Axioms and Theorems of Probability

MATH 3512, BCIT

Matrix Methods and Statistics for Geomatics

November 13, 2018

Introductory Concepts in Statistics I

- A population is the complete collection of all measurements or data that are being considered.
- A census is the collection of data from every member of the population.
- A sample is a subcollection of members selected from a population.

Introductory Concepts in Statistics I

- A population is the complete collection of all measurements or data that are being considered.
- A census is the collection of data from every member of the population.
- A sample is a subcollection of members selected from a population.

Introductory Concepts in Statistics I

- A population is the complete collection of all measurements or data that are being considered.
- A census is the collection of data from every member of the population.
- A sample is a subcollection of members selected from a population.

Introductory Concepts in Statistics II

- A voluntary response sample or self-selected sample is one in which the respondents themselves decide whether to be included.
- A random sample is one in which each member has the same probability of being selected.
- A stratified random sample is one in which random samples from subgroups are drawn and proportionally combined to form the complete sample.

Introductory Concepts in Statistics II

- A voluntary response sample or self-selected sample is one in which the respondents themselves decide whether to be included.
- A random sample is one in which each member has the same probability of being selected.
- A stratified random sample is one in which random samples from subgroups are drawn and proportionally combined to form the complete sample.

Introductory Concepts in Statistics II

- A voluntary response sample or self-selected sample is one in which the respondents themselves decide whether to be included.
- A random sample is one in which each member has the same probability of being selected.
- A stratified random sample is one in which random samples from subgroups are drawn and proportionally combined to form the complete sample.

Introductory Concepts in Statistics III

- A parameter is a numerical measurement describing some characteristic of the population.
- A statistic is a numerical measurement describing some characteristic of a sample.

Introductory Concepts in Statistics III

- A parameter is a numerical measurement describing some characteristic of the population.
- A statistic is a numerical measurement describing some characteristic of a sample.

Introductory Concepts in Statistics IV

- Quantitative (or numerical) data consist of numbers representing counts or measurements.
- Categorical (or qualitative) data consist of names or labels that are not numbers representing counts or measurements.

Introductory Concepts in Statistics IV

- Quantitative (or numerical) data consist of numbers representing counts or measurements.
- Categorical (or qualitative) data consist of names or labels that are not numbers representing counts or measurements.

Introductory Concepts in Statistics V

- Discrete data result when the data values are quantitative and the number of values is finite or countable.
- Continuous data result when the data values are quantitative and the number of values is infinite and not countable.

Here is an example for infinite discrete outcomes (this is rare). Roll a die until you roll a six. There are infinitely many ways to do this, but the data is not continuous.

Introductory Concepts in Statistics V

- Discrete data result when the data values are quantitative and the number of values is finite or countable.
- Continuous data result when the data values are quantitative and the number of values is infinite and not countable.

Here is an example for infinite discrete outcomes (this is rare). Roll a die until you roll a six. There are infinitely many ways to do this, but the data is not continuous.

Introductory Concepts in Statistics VI

- Blinding is when the subject doesn't know whether they are receiving a treatment or a placebo.
- The placebo effect occurs when an untreated subject reports an improvement in symptoms because of their participation in the study.
- An experiment is double-blind when it is blind and the experimenter also doesn't know whether they are applying a treatment or a placebo.

Introductory Concepts in Statistics VI

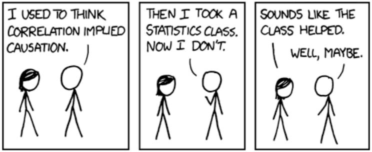
- Blinding is when the subject doesn't know whether they are receiving a treatment or a placebo.
- The placebo effect occurs when an untreated subject reports an improvement in symptoms because of their participation in the study.
- An experiment is double-blind when it is blind and the experimenter also doesn't know whether they are applying a treatment or a placebo.

Introductory Concepts in Statistics VI

- Blinding is when the subject doesn't know whether they are receiving a treatment or a placebo.
- The placebo effect occurs when an untreated subject reports an improvement in symptoms because of their participation in the study.
- An experiment is double-blind when it is blind and the experimenter also doesn't know whether they are applying a treatment or a placebo.

Correlation Does Not Imply Causation

Confounding occurs in an experiment when the investigators are not abloe to distinguish among the effects of different factors.

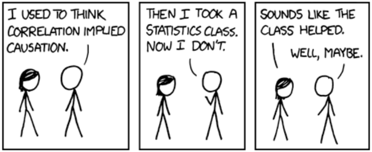


Examples:

1. astrological sign and IQ in elementary school
2. soft drinks and obesity
3. birth control pills and thrombosis

Correlation Does Not Imply Causation

Confounding occurs in an experiment when the investigators are not abloe to distinguish among the effects of different factors.

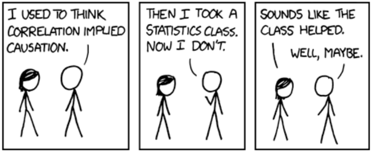


Examples:

1. astrological sign and IQ in elementary school
2. soft drinks and obesity
3. birth control pills and thrombosis

Correlation Does Not Imply Causation

Confounding occurs in an experiment when the investigators are not abloe to distinguish among the effects of different factors.



Examples:

1. astrological sign and IQ in elementary school
2. soft drinks and obesity
3. birth control pills and thrombosis

Characterizing Data Sets

The following serve to characterize data sets without listing the data:

- measures of centre
- measures of dispersion
- measures of position

Measures of Centre: Mean

$$\text{mean} = \frac{\sum x}{n} \tag{1}$$

where n is the number of data points in your quantitative data set and x is your data set. Mathematically speaking, x is an n-dimensional vector $x = (x_1, \ldots, x_n)$ and $\sum x$ means

$$\sum x = x_1 + \ldots x_n \tag{2}$$

Often, we write μ for the mean of a population and $\bar{x}$ for the mean of a sample.

Frequency Distributions

Often, data is provided in the form of a frequency distribution. For example, when I asked a class of statistics students about the number of countries they had visited in their lifetime, the response was as follows (given as an R command),

```
cn<-c(5,4,7,3,6,4,3,4,2,4,4,2,4,3,2,4,4)
```

A more intelligible way to display the data is to provide a frequency distribution.

```
> table(cn)
cn
2 3 4 5 6 7
3 3 8 1 1 1
```

There are 3 people who have been to 2 countries, 3 people who have been to 3 countries, 8 people who have been to 4 countries, 1 person who has been to 5 countries, and so on.

If you have a frequency distribution (usually of a sample, so we will call the mean $\bar{x}$), the mean is

$$\bar{x} = \frac{\sum (f \cdot x)}{\sum f} \tag{3}$$

For the example in the last slide,

$$\bar{x} = \frac{2 \cdot 3 + 3 \cdot 3 + 4 \cdot 8 + 5 \cdot 1 + 6 \cdot 1 + 7 \cdot 1}{3 + 3 + 8 + 1 + 1 + 1} \approx 3.82 \tag{4}$$

In R, you can also simply use the command `mean`. Notice that `mean(x)` and `sum(x)/length(x)` will give you the same number.

Exercises

Exercise 1: Find the mean of the following five counts for Chips Ahoy chocolate chip cookies: 22 chips, 22, chips, 26 chips, 24 chips, and 23 chips.

Exercise 2: Anne measures the temperature in her walk-in freezer. She measures -23°C once; -22°C 31 times; -21°C 13 times; -20°C 7 times; -18°C twice. What is the mean temperature given this dataset?

Measures of Centre: Median

The median is the value in the middle. If there is an even number of data points, the median is the mean of the two data points in the middle. To find the median, sort the data points. For example, the numbers of countries visited are

2	2	2	3	3	3	4	4	4	4	4	4	4	4	5	6	7

The value in the middle is the number 4, which is also the median of the data. It is quite similar to the mean, which is approximately 3.82.

Difference Between Mean and Median

Imagine we had one more student in the class who was a world traveler. She had visited 112 countries! The mean is now

$$\bar{x} = \frac{\sum x}{n} = \frac{177}{18} \approx 9.83 \tag{5}$$

A mean of 9.83 is no longer a good summary of the data. Let's see if the median does better.

2	2	...	4	4	4	4	4	...	6	7	112

The new median is $(4+4)/2 = 4$, which is a much better summary of the data, pretty much ignoring the outlier.

Measures of Dispersion: Motivation

Have a look at these two different data sets.

```
x1<-c(12,12,12,12,12,12,11,12,12,13,12,12,12,12)
```

and

```
x2<-c(15,10,14,7,17,15,11,18,12,12,15,9,7,6)
```

The mean of both data sets is 12. The median of both data sets is 12. However, something about these two data sets is different. `x2` is more dispersed than `x1`, which means that the data is spread out more. There is more variation in `x2`. We try to capture this variation by finding measures of dispersion.

Measures of Dispersion: Absolute Value of Deviation

The difference between data points and the mean always sums to zero! That is not helpful. If we want to make this measure of dispersion more useful, we need to sum the absolute value of deviation

$$\text{mean absolute deviation } = \frac{\sum |x - \bar{x}|}{n} \tag{6}$$

For our examples `x1` and `x2`, the mean absolute deviations are 2 and 44, respectively. Although at first glance this measure of dispersion looks useful, it makes for very complicated calculations that can be simplified by choosing a different way to make all the distances between data points and mean positive: not the absolute value, but the square of the distance.

Measures of Dispersion: Variance

The variance is calculated as follows,

$$\text{variance of a population } = \sigma^2 = \frac{\sum (x-\mu)^2}{n} \tag{7}$$

Something odd happens when we take the variance of a sample. If we were to use equation (7) to calculate the sample variance for all possible samples of a population, the mean of these sample variance would not equal the population variance. This means that in this case the sample variance would be a biased estimator of the population variance. We don't want that! To correct for this problem and define a sample variance which is an unbiased estimator of the population variance, we introduce Bessel's correction and define

$$\text{variance of a sample } = s^2 = \frac{\sum (x-\bar{x})^2}{n-1} \tag{8}$$

Measures of Dispersion: Standard Deviation

One disadvantage of the variance is that it is not an intuitive measurement of dispersion. If we take the square root of the variance, then we get something similar to the absolute value of deviation, which tells us approximately how far on average the data points are from the mean. We call this measurement the standard deviation

$$\text{standard deviation of a population } = \sigma = \sqrt{\frac{\sum (x-\mu)^2}{n}} \tag{9}$$

$$\text{standard deviation of a sample } = s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} \tag{10}$$

Why we still sometimes prefer the variance will become clear on the next slide. The standard deviation, whether with or without Bessel's Correction, is a biased estimator!

Bessel's Correction

	$\sqrt{\frac{\sum(x-\bar{x})^2}{n}}$	$\sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$	$\frac{\sum(x-\bar{x})^2}{n}$	$\frac{\sum(x-\bar{x})^2}{n-1}$
2 and 2	0.00	0.00	0.00	0.00
2 and 3	0.50	0.71	0.25	0.50
2 and 8	3.00	4.24	9.00	18.00
3 and 2	0.50	0.71	0.25	0.50
3 and 3	0.00	0.00	0.00	0.00
3 and 8	2.50	3.54	6.25	12.50
8 and 2	3.00	4.24	9.00	18.00
8 and 3	2.50	3.54	6.25	12.50
8 and 8	0.00	0.00	0.00	0.00
mean	1.33	1.89	3.44	6.89

The population standard deviation is:

$$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{n}} = 2.62$$

The population variance is:

$$\sigma^2 = \frac{\sum(x-\mu)^2}{n} = 6.89$$

Calculating the Variance I

It is easiest to calculate the variance using statistical software. In R Studio, for example,

```
> var(x1)
[1] 0.1538462
> var(x2)
[1] 14.76923
```

and

```
> sd(x1)
[1] 0.3922323
> sd(x2)
[1] 3.843076
```

Calculating the Variance II

When you do have to calculate the variance by hand, it is helpful to use the following shortcut formula,

$$s^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} \tag{11}$$

because you do not have to keep entering the mean, which may contain numerous significant digits.

Variance for a Frequency Distribution I

Consider the data set `x3`

`4,4,2,2,4,3,3,3,1,4,4,1,2,2,2,4,2,1,2,1,1,2,3,3,2,3,3,3`

We can summarize the data in a frequency distribution

```
> table(x3)
x3
1 2 3 4
5 9 8 6
```

Variance for a Frequency Distribution II

Remember that in equation (3) we calculated the mean using a formula for the frequency distribution,

$$\bar{x} = \frac{\sum (f \cdot x)}{\sum f} \tag{12}$$

We can do the same for the variance and the standard deviation,

$$s^2 = \frac{\sum \left(f \cdot x^2\right) - \frac{(\sum f \cdot x)^2}{\sum f}}{\sum f - 1} \tag{13}$$

Remember that $n = \sum f$.

Definitions

event An event is any collection of results or outcomes of a procedure.

sample space The sample space for a procedure consists of all possible simple events. That is, the sample space consists of all outcomes that cannot be broken down any further. The symbol for the sample space is Ω.

complement The complement of event A is $\neg A$ and consists of all outcomes in which A does not occur.

Logic and Sets I

1. $A \vee B$ is the event "either A or B happens."
2. $A \wedge B$ is the event "both A and B happens."
3. $\neg A$ is the event "A does not happens."
4. Ω and $\emptyset$ are events; they are called 'tautology' and 'contradiction,' respectively.

Logic and Sets I

1. $A \vee B$ is the event "either A or B happens."
2. $A \wedge B$ is the event "both A and B happens."
3. $\neg A$ is the event "A does not happens."
4. Ω and $\emptyset$ are events; they are called 'tautology' and 'contradiction,' respectively.

Logic and Sets I

1. $A \vee B$ is the event "either A or B happens."
2. $A \wedge B$ is the event "both A and B happens."
3. $\neg A$ is the event "A does not happens."
4. Ω and $\emptyset$ are events; they are called 'tautology' and 'contradiction,' respectively.

Logic and Sets I

1. $A \vee B$ is the event "either A or B happens."
2. $A \wedge B$ is the event "both A and B happens."
3. $\neg A$ is the event "A does not happens."
4. Ω and $\emptyset$ are events; they are called 'tautology' and 'contradiction,' respectively.

Logic and Sets II

The logical statement $A \vee B$ corresponds to the union of sets $A \cup B$ if A and B are understood as sets of simple events.

The logical statement $A \wedge B$ corresponds to the intersection of sets $A \cap B$ if A and B are understood as sets of simple events.

Events A and B are disjoint (or mutually exclusive) if they cannot occur together. In set theory, we can express this by saying that they are disjoint if and only if $A \cap B = \emptyset$.
Think of dice rolls as an example. $\Omega = \{1, 2, 3, 4, 5, 6\}$. Event A may be $\{1, 2, 3\}$, and event B may be $\{2, 4, 6\}$. What, then, are events $A \cup B$ and $A \cap B$?

Definition of Probability

Let Ω be a set of simple events. An event A is then a subset of Ω. A function P from the collection of all these subsets (sometimes called the power set of Ω) to the real numbers is a probability function if the following three conditions are fulfilled.

1. $P(A) \geq 0$ for all events A.
2. $P(\Omega) = 1$.
3. $P(A \cup B) = P(A) + P(B)$ for any collection of disjoint events A, B.

Definition of Probability

Let Ω be a set of simple events. An event A is then a subset of Ω. A function P from the collection of all these subsets (sometimes called the power set of Ω) to the real numbers is a probability function if the following three conditions are fulfilled.

1. $P(A) \geq 0$ for all events A.
2. $P(\Omega) = 1$.
3. $P(A \cup B) = P(A) + P(B)$ for any collection of disjoint events A, B.

Let Ω be a set of simple events. An event A is then a subset of Ω. A function P from the collection of all these subsets (sometimes called the power set of Ω) to the real numbers is a probability function if the following three conditions are fulfilled.

1. $P(A) \geq 0$ for all events A.
2. $P(\Omega) = 1$.
3. $P(A \cup B) = P(A) + P(B)$ for any collection of disjoint events A, B.

Basic Theorems of Probability

Here are some basic theorems that follow from the conditions.

Rule of Complementary Events

$P(\neg A) = 1 - P(A)$ for all events A

This immediately implies that $P(\emptyset) = 0$ since $\emptyset = \neg\Omega$.

Addition Rule

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Conditional probability of A conditional on B is defined as follows,

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{14}$$

This theorem follows immediately,

Multiplication Rule

$P(A \cap B) = P(A) \cdot P(B|A)$

Two events A and B are independent if and only if $P(A \cap B) = P(A) \cdot P(B)$. Given the multiplication rule, this is equivalent to saying that $P(A|B) = P(A)$ and $P(B|A)P(B)$.

Rules

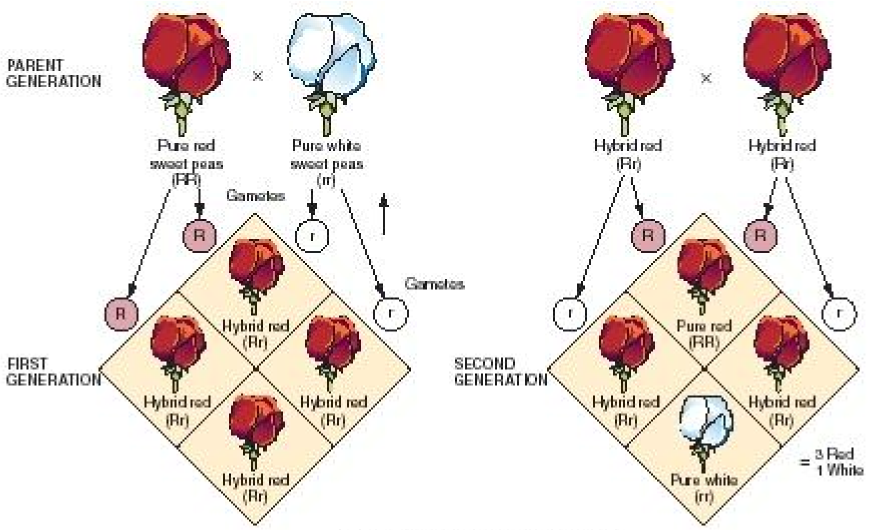
Addition Rule

When A and B are *disjoint*, then $P(A \cup B) = P(A) + P(B)$; otherwise use $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Multiplication Rule

When A and B are *independent*, then $P(A \cap B) = P(A) \cdot P(B)$; otherwise use $P(A \cap B) = P(A|B) \cdot P(B)$

Mendel's Law of Separation



Mendel's First Law: The Law of Segregation

Exercise 3: Your friend tosses two coins. You don't see the coins, but your friend tells you that at least one of them landed heads. What is the probability that they both landed heads?
Exercise 4: Alice and Branden have brown eyes. Their son Joel has blue eyes. What is the probability that their next child will have blue eyes as well?
Exercise 5: In a sample of 207 adults, 43 are smokers. What is the probability of choosing a person at random who is a smoker?

Exercise 6: A game show host asks you a multiple choice question with four answers A, B, C, and D. If you make a random guess, what is your probability of getting the correct answer?

Exercise 7: In a country far away, all parents want to have girls. The probability of having a girl is 50%. All parents have boys until they have a girl. What would you expect to be the proportion of girls in that country?

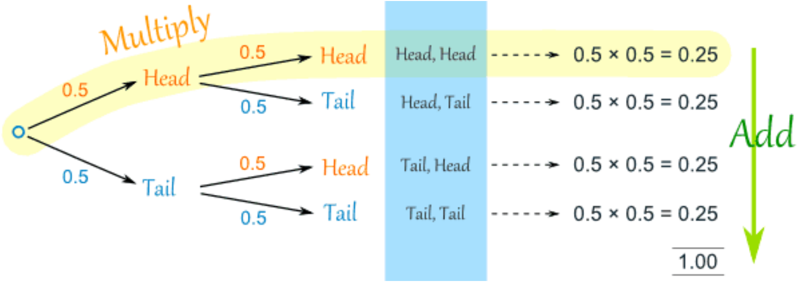
Exercise 8: The government found out that 102 out of 810 luggage scales at the airport are defective. If you choose 2 luggage scales at random *with replacement*, what is the probability that they are both defective? If you choose 2 luggage scales at random *without replacement*, what is the probability that they are both defective?

Exercise 9: The probability that BCIT hires a person on a particular weekday is the same as any other weekday. What is the probability that two randomly selected employees were both hired on a Monday? What is the probability that two randomly selected employees were both hired on the same weekday?
Exercise 10: In a group of people, 492 would choose a window seat on an airplane, 8 would choose a middle seat, and 306 would choose an aisle seat. What is the probability of randomly choosing a person who would not choose a middle seat? What is the probability of randomly choosing two people who would not choose a middle seat? What is the probability of randomly choosing twenty-five people who would not choose a middle seat?

Exercise 11: What is the probability of rolling a sum of 9 on two dice rolls?

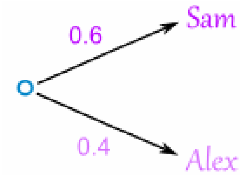
Exercise 12: What is the probability of having two girls and three boys when there are five children and the probability of having a boy is 50%?

You can use independence and mutual exclusion to draw tree diagrams.

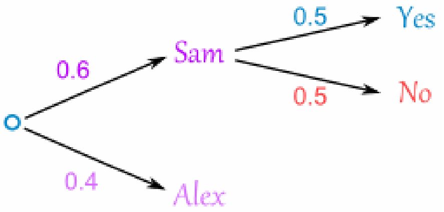


Exercise 13: You have two coaches, Sam and Alex. When Sam coaches the team, your proability of being the goalkeeper is 50%. When Alex coaches the team, your proability of being the goalkeeper is 30%. The probability that Sam (rather than Alex) will coach your team today is 60%. What is the probability that you will be goalkeeper?

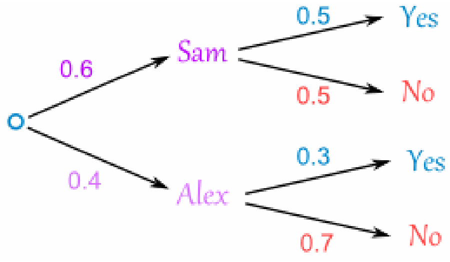
Sam and Alex I



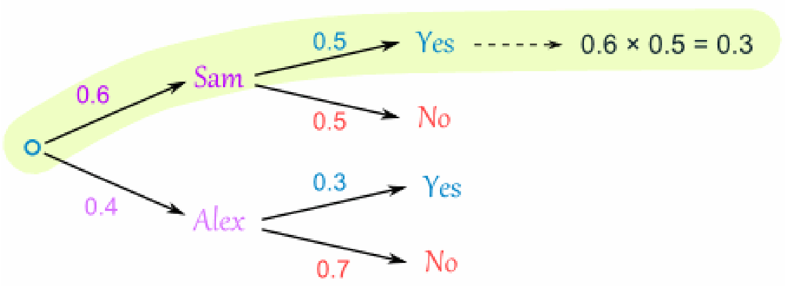
Sam and Alex II



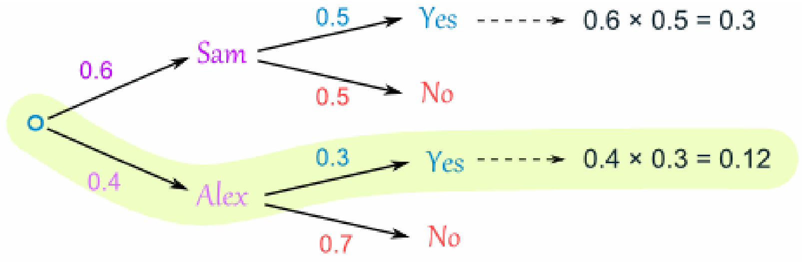
Sam and Alex III



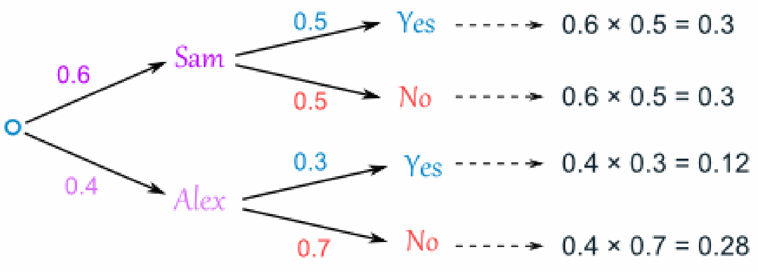
Sam and Alex IV



Sam and Alex V



Sam and Alex VI

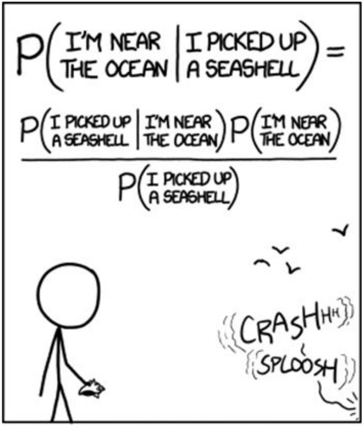


How to Solve Probability Problems

In summary, here are some strategies to solve probability problems.

1. Count simple events. If the simple events are all equally probable, then the probability of event A is the number of simple events in A divided by the total number of simple events, so $P(A) = \#A/\#\Omega$.
2. Make sure to watch for independence and mutual exclusion. Whenever events are independent or mutually exclusive (disjoint), you can use $P(A \cap B) = P(A)P(B)$ or $P(A \cup B) = P(A) + P(B)$, respectively.
3. If events are not mutually exclusive, you can use the addition rule $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
4. If events are not independent, you can use conditional probabilities in $P(A \cap B) = P(A)P(B|A)$.
5. If you are dealing with events that are independent and mutually exclusive, it is often useful to draw a tree diagram.

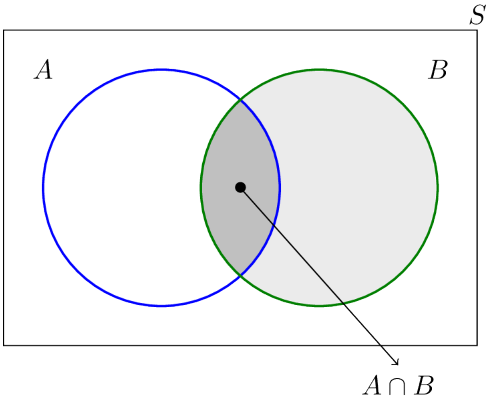
xkcd on Bayes' Formula



STATISTICALLY SPEAKING, IF YOU PICK UP A SEASHELL AND *DON'T* HOLD IT TO YOUR EAR, YOU CAN PROBABLY HEAR THE OCEAN.

Conditional Probability

Let's remember what conditional probability means.



$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Multiplication Rule

Remember the thief who wants to crack the four-digit PIN of a bank card. Let A be the event that she successfully cracks the PIN. If A_1 is the event that she succeeds on her first attempt (and so on for A_2 and A_3), then

$$P(A) = P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) = 0.003 \quad (15)$$

because A_1, A_2, A_3 are disjoint. We are assuming that her attempts happen without replacement. Therefore, A_1, A_2, A_3 are not independent, and the correct application of the multiplication rule is

$$P(A) = 1 - P(\neg A) =$$

$$1 - P(\neg A_1) \cdot P(\neg A_2 | \neg A_1) \cdot P(\neg A_3 | \neg A_1 \cap \neg A_2) = 0.003 \quad (16)$$

Law of Total Probability

It is often easier to calculate conditional probabilities than unconditional probabilities. To express one by the other use the law of total probability,

$$P(A) = P(A|B)P(B) + P(A|\neg B)P(\neg B) \tag{17}$$

This formula also applies when you split up B into three or more disjoint subsets that exhaust B. It follows from set theory.

Example: Suppose that two factories supply light bulbs to the market. Factory X's bulbs work for over 5000 hours in 99% of cases, whereas factory Y's bulbs work for over 5000 hours in 95% of cases. It is known that factory X supplies 60% of the total bulbs available. What is the chance that a purchased bulb will work for longer than 5000 hours?

Law of Total Probability

Example: Suppose that two factories supply light bulbs to the market. Factory X's bulbs work for over 5000 hours in 99% of cases, whereas factory Y's bulbs work for over 5000 hours in 95% of cases. It is known that factory X supplies 60% of the total bulbs available. What is the chance that a purchased bulb will work for longer than 5000 hours?

Let X be the event that the light bulb is from factory X. Let F be the event that the bulb will work for longer than 5000 hours. Then

$$P(F) = P(F|X)P(X) + P(F|\neg X)P(\neg X) =$$

$$0.99 \cdot 0.60 + 0.95 \cdot 0.40 = 0.974 \tag{18}$$

Law of Total Probability Exercises I

What is the probability that the second card in a conventional deck of cards is an ace?

Law of Total Probability Exercises II

Suppose we have two hats: one has 4 red balls and 7 green balls, the other has 11 red and 5 green. We toss an unfair coin (60/40 for heads), if heads, pick a random ball from the first hat, if tails from the second. What is the probability of getting a red ball?

Law of Total Probability Exercises III

You have three bags that each contain 100 marbles:

- Bag 1 has 75 red and 25 blue marbles
- Bag 2 has 60 red and 40 blue marbles
- Bag 3 has 45 red and 55 blue marbles

I choose one of the bags at random and then pick a marble from the chosen bag, also at random. What is the probability that the chosen marble is red?

Some Interesting Cases

A group of police officers have breathalyzers displaying false drunkenness in 5% of the cases in which the driver is sober. However, the breathalyzers never fail to detect a truly drunk person. One in a thousand drivers is driving drunk. Suppose the police officers then stop a driver at random, and force the driver to take a breathalyzer test. It indicates that the driver is drunk. We assume you don't know anything else about him or her. How high is the probability he or she really is drunk?

Some Interesting Cases

A room is full of engineers and lawyers (most of them are lawyers, 90%). The probability that an engineer enjoyed physics in school is 80%. The probability that a lawyer enjoyed physics in school is 30%. You ask someone in the room whether they enjoyed physics, and the answer is yes. Should you bet that this person is a lawyer, or should you bet that she is an engineer?

Some Interesting Cases

You have a million food items, of which 1 in 1000 is contaminated. You have a contamination test with a 2% false positive rate and a 0.5% false negative rate. A food item tests positive for contamination. What is the probability that it is contaminated?

Some Interesting Cases

In a city of 1 million inhabitants let there be 100 terrorists and 999,900 non-terrorists. To simplify the example, it is assumed that all people present in the city are inhabitants. Thus, the base rate probability of a randomly selected inhabitant of the city being a terrorist is 0.0001, and the base rate probability of that same inhabitant being a non-terrorist is 0.9999. In an attempt to catch the terrorists, the city installs an alarm system with a surveillance camera and automatic facial recognition software.

The software has two failure rates of 1%:

- The false negative rate: If the camera scans a terrorist, a bell will ring 99% of the time, and it will fail to ring 1% of the time.
- The false positive rate: If the camera scans a non-terrorist, a bell will not ring 99% of the time, but it will ring 1% of the time.

Suppose now that an inhabitant triggers the alarm. What is the chance that the person is a terrorist?

Bayes' Formula

Consider the definition of conditional probability,

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \tag{19}$$

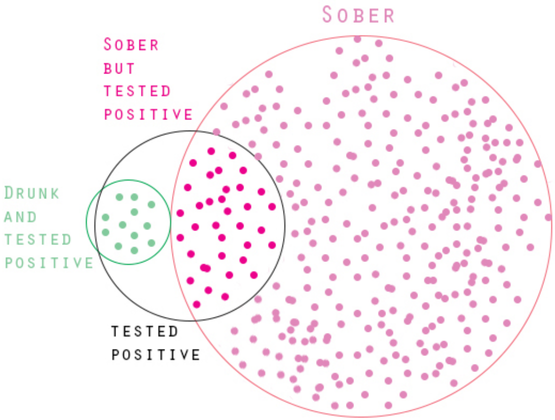
Now notice that $P(B \cap A) = P(A \cap B) = P(B)P(A|B)$. That means that

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)} \tag{20}$$

By the law of total probability we can replace the denominator to give us Bayes' Formula

$$P(B|A) = \frac{P(B)P(A|B)}{P(A|B)P(B) + P(A|\neg B)P(\neg B)} \tag{21}$$

Base Rate Fallacy Diagram



Base Rate Fallacy Example

Let 100 out of 100,000 people have a disease. The test for this disease has a 5% false positive rate and a 5% false negative rate. If you test positive for this disease, what is your probability of actually having the disease. Consider the following contingency table and then apply Bayes' formula.

		Have Disease	
		Yes	No
Test Results	Positive	95	4,995
	Negative	5	94,905

Contingency Tables

Event	Event: B_1	Event: B_2	Total
A_1	$P(A_1 \text{ and } B_1)$	$P(A_1 \text{ and } B_2)$	$P(A_1)$
A_2	$P(A_2 \text{ and } B_1)$	$P(A_2 \text{ and } B_2)$	$P(A_2)$
Total	$P(B_1)$	$P(B_2)$	1

Joint Probabilities

Marginal (Simple) Probabilities

Prison and Plea

Here is a contingency table:

	Guilty Plea	Plea of Not Guilty
Sentenced to Prison	392	58
Not Sentenced to Prison	564	14

Prison and Plea

Answer the following questions:

1. Find the probability of a randomly selected subject being sentenced to prison.
2. Find the probability of being sentenced to prison, given that the subject entered a plea of guilty.
3. Find the probability of being sentenced to prison, given that the subject entered a plea of not guilty.
4. Find the probability of a randomly selected subject being sentenced to prison or entering a plea of guilty.

Prison and Plea

Answer the following questions:

5. If two subjects are randomly selected, find the probability that they were both sentenced to prison.
6. If two subjects are randomly selected, find the probability that they both entered pleas of not guilty.
7. Find the probability of a randomly selected subject being entering a plea of not guilty or not being sentenced to prison.
8. Find the probability of a randomly selected subject being sentenced to prison and entering a plea of guilty.
9. Find the probability of a randomly selected subject not being sentenced to prison and not entering a plea of guilty.

Exercises I

Three urns contain respectively 1 white and 2 black balls; 3 white and 1 black ball; 2 white and 3 black balls. One ball is taken from each urn. What is the probability that among the balls drawn there are 2 white and 1 black?

Exercises I

Three urns contain respectively 1 white and 2 black balls; 3 white and 1 black ball; 2 white and 3 black balls. One ball is taken from each urn. What is the probability that among the balls drawn there are 2 white and 1 black? Answer: 23/60

Exercises II

A student has a box containing 25 computer disks, of which 15 are blank and 10 are not. She randomly selects disks one by one and examines each one, terminating the process only when she finds a blank disk. What is the probability that she must examine at least two disks?

Exercises II

A student has a box containing 25 computer disks, of which 15 are blank and 10 are not. She randomly selects disks one by one and examines each one, terminating the process only when she finds a blank disk. What is the probability that she must examine at least two disks? Answer: 40%

Exercises III

There are five faculty members in a certain academic department. These individuals have 3, 6, 7, 10, and 14 years of teaching experience, respectively. Two of these individuals are randomly selected to serve on a committee. What is the probability that they have at least 15 years of teaching experience?

Exercises III

There are five faculty members in a certain academic department. These individuals have 3, 6, 7, 10, and 14 years of teaching experience, respectively. Two of these individuals are randomly selected to serve on a committee. What is the probability that they have at least 15 years of teaching experience? Answer: 60%

Exercises IV

Suppose three cards are selected from a well-mixed deck without replacement.

1. What is the probability that all three are hearts?
2. What is the probability that all three are from the same suit?
3. If five cards are dealt from a randomized deck, determine the probability that they are all hearts.

Exercises IV

Suppose three cards are selected from a well-mixed deck without replacement.

1. What is the probability that all three are hearts?
2. What is the probability that all three are from the same suit?
3. If five cards are dealt from a randomized deck, determine the probability that they are all hearts.

Exercises IV

Suppose three cards are selected from a well-mixed deck without replacement.

1. What is the probability that all three are hearts?
2. What is the probability that all three are from the same suit?
3. If five cards are dealt from a randomized deck, determine the probability that they are all hearts.

Exercises IV

Suppose three cards are selected from a well-mixed deck without replacement.

1. What is the probability that all three are hearts? Answer: 1.29%
2. What is the probability that all three are from the same suit? Answer: 5.18%
3. If five cards are dealt from a randomized deck, determine the probability that they are all hearts. Answer: 0.0495%

Exercises V

A tennis coach has brought out 12 tubes of Penn balls and 8 tubes of Wilson balls for his class. If 5 tubes are randomly selected, what is the probability that all 5 are of the same brand?

Exercises V

A tennis coach has brought out 12 tubes of Penn balls and 8 tubes of Wilson balls for his class. If 5 tubes are randomly selected, what is the probability that all 5 are of the same brand? Answer: 5.47%

In Orange County, 51% of the adults are males. (It doesn't take too much advanced mathematics to deduce that the other 49% are females.) One adult is randomly selected for a survey involving credit card usage.

a. Find the prior probability that the selected person is a male.

b. It is later learned that the selected survey subject was smoking a cigar. Also, 9.5% of males smoke cigars, whereas 1.7% of females smoke cigars (based on data from the Substance Abuse and Mental Health Services Administration). Use this additional information to find the probability that the selected subject is a male.

In Orange County, 51% of the adults are males. (It doesn't take too much advanced mathematics to deduce that the other 49% are females.) One adult is randomly selected for a survey involving credit card usage.

a. Find the prior probability that the selected person is a male.

b. It is later learned that the selected survey subject was smoking a cigar. Also, 9.5% of males smoke cigars, whereas 1.7% of females smoke cigars (based on data from the Substance Abuse and Mental Health Services Administration). Use this additional information to find the probability that the selected subject is a male.

Answer: (a.) 51% (b.) 85.3%

Exercises VII

7. Pleas and Sentences In a study of pleas and prison sentences, it is found that 45% of the subjects studied were sent to prison. Among those sent to prison, 40% chose to plead guilty. Among those not sent to prison, 55% chose to plead guilty.

a. If one of the study subjects is randomly selected, find the probability of getting someone who was not sent to prison.

b. If a study subject is randomly selected and it is then found that the subject entered a guilty plea, find the probability that this person was not sent to prison.

7. Pleas and Sentences In a study of pleas and prison sentences, it is found that 45% of the subjects studied were sent to prison. Among those sent to prison, 40% chose to plead guilty. Among those not sent to prison, 55% chose to plead guilty.

a. If one of the study subjects is randomly selected, find the probability of getting someone who was not sent to prison.

b. If a study subject is randomly selected and it is then found that the subject entered a guilty plea, find the probability that this person was not sent to prison.

Answer: (a.) 55% (b.) 62.7%

13. Biased Coin In an article about confusion of eyewitnesses, John Allen Paulos cites the problem of three coins, one of which is biased so that it turns up heads 75% of the time. If you randomly select one of the coins, toss it three times, and obtain three heads, what is the probability that this is the biased coin?

Exercises VIII

13. Biased Coin In an article about confusion of eyewitnesses, John Allen Paulos cites the problem of three coins, one of which is biased so that it turns up heads 75% of the time. If you randomly select one of the coins, toss it three times, and obtain three heads, what is the probability that this is the biased coin?

Answer: 62.8%

Probability Distributions: Concepts

Here are some definitions.

random variable A random variable is a variable (typically represented by X) that has a single numerical value, determined by chance, for each outcome of a procedure.

probability distribution A probability distribution is a description that gives the probability for each value of the random variable. It is often expressed in the format of a table, formula, or graph.

Probability Distributions: Discrete and Continuous

discrete random variable A discrete random variable has a collection of values that is finite or countable.

continuous random variable A continuous random variable has infinitely many values, and the collection of values is not countable.

countability This is best explained by example: the integers are countable, but the real numbers are not.

Discrete Probability Distributions

If there are a finite number of outcomes $X = a_k$ for $k = 1, \ldots, n$, we can list the values of $P(X = a_k)$ in a table.

Example 1: Coin Toss. let $X = 1$ for heads and $X = 0$ for tails. Then

Event	Probability
$X = 1$ or H	0.50
$X = 0$ or T	0.50

When all the probabilities are equal, we call the probability distribution a uniform distribution.

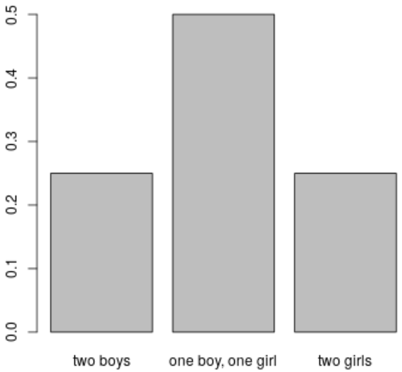
Non-Uniform Discrete Probability Distributions

Some distribution probability distributions are not uniform.

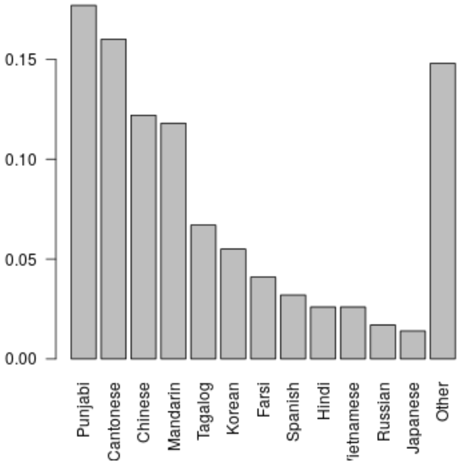
Example 2: Number of Male Children. Consider the two-child family. If X is the random variable corresponding to the number of boys in the family, then the probability distribution table looks as follows (assuming that the probability distribution for one child is uniform).

Discrete Probability Distribution Graphs I

Event	Probability
$X = 2$ or two boys	0.25
$X = 1$ or one boy, one girl	0.50
$X = 0$ or two girls	0.25



Example 3: Immigrant Languages. Here is the probability distribution for a randomly selected "Vancouverite" (Greater Vancouver) to speak a certain immigrant language at home.



Mean and Variance Formulas

There is a sense in which a probability distribution together with its associated random variable correspond to a population and the property which the random variable picks out. In this spirit, let us define a mean and a variance for a probability distribution.

$$\mu = \sum X \cdot P(X) \tag{22}$$

$$\sigma^2 = \sum (X - \mu)^2 \cdot P(X) \tag{23}$$

$$\sigma^2 = \sum (X^2 \cdot P(X)) - \mu^2 \tag{24}$$

$$\sigma = \sqrt{\sum (X^2 \cdot P(X)) - \mu^2} \tag{25}$$

Die Roll Example

Example 4: Fair Die Roll. Think of rolling a fair die many times. The probability distribution is uniform. The mean is

$$\mu = \sum X \cdot P(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

We also call this number the expectation EX of the random variable X. Although you would never expect a die roll to result in "3.5," you would expect the mean of many die rolls to be close to this number. The expected number of boys for one birth is $EX = 0.5$.

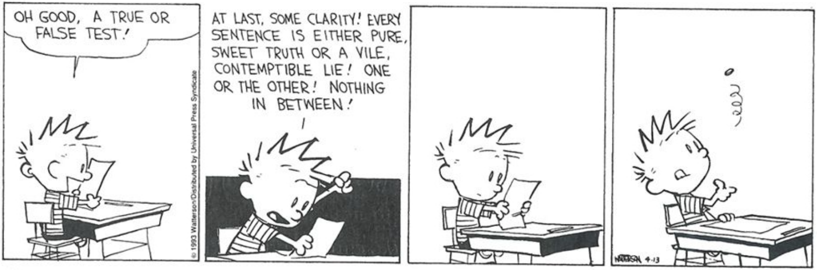
The Binomial Probability Distribution

A binomial probability distribution results from a procedure that meets the following requirements.

1. The procedure has a fixed number of trials.
2. The trials must be independent.
3. The outcomes of a trial are binary, i.e. there are only two possible outcomes.
4. The probability of the two outcomes remains constant.

The number of trials is usually labeled n, the two outcomes are called success and failure, and their probabilities on one trial are p and $1 - p$. The random variable keeps track of the number of successes. If, for example, there are 10 trials, then $P(X = 4)$ is the probability of 4 successes out of 10. The number of successes is often labeled x, and we are usually interested in $P(X = x)$.

Calvin on the Binomial Distribution



The Binomial Probability Formula

If n, p, x are as described on the previous slide, then

$$P(X = x) = \binom{n}{x} \cdot p^x \cdot (1 - p)^{n-x} \tag{26}$$

Exercises for the Binomial Distribution I

Exercise 14: If you randomly guess on a multiple choice test with four possible answers, what is your probability of getting strictly more than 50% of questions right when there are six questions?

Exercises for the Binomial Distribution II

Exercise 15: The incidence of blue eyes in the population is 12%. In a room with 20 randomly selected people, what is the probability of having three or more people with blue eyes? What is the probability of having strictly fewer than five people with blue eyes?

> *Strictly speaking, the binomial probabilities are only approximate because the selection happens without replacement. If the population is large from which the sample is drawn, then you are allowed to ignore this.*

Exercise 16: Here is the distribution of blood types in Canada.

	O	A	B	AB
Positive	0.390	0.360	0.076	0.025
Negative	0.070	0.060	0.014	0.005

(a) What is the probability of being rhesus factor positive for someone of blood type "A"?

(b) If you meet four randomly selected Canadians, what is the probability that two of them are "O" positive?

(c) In a room with twelve randomly selected Canadians, what is the probability that there are strictly fewer than three people with blood type "B"?

Exercise 17: 80.5% of US flights arrive on time. For twelve randomly selected flights, what is the probability that exactly ten of them are on time? What is the probability that between two and four of them are not on time?

Mean and Variance for the Binomial Distribution

There are formulas for the mean and variance of the binomial distribution. Especially the formula for the mean makes immediate sense:

Formulas

mean	μ	$=$	np
variance	σ^2	$=$	npq
standard deviation	σ	$=$	$\sqrt{npq}$

It is a useful rule of thumb to remember that it is unlikely ($< 5\%$) that x is outside of the interval from $\mu - 2\sigma$ to $\mu + 2\sigma$.

Exercises for the Binomial Distribution V

Exercise 18: What is the rule-of-thumb 95% interval for the following binomial procedures:

1. flipping a fair coin 15 times
2. answering 60 multiple choice questions with four possible answers for each question, where the probability of getting the right answer is 80%
3. randomly answering 60 multiple choice questions with four possible answers for each question
4. The number of "O" positive blood types in a crowd of 100 Canadians.

Exercises for the Binomial Distribution V

Exercise 19: What is the rule-of-thumb 95% interval for the following binomial procedures:

1. flipping a fair coin 15 times
2. answering 60 multiple choice questions with four possible answers for each question, where the probability of getting the right answer is 80%
3. randomly answering 60 multiple choice questions with four possible answers for each question
4. The number of "O" positive blood types in a crowd of 100 Canadians.

Exercise 20: What is the rule-of-thumb 95% interval for the following binomial procedures:

1. flipping a fair coin 15 times
2. answering 60 multiple choice questions with four possible answers for each question, where the probability of getting the right answer is 80%
3. randomly answering 60 multiple choice questions with four possible answers for each question
4. The number of "O" positive blood types in a crowd of 100 Canadians.

Exercise 21: What is the rule-of-thumb 95% interval for the following binomial procedures:

1. flipping a fair coin 15 times
2. answering 60 multiple choice questions with four possible answers for each question, where the probability of getting the right answer is 80%
3. randomly answering 60 multiple choice questions with four possible answers for each question
4. The number of "O" positive blood types in a crowd of 100 Canadians.

Exercise 22: Based on observed males using public restrooms, 85% of adult males wash their hands in a public restroom (based on data from the American Society for Microbiology and the American Cleaning Institute). In a survey of 523 adult males, 518 reported that they wash their hands in a public restroom. Assuming that the 85% observed rate is correct, find the probability that among 523 randomly selected adult males, 518 or more wash their hands in a public restroom. What do you conclude?

Exercise 23: In a survey of 1002 people, 701 said that they voted in a recent presidential election (based on data from ICR Research Group). Voting records show that 61% of eligible voters actually did vote. Given that 61% of eligible voters actually did vote, find the probability that among 1002 randomly selected eligible voters, at least 701 actually did vote. What does the result suggest?

Exercise 24: In a study of 420,095 cell phone users in Denmark, it was found that 135 developed cancer of the brain or nervous system. Assuming that the use of cell phones has no effect on developing such cancers, there is a 0.000340 probability of a person developing cancer of the brain or nervous system. We therefore expect about 143 cases of such cancers in a group of 420,095 randomly selected people. Estimate the probability of 135 or fewer cases of such cancers in a group of 420,095 people. What do these results suggest about media reports that cell phones cause cancer of the brain or nervous system?

Exercise 25: Based on a recent Harris Interactive survey, 20% of adults in the United States smoke. In a survey of 50 statistics students, it is found that six of them smoke. Find the probability that should be used for determining whether the 20% rate is correct for statistics students. What do you conclude?

Exercise 26: Online TV In a Comcast survey of 1000 adults, 17% said that they watch prime-time TV online. If we assume that 20% of adults watch prime-time TV online, find the probability that should be used to determine whether the 20% rate is correct or whether it should be lower than 20%? What do you conclude?

Exercise 27: Internet Access Of U.S. households, 67% have Internet access (based on data from the Census Bureau). In a random sample of 250 households, 70% are found to have Internet access. Find the probability that should be used to determine whether the 67% rate is too low. What do you conclude?

End of Lesson

Next Lesson: Central Limit Theorem