compared, and analyzed to see the differences between researched methods.
- (Pseudo) Random numbers are generated by computer algorithm (e.g Mersenne-Twister).
- To re-obtain the results, we pass <u>seed</u> into the random number generators.

Random Number Generator

- Random number generators is an essential part in computer simulation.
- By using several randomly generated numbers, we can expect different results from our experiments.
- The results then can be examined, compared, and analyzed to see the differences between researched methods.
- (Pseudo) Random numbers are generated by computer algorithm (e.g Mersenne-Twister).
- To re-obtain the results, we pass <u>seed</u> into the random number generators.
- Programming languages commonly provide random number generators.

Random Number Generator

- Random number generators is an essential part in computer simulation.
- By using several randomly generated numbers, we can expect different results from our experiments.
- The results then can be examined, compared, and analyzed to see the differences between researched methods.
- (Pseudo) Random numbers are generated by computer algorithm (e.g Mersenne-Twister).
- To re-obtain the results, we pass <u>seed</u> into the random number generators.
- Programming languages commonly provide random number generators.
- In today's lecture we are going to use R to start exploring RNG.

Random Numbers in R

- R could produce random numbers with defined probability distribution.

Random Numbers in R

- R could produce random numbers with defined probability distribution.
- Important things we should keep in mind are:

Random Numbers in R

- R could produce random numbers with defined probability distribution.
- Important things we should keep in mind are:
  - The functions to produce random numbers are started with `r` (e.g. `runif()`, `rnorm()`).

Random Numbers in R

- R could produce random numbers with defined probability distribution.
- Important things we should keep in mind are:
  - The functions to produce random numbers are started with `r` (e.g. `runif()`, `rnorm()`).
  - The theoretical densities can be produced by the functions that started with `d` (e.g. `dnorm()`).

Random Numbers in R

- R could produce random numbers with defined probability distribution.
- Important things we should keep in mind are:
  - The functions to produce random numbers are started with `r` (e.g. `runif(), rnorm()`).
  - The theoretical densities can be produced by the functions that started with `d` (e.g. `dnorm()`).
  - The cumulative distribution functions are produced by the functions that started with `p` (e.g. `pnorm()`).

Random Numbers in R

- R could produce random numbers with defined probability distribution.
- Important things we should keep in mind are:
  - The functions to produce random numbers are started with `r` (e.g. `runif()`, `rnorm()`).
  - The theoretical densities can be produced by the functions that started with `d` (e.g. `dnorm()`).
  - The cumulative distribution functions are produced by the functions that started with `p` (e.g. `pnorm()`).
- Remember that we can plot the frequency and/or the distribution in a histogram, density plot, boxplot, and stripchart.

Outline

Generating Normally Distributed Random Numbers in R

- Normally distributed random numbers in R are generated using the `rnorm()` function.

Generating Normally Distributed Random Numbers in R

- Normally distributed random numbers in R are generated using the `rnorm()` function.

```
rnorm(n = 5, mean = 2.5, sd = 1)

## [1] 2.3 1.0 2.0 1.5 1.8
```

Generating Normally Distributed Random Numbers in R

- Normally distributed random numbers in R are generated using the `rnorm()` function.

```
rnorm(n = 5, mean = 2.5, sd = 1)

## [1] 2.3 1.0 2.0 1.5 1.8
```

- The mandatory parameters are:

Generating Normally Distributed Random Numbers in R

- Normally distributed random numbers in R are generated using the `rnorm()` function.

```r
rnorm(n = 5, mean = 2.5, sd = 1)

## [1] 2.3 1.0 2.0 1.5 1.8
```

- The mandatory parameters are:
  - Number of generated numbers (`n`)

Generating Normally Distributed Random Numbers in R

- Normally distributed random numbers in R are generated using the `rnorm()` function.

```
rnorm(n = 5, mean = 2.5, sd = 1)

## [1] 2.3 1.0 2.0 1.5 1.8
```

- The mandatory parameters are:
    - Number of generated numbers (`n`)
    - Mean of the generated numbers (`mean`)

Generating Normally Distributed Random Numbers in R

- Normally distributed random numbers in R are generated using the `rnorm()` function.

```
rnorm(n = 5, mean = 2.5, sd = 1)

## [1] 2.3 1.0 2.0 1.5 1.8
```

- The mandatory parameters are:
  - Number of generated numbers (`n`)
  - Mean of the generated numbers (`mean`)
  - Standard deviation of the generated numbers (`sd`)

Generating Normally Distributed Random Numbers in R

- Normally distributed random numbers in R are generated using the `rnorm()` function.

```
rnorm(n = 5, mean = 2.5, sd = 1)

## [1] 2.3 1.0 2.0 1.5 1.8
```

- The mandatory parameters are:
  - Number of generated numbers (`n`)
  - Mean of the generated numbers (`mean`)
  - Standard deviation of the generated numbers (`sd`)
- We can have different results by using different seeds.

Generating Normally Distributed Random Numbers in R

- Normally distributed random numbers in R are generated using the `rnorm()` function.

```
rnorm(n = 5, mean = 2.5, sd = 1)

## [1] 2.3 1.0 2.0 1.5 1.8
```

- The mandatory parameters are:
  - Number of generated numbers (`n`)
  - Mean of the generated numbers (`mean`)
  - Standard deviation of the generated numbers (`sd`)
- We can have different results by using different seeds.

```
set.seed(1)
rnorm(n = 5, mean = 2.5, sd = 1)

## [1] 1.9 2.7 1.7 4.1 2.8

set.seed(2)
rnorm(n = 5, mean = 2.5, sd = 1)

## [1] 1.6 2.7 4.1 1.4 2.4
```

Generating Normally Distributed Random Numbers in R

- On the other hand, we can retain the same same results with the same seed.

Generating Normally Distributed Random Numbers in R

- On the other hand, we can retain the same same results with the same seed.

```
set.seed(1)
rnorm(n = 5, mean = 2.5, sd = 1)

## [1] 1.9 2.7 1.7 4.1 2.8

set.seed(2)
rnorm(n = 5, mean = 2.5, sd = 1)

## [1] 1.6 2.7 4.1 1.4 2.4

set.seed(1)
rnorm(n = 5, mean = 2.5, sd = 1)

## [1] 1.9 2.7 1.7 4.1 2.8
```

Generating Normally Distributed Random Numbers in R

- More examples:

Generating Normally Distributed Random Numbers in R

- More examples:

```
rnorm(n = 5, mean = 2.5, sd = 1)

## [1] 1.7 3.0 3.2 3.1 2.2

rnorm(n = 10, mean = 100, sd = 1)

##  [1] 102 100  99  98 101 100 100 101 101 101

rnorm(n = 20, mean = 100, sd = 50)

##  [1] 145.95 139.11 103.73   0.53 130.99  97.19  92.21  26
##  [9]  76.09 120.90
##  [ reached getOption("max.print") -- omitted 10 entries ]
```

Generating Normally Distributed Random Numbers in R

- More examples:

```
rnorm(n = 5, mean = 2.5, sd = 1)

## [1] 1.7 3.0 3.2 3.1 2.2

rnorm(n = 10, mean = 100, sd = 1)

##  [1] 102 100  99  98 101 100 100 101 101 101

rnorm(n = 20, mean = 100, sd = 50)

##  [1] 145.95 139.11 103.73   0.53 130.99  97.19  92.21  26
##  [9]  76.09 120.90
##  [ reached getOption("max.print") -- omitted 10 entries ]
```

- As usual, we can assign the output to a variable:

Generating Normally Distributed Random Numbers in R

- More examples:

```
rnorm(n = 5, mean = 2.5, sd = 1)
```

```
## [1] 1.7 3.0 3.2 3.1 2.2
```

```
rnorm(n = 10, mean = 100, sd = 1)
```

```
##  [1] 102 100  99  98 101 100 100 101 101 101
```

```
rnorm(n = 20, mean = 100, sd = 50)
```

```
##  [1] 145.95 139.11 103.73   0.53 130.99  97.19  92.21  26
##  [9]  76.09 120.90
##  [ reached getOption("max.print") -- omitted 10 entries ]
```

- As usual, we can assign the output to a variable:

```
norm.dist.1000 <- rnorm(n = 1000, mean = 2.5, sd = 1)
norm.dist.1000
```

```
##  [1] 2.3 2.2 3.2 3.1 1.8 1.8 2.9 3.3 2.4 3.4
##  [ reached getOption("max.print") -- omitted 990 entries
```

Generating Normally Distributed Random Numbers in R

- Remember that we can plot the distribution in a histogram.

Generating Normally Distributed Random Numbers in R

- Remember that we can plot the distribution in a histogram.
- Let us create a histogram for **norm.dist.1000** data.

Generating Normally Distributed Random Numbers in R

- Remember that we can plot the distribution in a histogram.
- Let us create a histogram for **norm.dist.1000** data.

```r
hist(norm.dist.1000, col = "pink", main = NULL)
```

Generating Normally Distributed Random Numbers in R

- Remember that we can plot the distribution in a histogram.
- Let us create a histogram for **norm.dist.1000** data.

```
hist(norm.dist.1000, col = "pink", main = NULL)
```

- The produced plot is shown in Figure 18

Generating Normally Distributed Random Numbers in R

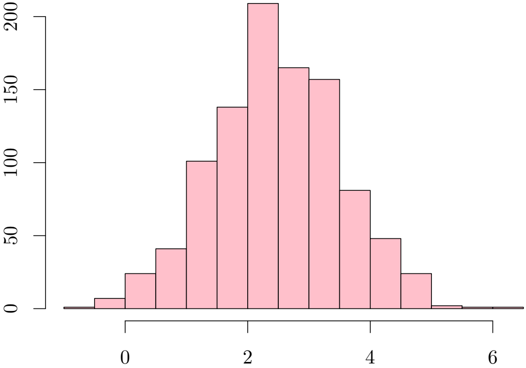

Figure 18 : Frequency histogram for 1000 randomly generated numbers ($\sigma = 1; \mu = 2.5$)

Generating Normally Distributed Random Numbers in R

- Let us generate more of normally distributed random number with $\sigma = 1; \mu = 2.5$ and plot its frequency histogram (Figure 19).

Generating Normally Distributed Random Numbers in R

- Let us generate more of normally distributed random number with $\sigma = 1; \mu = 2.5$ and plot its frequency histogram (Figure 19).

```
norm.dist.100000 <- rnorm(n = 1e+05, mean = 2.5, sd = 1)
hist(norm.dist.100000, col = "navyblue", main = NULL)
```

Generating Normally Distributed Random Numbers in R

- Let us generate more of normally distributed random number with $\sigma = 1; \mu = 2.5$ and plot its frequency histogram (Figure 19).

```
norm.dist.100000 <- rnorm(n = 1e+05, mean = 2.5, sd = 1)
hist(norm.dist.100000, col = "navyblue", main = NULL)
```

- We can 'play' with the breaks in histogram by passing the `break=` parameter (Figure 20).

Generating Normally Distributed Random Numbers in R

- Let us generate more of normally distributed random number with $\sigma = 1; \mu = 2.5$ and plot its frequency histogram (Figure 19).

```
norm.dist.100000 <- rnorm(n = 1e+05, mean = 2.5, sd = 1)
hist(norm.dist.100000, col = "navyblue", main = NULL)
```

- We can 'play' with the breaks in histogram by passing the `break=` parameter (Figure 20).

```
breaks <- seq(
        from = min(norm.dist.100000)-0.1,
        to = max(norm.dist.100000)+0.1,
        by = 0.1
)
hist(norm.dist.100000, col='lightblue', main = NULL, breaks = break
```

Generating Normally Distributed Random Numbers in R

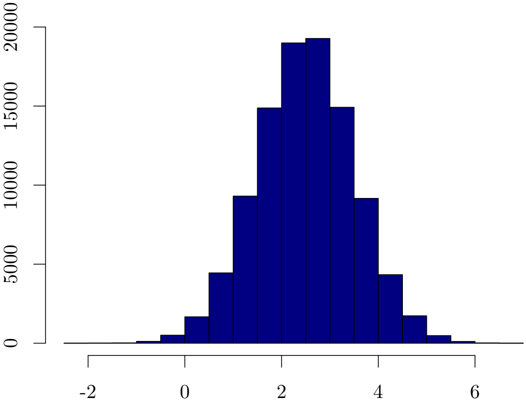

Figure 19 : Frequency histogram for 100,000 randomly generated numbers ($\sigma = 1; \mu = 2.5$)

Generating Normally Distributed Random Numbers in R

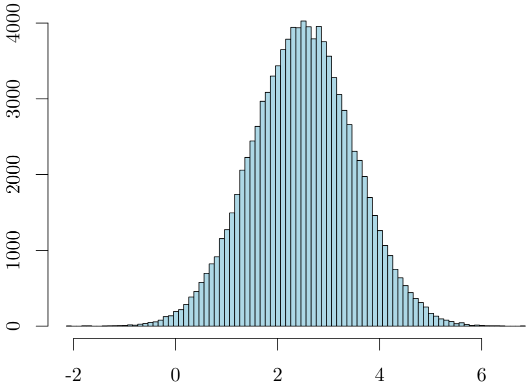

Figure 20 : Frequency histogram for 100,000 randomly generated numbers ($\sigma = 1; \mu = 2.5$). The bar width is 0.1.

Add boxplot and stripchart

```
nf <- layout(mat = matrix(c(1, 2), 2, 1, byrow = T), height = c(1,
    2))
old.mar <- par(mar = c(4, 4, 0, 1) + 0.1)
boxplot(norm.dist.100000, horizontal = T, outline = T, frame = F,
    col = "lightblue")
hist(norm.dist.100000, col = "lightblue", main = NULL, breaks = breaks)
stripchart(norm.dist.100000, col = "navyblue", pch = 19, method = "stac
    add = T)
par(old.mar)
```

Generating Normally Distributed Random Numbers in R

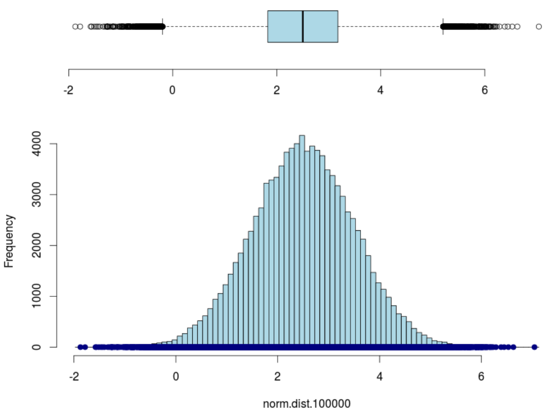

Figure 21 : A histogram with boxplot and stripchart of 100,000 randomly generated numbers ($\sigma = 1; \mu = 2.5$)

Generating Normally Distributed Random Numbers in R

- As described earlier, the `dnorm()` function is used to calculate the theoretical density of the data.

Generating Normally Distributed Random Numbers in R

- As described earlier, the `dnorm()` function is used to calculate the theoretical density of the data.

```
dnorm(norm.dist.100000)
```

```
##  [1] 4.8e-03 8.4e-03 1.3e-01 1.2e-01 2.3e-05 1.3e-03 2.1e-03
##  [8] 1.2e-02 4.6e-03 1.2e-02
##  [ reached getOption("max.print") -- omitted 99990 entries ]
```

Generating Normally Distributed Random Numbers in R

- As described earlier, the `dnorm()` function is used to calculate the theoretical density of the data.

```
dnorm(norm.dist.100000)
```

```
##  [1] 4.8e-03 8.4e-03 1.3e-01 1.2e-01 2.3e-05 1.3e-03 2.1e-03
##  [8] 1.2e-02 4.6e-03 1.2e-02
##  [ reached getOption("max.print") -- omitted 99990 entries ]
```

- By utilizing the output of `dnorm()`, we can plot the Probability Density Function (PDF).

Generating Normally Distributed Random Numbers in R

- As described earlier, the `dnorm()` function is used to calculate the theoretical density of the data.

```
dnorm(norm.dist.100000)

##  [1] 4.8e-03 8.4e-03 1.3e-01 1.2e-01 2.3e-05 1.3e-03 2.1e-03
##  [8] 1.2e-02 4.6e-03 1.2e-02
##  [ reached getOption("max.print") -- omitted 99990 entries ]
```

- By utilizing the output of `dnorm()`, we can plot the Probability Density Function (PDF).

```
hist(norm.dist.100000, col='lightblue', main = NULL, probability=T)
curve(
        dnorm(x, mean(norm.dist.100000), sd(norm.dist.100000)),
        add=T,
        col='red',
        lwd=4
)
```

Generating Normally Distributed Random Numbers in R

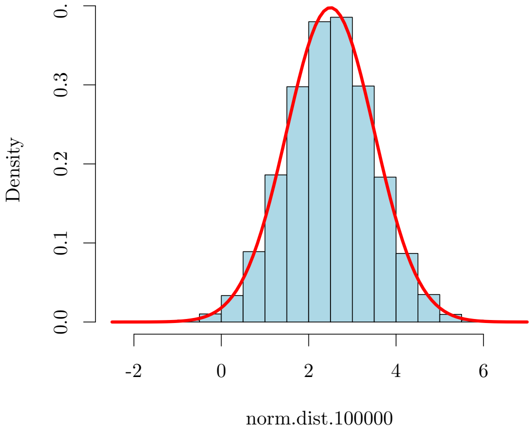

Figure 22 : Probability histogram with PDF plot for 100,000 randomly generated numbers ($\sigma = 1; \mu = 2.5$).

Generating Normally Distributed Random Numbers in R

- We could make use the `pnorm()` function to have the probability distribution

Generating Normally Distributed Random Numbers in R

- We could make use the `pnorm()` function to have the probability distribution

```
pnorm(norm.dist.100000)

##  [1] 1.00 1.00 0.94 0.94 1.00 1.00 1.00 1.00 1.00 1.00
##  [ reached getOption("max.print") -- omitted 99990 entries ]
```

Generating Normally Distributed Random Numbers in R

- We could make use the `pnorm()` function to have the probability distribution

```
pnorm(norm.dist.100000)

##  [1] 1.00 1.00 0.94 0.94 1.00 1.00 1.00 1.00 1.00 1.00
##  [ reached getOption("max.print") -- omitted 99990 entries ]
```

- By utilizing the output of `pnorm()`, we can plot the Cumulative Distribution Function (CDF).

Generating Normally Distributed Random Numbers in R

- We could make use the `pnorm()` function to have the probability distribution

```
pnorm(norm.dist.100000)

##  [1] 1.00 1.00 0.94 0.94 1.00 1.00 1.00 1.00 1.00 1.00
##  [ reached getOption("max.print") -- omitted 99990 entries ]
```

- By utilizing the output of `pnorm()`, we can plot the Cumulative Distribution Function (CDF).

```
curve(
        pnorm(
                x,
                mean(norm.dist.100000),
                sd(norm.dist.100000)
        ),
        min(norm.dist.100000),
        max(norm.dist.100000),
        xaxs='i', yaxs='i',
        ylab='Fn(x)',
        col='navyblue'
)
```

Generating Normally Distributed Random Numbers in R

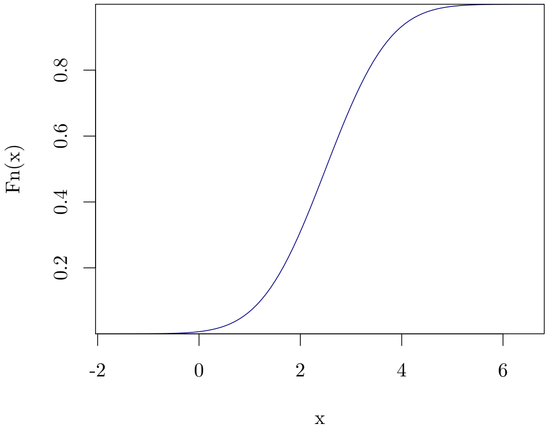

Figure 23 : CDF plot for 100,000 randomly generated numbers ($\sigma = 1; \mu = 2.5$).

Outline

Generating Uniformly Distributed Random Numbers in R

- The pattern of functions in the normal distribution family is the same for all types of distribution in R.

Generating Uniformly Distributed Random Numbers in R

- The pattern of functions in the normal distribution family is the same for all types of distribution in R.
- Let us preserve the range of data from our random normal example and create a uniformly distributed numbers.

Generating Uniformly Distributed Random Numbers in R

- The pattern of functions in the normal distribution family is the same for all types of distribution in R.
- Let us preserve the range of data from our random normal example and create a uniformly distributed numbers.

```
unif.dist.100000 <- runif(
        n=length(norm.dist.100000),
        min=min(norm.dist.100000),
        max=max(norm.dist.100000)
)
unif.dist.100000
```

```
##  [1]  6.079  3.931  6.360 -1.653  4.572  1.171  4.440  4.231
##  [9] -0.883  0.018
##  [ reached getOption("max.print") -- omitted 99990 entries ]
```

Generating Uniformly Distributed Random Numbers in R

- The pattern of functions in the normal distribution family is the same for all types of distribution in R.
- Let us preserve the range of data from our random normal example and create a uniformly distributed numbers.

```
unif.dist.100000 <- runif(
        n=length(norm.dist.100000),
        min=min(norm.dist.100000),
        max=max(norm.dist.100000)
)
unif.dist.100000
```

```
##  [1]  6.079  3.931  6.360 -1.653  4.572  1.171  4.440  4.231
##  [9] -0.883  0.018
##  [ reached getOption("max.print") -- omitted 99990 entries ]
```

- The mandatory parameters are:

Generating Uniformly Distributed Random Numbers in R

- The pattern of functions in the normal distribution family is the same for all types of distribution in R.
- Let us preserve the range of data from our random normal example and create a uniformly distributed numbers.

```r
unif.dist.100000 <- runif(
        n=length(norm.dist.100000),
        min=min(norm.dist.100000),
        max=max(norm.dist.100000)
)
unif.dist.100000
```

```
##  [1]  6.079  3.931  6.360 -1.653  4.572  1.171  4.440  4.231
##  [9] -0.883  0.018
##  [ reached getOption("max.print") -- omitted 99990 entries ]
```

- The mandatory parameters are:
    - Number of generated numbers (`n`)

Generating Uniformly Distributed Random Numbers in R

- The pattern of functions in the normal distribution family is the same for all types of distribution in R.
- Let us preserve the range of data from our random normal example and create a uniformly distributed numbers.

```
unif.dist.100000 <- runif(
        n=length(norm.dist.100000),
        min=min(norm.dist.100000),
        max=max(norm.dist.100000)
)
unif.dist.100000
```

```
##  [1]  6.079  3.931  6.360 -1.653  4.572  1.171  4.440  4.231
##  [9] -0.883  0.018
##  [ reached getOption("max.print") -- omitted 99990 entries ]
```

- The mandatory parameters are:
  - Number of generated numbers (`n`)
  - The lower and upper limit (`min, max`)

Generating Uniformly Distributed Random Numbers in R

As we did earlier, let us plot the histogram, along with the PDF. Do not forget that this time we use `dunif()` instead of `dnorm()`.

```
hist(
        unif.dist.100000,
        col='lightgreen',
        main = NULL,
        probability=T,
        breaks = breaks
)
curve(
        dunif(
                x,
                min=min(norm.dist.100000),
                max=max(norm.dist.100000)
        ),
        add=T,
        col='red',
        lwd=4
)
```

Generating Uniformly Distributed Random Numbers in R

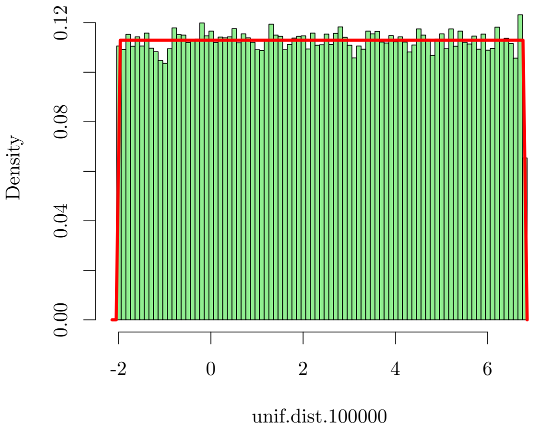

Figure 24 : Probability histogram with PDF plot for 100,000 randomly generated numbers with uniform distribution.

Generating Uniformly Distributed Random Numbers in R

And at last, the CDF by using `punif()`.

```
curve(
        punif(
                x,
                min=min(norm.dist.100000),
                max=max(norm.dist.100000)
        ),
        min(norm.dist.100000),
        max(norm.dist.100000),
        xaxs='i', yaxs='i',
        ylab='Fn(x)',
        col='chartreuse4'
)
```

Generating Uniformly Distributed Random Numbers in R

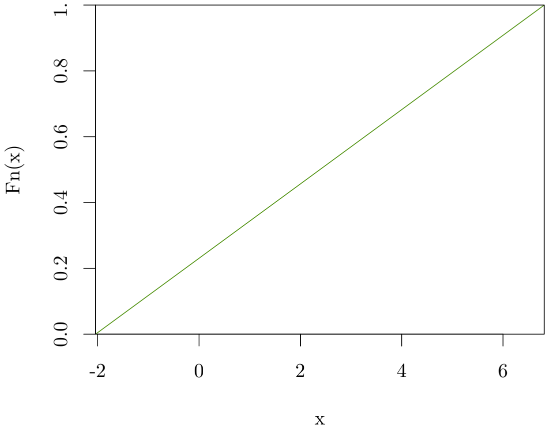

Figure 25 : CDF plot for 100,000 randomly generated numbers with uniform distribution.

Outline

Exercise 5

- Use your Student Number as seed for randomization.

Exercise 5

- Use your Student Number as seed for randomization.
- Create a numeric vector of 1 million numbers of

Exercise 5

- Use your Student Number as seed for randomization.
- Create a numeric vector of 1 million numbers of
  - Normal distribution

Exercise 5

- Use your Student Number as seed for randomization.
- Create a numeric vector of 1 million numbers of
  - Normal distribution
  - Uniform distribution

Exercise 5

- Use your Student Number as seed for randomization.
- Create a numeric vector of 1 million numbers of
  - Normal distribution
  - Uniform distribution
- Plot histogram, boxplot, stripchart, PDF, and CDF for each distribution.

Exercise 5

- Use your Student Number as seed for randomization.
- Create a numeric vector of 1 million numbers of
  - Normal distribution
  - Uniform distribution
- Plot histogram, boxplot, stripchart, PDF, and CDF for each distribution.
- Repeat the whole process with binomial distribution, geometric distribution, and exponential distribution.

Exercise 5

- Use your Student Number as seed for randomization.
- Create a numeric vector of 1 million numbers of
  - Normal distribution
  - Uniform distribution
- Plot histogram, boxplot, stripchart, PDF, and CDF for each distribution.
- Repeat the whole process with binomial distribution, geometric distribution, and exponential distribution.
- Compare the shape of the plots from all types of distribution.

Exercise 5

- Use your Student Number as seed for randomization.
- Create a numeric vector of 1 million numbers of
  - Normal distribution
  - Uniform distribution
- Plot histogram, boxplot, stripchart, PDF, and CDF for each distribution.
- Repeat the whole process with binomial distribution, geometric distribution, and exponential distribution.
- Compare the shape of the plots from all types of distribution.
- Create a plot with histogram from all distributions[8]. Repeat for boxplot, CDF and PDF.

Exercise 5

- Use your Student Number as seed for randomization.
- Create a numeric vector of 1 million numbers of
  - Normal distribution
  - Uniform distribution
- Plot histogram, boxplot, stripchart, PDF, and CDF for each distribution.
- Repeat the whole process with binomial distribution, geometric distribution, and exponential distribution.
- Compare the shape of the plots from all types of distribution.
- Create a plot with histogram from all distributions[8]. Repeat for boxplot, CDF and PDF.

[8] Hint: Explore the R's graphical paramaters, especially `mfrow`.

Next...

- Inference Statistics, numerical data.

Next...

- Inference Statistics, numerical data.
- For preparation read Chapters 4 and 5 from [2].

Outline

References I

[1] P. S. Mann, Introductory Statistics, 7th ed. NJ: John Wiley & Sons, Inc., 2010.

[2] D. Diez, C. Barr, and M. Çetinkaya-Rundel, OpenIntro Statistics. OpenIntro, Incorporated, 2015.