Relevant videos for practicing for the Mathematics entrance test for the LLM Master in Law and Finance programme

- https://www.youtube.com/playlist?list=PLM-mb7IpX4moTbG2Pe96z2JcQcmIeFYIr
  A playlist with mathematics test training videos

- https://www.youtube.com/watch?v=mk8tOD0t8M0
  A very simple video, using numerical examples to explain Mode, Median, Mean, Range, and Standard Deviation

- https://www.youtube.com/watch?v=qqOyy_NjflU
  A simple video showing you how to calculate Standard Deviation and Variance, providing a numerical example with 6 observations. (This video uses a sample variance formula)

- https://www.youtube.com/watch?v=sOb9b_AtwDg
  This video can help you distinguish sample and population variance. More theories are also explained in this video, telling you what variance actually is

- https://www.khanacademy.org/math/probability/probability-geometry#probability-basics
  This video shows the basics of probabilities

- https://www.youtube.com/watch?v=OvTEhNL96v0
  This video shows information about Expected Value and Variance of Discrete Random Variables with numerical examples

- https://www.khanacademy.org/math/ap-statistics/random-variables-ap/discrete-random-variables/v/variance-and-standard-deviation-of-a-discrete-random-variable
  This video shows a numerical example of calculating variance and a standard deviation of a discrete random variable (5 variables)

Below you will find “Introduction to statistics and probability theory” slides for math test preparation purposes

Also note the *Trial Exam* on our website which is of course strongly recommended in order to practice for the math test

Introduction to statistics and probability theory

This set of notes is based on Chapters 1,4, 6, 7 of *Keller, Gerald, Statistics for Management and Economics,* 9th edition. ISBN: 978-1111527327.

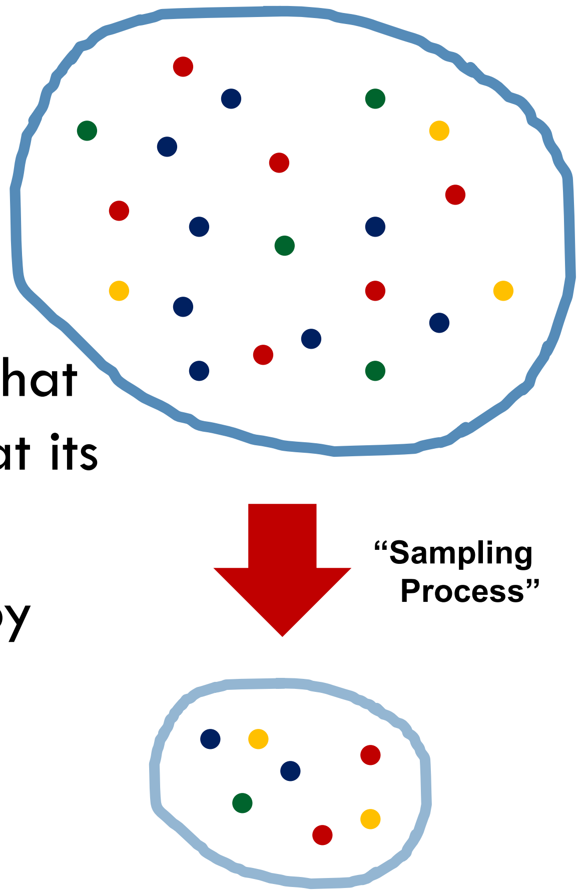
SOME BASICS

Some Basics

- **1. Some Basics**
  - 1.1 Statistics vs. Statistic
    - **Statistics** as a field of mathematics dealing with the collection, explanation and interpretation of data.
    - **A statistic** is a measure that tries to capture some information about the data set.
      - Examples: mean, median, standard deviation, correlation, covariance, sample estimate of the population mean, …
      - (The plural form of statistic is also called “statistics”!)

Some Basics

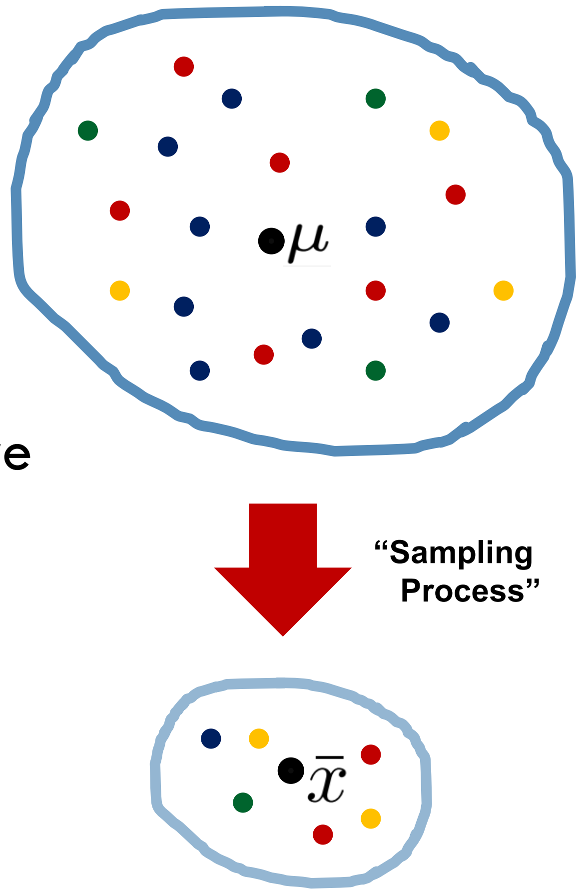
- 1. Some Basics
  - 1.2 Population vs. Sample
    - The **population** of a study is the group of all items of interest in that study.
    - A **sample** is a subset of the population that is studied. (A desired characteristic is that its data was obtained **randomly**.)
    - The "**sampling process**" is the method by which we select observations from the population to arrive at a sample.
      - Typically some form of random sampling.



Some Basics

- 1. Some Basics
  - 1.3 Parameters vs. Statistics
    - Descriptive measures about the population (e.g., the population mean) are called **parameters.**
    - Descriptive measures about a sample are called **statistics.**
      - We typically use latin-based letters for sample statistics.
      - Example: the (sample) mean

$\bar{x}$ (pronounced “x-bar”)



Some Basics

- 1. **Some Basics**
  - 1.4 Data Types

    We differentiate between the following data types:

    - **1. INTERVAL or QUANTITATIVE data**
      - Real numbers
      - Distances between values with intrinsic meaning.
      - Examples:
        - On previous slide: Time in seconds

          (distance meaningful: 60 seconds twice as long as 30 seconds)
        - Height of students
        - Length of cars
        - …

Some Basics

- **1. Some Basics**
  - 1.4 Data Types

    We differentiate between the following data types:

    - **1. INTERVAL or QUANTITATIVE data**
    - **2. ORDINAL data**
      - An ordered ranking among data exists.
      - Distances do no have intrinsic meaning.
      - Ex. on previous slide : Song rating: {bad, average, good, excellent}
        - We can assign numbers to each value, but we need to maintain the order, e.g., bad=1, average=2, good=3, excellent=4.
        - Also possible : bad=0, average=23; good=34; excellent=100; distances between values do not matter.

Some Basics

- **1. Some Basics**
  - 1.4 Data Types

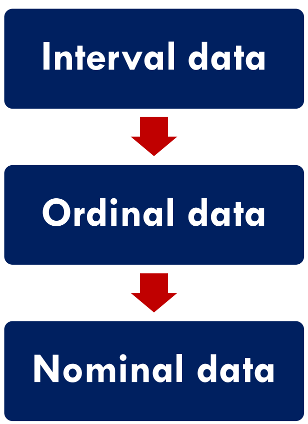
    We differentiate between the following data types:
    - **1. INTERVAL or QUANTITATIVE data**
    - **2. ORDINAL data**
    - **3. NOMINAL or CATEGORICAL data**
      - Values have no order, nor any intrinsic numerical value; any number can be applied to represent a value.
      - Examples:
        - On previous slide: Artist, Album, Genre
        - Family status = {single, married, divorced, widowed}
        - Gender = {male, female} → No difference b/w {male=0, female=1} or having {male=1, female=0}

Some Basics

- **1. Some Basics**
  - 1.4 Data Types

    In fact there is a **hierarchy among data types:**

    - Example: Exam scores



- Exam scores (interval data) is often compressed into letter grades (ordinal data): 94 points = A
- Letter grades (ordinal data) can be further compressed into simple pass/fail categories (nominal data): A = Passed.
- Note: Moving from a higher level to the lower level, we lose information, which we cannot undo, so going back from a lower level to a higher level is not any more possible:
  - Student had a B → How many points did she have?
  - Student passed → Which letter grade/score did he have?

DESCRIPTIVE STATISTICS

Descriptive Statistics

- 1. What is Descriptive Statistics?
  - **Question:**
    - Suppose you have some data set and you would like to tell someone about it without having to show him each single data point. What do you do?
    - (Or, the data set is simply so large that you get a headache from looking at thousands of values.)
  - **Answer:**
    - You may try to come up with some summary statistics that somehow **describe** the data set.

Simply put, all that is descriptive statistics.

Descriptive Statistics

- 1. What is Descriptive Statistics?
  - Examples:
    - **Measures of Central Location**
      - Mean (arithmetic, geometric), Mode, Median
    - **Measures of Dispersion** (or Spread, Variability)
      - Range, Interquartile Range, Mean Absolute Deviation (MAD)
      - Standard Deviation, Variance
      - Coefficient of Variation
    - **Measures of Relative Standing**
      - Percentiles (Quartiles, Quintiles, Deciles)
      - z-Score
    - **Measures of Linear Relationships**
      - Covariance, Correlation

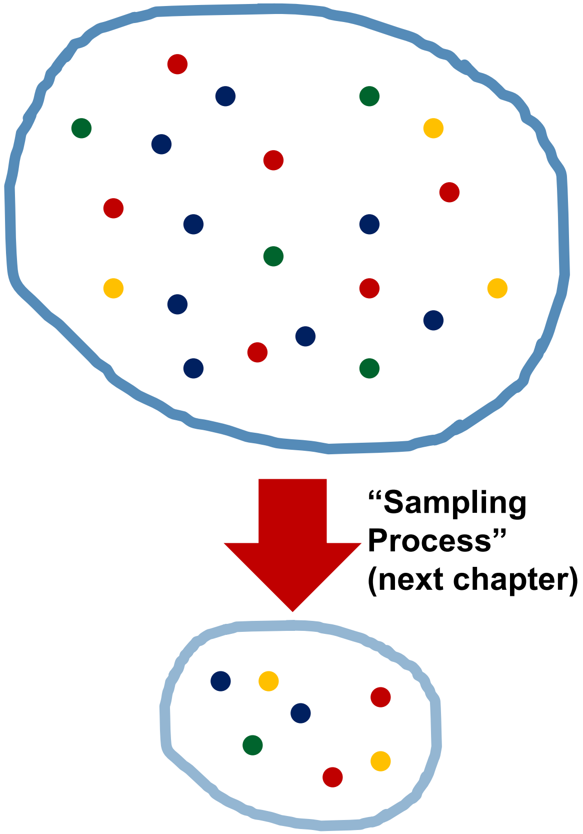
These are the most commonly used ones.

There is nothing like a "right" vs. "wrong" measure. You can also come up with your own new measures if you like to that seem better to you!

Descriptive Statistics

- 1. **What is Descriptive Statistics?**
  - Population versus Sample
    - The **population** of a study is the group of all items of interest in that study.
    - A **sample** is a subset of the population that is studied.
    - Note that the mean of the sample of course differs from the mean of the population – in fact, every time we take a sample, its mean will slightly be different.

      → We need to differentiate between the sample mean and population mean.



Descriptive Statistics

- **1. What is Descriptive Statistics?**
  - Take-away point thus far:
    - Since population parameters are different from sample statistics, there are (sometimes) different formulas for computing population versus sample statistics!

Descriptive Statistics

- **1. Let's work with an example:**
  - Suppose we have the following data set: a random **sample** of 15 final scores of last semester:
    - 859, 1018, 422, 813, 823, 912, 824, 643, 1013, 874, 929, 912, 655, 778, 629
    - **Q1: What sort of measures would you want about those scores to describe this data set?**
    - **Q2: What sort of charts would be helpful?**

Descriptive Statistics

- 2. Measures of Central Location
  - (A) Arithmetic Mean

Sample: $$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Population: $$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

Little "n" is used for sample sizes and capital "N" for the population sizes.

An aside:

Σ (which is the Greek capital letter of "Sigma") is the **summation sign** used in math to sum up several terms. "*i*" is the index variable running from 1 to "n" (n would be in our case 14).

$$\sum_{i=1}^{3} x_i = x_1 + x_2 + x_3$$

We will see the Sigma summation sign over and over again; make sure you understand it!

Descriptive Statistics

- 2. Measures of Central Location
  - (A) Arithmetic Mean

Our Sample: 859, 1018, 422, 813, 823, 912, 824, 643, 1013, 874, 929, 912, 655, 778, 629.

$$\begin{aligned}\bar{x} &= \frac{\sum_{i=1}^{n} x_i}{n} \\ &= \frac{\sum_{i=1}^{15} x_i}{15} \\ &= \frac{x_1 + x_2 + x_3 + \ldots + x_{14} + x_{15}}{15} \\ &= \frac{859 + 1,018 + 422 + \ldots + 778 + 629}{15} = 806.9\end{aligned}$$

Descriptive Statistics

- **2. Measures of Central Location**
  - **(A) Geometric Mean**
    - The arithmetic mean does not work in examples where we deal with growth rates or, generally, some rate of change.
    - Example:
      - Suppose you invest $1,000 for 2 years.
      - In 1st year it grows 100% to $2,000. ($R_1$=100%)
      - In 2nd year it suffers a 50% loss, back to $1,000. ($R_2$=-50%)

Arithmetic mean = $(R_1 + R_2)/2 = (100+(-50))/2 = 25\%$

But investment started with $1,000 is again at $1,000 → 0% growth!

Descriptive Statistics

- 2. Measures of Central Location
  - (A) Geometric Mean

$$(1+R_g)^n = (1+R_1)*(1+R_2)*...*(1+R_n)$$

$$\Leftrightarrow R_g = ((1+R_1)*(1+R_2)*...*(1+R_n))^{\frac{1}{n}} - 1$$

- R_g = ((1+100%)*(1-50%))^(1/2) -1 = **0**
- In Excel, use 1+R for growth rates and then use function "=GEOMEAN(cell area)-1" to compute the average growth value.

Descriptive Statistics

- **2. Measures of Central Location**
  - **(B) Median**

The "middle value" in the ordered data set, i.e. the value at which same number of data points above and below.

Step 1: Sort data from small to large values.

Step 2: Find middle value:

If odd number of data points, then median is at position (n+1)/2. If even number, then take the (arithmetic) mean of the two "middle" numbers.

The median only exists for interval or ordinal data (because nominal data such as "color of the eyes" has no intrinsic ordering and cannot be sorted).

Descriptive Statistics

- 2. Measures of Central Location
  - (B) Median

Formally:

Sample:

$$\tilde{x} = \begin{cases} X_{(n+1)/2} & \text{if } n \text{ is odd} \\ \frac{1}{2}(X_{n/2} + X_{n/2+1}) & \text{if } n \text{ is even} \end{cases}$$

Population:

$$\mu_{1/2} = x \text{ such that} D(x) = 1/2$$

where D(.) is the cumulative distribution function*

* That means that half of the values of the data set are below the value x. More on this in chapter 6.

Descriptive Statistics

- **2. Measures of Central Location**
  - **(B) Median**

In our sample:

Step 1: Sort the data →{422, 629, 643, 655, 778, 813, 823, 824, 859, 874, 912, 912, 929, 1013, 1018}.

Step 2: Find middle value:

15 observations

→ odd number of data points

→ median is at position $(n+1)/2 = 8$

→ 8^{th} value in data set = 824 = Median

Descriptive Statistics

- **3. Measures of Dispersion**
  - **(B) MAD, Variance & Standard Deviation**
    - More commonly used are MAD, Variance and Standard Deviation.
    - To derive their formulas, lean back and think for a minute how *you* would create a measure that somehow represented the variation in a data set.

Descriptive Statistics

- 3. Measures of Dispersion
  - (B) MAD, Variance & Standard Deviation
    - One intuitive way could be to measure the spread of the data as the average distance of the data points from the center of the data set.
    - In math terms that would be:

$$\frac{\sum_{i=1}^{N}(x_i - m)}{N}$$

1. Numerator sums up all the distances between each observation the center.
2. The denominator divides by the number of observations to get the average distance.

"m" is some measure of center – could be the arithmetic mean or the median.

Descriptive Statistics

- **3. Measures of Dispersion**
  - **(B) MAD, Variance & Standard Deviation**
    - Suppose we use the arithmetic mean of the population (μ) as the measure of center, so:

$$\frac{\sum_{i=1}^{N}(x_i - \mu)}{N}$$

Descriptive Statistics

- 3. Measures of Dispersion
  - (B) MAD, Variance & Standard Deviation
    - Suppose we use the arithmetic mean of the population (μ) as the measure of center, so:

$$\frac{\sum_{i=1}^{N}(x_i - \mu)}{N}$$

	x	(x - mean)
	10	-10
	20	0
	30	10
sum	60	0
divided by 3	20	

Issue: The numerator by definition is 0 if we sum up the differences between each value to the overall mean.

Problem: the distances to values below the mean and to values above the mean cancel each other out!

← Ex: data set {10,20,30}.

Descriptive Statistics

- 3. **Measures of Dispersion**
  - **(B) MAD, Variance & Standard Deviation**
    - 2 solutions to get rid of the problem of distances canceling one another out:
      - 1. We add up the absolute values of the distances.

$$\frac{\sum_{i=1}^{N} |x_i - \mu|}{N}$$

= Mean absolute deviation (MAD)* of the population

      - 2. We add up the squared distances:

$$\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

=Population Variance

* There is a variationsof this formula (the median absolute deviation), where the median is used instead of μ.

Descriptive Statistics

- **3. Measures of Dispersion**
  - **(B) MAD, Variance & Standard Deviation**
    - The mean absolute distance (MAD) here: 6.67 units (e.g., dollar per hour salary)
    - The variance here: 66.67 units (*squared* dollars!!)

We cannot interpret squared dollars easily, so let's take the square root to get back the variation in dollars.

	x	(x - mean)	|x-mean|	(x-mean)^2
	10	-10	10	100
	20	0	0	0
	30	10	10	100
sum	60	0	20	200
divided by 3	20		6.67	66.67
			MAD	Variance

Descriptive Statistics

- 3. Measures of Dispersion
  - (B) MAD, Variance & Standard Deviation
    - The square root of the population variance is called “**standard deviation**” (called, “sigma” σ):

$$\sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{N}}$$

= Population standard deviation.
(we have “readable” units again).

	x	(x - mean)	|x-mean|	(x-mean)^2	
	10	-10	10	100	
	20	0	0	0	
	30	10	10	100	
sum	60	0	20	200	
divided by 3	20		6.67	66.67	8.16
			MAD	Variance	Std Dev

Descriptive Statistics

- **3. Measures of Dispersion**
  - **(B) MAD, Variance & Standard Deviation**
    - Recall that we said earlier that sample statistics are different from population statistics – this is the case for the variance/standard deviation:

	Sample	Population
Variance	$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$	$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$
Standard Deviation	$s = \sqrt{s^2}$	$\sigma = \sqrt{\sigma^2}$

Descriptive Statistics

- 3. Measures of Dispersion
  - (B) MAD, Variance & Standard Deviation
    - How to compute the MAD, Variance and Std Deviation?

1. By hand (if data set not too large), or
2. Variance and Std Deviation using Excel

i	x_i		Statistic		Formula Used
1	10		Population variance	66.67	=VARP(F4:F6)
2	20		Sample variance	100.00	=VAR(F4:F6)
3	30				
			Population std deviation	8.16	=STDEVP(F4:F6)
			Sample std deviation	10.00	=STDEV(F4:F6)

Excel formula for MAD is "**AVEDEV(*range*)**"

3. With your calculator (differs by model).

Descriptive Statistics

- **3. Measures of Dispersion**
  - **(B) Variance & Standard Deviation**

    Our Sample: 859, 1018, 422, 813, 823, 912, 824, 643, 1013, 874, 929, 912, 655, 778, 629.

    Calculate the variance and standard deviation of this sample

$$s^2 = \frac{(859-806.9)^2 + (1018-806.9)^2 + \cdots + (778-806.9)^2 + (629-806.9)^2}{14} = \frac{362414.95}{14} = 25886.78$$

$$s = \sqrt{s^2} = \sqrt{25886.78} = 160.89$$

Descriptive Statistics

- 3. Measures of Dispersion
  - (C) Coefficient of Variation
    - Is a std dev of 161 points among final scores large? The standard deviation by itself cannot be interpreted when the magnitude of the variable under question is unknown.
    - The Coefficient of Variation simply scales the standard deviation by the mean:

<u>Population</u>

$$\mathrm{CV} = \frac{\sigma}{\mu}$$

<u>Sample</u>

$$\mathrm{cv} = \frac{s}{\bar{x}}$$

Descriptive Statistics

- 3. Measures of Dispersion
  - **(C) Coefficient of Variation**
    - For our sample of final scores of last year:

$$\text{cv} = \frac{s}{\bar{x}} = \frac{160.89}{806.93} = 0.199$$

The variation among the final scores is about 20% of the mean score. This is not a large variation.

    - (+) While std dev across different data sets cannot be compared, the coefficient of variation can be.
    - (-) If the mean is close to 0, then CV explodes, rendering the coefficient of variation meaningless.

PROBABILITY

Probability

- 1.1 **Basic Terminology: Random Experiment**
  - A **random experiment** is any process that leads to one or several basic outcomes (or a list thereof).*
  - Examples:
    - 1. Flipping a coin
    - 2. PA Lottery "Treasure Hunt"
    - 3. Playing football/baseball
    - 4. Predicting short-time stock price movements

An EXPERIMENT is called "random" if its RESULT cannot be predicted with certainty.

* **Basic outcomes need to be mutually exclusive**. (If we consider the weather as a random experiment, then for example "sunny" and "rainy" could not be basic outcomes as both can occur simultaneously.)

Probability

- 1.1 **Basic Terminology: Random Variable**
  - A **random variable (RV)** is a variable that takes on the value of the outcome of an experiment.
    - (We usually use X, Y, Z as the letter to designate a RV.)
  - Examples:
    - 1. Let X be the RV for flipping 1 coin → X can take on the value "Heads" or "Tails"
    - 2. Let Y be the RV for the PA Lottery "Treasury Hunt" → Y is a set of 5 numbers, each of which between 1 to 30.

A VARIABLE is called "random" if its VALUE cannot be predicted with certainty.

Probability

- ## 1.1 Basic Terminology: Sample Space
  - The **sample space** of a random experiment is the list of all possible basic outcomes:

$$S = \{O_1, O_2, \ldots, O_k\}$$

  - Examples:

countable

- 1. Flipping coins { Heads, Tails }
- 2. PA Lottery "Treasure Hunt" { (1,2,3,4,5), …, (26,27,28,29,30) }

uncountable

- 3. Football scores { …, (24:14), … }
- 4. Batting averages { …, 0.25, …, 0.39, … }

Probability

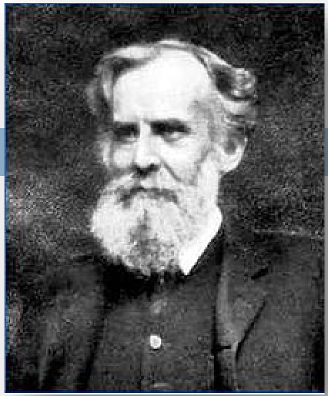
- **1.1 Basic Terminology: Probability**
  - A probability is the chance of one (or several) outcome(s) of the sample space occurring.
  - A valid probability needs to satisfy 2 requirements:
    - 1. Needs to lie between 0% and 100%: $0 \leq P(O_i) \leq 1$
    - 2. The probabilities of all outcomes in a sample space need to add up to exactly 1 (=100%).

$$\sum_{i=1}^{k} P(O_i) = 1$$

Probability

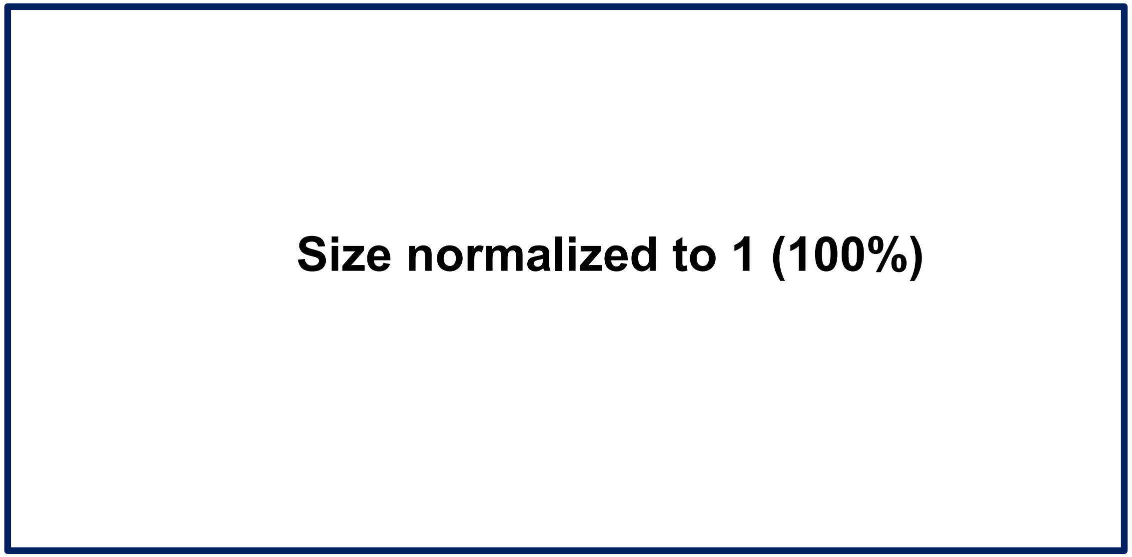
- **1.1 Basic Terminology: Events**
  - An **event** is a set of one or more outcomes from the sample space. The probability of an event is the sum of probabilities of all outcomes of the set.
  - Example:
    - Experiment: Rolling a die once
    - Sample Space: {1,2,3,4,5,6}
    - Probability of any number (by classical approach): 1/6.
    - Event: Rolling an odd number with a die = {1, 3, 5}
    - Pr (odd number) = 1/6 + 1/6 + 1/6 = ½ = 50%.

Probability



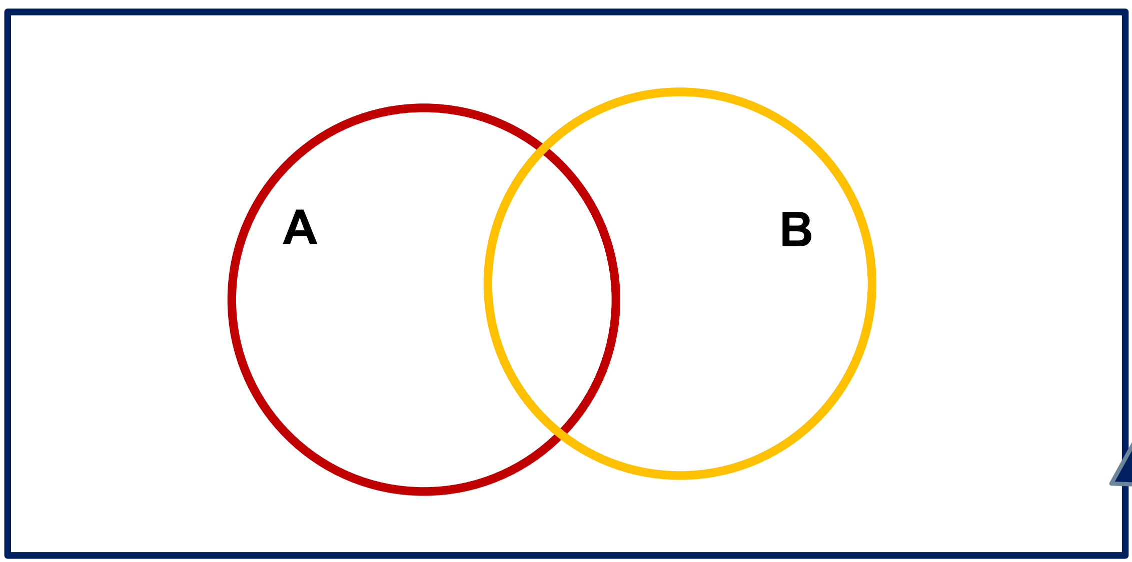
John Venn
(Brit, 1834-1923)

- **1.2.1 Venn Diagrams**
  - A Venn diagram is a fancy name for a box that represents the sample space.
  - Helps in exercises with probability rules/operators.



Probability

- 1.2.1 Venn Diagrams
  - Regions within the box represent events with their area equal to the probabilities of those events.
    - We typically use circles for generic events (but we don't have to, see next slide)
    - The area of the event represents its probability.



How does the Venn diagram look like for the random experiment "Rolling a die once"?

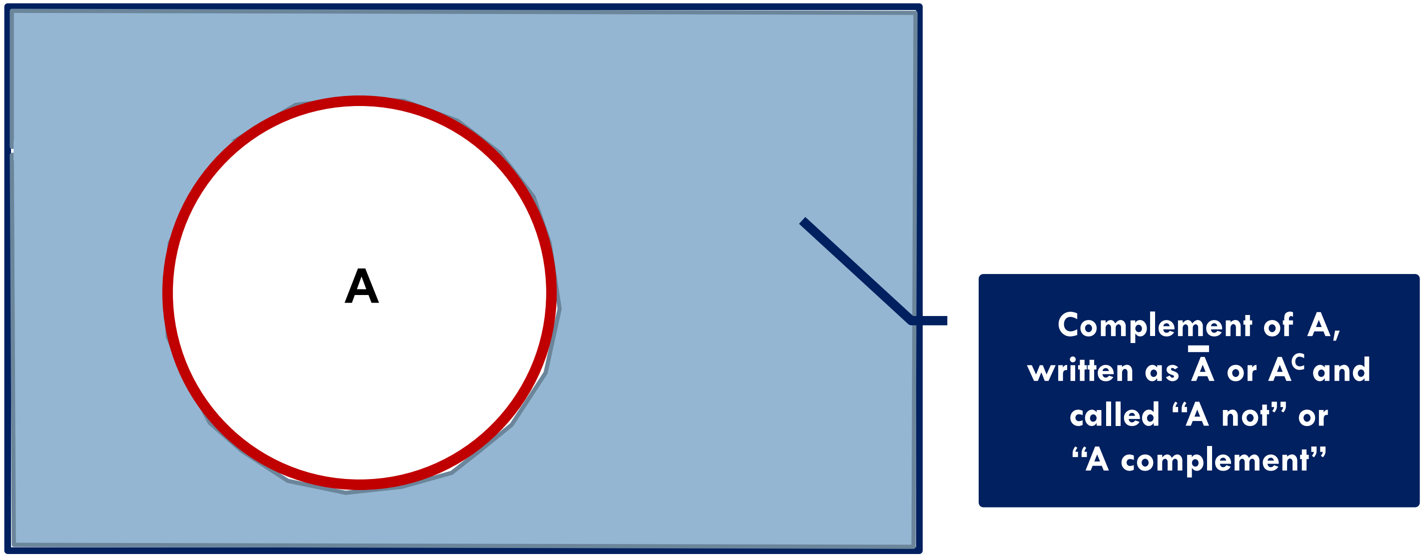
Probability

- **1.2.1 Venn Diagrams**
  - Example: “Rolling a die once”
    - Each event has the same probability to occur (classical approach), so the areas are of same size.

Rolling a 1 Pr (“1”) = 1/6=16.7%	Rolling a 2: Pr (“2”) = 16.7%	Rolling a 3 Pr (“3”) = 16.7%
Rolling a 4 Pr (“4”)=16.7%	Rolling a 5 Pr (“5”) = 16.7%	Rolling a 6 Pr (“6”) = 16.7%

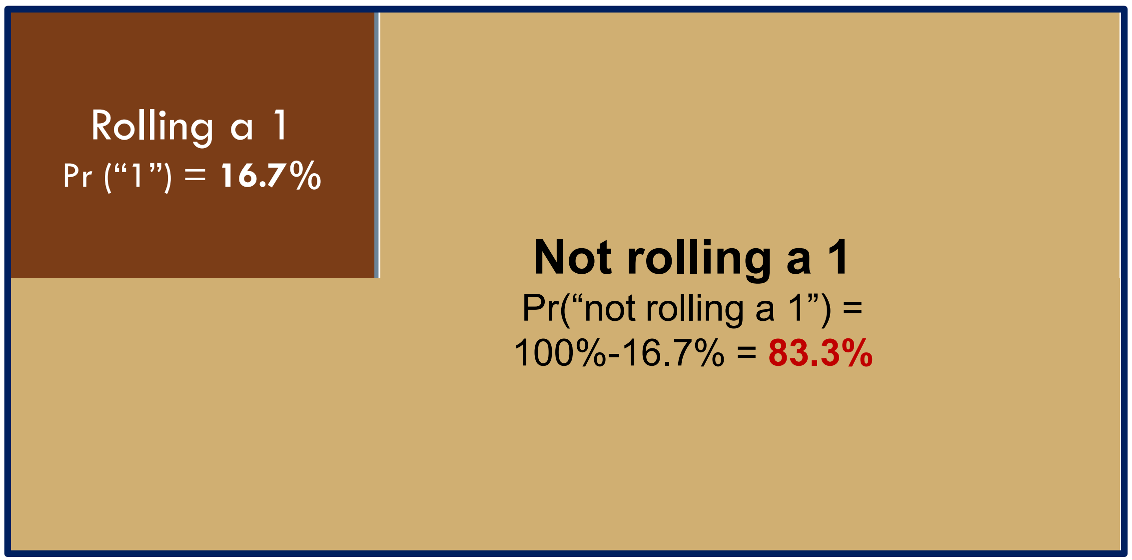
Probability

- **1.2.1 Venn Diagrams**
  - The "outside" of an event is called the **complement** of an event.



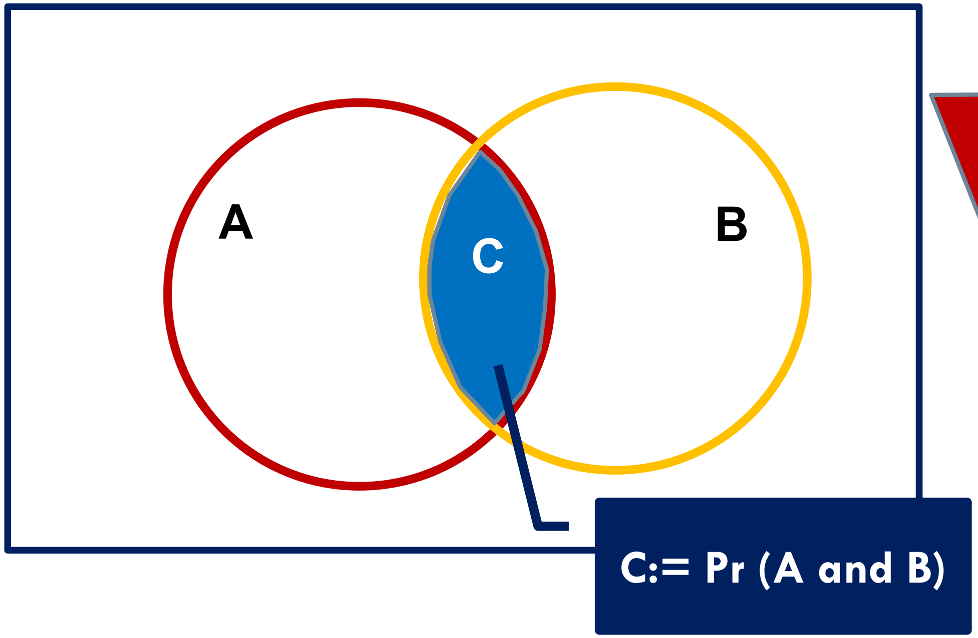
Probability

- **1.2.1 Venn Diagrams**
  - Example: Rolling a die
    - Event: If A = "Rolling a 1", then A^C="**Not** rolling a 1"



Probability

- **1.2.1 Venn Diagrams**
  - The intersection of 2 events A and B is the event when both events occur simultaneously (call it "C").
    - Note: A and B are *not* basic outcomes, so events can occur simultaneously.



Say, A=rolling an odd number; B=rolling a number > 3.

Pr (A)=1/2
Pr(B)=1/2

C=rolling a 5
Pr (C) =1/6

Probability

- **1.2.2 Contingency Tables**
  - Suppose we are interested if gender discrimination occurs at a company and have the following 5-year personnel data set of 200 workers with their gender & promotions.

1. Female; Promoted	…	198. Male; Not promoted
2. Male; Promoted	…	199. Male; Promoted
3. Female; Not promoted	…	200. Female; Not promoted

- **Step 1:** We can create an absolute frequency table.

	Promoted	Not Promoted
Female	6	24
Male	34	136

Probability

- **1.2.2 Contingency Tables**
  - Ex: Gender discrimination at a company
    - **Step 2:** We transfer the absolute frequency contingency table into a relative frequency contingency table.
      - (We just learned to do this in Excel in Chapter 3, using Pivot Tables.)

	Promoted (B_1)	Not Promoted (B_2)
Female (A_1)	0.03	0.12
Male (A_2)	0.17	0.68

Probability

- 1.2.2 Contingency Tables
  - Ex: Gender discrimination at a company.

	Promoted (B_1)	Not Promoted (B_2)	
Female (A_1)	0.03	0.12	**0.15**
Male (A_2)	0.17	0.68	**0.85**
	0.20	**0.80**	**1.00**

Joint Probabilities

Marginal Probabilities

Marginal Probabilities = the sum of rows/columns

Has to add up to 1 (=100%)

Probability

- **1.2.2 Joint and Marginal Probabilities**
  - Marginal Probabilities = Sum of Rows or Columns
    - Pr (Female) = Pr (Female AND Promoted) + Pr (Female AND Not Promoted) = 0.03 + 0.12 = 0.15

	Promoted (B_1)	Not Promoted (B_2)	
Female (A_1)	0.03	0.12	**0.15**
Male (A_2)	0.17	0.68	**0.85**
	0.20	**0.80**	**1.00**

Probability

- **1.2.2 Joint and Marginal Probabilities**
  - The joint probabilities are obtained by multiplication of marginal probabilities:*
    - Pr (Promoted AND Female) = 0.20 * 0.15
      = Pr (Promoted)*Pr (Female)
      = 0.03

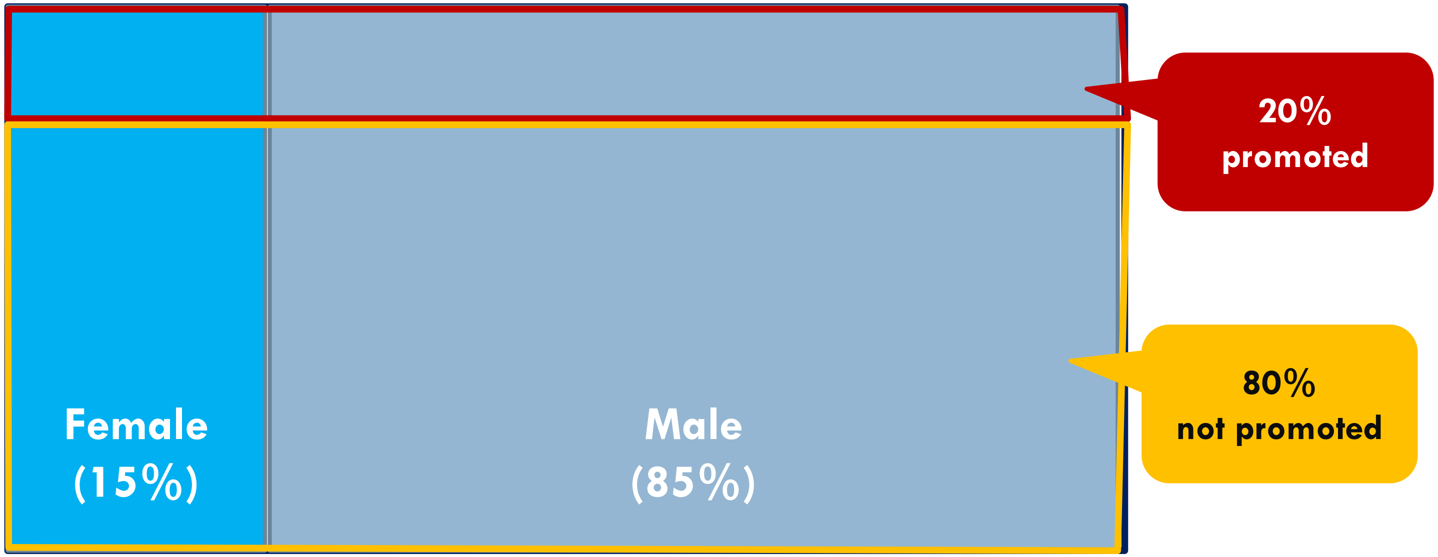
	Promoted (B_1)	Not Promoted (B_2)	
Female (A_1)	0.03	0.12	**0.15**
Male (A_2)	0.17	0.68	**0.85**
	0.20	**0.80**	**1.00**

$$Pr(J = j \cup K = k) = Pr(J = j) * Pr(K = k)$$

* **This only works if the events are independent** – more on this below!

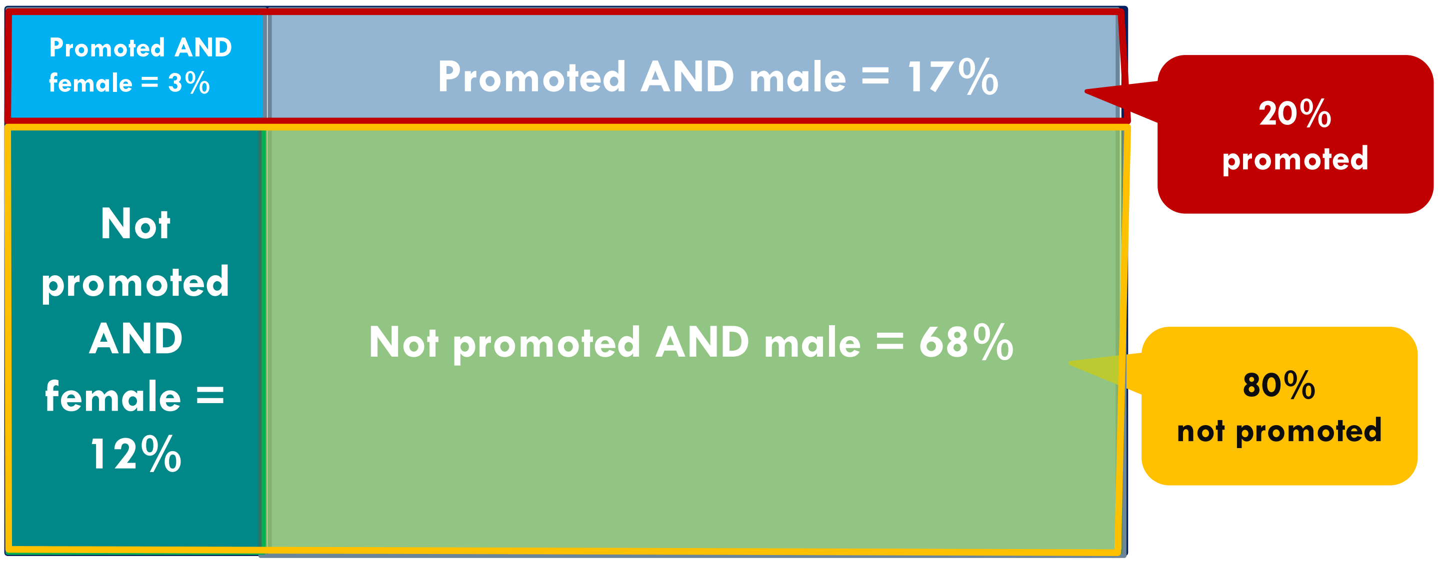
Probability

- **1.2.2 Joint and Marginal Probabilities**
  - How does the Venn diagram look like for our gender discrimination example?
    - 1. Marginal Probabilities



Probability

- 1.2.2 **Joint and Marginal Probabilities**
  - How does the Venn diagram look like for our gender discrimination example?
    - 2. Joint Probabilities



Probability

1.2.3 Conditional Probabilities

- To get back to our discrimination example:

	Promoted	Not Promoted	
Female	0.03	0.12	**0.15**
Male	0.17	0.68	**0.85**
	0.20	**0.80**	**1.00**

- To answer our question whether there is discrimination:

  Shall we simply compare joint probabilities?

  - "17% versus 3% of all 20% promotions go to men. So men are clearly favored." ?

Probability

- **1.2.3 Conditional Probabilities**
  - Discrimination example:

	Promoted	Not Promoted	
Female	0.03	0.12	**0.15**
Male	0.17	0.68	**0.85**
	0.20	**0.80**	**1.00**

- **No**, because 85% of all workers are male, naturally we would expect a higher share of promotions to go to men given that there are so many more men in the company!

Probability

- ## 1.2.3 Conditional Probabilities
  - *Instead, we should compare:*
    - What is the probability of being …

      male $\Pr(\text{Male}|\text{Promoted})$

      female $\Pr(\text{Female}|\text{Promoted})$

      <u>given that</u> one is promoted?
    - If there was no discrimination, we would expect the that men are promoted 85% of the time (as they are constitute 85% of the workers) and women 15%.
    - In other words: if there was no discrimination we would expect that "being promoted" is **independent from** gender "male/female".

Conditional probabilities

So how do we compute conditional probabilities?

Probability

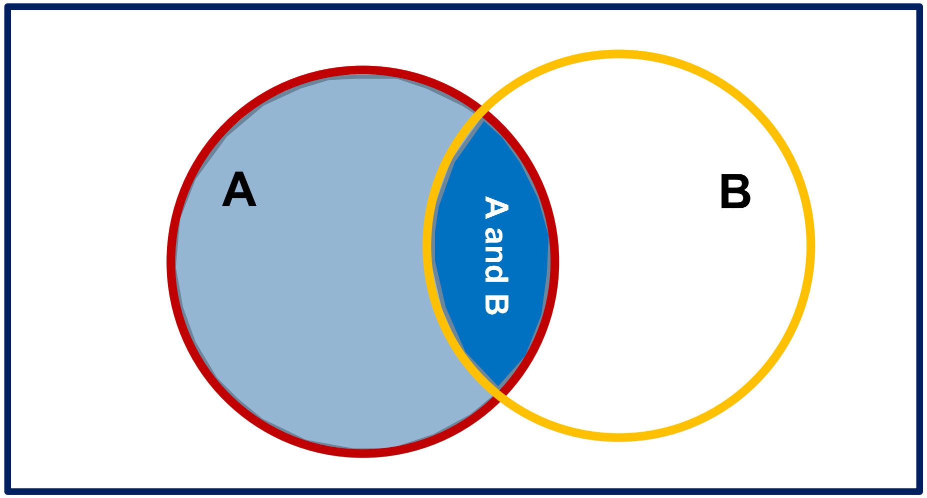
- **1.2.3 Conditional Probabilities**
  - Using the Venn Diagram to find the conditional prob.:

Pr (B|A) = Pr (B GIVEN that A is true)

= Pr (area of B GIVEN that we are "inside" A)

= Pr (dark blue as a share of light blue area)

= Pr (A and B) / Pr (A)



Probability

1.2.3 Conditional Probabilities

- Gender discrimination example:
  - Let event A=Gender, event B=Promoted and simply plug in:

$$\Pr\left(\text{Male}|\text{Promoted}\right) = \frac{\Pr\left(\text{Male AND Promoted}\right)}{\Pr\left(\text{Promoted}\right)} = \frac{0.17}{0.20} = 0.85$$

$$\Pr\left(\text{Female}|\text{Promoted}\right) = \frac{\Pr\left(\text{Female AND Promoted}\right)}{\Pr\left(\text{Promoted}\right)} = \frac{0.03}{0.20} = 0.15$$

	Promoted (B)	Not Promoted (B^C)	
Female (A)	0.03	0.12	**0.15**
Male (A^C)	0.17	0.68	**0.85**
	0.20	**0.80**	**1.00**

85% and 15% coincides with the share of male and female workers.
→No evidence of gender discrimi-nation.

Probability

- 1.2.3 **Conditional Probabilities**
  - What we just showed: the event "male/female" is **independent from** the event "promoted/not promoted"!

$$\Pr\ (\text{Female}|\text{Promoted}) = 0.15 = \Pr(\text{Female})$$

$$\Pr\ (\text{Male}|\text{Promoted}) = 0.85 = \Pr(\text{Male})$$

	Promoted (B)	Not Promoted (B^C)	
Female (A)	0.03	0.12	**0.15**
Male (A^C)	0.17	0.68	**0.85**
	0.20	**0.80**	**1.00**

Probability

- **1.2.3 Conditional Probabilities**
  - Definition: Two events are said to be **independent** if

$$\Pr(A|B) = P(A)$$

(or since the choice of the letters are arbitrary:

$$\Pr(B|A) = P(B)$$)

In our example:

- $\Pr(A|B) = \Pr(\text{Male}|\text{Promoted}) = 0.85$

$\Pr(A) = \Pr(\text{Male}) = 0.85$

Probability

- **1.2.3 Conditional Probabilities**
  - How would **no independence** look like in our example?

	Promoted (B)	Not Promoted (B^C)	
Female (A)	~~0.03~~ 0.07	0.12	**0.15**
Male (A^C)	0.17	0.68	**0.85**
	0.20	**0.80**	**1.00**

Change just one joint probability…

Probability

1.2.3 Conditional Probabilities

- How would **no independence** look like in our example?

	Promoted (B)	Not Promoted (B^C)	
Female (A)	~~0.03~~ 0.07	~~0.12~~ 0.08	**0.15**
Male (A^C)	~~0.17~~ 0.13	~~0.68~~ 0.72	**0.85**
	0.20	**0.80**	**1.00**

Then, to maintain the marginal probabilities, all the other joint probabilities need to change accordingly.

Probability

- 1.2.3 Conditional Probabilities
  - How would **no independence** look like in our example?

	Promoted (B)	Not Promoted (B^C)	
Female (A)	0.07	0.08	**0.15**
Male (A^C)	0.13	0.72	**0.85**
	0.20	**0.80**	**1.00**

$$\Pr(A|B) = \Pr(\text{Female}|\text{Promoted})$$

$$= \frac{\Pr(\text{Female AND Promoted})}{\Pr(\text{Promoted})}$$

$$= \frac{0.07}{0.20} = 0.35 \neq 0.15 = \Pr(\text{Female}) = \Pr(A)$$

Now women receive 35%of all promotions even though they are just 15% of the workers.
→ No independence any longer!

Probability

- **1.2.3 Conditional Probabilities**
  - How would **no independence** look like in our example?

	Promoted (B)	Not Promoted (B^C)	
Female (A)	0.07	0.08	**0.15**
Male (A^C)	0.13	0.72	**0.85**
	0.20	**0.80**	**1.00**

Aside: we can see that the joint probability formula on slide 24:
P(A and B) = P(A) * P(B)
<u>only works if both events A and B are independent from one another!</u>
0.20*0.15=0.03 ≠0.07

$$\text{...Female}|\text{Promoted})$$

$$\frac{\text{...Female AND Promoted})}{\Pr(\text{Promoted})}$$

$$= 0.35 \neq 0.15 = \Pr(\text{Female}) = \Pr(A)$$

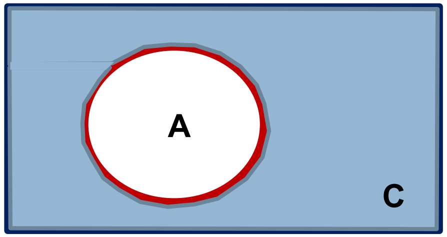
Now women receive 35%of all promotions even though they are just 15% of the workers.
→ No independence any longer!

Probability

- 1.3 <u>Probability Operators</u> and Probability Rules

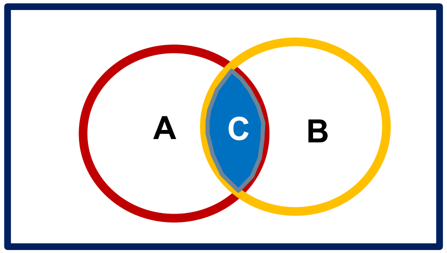
  - The Complement Operator ("**not**"):

$$C = \bar{A} = A^C$$



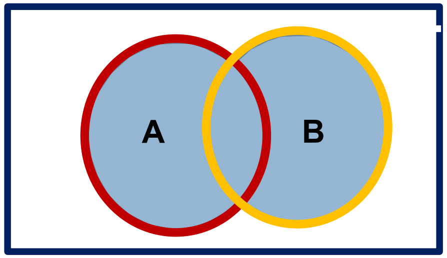
  - The Intersection Operator ("**and**")

$$C = A \cap B$$



  - The Union Operator ("**or**")

$$C = A \cup B$$



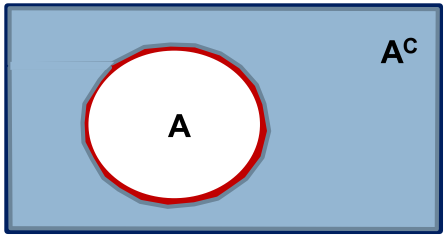
Probability

- 1.3 Probability Operators and **Probability Rules**
  - 1. The Complement Rule
    - An event and its complement event has to sum up to 1.
    - Example:
      - Say, A = "Being in New York City on Sunday"

        → A^C = "NOT being in New York City on Sunday"

        The probability of "being in NYC" plus the probability of "not being in NYC" has to add up to 100%.
    - Thus:

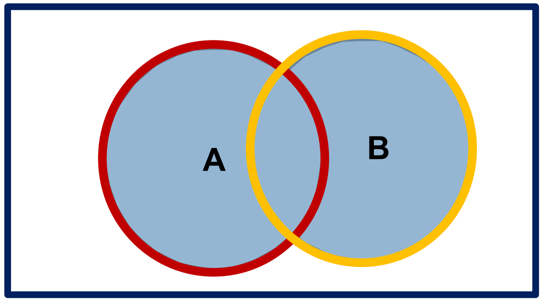
$$\Pr(A^C) = 1 - \Pr(A)$$



- **1.3 Probability Operators and <u>Probability Rules</u>**
  - 2. The Addition Rule
    - The probability that event A **or** event B (or both) occur:

$$\Pr(\text{A or B}) = \Pr(\text{A}) + \Pr(\text{B}) - \Pr(\text{A and B})$$

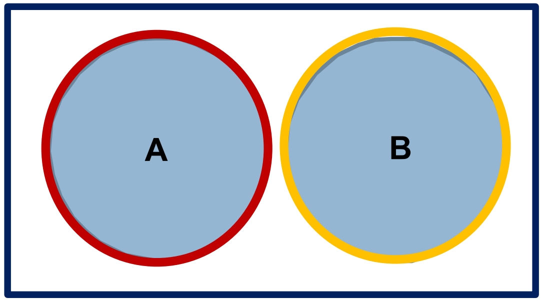
$$\Leftrightarrow \Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$



To avoid to double-count the overlapping part!
(Because that overlap is part of both events A and B.)

Probability

- **1.3 Probability Operators and Probability Rules**
  - 2. The Addition Rule for Mutually Exclusive Events
    - Two events are said to be mutually exclusive, if both cannot be true simultaneously.
    - Example:
      - Event A = "I am in New York."
      - Event B = "I am in Pittsburgh."



- Addition Rule for Mutually Exclusive Events:

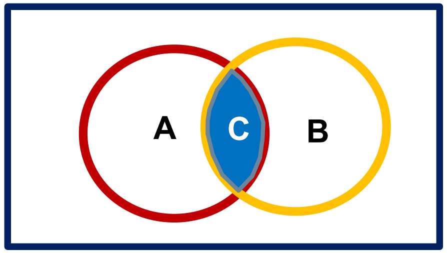
$$\underline{\Pr(\text{A or B})} = \Pr(\text{A}) + \Pr(\text{B}) - \underbrace{\text{P}(\text{A and B})}_{=0} = \underline{\Pr(\text{A}) + \Pr(\text{B})}$$

Probability

- **1.3 Probability Operators and Probability Rules**
  - 3. The Multiplication Rule
    - Used to compute the joined probability of 2 events. It is based on the conditional probability formula which we have explained earlier:

$$\Pr(A|\ B) = \frac{\Pr(A \text{ and } B)}{\Pr(B)}$$

$$\Leftrightarrow \Pr(A \text{ and } B) = \Pr(A|\ B)\Pr(B)$$



- Or, rewriting the last line:

$$\Pr(A \text{ and } B) = \Pr(B \text{ and } A) = \Pr(B \mid A)\ P(A)$$

Probability

- **1.3 Probability Operators and <u>Probability Rules</u>**
  - 3. The Multiplication Rule <u>for Independent Events</u>
    - We have seen earlier that for 2 independent events we have:

$$\Pr(A|B) = P(A) \quad \text{and} \quad \Pr(B|A) = P(B)$$

    - Therefore, when 2 events are independent:

$$\underline{\Pr(A \text{ and } B)} = \Pr(A|B)\Pr(B) = \underline{\Pr(A)\Pr(B)}$$

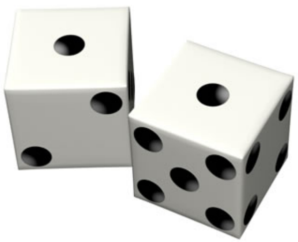
      - (that's what we used on slide 23.)

DISCRETE PROBABILITY DISTRIBUTION

Discrete Prob. Distribution

- **2.1 Random Variables (RVs)**
  - An RV is a function that assigns a number to each outcome of an experiment. (We usually use capital letters X,Y or Z to refer to a RV.)
  - A **discrete** RV is one where the outcomes are countable. An RV where the outcomes are uncountable is called **continuous.**
    - Discrete RV: The sum of rolling 2 dice.
    - Continuous RV: The time to finish a problem set.
      - (for any 2 point of time there is a some point in time in between.)

Discrete Prob. Distribution



- 2.1 Random Variables (RVs)
  - Example 1: Sum of 2 dice

1st die → ↓ 2nd die	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

A random variable is a function or rule that assigns a number to each outcome in the sample space of an experiment.

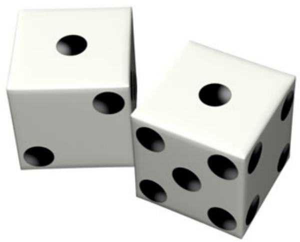
Here:

"Rolling two ones" → X=2
"Rolling a 1 and 4" → X=5
"Rolling a 2 and 3" → X=5

In some experiments …

- **the RV can take on the same value for several outcomes**
- **the outcomes themselves are numbers (e.g., return on an invest-ment) → in those cases the value of the RV simply is the numerical event itself.**

Discrete Prob. Distribution



2.1 Random Variables (RVs)

- Example 1: Sum of 2 dice

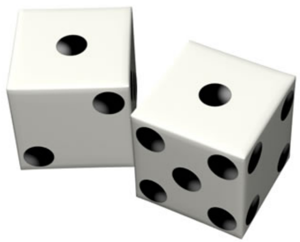
1st die → / ↓ 2nd die	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

36 possible sum outcomes

X =

2: 1
3: 2
4: 3
5: 4
6: 5
7: 6
8: 5
9: 4
10:3
11:2
12:1

Discrete Prob. Distribution



2.1 Random Variables (RVs)

- The probability of an RV's outcome is written as P(X=x) or simply as P(x).

X	# of occurences	P(x)
2	1	1/36
3	2	2/36
4	3	3/36
5	4	4/36
6	5	5/36
7	6	6/36
8	5	5/36
9	4	4/36
10	3	3/36
11	2	2/36
12	1	1/36
SUM	36	1

This is the probability distribution of the discrete random variable X.

(We obtained in this example the probability distribution by using the "classical approach" to assign probabilities: assuming the dice are fair, each outcome is equally likely.)

→ P(X=7) = P(7) = 6/36.

→ P(X=12) = P(12) = 1/36.

Discrete Prob. Distribution

- **2.1 Random Variables (RVs)**
  - Example 2:
    - Let X = the number of boys born to a family with 3 kids.
    - Suppose getting a boy has a probability of 0.52.
    - → Find the RV and its probability distribution.
      (Hint: A probability tree can help.)

Probability from relative frequency approach.

Discrete Prob. Distribution

- 2.1 Random Variables (RVs)
  - Example 2:
    - Let X = the number of boys born to a family with 3 kids.
    - Suppose getting a boy has a probability of 0.52.
    - → Find the RV and its probability distribution.

      (Hint: A probability tree can help.)

Probability from relative frequency approach.

Random Variable

X = {
- 0: 1 (GGG)
- 1: 3 (BGG, GBG, GGB)
- 2: 3 (BBG, BGB, GBB)
- 3: 1 (BBB)

Probability distribution

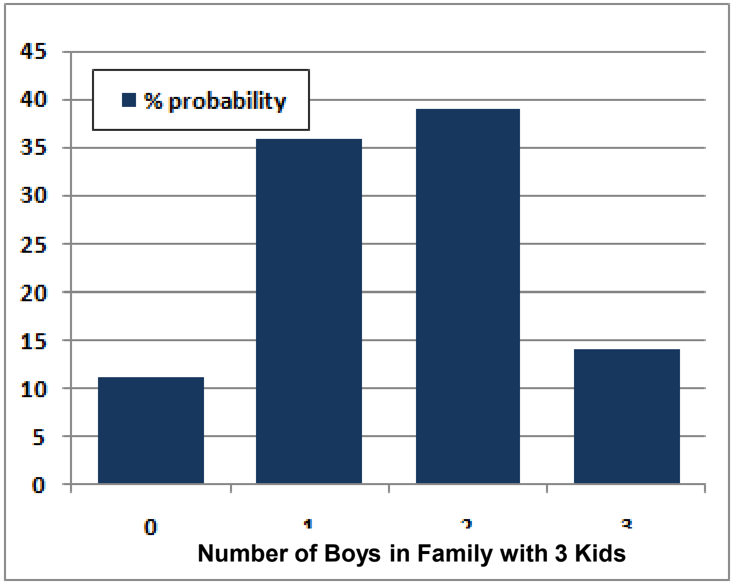
P(x) = {
- 0: 1*0.48^3 = 11.1%
- 1: 3*0.52^1*0.48^2= 35.9%
- 2: 3*0.52^2*0.48^1= 38.9%
- 3: 1*0.52^3 = 14.1%

Sum: 100%

Number of branches * probability of each branch in the tree.

Discrete Prob. Distribution

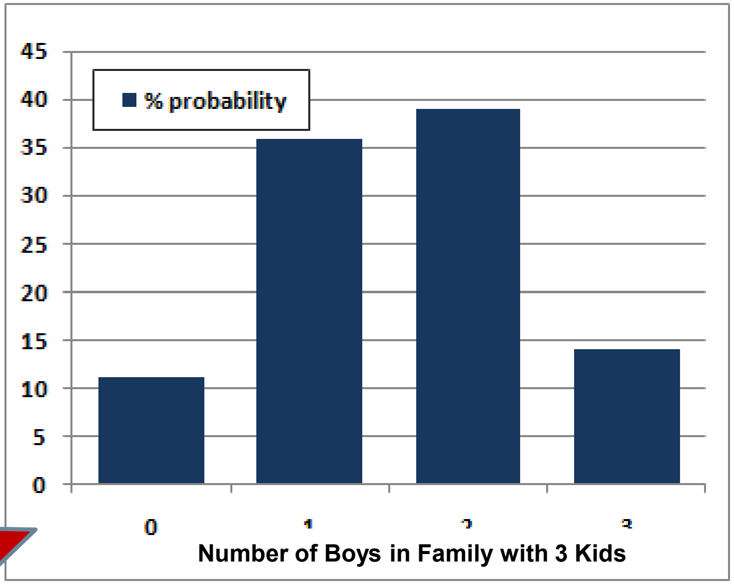
- **2.2 Probability Distributions**
  - Just as any other distribution, a probability distribution can be displayed with a relative frequency bar chart and be described via descriptive statistics.



Discrete Prob. Distribution

- **2.2 Probability Distributions**
  - Just as any other distribution, a probability distribution can be displayed with a relative frequency bar chart and be described via descriptive statistics.
  - The **average** of a probability distribution is called the **"expected value"**.

$$E(X) = \sum_{\text{all x}} \{xP(x)\} = \mu$$



Here: 0*0.111 + 1*0.359 + 2*0.389 + 3*0.141 = 1.56

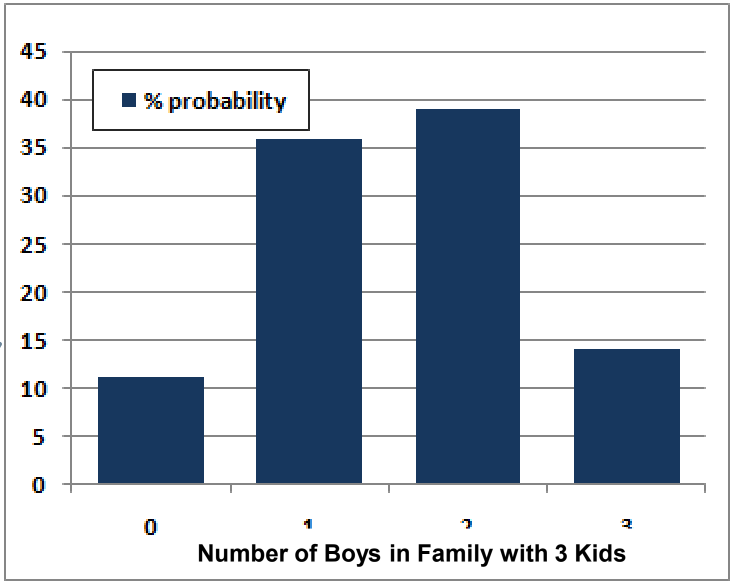
Discrete Prob. Distribution

- **2.2 Probability Distributions**
  - The spread of a probability distribution is called (just as before) **variance** and **standard deviation.**

$$V(X) = \sum_{\text{all } x} \{(x - E(X))^2 P(x)\} = \sigma^2$$

$$\text{StdDev}(X) = \sqrt{\sigma^2} = \sigma$$

Var (# of boys)=
(0-1.56)^2*0.11 +
(1-1.56)^2*0.359 +
(2-1.56)^2*0.389 +
(3-1.56)^2*0.141
= 0.745
StdDev (# of boys) =
= sqrt(Var (#of boys))
= sqrt(0.745) = 0.863



Discrete Prob. Distribution

- **2.2 Probability Distributions**
  - Laws of Expected Value

$$E(c) = c$$

- The average of a constant is just the constant itself.

$$E(X + c) = E(X) + c$$

- If you add to each value that the RV returns a constant, the whole distribution just shifts. Hence the average itself will also just be shifted by the constant c.

$$E(c * X) = c * E(X)$$

- Multiplying each value that the RV returns by a constant, will simply shift the whole distribution (and increase its spread). Hence the average itself will also just shift by that factor.

Discrete Prob. Distribution

- 2.2 Probability Distributions
  - Laws of the Variance

$$V(c) = 0$$

- A constant does not have any variance/spread (by definition).

$$V(X + c) = V(X)$$

- If you add to each value that an RV returns a constant, the whole distribution just shifts. The spread however remains unchanged.

$$V(c * X) = c^2 * V(X)$$

- Multiplying each value in a distribution by a constant, will shift the distribution by the factor to the left or right *while increasing its spread*. As the variance is the avg. *squared* deviation from the center, constant c becomes c^2.

Discrete Prob. Distribution

- **2.2 Probability Distributions**
  - Examples 1 & 2:
    - $E(4X+5) =$ // let Y=4X and re-substitute it later

      $E(Y+5) = E(Y)+5 = E(4X)+5 =$ **4*E(X)+5**

    - $V(2X+3) =$ // let Y=2X and re-substitute it later

      $V(Y+3) = V(Y) = V(2X) =$ **4*V(X)**

Discrete Prob. Distribution

- **2.2 Probability Distributions**
  - Example 3:
    - Suppose a slot machine costs \$1 per game and works as follows: with 30% probability you win \$3, with 70% probability you win nothing (\$0). (In both cases you still had to pay \$1 to play the slot machine.)
    - Q1: What is your expected payoff in the long run?

Discrete Prob. Distribution

- **2.2 Probability Distributions**
  - Example 3:
    - Suppose a slot machine costs \$1 per game and works as follows: with 30% probability you win \$3, with 70% probability you win nothing (\$0). (In both cases you still had to pay \$1 to play the slot machine.)
    - Q1: What is your expected payoff in the long run?
      - E(win of machine) = 0.3*\$3 + 0.7*\$0 = \$0.90
      - You still pay \$1 to play the game → \$0.90 -\$1 = -\$0.10
      - In the long run you will lose on avg. 10 cents per game.

Discrete Prob. Distribution

- **2.2 Probability Distributions**
  - Example 4:
    - Suppose that the return, r, of an asset can take one of three values:
      - r=4 with probability 0.2
      - r=5 with probability 0.5
      - r=6 with probability 0.3
    - Calculate the expected value and the variance of the return of the asset
    - E[r]=0.2*4+0.5*5+0.3*6=5.1
    - Var[r]=0.2*(4-5.1)^2+0.5*(5-5.1)^2+0.3*(6-5.1)^2 =0.2*1.21+0.5*0.01+0.3*0.81=0.242+0.005+0.243=0.49

Discrete Prob. Distribution

- **2.2 Probability Distributions**
  - Example 5:
    - Suppose that the return, r, of an asset can take one of three values:
      - r=-2 with probability 0.3
      - r=1 with probability 0.5
      - r=5 with probability 0.2
    - Calculate the expected value and the variance of the return of the asset
    - E[r]=0.3*(-2)+0.5*1+0.2*5=0,9
    - Var[r]=0.3*(-2-0.9)^2+0.5*(1-0.9)^2+0.2*(5-0.9)^2 =0.3*8.41+0.5*0.01+0.2*16.81 =2.523+0.005+3.362=5.89