Statistics Part-1

By Roshan Sharma

Let’s get Started

Topics

To Learn and Understand!

- Graphing Distributions
- Summarizing Distributions
- Bivariate Data
- Probability

Let’s get Started

Introduction

Statistics is a subfield of mathematics. It refers to a collection of methods for working with data and using data to answer questions.

It is because the field is comprised of a grab bag of methods for working with data that it can seem large and amorphous to beginners. It can be hard to see the line between methods that belong to statistics and methods that belong to other fields of study.

When it comes to the statistical tools that we use in practice, it can be helpful to divide the field of statistics into two large groups of methods: descriptive statistics for summarizing data, and inferential statistics for drawing conclusions from samples of data.

- **Descriptive Statistics**: Descriptive statistics refer to methods for summarizing raw observations into information that we can understand and share.
- **Inferential Statistics**: Inferential statistics is a fancy name for methods that aid in quantifying properties of the domain or population from a smaller set of obtained observations called a sample.

Descriptive Vs Inferential Statistics

Descriptive statistics are just descriptive. They do not involve generalizing beyond the data at hand. Generalizing from our data to another set of cases is the business of inferential statistics

Inferential Statistics Involves the Study of Fiding and studying the patterns and analyzing the behaviour to Understand the the Data in depth and come up with Strong Inferential Information.

Population vs Samples

In statistics, we often rely on a sample that is, a small subset of a larger set of data to draw inferences about the larger set. The larger set is known as the population from which the sample is drawn.

Example: You have been hired by the National Election Commission to examine how the American people feel about the fairness of the voting procedures in the U.S. Who will you ask?

It is not practical to ask every single American how he or she feels about the fairness of the voting procedures. Instead, we query a relatively small number of Americans, and draw inferences about the entire country from their responses.

Simple Random Sampling

Simple Random sampling requires every member of the population to have an equal chance of being selected into the sample. In addition, the selection of one member must be independent of the selection of every other member. That is, picking one member from the population must not increase or decrease the probability of picking any other member (relative to the others). In this sense, we can say that simple random sampling chooses a sample by pure chance

Example

A research scientist is interested in studying the experiences of twins raised together versus those raised apart. She obtains a list of twins from the National Twin Registry, and selects two subsets of individuals for her study. First, she chooses all those in the registry whose last name begins with Z. Then she turns to all those whose last name begins with B. Because there are so many names that start with B, however, our researcher decides to incorporate only every other name into her sample. Finally, she mails out a survey and compares characteristics of twins raised apart versus together.

Sample Size

Recall that the definition of a random sample is a sample in which every member of the population has an equal chance of being selected. This means that the sampling procedure rather than the results of the procedure define what it means for a sample to be random. Random samples, especially if the sample size is small, are not necessarily representative of the entire population. For example, if a random sample of 20 subjects were taken from a population with an equal number of males and females, there would be a nontrivial probability (0.06) that 70% or more of the sample would be female.

Complex Sampling

Sometimes it is not feasible to build a sample using simple random sampling. To see the problem, consider the fact that both Dallas and Houston are competing to be hosts of the 2012 Olympics. Imagine that you are hired to assess whether most Texans prefer Houston to Dallas as the host, or the reverse. Given the impracticality of obtaining the opinion of every single Texan, you must construct a sample of the Texas population. But now notice how difficult it would be to proceed by simple random sampling. For example, how will you contact those individuals who don't vote and don't have a phone? Even among people you find in the telephone book, how can you identify those who have just relocated to California (and had no reason to inform you of their move)? What do you do about the fact that since the beginning of the study, an additional 4,212 people took up residence in the state of Texas? As you can see, it is sometimes very difficult to develop a truly random procedure. For this reason, other kinds of sampling techniques have been devised. We now discuss two of them.

There are two types of Complex Sampling

- Random Assignment
- Stratified Sampling

Random Assignment

In experimental research, populations are often hypothetical. For example, in an experiment comparing the effectiveness of a new antidepressant drug with a placebo, there is no actual population of individuals taking the drug. In this case, a specified population of people with some degree of depression is defined and a random sample is taken from this population. The sample is then randomly divided into two groups; one group is assigned to the treatment condition (drug) and the other group is assigned to the control condition (placebo). This random division of the sample into two groups is called **random assignment**.

Stratified Sampling

Since simple random sampling often does not ensure a representative sample, a sampling method called stratified random sampling is sometimes used to make the sample more representative of the population. This method can be used if the population has a number of distinct “strata” or groups. In stratified sampling, you first identify members of your sample who belong to each group. Then you randomly sample from each of those subgroups in such a way that the sizes of the subgroups in the sample are proportional to their sizes in the population.

Variables

Independent vs Dependent Variables

Variables are properties or characteristics of some event, object, or person that can take on different values or amounts. When conducting research, experimenters often manipulate variables. For example, an experimenter might compare the effectiveness of four types of antidepressants. In this case, the variable is "type of antidepressant." When a variable is manipulated by an experimenter, it is called an independent variable. The experiment seeks to determine the effect of the independent variable on relief from depression. In this example, relief from depression is called a dependent variable. In general, the independent variable is manipulated by the experimenter and its effects on the dependent variable are measured.

Qualitative vs Quantitative Variables

An important distinction between variables is between qualitative variables and quantitative variables. Qualitative variables are those that express a qualitative attribute such as hair color, eye color, religion, favorite movie, gender, and so on. The values of a qualitative variable do not imply a numerical ordering. Values of the variable "religion" differ qualitatively; no ordering of religions is implied. Qualitative variables are sometimes referred to as categorical variables. Quantitative variables are those variables that are measured in terms of numbers. Some examples of quantitative variables are height, weight, and shoe size

Discrete vs Continuous Variables

Variables such as number of children in a household are called discrete variables since the possible scores are discrete points on the scale. For example, a household could have three children or six children, but not 4.53 children. Other variables such as "time to respond to a question" are continuous variables since the scale is continuous and not made up of discrete steps. The response time could be 1.64 seconds, or it could be 1.64237123922121 seconds. Of course, the practicalities of measurement preclude most measured variables from being truly continuous.

Percentiles

There is no universally accepted definition of a percentile. Using the 65th percentile as an example, the 65th percentile can be defined as the lowest score that is greater than 65% of the scores. This is the way we defined it above and we will call this "Definition 1." The 65th percentile can also be defined as the smallest score that is greater than or equal to 65% of the scores. This we will call "Definition 2."

The first step is to compute the rank (R) of the 25th percentile. This is done using the following formula:

$$R = P/100 *(N+1)$$

Example: You are the fourth tallest person in a group of 20

80% of people are shorter than you:

That means you are at the **80th percentile**.

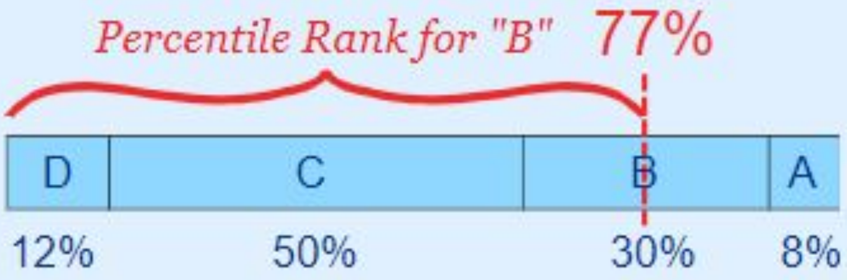
If your height is 1.85m then "1.85m" is the 80th percentile height in that group.

Example: You Score a B!

In the test 12% got D, 50% got C, 30% got B and 8% got A

You got a B, so add up

- all the 12% that got D,
- all the 50% that got C,
- half of the 30% that got B,



for a total percentile of 12% + 50% + 15% = **77%**

In other words you did "as well or better than 77% of the class"

(Why take half of B? Because you shouldn't imagine you got the "Best B", or the "Worst B", just an average B.)

Deciles

Example: (continued)

You are at the **8th decile** (the 80th percentile).

Quartiles

The Quartiles also divide the data into divisions of 25%, so:

- Quartile 1 (Q1) can be called the **25th percentile**
- Quartile 2 (Q2) can be called the **50th percentile**
- Quartile 3 (Q3) can be called the **75th percentile**

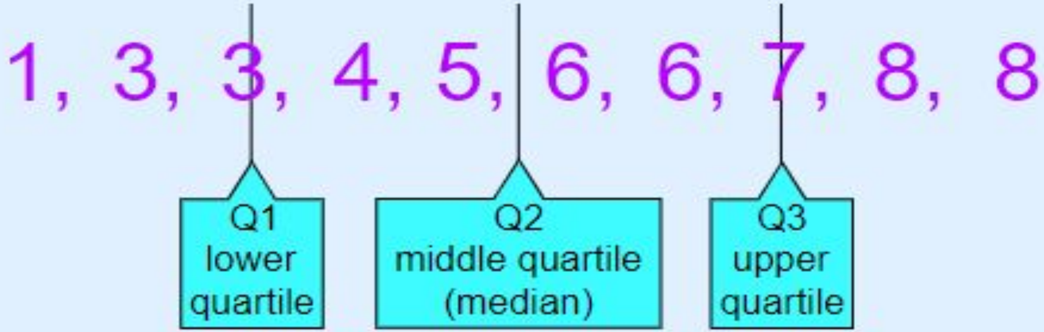
Example: (continued)

For **1, 3, 3, 4, 5, 6, 6, 7, 8, 8**:

- The 25th percentile = **3**
- The 50th percentile = **5.5**
- The 75th percentile = **7**

Example: 1, 3, 3, 4, 5, 6, 6, 7, 8, 8

The numbers are in order. Cut the list into quarters:



In this case Quartile 2 is half way between 5 and 6:

$$Q2 = (5+6)/2 = \mathbf{5.5}$$

And the result is:

- Quartile 1 (Q1) = **3**
- Quartile 2 (Q2) = **5.5**
- Quartile 3 (Q3) = **7**

Let’s Solve Some Problems

Example: Shopping

A total of 10,000 people visited the shopping mall over 12 hours:

Time (hours)	People
0	0
2	350
4	1100
6	2400
8	6500
10	8850
12	10,000

a) Estimate the 30th percentile (when 30% of the visitors had arrived).

b) Estimate what percentile of visitors had arrived after 11 hours.

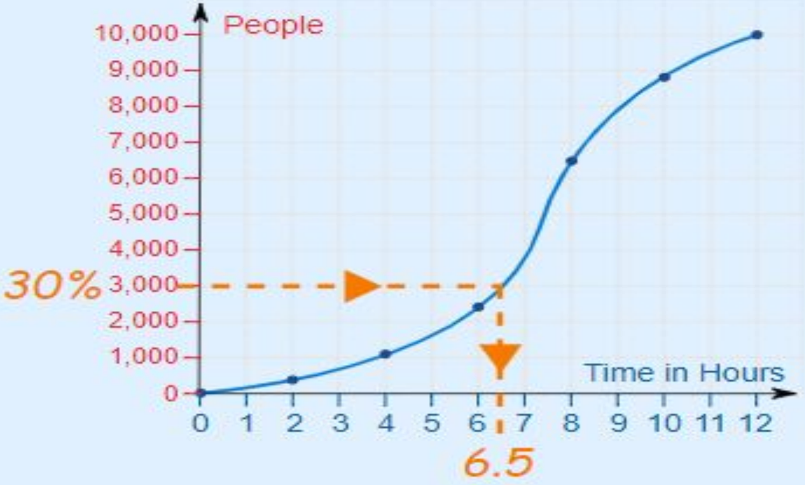
Step 1:

First draw a line graph of the data: plot the points and join them with a smooth curve:

Step 2:

a) The 30th percentile occurs when the visits reach 3,000.

Draw a line horizontally across from 3,000 until you hit the curve, then draw a line vertically downwards to read off the time on the horizontal axis:



So the **30th percentile** occurs after about **6.5 hours**.

Step 3:

b) To estimate the percentile of visits after 11 hours: draw a line vertically up from 11 until you hit the curve, then draw a line horizontally across to read off the population on the vertical axis:

So the visits at **11 hours** were about 9,500, which is the **95th percentile**.

Nominal Scale: When measuring using a nominal scale, one simply names or categorizes responses. Gender, handedness, favorite color, and religion are examples of variables measured on a nominal scale.

Ordinal Scale: A researcher wishing to measure consumers' satisfaction with their microwave ovens might ask them to specify their feelings as either “very dissatisfied,” “somewhat dissatisfied,” “somewhat satisfied,” or “very satisfied.” The items in this scale are ordered, ranging from least to most satisfied. This is what distinguishes ordinal from nominal scales.

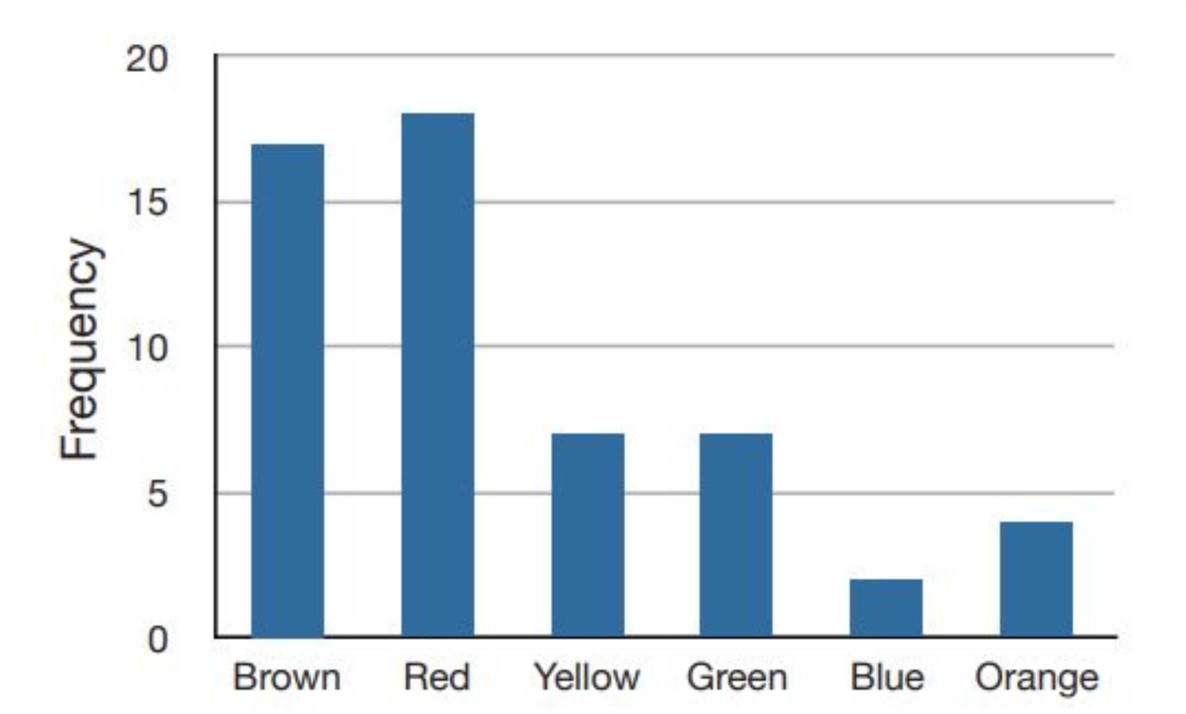
Interval Scales: Interval scales are numerical scales in which intervals have the same interpretation throughout. As an example, consider the Fahrenheit scale of temperature. The difference between 30 degrees and 40 degrees represents the same temperature difference as the difference between 80 degrees and 90 degrees.

Ratio Scales: The ratio scale of measurement is the most informative scale. It is an interval scale with the additional property that its zero position indicates the absence of the quantity being measured.

Distributions

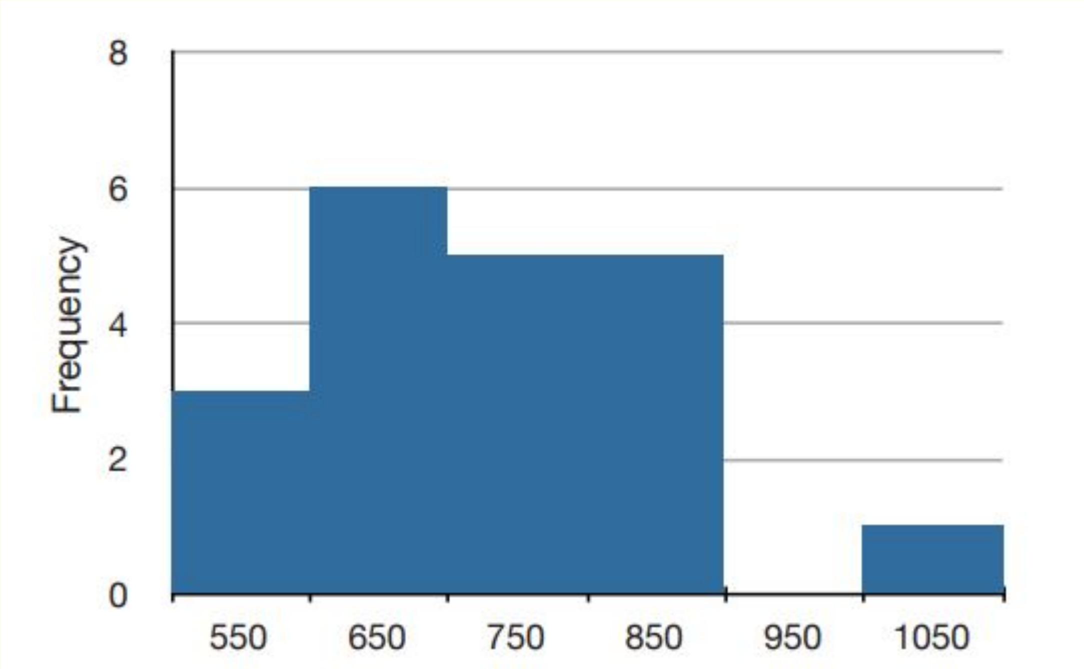
- Distribution for Discrete Variables
- Distributions for Continuous Variables
- Probability Densities
- Shapes of Distributions

Discrete Variables



Distributions for Discrete Variables can be best showcased by Bar Charts. This Bar Chart Depicts the Values for each of the Variables present in the data.

Continuous Variables



Distributions for Continuous Variables can be best showcased by Distribution Charts. This Chart Depicts the Values of the Range of the Continuous Variables present in the data.

Probability Density

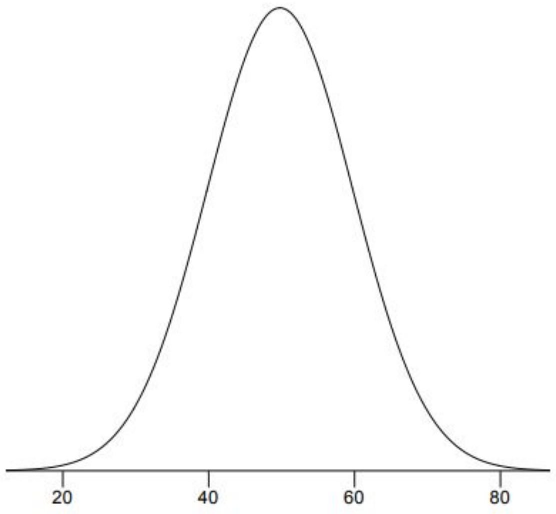


Figure 4. A normal distribution.

we plot the distribution for the continuous variable of time. Distributions for continuous variables are called continuous distributions. They also carry the fancier name probability density. Some probability densities have particular importance in statistics. A very important one is shaped like a bell, and called the normal distribution. Many naturally-occurring phenomena can be approximated surprisingly well by this distribution. It will serve to illustrate some features of all continuous distributions.

Insights from Bell Curve

- Should keep the following ideas in mind about the curve that describes a continuous distribution (like the normal distribution).
    - First, the area under the curve equals 1.
    - Second, the probability of any exact value of X is 0.
    - Finally, the area under the curve and bounded between two given points on the X-axis is the probability that a number chosen at random will fall between the two points.

Shapes of Distributions

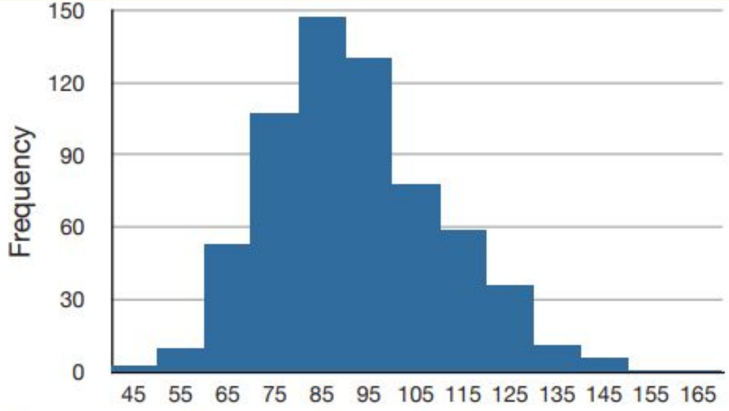


Figure 5. A distribution with a positive skew.

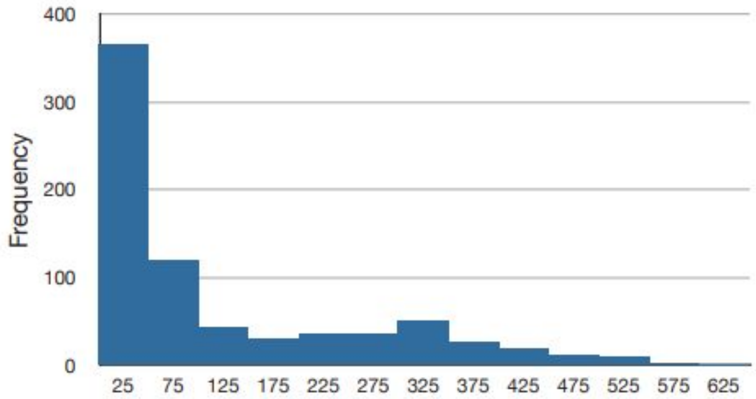


Figure 6. A distribution with a very large positive skew.

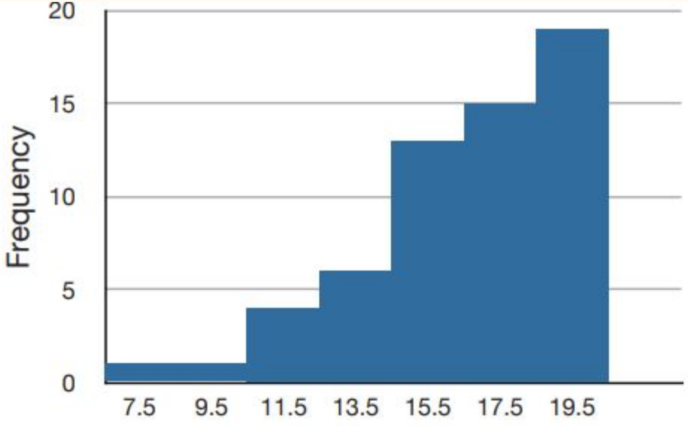


Figure 8. A distribution with negative skew. This histogram shows the

Logarithms

Rule 1: $\log_b(M \cdot N) = \log_b M + \log_b N$

Rule 2: $\log_b\left(\frac{M}{N}\right) = \log_b M - \log_b N$

Rule 3: $\log_b\left(M^k\right) = k \cdot \log_b M$

Rule 4: $\log_b(1) = 0$

Rule 5: $\log_b(b) = 1$

Rule 6: $\log_b\left(b^k\right) = k$

Rule 7: $b^{\log_b(k)} = k$

Where: $b > 1$, and M, N and k can be any real numbers

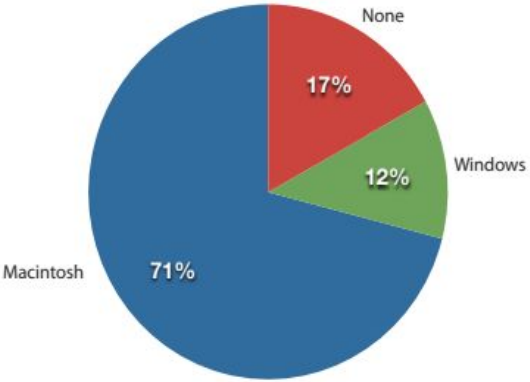
but M and N must be positive!

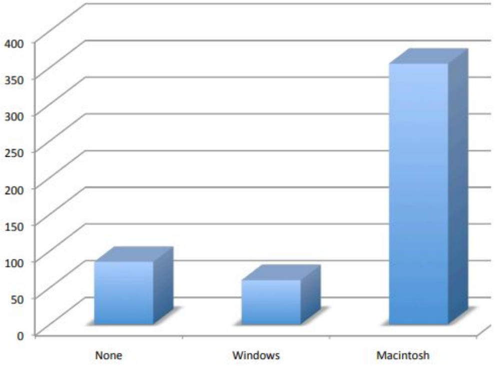
Graphing Distributions

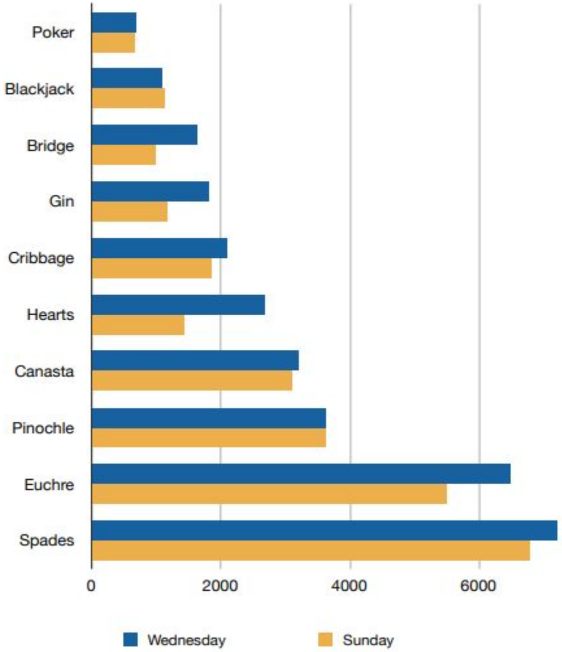
Graphing Qualitative Variables

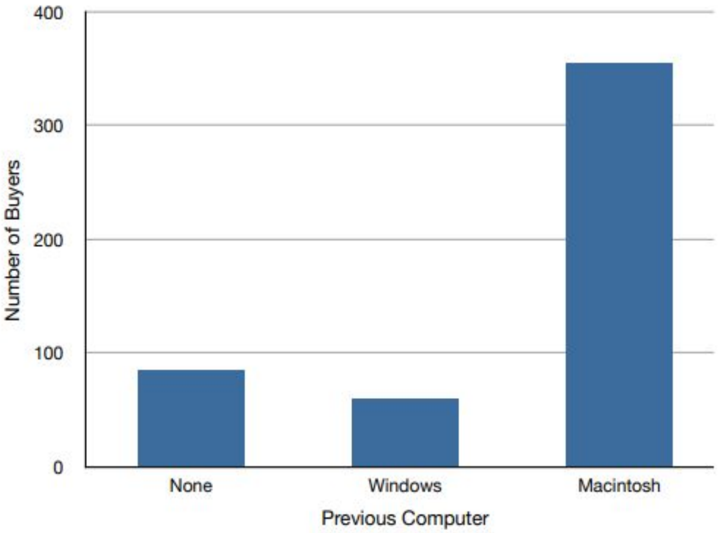
- Frequency Tables
- Pie Charts
- Bar Charts
- Grouped Bar Charts
- Stacked Bar Charts
- 3D Bar Charts
- Line Charts
- Pictorial Charts/Bubble Charts

Charts for Qualitative Variables







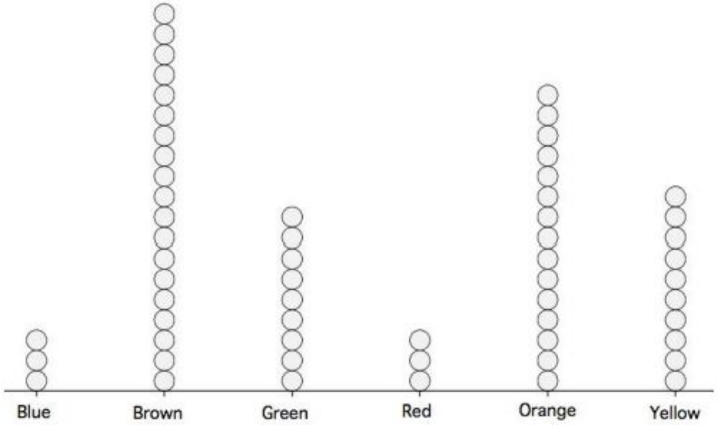


Previous Ownership	Frequency	Relative Frequency
None	85	0.17
Windows	60	0.12
Macintosh	355	0.71
Total	500	1

Graphing Quantitative Variables

- Stem and Leaf Display
- Histograms
- Frequency Polygons
- Box Plots
- Bar Charts
- Line Graphs
- Dot Plots

Charts for Quantitative Variables



3|2337
2|001112223889
1|2244456888899
0|69

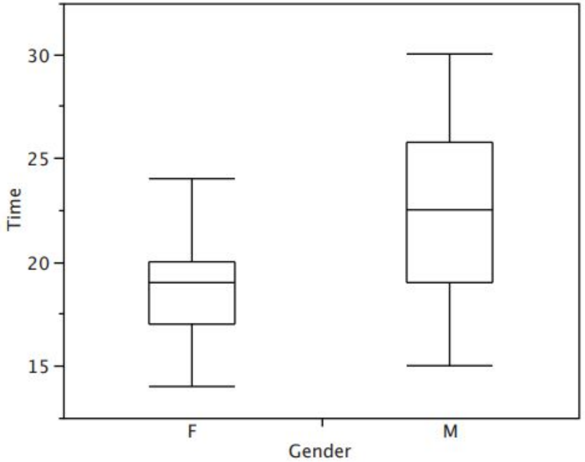


Figure 2. The box plots with the whiskers drawn.

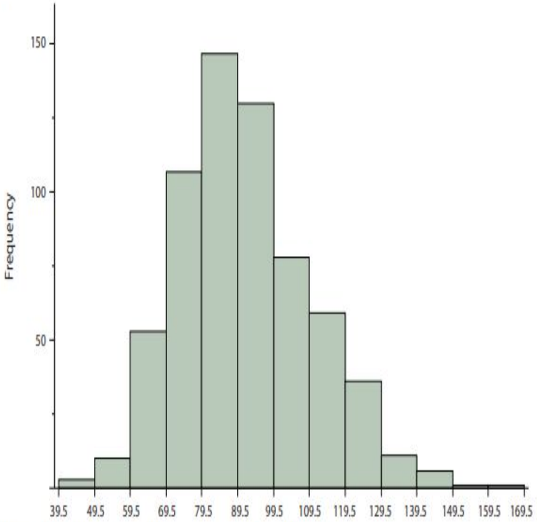


Figure 1. Histogram of scores on a psychology test.

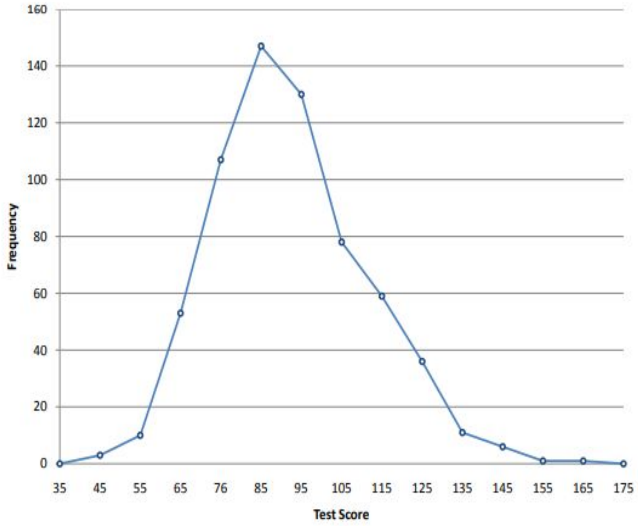


Figure 1. Frequency polygon for the psychology test scores.

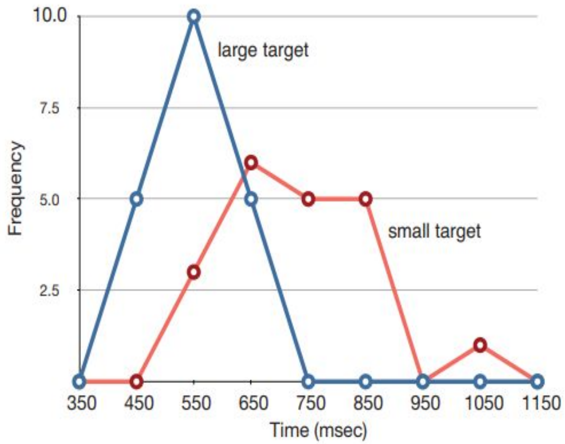


Figure 3. Overlaid frequency polygons.

Measures of Central Tendency

Arithmetic Mean: The arithmetic mean is the most common measure of central tendency. It is simply the sum of the numbers divided by the number of numbers. The symbol "μ" is used for the mean of a population. The symbol "M" is used for the mean of a sample. The formula for μ is shown below:

$$\mu = \frac{\sum X}{N}$$

where ΣX is the sum of all the numbers in the population and N is the number of numbers in the population.

Median

The median is also a frequently used measure of central tendency. The median is the midpoint of a distribution: the same number of scores is above the median as below it.

When there is an odd number of numbers, the median is simply the middle number. For example, the median of 2, 4, and 7 is 4. When there is an even number of numbers, the median is the mean of the two middle numbers. Thus, the median of the numbers 2, 4, 7, 12 is:

$$\frac{(4+7)}{2} = 5.5$$

Mode

The mode is the most frequently occurring value. For the data in Table 1, the mode is 18 since more teams (4) had 18 touchdown passes than any other number of touchdown passes. With continuous data, such as response time measured to many decimals, the frequency of each value is one since no two scores will be exactly the same (see discussion of continuous variables). Therefore the mode of continuous data is normally computed from a grouped frequency distribution.

Table 1. Number of touchdown passes.

37, 33, 33, 32, 29, 28,
28, 23, 22, 22, 22, 21,
21, 21, 20, 20, 19, 19,
18, 18, 18, 18, 16, 15,
14, 14, 14, 12, 12, 9, 6

Range

The range is the simplest measure of variability to calculate, and one you have probably encountered many times in your life. The range is simply the highest score minus the lowest score. Let's take a few examples. What is the range of the following group of numbers: 10, 2, 5, 6, 7, 3, 4? Well, the highest number is 10, and the lowest number is 2, so 10 - 2 = 8. The range is 8.

Interquartile Range

The interquartile range (IQR) is the range of the middle 50% of the scores in a distribution.

It is computed as follows: IQR = 75th percentile - 25th percentile

Steps:

- Step 1: **Put the numbers in order.**
  1, 2, 5, 6, 7, 9, 12, 15, 18, 19, 27.

- Step 2: **Find the median.**
  1, 2, 5, 6, 7, **9**, 12, 15, 18, 19, 27.

- Step 3: **Place parentheses around the numbers above and below the median.**
  Not necessary **statistically**, but it makes Q1 and Q3 easier to spot.
  (1, 2, 5, 6, 7), 9, (12, 15, 18, 19, 27).

- Step 4: **Find Q1 and Q3**
  Think of Q1 as a median in the lower half of the data and think of Q3 as a median for the upper half of data.
  (1, 2, **5**, 6, 7), **9**, (12, 15, **18**, 19, 27). Q1 = 5 and Q3 = 18.

- Step 5: **Subtract Q1 from Q3 to find the interquartile range.**
  18 – 5 = 13.

What if I Have an Even Set of Numbers?

Sample question: Find the IQR for the following data set: 3, 5, 7, 8, 9, 11, 15, 16, 20, 21.

- Step 1: **Put the numbers in order.**
  3, 5, 7, 8, 9, 11, 15, 16, 20, 21.

- Step 2: **Make a mark in the center of the data:**
  3, 5, 7, 8, 9, | 11, 15, 16, 20, 21.

- Step 3: **Place parentheses around the numbers above and below the mark you made in Step 2–it makes Q1 and Q3 easier to spot.**
  (3, 5, 7, 8, 9), | (11, 15, 16, 20, 21).

- Step 4: **Find Q1 and Q3**
  Q1 is the median (the middle) of the lower half of the data, and Q3 is the median (the middle) of the upper half of the data.
  (3, 5, **7**, 8, 9), | (11, 15, **16,** 20, 21). Q1 = 7 and Q3 = 16.

- Step 5: **Subtract Q1 from Q3.**
  16 – 7 = 9.
  This is your IQR.

Variance

Variability can also be defined in terms of how close the scores in the distribution are to the middle of the distribution. Using the mean as the measure of the middle of the distribution, the variance is defined as the average squared difference of the scores from the mean.

$$\sigma^2 = \frac{\Sigma(X-\mu)^2}{N}$$

where σ2 is the variance, μ is the mean, and N is the number of numbers.

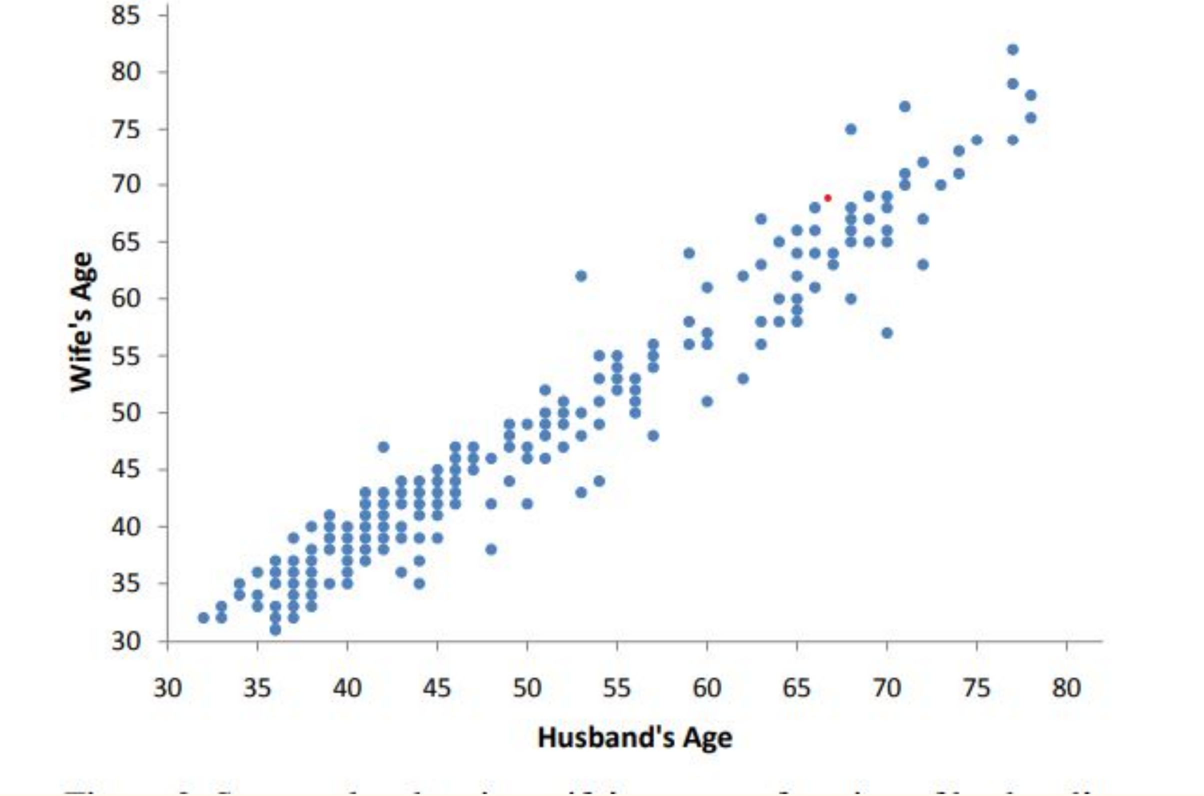
Standard Deviation

The standard deviation is simply the square root of the variance. The standard deviation is an especially useful measure of variability when the distribution is normal or approximately normal because the proportion of the distribution within a given number of standard deviations from the mean can be calculated.

For example, 68% of the distribution is within one standard deviation of the mean and approximately 95% of the distribution is within two standard deviations of the mean. Therefore, if you had a normal distribution with a mean of 50 and a standard deviation of 10, then 68% of the distribution would be between 50 - 10 = 40 and 50 +10 =60. Similarly, about 95% of the distribution would be between 50 - 2 x 10 = 30 and 50 + 2 x 10 = 70.

Bivariate Data

Scatter Plot to Analyze Bivariate Data



Pearson Correlation

The Pearson product-moment correlation coefficient is a measure of the strength of the linear relationship between two variables. It is referred to as Pearson's correlation or simply as the correlation coefficient. If the relationship between the variables is not linear, then the correlation coefficient does not adequately represent the strength of the relationship between the variables.

The symbol for Pearson's correlation is "ρ" when it is measured in the population and "r" when it is measured in a sample. Because we will be dealing almost exclusively with samples, we will use r to represent Pearson's correlation unless otherwise noted.

Pearson's r can range from -1 to 1. An r of -1 indicates a perfect negative linear relationship between variables, an r of 0 indicates no linear relationship between variables, and an r of 1 indicates a perfect positive linear relationship between variables. Figure 1 shows a scatter plot for which r = 1.

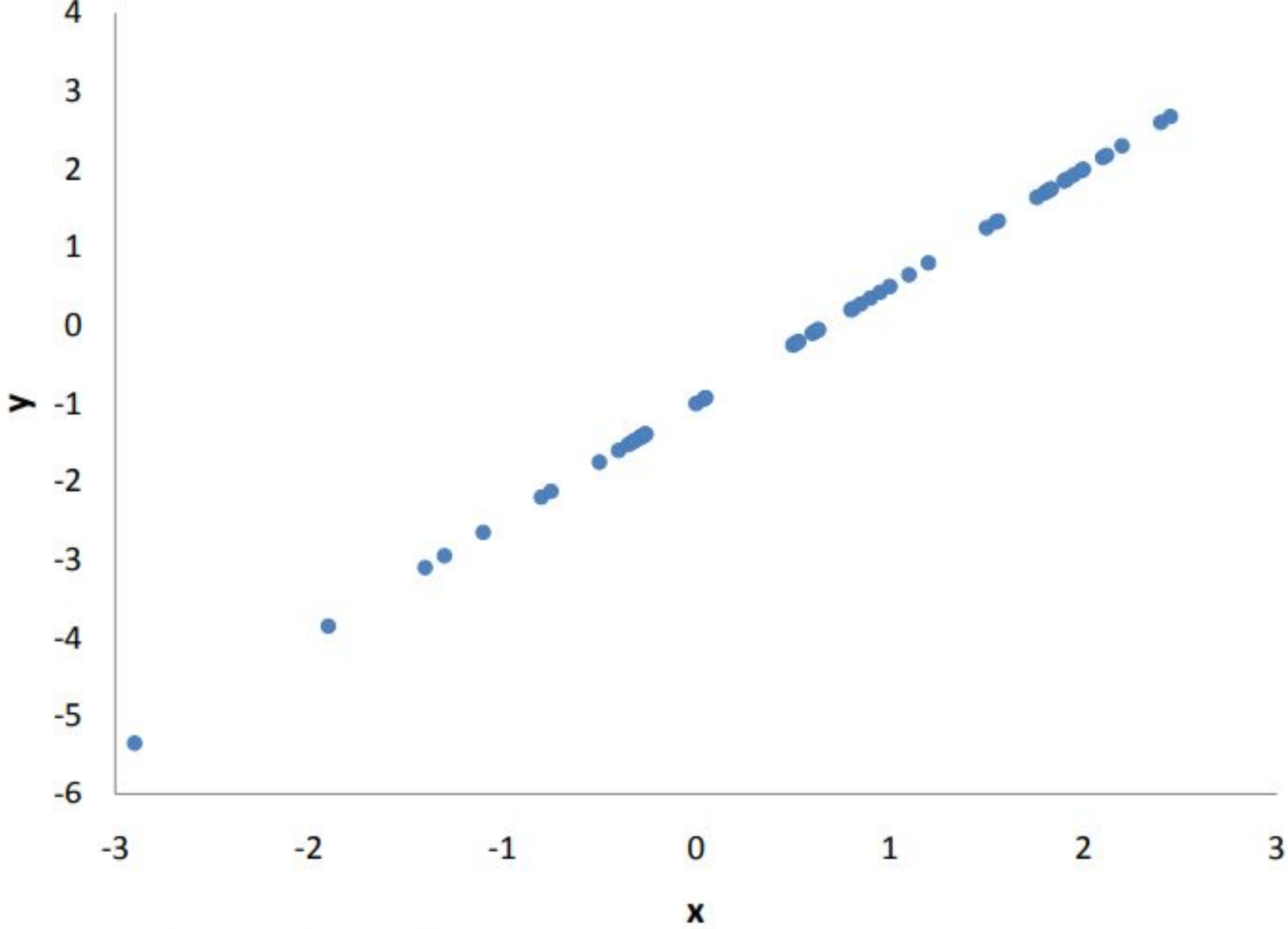


Figure 1. A perfect linear relationship, r = 1.

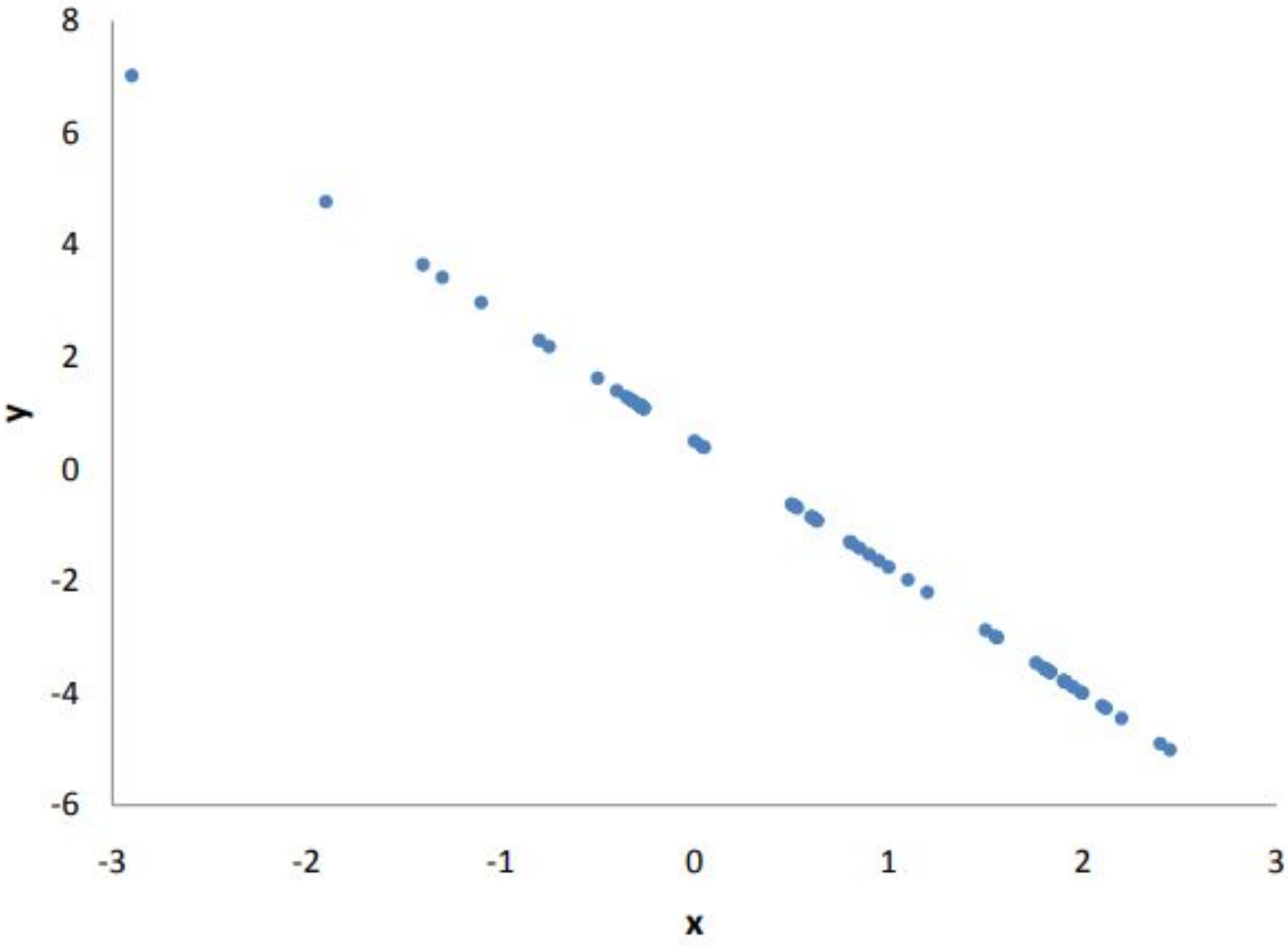


Figure 2. A perfect negative linear relationship, r = -1.

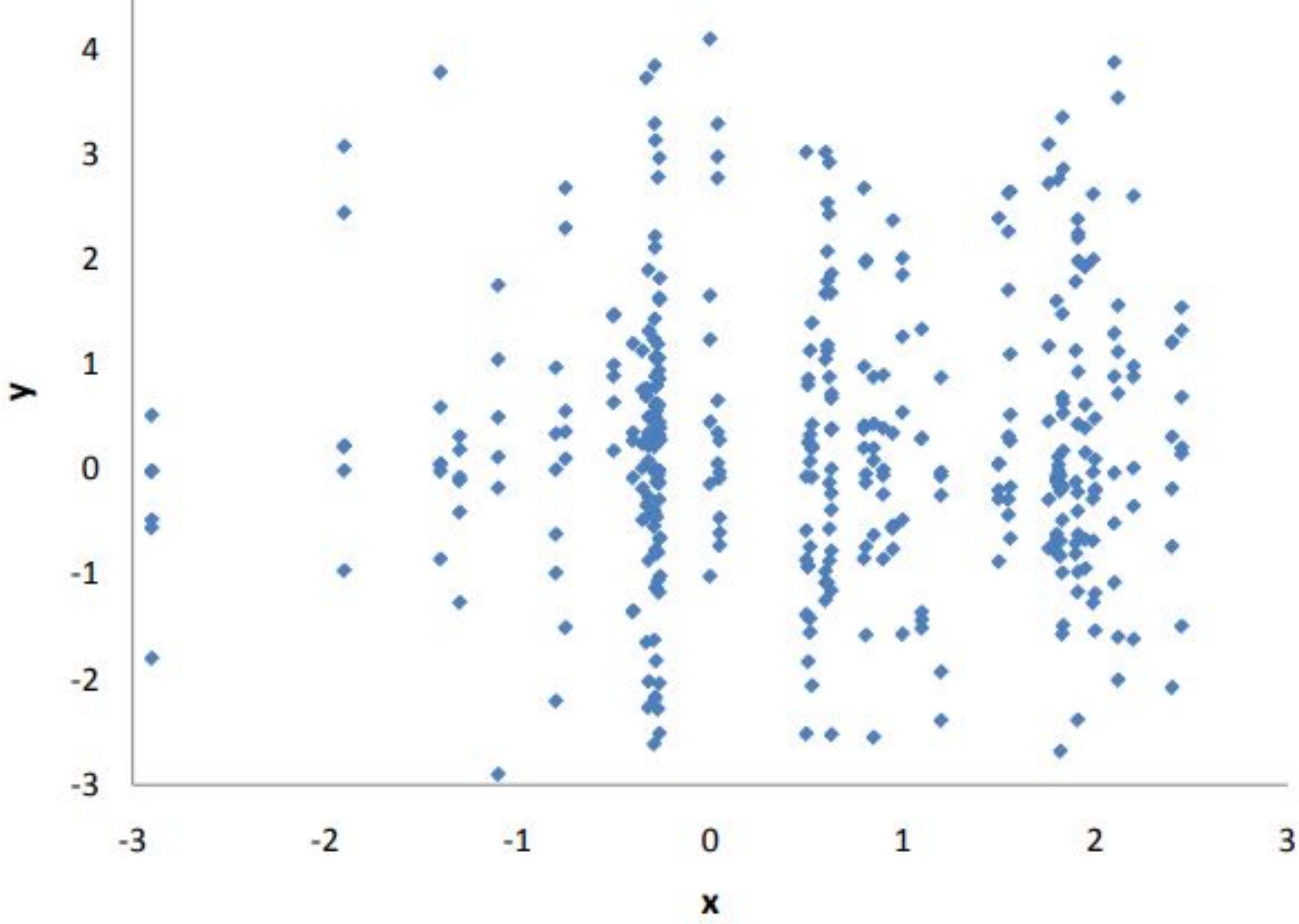


Figure 3. A scatter plot for which r = 0. Notice that there is no relationship between X and Y.

Properties of Pearson's r

- A basic property of Pearson's r is that its possible range is from -1 to 1. A correlation of -1 means a perfect negative linear relationship, a correlation of 0 means no linear relationship, and a correlation of 1 means a perfect positive linear relationship.
- Pearson's correlation is symmetric in the sense that the correlation of X with Y is the same as the correlation of Y with X.
- A critical property of Pearson's r is that it is unaffected by linear transformations. This means that multiplying a variable by a constant and/or adding a constant does not change the correlation of that variable with other variables.

Computing Pearson's r

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

Table 1. Calculation of r.

	X	Y	x	y	xy	x^2	y^2
	1	4	-3	-5	15	9	25
	3	6	-1	-3	3	1	9
	5	10	1	1	1	1	1
	5	12	1	3	3	1	9
	6	13	2	4	8	4	16
Total	20	45	0	0	30	16	60
Mean	4	9	0	0	6		

$$r = \frac{30}{\sqrt{(16)(60)}} = \frac{30}{\sqrt{960}}$$

$$= \frac{30}{30.984} = 0.968$$

Probability

Introduction

Inferential statistics is built on the foundation of **probability theory,** and has been remarkably successful in guiding opinion about the conclusions to be drawn from data. Yet (paradoxically) the very idea of probability has been plagued by controversy from the beginning of the subject to the present day. In this section we provide a glimpse of the debate about the interpretation of the probability concept. One conception of probability is drawn from the idea of symmetrical outcomes.

Probability of a Single Event

$$probability = \frac{Number\ of\ favorable\ outcomes}{Number\ of\ possible\ equally - likely\ outcomes}$$

If you roll a six-sided die, there are six possible outcomes, and each of these outcomes is equally likely. A six is as likely to come up as a three, and likewise for the other four sides of the die. What, then, is the probability that a one will come up? Since there are six possible outcomes, the probability is 1/6.

Probability of two or more Events

Events A and B are independent events if the probability of Event B occurring is the same whether or not Event A occurs. Let's take a simple example. A fair coin is tossed two times. The probability that a head comes up on the second toss is 1/2 regardless of whether or not a head came up on the first toss.

- The two events are (1) first toss is a head and (2) second toss is a head. So these events are independent.

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

where P(A and B) is the probability of events A and B both occurring, P(A) is the probability of event A occurring, and P(B) is the probability of event B occurring.

If you flip a coin twice, what is the probability that it will come up heads both times? Event A is that the coin comes up heads on the first flip and Event B is that the coin comes up heads on the second flip. Since both P(A) and P(B) equal 1/2, the probability that both events occur is

$$1/2 \times 1/2 = 1/4$$

If Events are Independent

If Events A and B are independent,

The probability that either Event A or Event B occurs is:

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

When we say “A or B occurs” we include three possibilities:

1. A occurs and B does not occur

2. B occurs and A does not occur

3. Both A and B occur

Now for some examples. If you flip a coin two times, what is the probability that you will get a head on the first flip or a head on the second flip (or both)? Letting Event A be a head on the first flip and Event B be a head on the second flip, then P(A) = 1/2, P(B) = 1/2, and P(A and B) = 1/4. Therefore,

$$P(A \text{ or } B) = 1/2 + 1/2 - 1/4 = 3/4.$$

If you throw a six-sided die and then flip a coin, what is the probability that you will get either a 6 on the die or a head on the coin flip (or both)? Using the formula,

$$\begin{aligned} P(6 \text{ or head}) &= P(6) + P(\text{head}) - P(6 \text{ and head}) \\ &= (1/6) + (1/2) - (1/6)(1/2) \\ &= 7/12 \end{aligned}$$

Conditional Probabilities

Often it is required to compute the probability of an event given that another event has occurred. For example, what is the probability that two cards drawn at random from a deck of playing cards will both be aces? It might seem that you could use the formula for the probability of two independent events and simply multiply $4/52 \times 4/52 = 1/169$. This would be incorrect, however, because the two events are not independent. If the first card drawn is an ace, then the probability that the second card is also an ace would be lower because there would only be three aces left in the deck.

Once the first card chosen is an ace, the probability that the second card chosen is also an ace is called the conditional probability of drawing an ace.

In this case, the "condition" is that the first card is an ace. Symbolically, we write this as:

P(ace on second draw | an ace on the first draw)

The probability that an ace is drawn on the second draw given that an ace was drawn on the first draw.

What is this probability? Since after an ace is drawn on the first draw, there are 3 aces out of 51 total cards left.

This means that the probability that one of these aces will be drawn is 3/51 = 1/17.

```
If Events A and B are not independent, then
P(A and B) = P(A) x P(B|A).
```

Applying this to the problem of two aces, the probability of drawing two aces from a deck is

$4/52 \times 3/51 = 1/221$.

Multiplication Rule

Imagine a small restaurant whose menu has 3 soups, 6 entrées, and 4 desserts. How many possible meals are there? The answer is calculated by multiplying the numbers to get 3 x 6 x 4 = 72. You can think of it as first there is a choice among 3 soups. Then, for each of these choices there is a choice among 6 entrées resulting in 3 x 6 = 18 possibilities. Then, for each of these 18 possibilities there are 4 possible desserts yielding 18 x 4 = 72 total possibilities.

Permutations

Suppose that there were four pieces of candy (red, yellow, green, and brown) and you were only going to pick up exactly two pieces.

How many ways are there of Number First Second Third 1 red yellow green 2 red green yellow 3 yellow red green 4 yellow green red 5 green red yellow 6 green yellow red picking up two pieces?

Table lists all the possibilities.

The first choice can be any of the four colors. For each of these 4 first choices there are 3 second choices.

Therefore there are 4 x 3 = 12 possibilities.

$$_{n}P_{r} = \frac{n!}{(n-r)!}$$

Number	First	Second
1	red	yellow
2	red	green
3	red	brown
4	yellow	red
5	yellow	green
6	yellow	brown
7	green	red
8	green	yellow
9	green	brown
10	brown	red
11	brown	yellow
12	brown	green

where nPr is the number of permutations of n things taken r at a time. In other words, it is the number of ways r things can be selected from a group of n things. In this case.

$$ {}_{4}P_{2} = \frac{4!}{(4-2)!} = \frac{4 \times 3 \times 2 \times 1}{2 \times 1} = 12 $$

It is important to note that order counts in permutations. That is, choosing red and then yellow is counted separately from choosing yellow and then red. Therefore permutations refer to the number of ways of choosing rather than the number of possible outcomes. When order of choice is not considered, the formula for combinations is used.

Combinations

Now suppose that you were not concerned with the way the pieces of candy were chosen but only in the final choices. In other words, how many different combinations of two pieces could you end up with?

$$ {}_{n}C_{r} = \frac{n!}{(n-r)!\,r!} $$

Thank you