ECE5984 – Applications of Machine Learning
Lecture 3 – Review of Statistics

Creed Jones, PhD

Course Updates

- Graduate Teaching Assistant – Ashley Smith
  - Office hours: Tuesday 10 AM to noon, Thursday 11 AM to 1 PM
  - Office hours by zoom: ID #4095579468
- Quiz 1 will be THIS Thursday, January 27
  - On lectures 1-3
  - Must be taken between 12 noon and 6 PM
  - 20 minute time limit
- Homework 1 will be posted this week
  - Due on Tuesday, February 8
- If you send me an email on the course, please put "ECE5984" in the subject line!

Schedule note!!! The next class session will be asynchronous – video will be posted

- Due to some illness in my family, I have to be away for a few days
  - Maybe more
- Lecture on this Thursday, January 27, will be a video lecture that you can watch at any time
  - It should be posted by Wednesday night sometime
  - Sorry there won't be a chance for live questions – please use Piazza for any questions
- Watch for announcements on future class sessions
- Quiz 1 will still occur on Thursday

Today's Objectives

- Probability
- Conditional Probability
- Dependence
- Descriptive Statistics
- Covariance
- Correlation
- Covariance Matrix

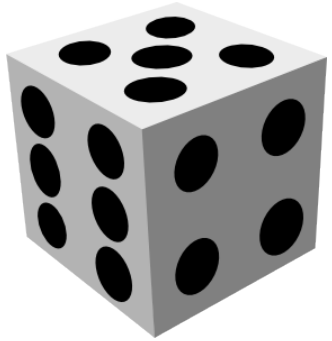
Probabilities are expressions of the likelihood of various outcomes of a *random process*

To say that a process is **random** means that when it takes place, one outcome from some set of outcomes is sure to occur, but it is impossible to predict with certainty which outcome that will be.

In case an experiment has finitely many outcomes and all outcomes are equally likely to occur, the *probability* of an event (set of outcomes) is just the ratio of the number of outcomes in the event to the total number of outcomes.

A single roll of one die is equiprobable

- If the die is truly cubic and weighted well ("fair"), then each number 1 through 6 is just as likely to come up as any other number
  - If we roll one die for a long time, we would expect a 1 to come up 1/6 of the time, a 2 to come up 1/6 of the time, etc.
- The probability of each number is equal to the others – this is called "equiprobable"



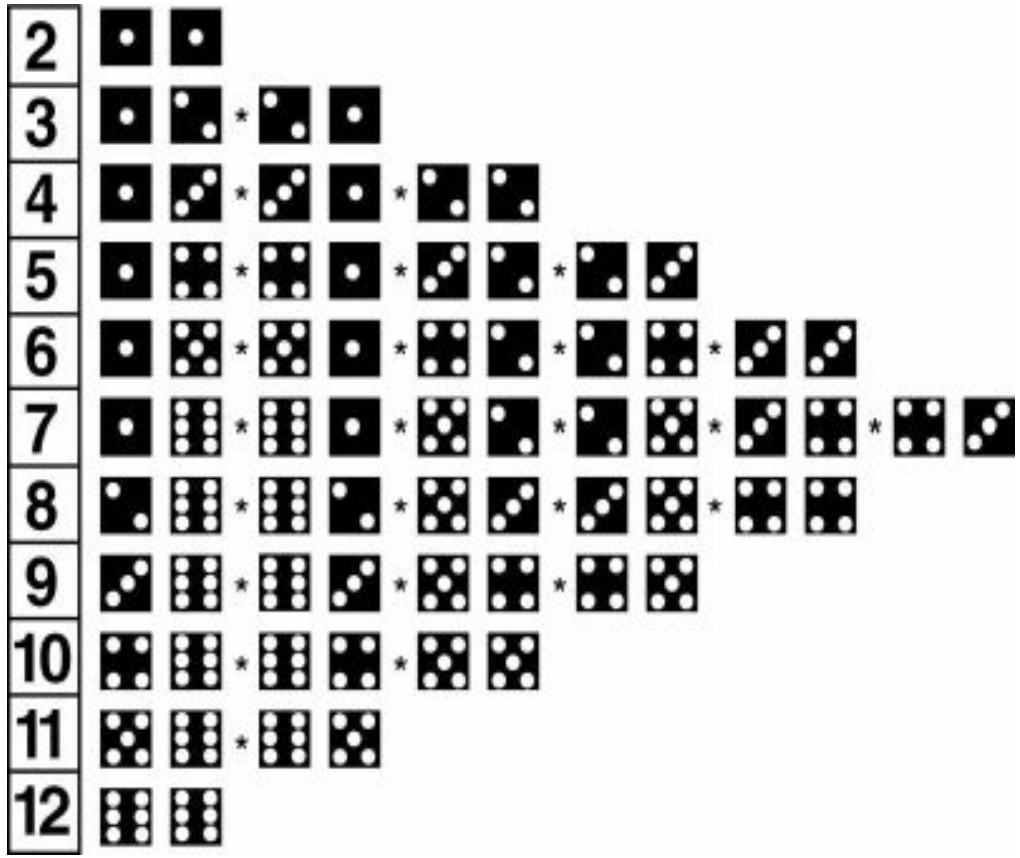
Experimental Probabilities

- The *probability* of an event is the likelihood that it will occur; written *p(event)*
  - probability of 1 means an event certainly will occur
  - probability of 0 means an event certainly won't occur
  - Higher probability events are more likely to happen
  - *p(sun will rise tomorrow)* ≈ 1

- While the probability is impossible to measure directly, because after the fact the event either occurred or it didn't, we can often estimate it by *repeated trials*
- Given a finite *sample space S* of size *N(S)*, the probability of an event *E* is $p(E) = \frac{N(E)}{N(S)}$

Probability of various dice totals – the sum of two dice is not equiprobable

$$p(0) = p(1) = {}^{0}/_{36} = 0.000000$$
$$p(2) = {}^{1}/_{36} = 0.027777$$
$$p(3) = {}^{2}/_{36} = 0.055555$$
$$p(4) = {}^{3}/_{36} = 0.083333$$
$$p(5) = {}^{4}/_{36} = 0.111111$$
$$p(6) = {}^{5}/_{36} = 0.138888$$
$$p(7) = {}^{6}/_{36} = 0.166667$$
$$p(8) = {}^{5}/_{36} = 0.138888$$
$$p(9) = {}^{4}/_{36} = 0.111111$$
$$p(10) = {}^{3}/_{36} = 0.083333$$
$$p(11) = {}^{2}/_{36} = 0.055555$$
$$p(12) = {}^{1}/_{36} = 0.027777$$



Some useful Probability Axioms

Probability Axioms

Let S be a sample space, A **probability function** P from the set of all events in S to the set of real numbers satisfies the following three axioms: For all events A and B in S,

1. $0 \leq P(A) \leq 1$
2. $P(\emptyset) = 0$ and $P(S) = 1$
3. If A and B are disjoint (that is, if $A \cap B = \emptyset$), then the probability of the union of A and B is

$$P(A \cup B) = P(A) + P(B).$$

- Probability ranges from 0 to 1
- P(impossible events) = 0, and p(certain events) = 1
- Probabilities of disjoint sets add to give the probability of their union

Probability of a General Union of Two Events

If S is any sample space and A and B are any events in S, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \qquad 9.8.2$$

- In general, probability of the union of two sets is the sum of their probabilities, minus the probability of their intersection.

Expected Value

- The expected value of an event that generates a real number is the sum of all possible values multiplied by their probabilities.
- In some sense, it's the "most likely" value to result, but in many cases it's a value that cannot actually occur.

• Definition

Suppose the possible outcomes of an experiment, or random process, are real numbers $a_1, a_2, a_3, \ldots, a_n$, which occur with probabilities $p_1, p_2, p_3, \ldots, p_n$. The **expected value** of the process is

$$\sum_{k=1}^{n} a_k p_k = a_1 p_1 + a_2 p_2 + a_3 p_3 + \cdots + a_n p_n.$$

Lottery

- Besides moral and ethical arguments against a lottery, they are a bad deal probabilistically.
- The Powerball lottery sells tickets for $2.
- The grand prize for this ticket is $1,000,000.
  - If you pay an extra $1 for your ticket, and pick another number or something, then you would win a much larger amount if you win the grand prize.
- The official literature says that the odds of winning the $1M prize are 1 in 5,153,632.65.
  - in 2018
- The expected value (considering only the $1M prize) is $\frac{\$1{,}000{,}000}{5{,}153{,}632.65} = \$0.194$, or an expected winnings of 19 cents on a $2 ticket.

Expected value of a Powerball ticket's winnings is $0.19

- Note that the expected value of the $1M prize drawing is about 19 cents per ticket.
- This is not the most likely amount to win; in fact, it's not possible to win 19 cents!

- This is the centroid of the probability-weighted space of possible outcomes, and shows that most of the possible winnings are far below the $2 break-even point (in fact, all but one are zero).

- This is how the expected value is used in decision-making.

Conditional probability is the probability of some event, given that we know some relevant information

- Imagine that we are walking down the street. The probability that the next person you meet will be a male is about 0.50.
  - $p(male) = 0.50$

- However, if we find out that an Engineering class just dismissed in a nearby building, the probability that the next pedestrian will be a male is more than 0.5.
  - $p(male, given\ that\ class\ dismissed) \approx 0.8$.

- Extra information helped refine our estimate of the probability

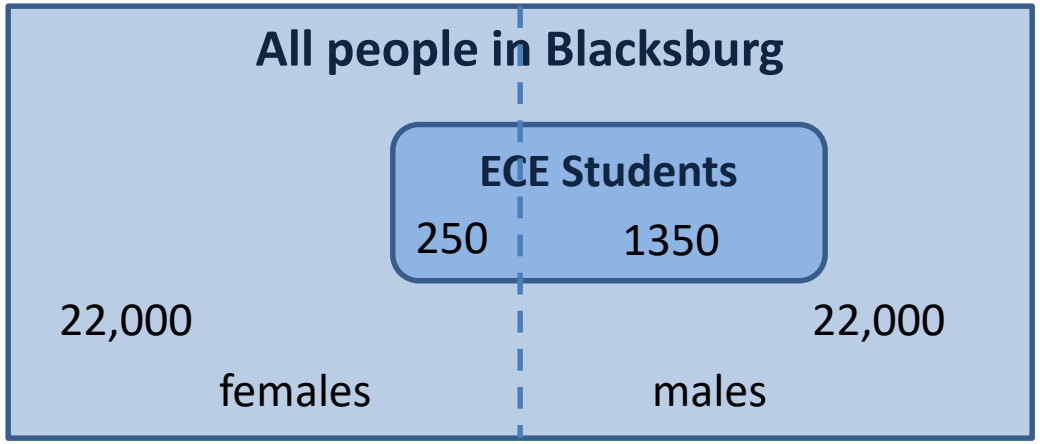
Conditional Probability of A if we know B is true is written $p(A|B)$

- The baseline probability of any pedestrian being a female is 0.5.

- The *conditional* probability of a pedestrian being a female, *given that an Engineering class just dismissed*, is less than that.
- Written $p(female \mid class\ let\ out)$

- The conditional probability of A given B (that is, the probability of A once we know that B has happened or is true), is written $p(A \mid B)$, and usually read as "the probability of A, given B".

Conditional Probability as a Venn Diagram

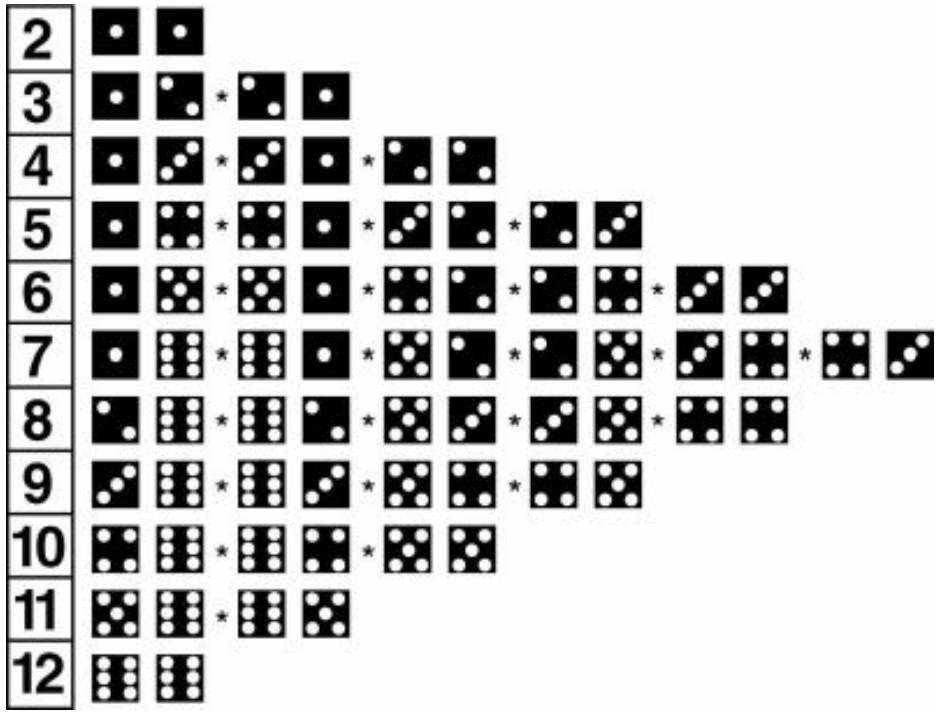


- $p(female) = \frac{22{,}000}{44{,}000} = 0.5$
- $p(female \mid ECE\ student) = \frac{250}{1600} = 0.156$

Conditional Look at Dice Rolls

We have seen that the probability of rolling two dice and getting a 5 is 4/36, or 0.11111.

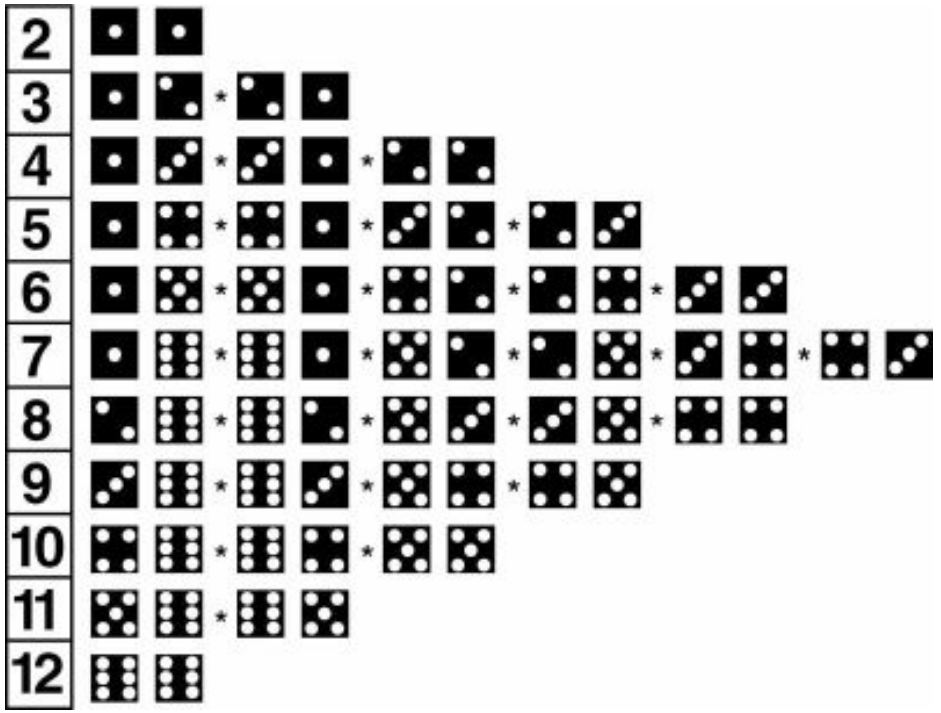
What is the probability of rolling a 5 if the first die has been rolled and came up a 2?



Conditional Look at Dice Rolls

We have seen that the probability of rolling two dice and getting a 5 is 4/36, or 0.11111.

What is the probability of rolling a 5 if the first die has been rolled and came up a 2?



If the first die came up a 2, then we must roll a 3 – so the odds are 1/6 = 0.166666.

$$p(5 \mid 2) = 0.16666$$

$$p(5 \mid 6) = p(5 \mid 5) = 0$$

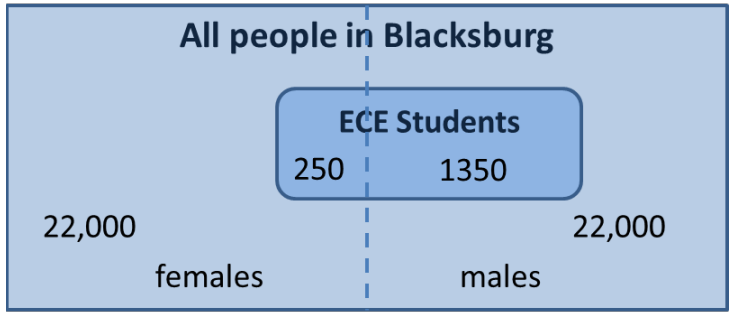
Conditional probability calculated from set intersection

• Definition

Let A and B be events in a sample space S. If $P(A) \neq 0$, then the **conditional probability of B given A**, denoted $P(B \mid A)$, is

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}. \qquad 9.9.1$$

- $p(female \mid ECE) = \frac{p(female \cap ECE)}{p(ECE)} = \frac{250/44000}{1600/44000} = 0.156$



- Another common form is $p(B \mid A)p(A) = p(A \cap B)$

It's crucial in assessing probabilities to understand when events are independent

- Conceptually, "Independent Events" are those for which the likelihood of one occurring is not affected by the occurrence of another.
- The probability of both occurring is just the product of each occurring "independently".

• Definition

If A and B are events in a sample space S, then A and B are **independent** if, and only if,

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

Events $A_1, A_2, A_3, \ldots, A_n$ in a sample space S are **mutually independent** if, and only if, the probability of the intersection of any subset of the events is the product of the probabilities of the events in the subset.

Independent Events

- Successive coin flips are independent.
  - The coin doesn't have a "memory" of past flips to influence future events.
- Don't confuse "fair" or "equiprobable" with "independent"
- Think about an unfair coin that lands heads 60% of the time
  - $p(heads) = 0.6$
  - Successive flips are still independent; $p(heads) = 0.6$ no matter what has come before.
- Successive human actions are generally <u>not</u> independent, though we sometimes make that simplifying assumption.

• Definition

Events $A_1, A_2, A_3, \ldots, A_n$ in a sample space S are **mutually independent** if, and only if, the probability of the intersection of any subset of the events is the product of the probabilities of the events in the subset.

Bayes' Theorem

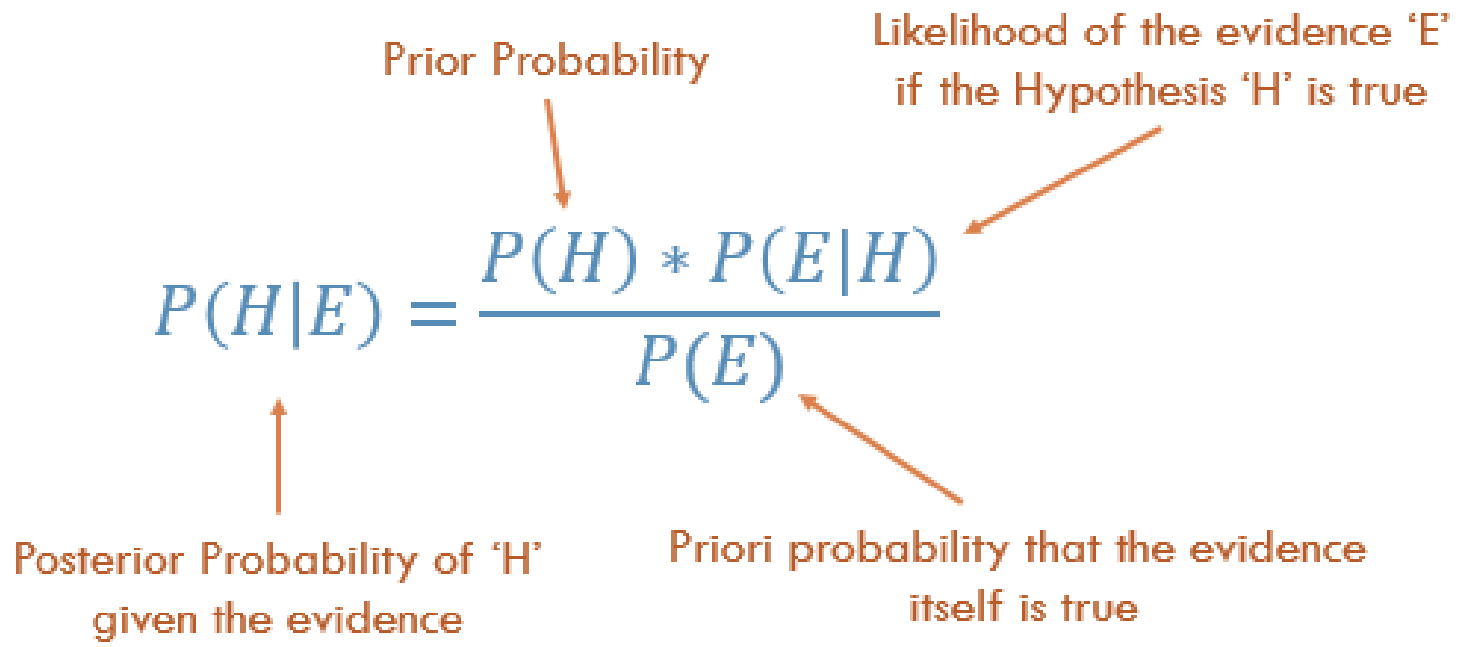
- Bayes' Theorem relates conditional probabilities –

> Suppose that a sample space S is a union of mutually disjoint events $B_1, B_2, B_3, \ldots, B_n$, suppose A is an event in S, and suppose A and all the B_1 have nonzero probabilities. If k is an integer with $1 \leq k \leq n$, then
>
> $$P(B_k \mid A) = \frac{P(A \mid B_k)P(B_k)}{P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \cdots + P(A \mid B_n)P(B_n)}$$

- A simpler form is:

$$p(B|A) = \frac{p(B)p(A|B)}{p(A)}$$

Bayes' theorem relates *prior* (before we have additional knowledge) and *posterior* (after we know more) probabilities



Bayes Theorem Applied

$$p(B|A) = \frac{p(B)p(A|B)}{p(A)}$$

$$p(pass|failed\ midterm) = \frac{p(pass)p(failed\ midterm|pass)}{p(failed\ midterm)} = \frac{0.9 \cdot 0.02}{0.1} = 0.18$$

$$p(tsunami) = \frac{p(earthquake)p(tsunami|earthquake)}{p(earthquake|tsunami)} = \frac{0.01 \cdot 0.6}{0.95} = 0.00631$$

Probabilities for Two Events, A,B

- **Marginal Probability** = The probability of an event not considering any other events. **P(A)**
- **Joint Probability** = The probability that two events happen at the same time. **P(A,B)**
- **Conditional Probability** = The probability that one event happens given that another event has happened. **P(A|B)**

Probabilities: Inherited Color Blindness*

- Inherited color blindness has different incidence rates in men and women. Women usually carry the defective gene and men usually inherit it.
- **Experiment: pick an individual at random from the population.**
  CB = has inherited color blindness
  MALE = gender, Not-Male = FEMALE
- **Marginal**: P(CB) = 2.75%
- P(MALE) = 50.0%
- **Joint**: P(CB and MALE) = 2.5%
  P(CB and FEMALE) = 0.25%
- **Conditional**: P(CB|MALE) = 5.0% (1 in 20 men)
  P(CB|FEMALE) = 0.5% (1 in 200 women)

*** There are several types of color blindness and large variation in the incidence across different demographic groups. These are broad averages that are roughly in the neighborhood of the true incidence for particular groups.**

Dependent Events

Random variables X and Y are dependent if $P_{XY}(X,Y) \neq P_X(X)P_Y(Y)$.

	Color Blind		
Gender	No	Yes	Total
Male	.475	.025	0.50
Female	.4975	.0025	0.50
Total	.9725	.0275	1.00

P(Color blind, Male) = .0250

P(Male) = .5000

P(Color blind) = .0275

P(Color blind) x P(Male)
= .0275 x .500 = .01375

.01375 is not equal to .025

Gender and color blindness are not independent.

Dependent Random Variables

- Random variables are dependent if the occurrence of one affects the probability distribution of the other.
- If P(Y|X) changes when X changes, then the variables are dependent.
- If P(Y|X) does not change when X changes, then the variables are independent.

Two Important Math Results

- For two random variables,
  P(X,Y) = P(X|Y) P(Y)
  P(Color blind, Male) = P(Color blind|Male)P(Male)
  = .05 x .5 = .025

- For two *independent* random variables,
  P(X,Y) = P(X) P(Y)
  P(Ace,Heart) = P(Ace) x P(Heart).
  (This does not work if they are not independent.)

So, P(X|Y) = P(X) if X and Y are independent

Independent Random Variables

One card is drawn randomly from a deck of 52 cards

	Ace		
Heart	**Yes=1**	**No=0**	**Total**
Yes=1	**1/52**	**12/52**	**13/52**
No=0	**3/52**	**36/52**	**39/52**
Total	**4/52**	**48/52**	**52/52**

P(Ace|Heart) = 1/13

P(Ace|Not-Heart) = 3/39 = 1/13

P(Ace) = 4/52 = 1/13

P(Ace) does not depend on whether the card is a heart or not.

P(Heart|Ace) = 1/4

P(Heart|Not-Ace) = 12/48 = 1/4

P(Heart) = 13/52 = 1/4

P(Heart) does not depend on whether the card is an ace or not.

A Theorem: For two independent random variables, P(X,Y) = P(X) P(Y)

P(Ace, Heart) = P(Ace)P(Heart) = 1/13 x 1/4 = 1/52

There are several widely used *descriptive statistics* to describe a sample from a population

- The central tendency or *mean* is the (commonly understood) average; it's considered to be typical of the sample

$$mean_X = \mu_X = \frac{1}{N}\sum_{i=1}^{N} x_n$$

  - The mean can be dramatically influenced by *outliers* (unusually large or small values)
- The *median* is the middle value when the sample values are ranked
  - It's less sensitive to outliers

- The *variance* is a measure of the amount of variation in the data (about the mean)

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

  - The standard deviation σ_X is often used; it's the square root of the variance (and has appropriate units)

An example of some simple descriptive statistics on a pair of samples (from two different populations, probably)

X	Y
1	0.10074
2	0.28912372
-3.5	0.00835797
2.71828	0.89713234
1.41421356	0.80416532
-1.73205081	0.68352129
4.2	0.0071702
-2	0.20708702
-6	0.91223393
0.00001	0.14566067
4	0.06174395
-5.25	0.1182755
3.14159	0.9260739
-2	0.00741946
4	0.07569365

Statistics	stat(X)	stat(Y)
Mean	0.13280285	0.349626595
Min	-6	0.007170203
Max	4.2	0.926073903
Range	10.2	0.9189037
Median	1	0.14566067
Mode	-2	#N/A
Variance	11.46047035	0.139748738
Std Deviation	3.385331646	0.373829825
Quartile 1	-2	0.061743946
Quartile 2	1	0.14566067
Quartile 3	3.14159	0.804165321

Measure of the Distribution Center: Mean (Average)

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1901	2.09	0.56	5.66	5.80	5.12	0.75	3.77	5.75	3.67	4.17	1.30	8.51
1902	2.13	3.32	5.47	2.92	2.42	4.54	4.66	4.65	5.83	5.59	1.27	4.27
1903	3.28	4.27	6.40	2.30	0.48	7.79	4.64	4.92	1.66	2.72	2.04	3.95
2000	3.00	3.40	3.82	4.14	4.26	7.99	6.88	5.40	5.36	2.29	2.83	4.24

$$\text{Mean for June} = \frac{.75 + 4.54 + 7.79 + \ldots + 7.99}{100} = \frac{377.76}{100} = 3.78$$

More generally,

$$\text{Mean}[\boldsymbol{x}] = \bar{\boldsymbol{x}} = \frac{x_1 + x_2 + \ldots + x_T}{T} = \frac{\sum_{t=1}^{T} x_t}{T}$$

where T = Total Number of Observations

x_1 = Value of the first observation (June 1901) = .75

x_2 = Value of the second observation (June 1902) = 4.54

x_3 = Value of the second observation (June 1903) = 7.79

$\vdots$

x_T = Value of the last (N^{th}) observation (June 2000) = 7.99

From Prof. Frank Westhoff
Amherst University

Calculating the Variance for 1951

Month	Precipitation	Mean	Deviation From Mean	Squared Deviation
Apr	3.63	3.47	3.63 – 3.47 = 0.16	0.0256
May	2.96	3.47	2.96 – 3.47 = –0.51	0.2601
Jun	3.05	3.47	3.05 – 3.47 = –0.42	0.1764
Jul	4.15	3.47	4.15 – 3.47 = 0.68	0.4624
Aug	3.56	3.47	3.56 – 3.47 = 0.09	0.0081
			Sum of Squared Deviations =	0.9326

$$\text{Variance} = \frac{\text{Sum of Squared Deviations}}{T} = \frac{0.9326}{5} = 0.1865$$

$$\text{Standard deviation} = \sqrt{\text{Variance}} = \sqrt{0.1865} = 0.43$$

Summary: In 1951, Variance = 0.1865 and Standard deviation = 0.43

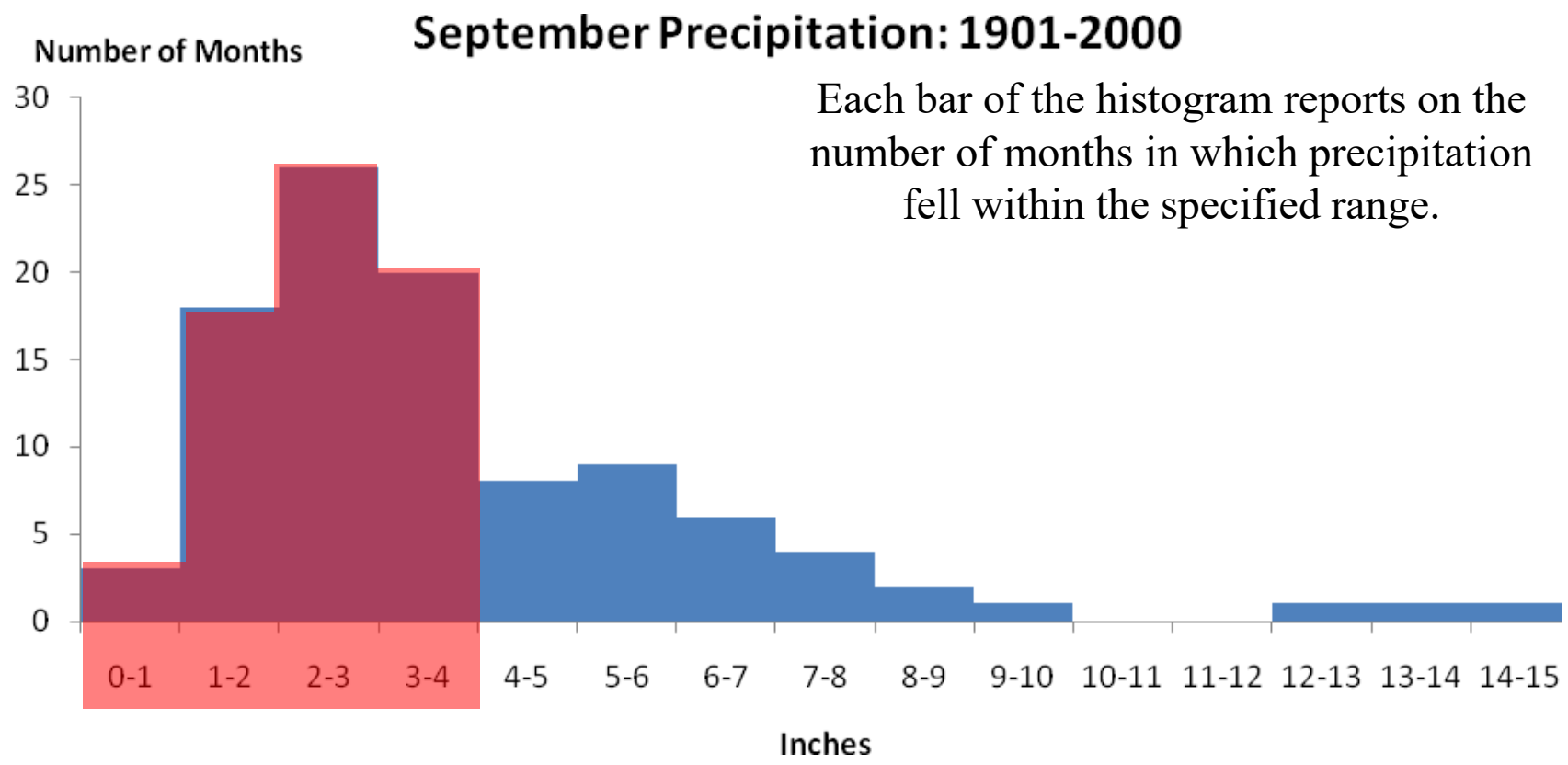
In 1998, Variance = 6.5256 and Standard deviation = 2.55

Small spread
↓
All deviations are small
↓
All squared deviations are small
↓
Variance small

Large spread
↓
Some deviations are large
↓
Some squared deviations are large
↓
Variance large

From Prof. Frank Westhoff
Amherst University

Histogram: Visual Illustration of a Variable's the Distribution of Values



Each bar of the histogram reports on the number of months in which precipitation fell within the specified range.

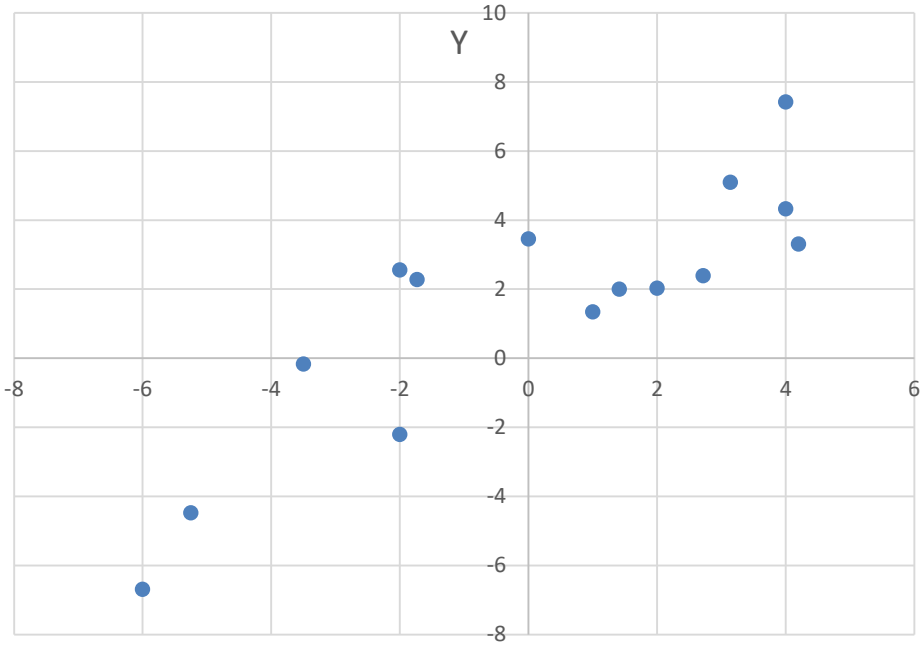
In 3 years, there was less than 1 inch of rain during September.
In 18 years, there was between 1 and 2 inches of rain during September.
In 26 years, there was between 2 and 3 inches of rain during September.
In 19 years, there was between 3 and 4 inches of rain during September.

From Prof. Frank Westhoff
Amherst University

Covariation and Expected Value

- Pick 10,325 **people** at random from the population. Predict how many will be color blind: 10,325 x .0275 = 284
- Pick 10,325 **MEN** at random from the population. Predict how many will be color blind: 10,325 x .05 = 516
- Pick 10,325 **WOMEN** at random from the population. Predict how many will be color blind: 10,325 x .005 = 52
- **The expected number of color blind people, given gender, depends on gender.**
- **Color Blindness covaries with Gender**

Covariances are the most common statistics to describe how multiple variables relate to each other



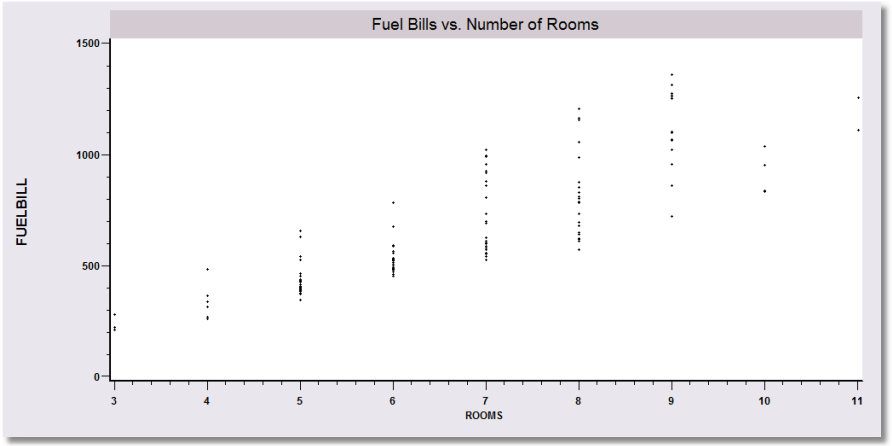
$$cov_{XY} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_X)(y_i - \mu_Y)$$

X	Y
1	1.34553764
2	2.03388004
-3.5	-0.16171495
2.71828	2.38965091
1.41421356	2.004445
-1.73205081	2.27773385
4.2	3.30622472
-2	2.55819525
-6	-6.68625651
0.00001	3.45906713
4	7.42686846
-5.25	-4.4685862
3.14159	5.10221025
-2	-2.20663644
4	4.32756045

Statistics	stat(X)	stat(Y)
Mean	0.13280285	1.51387864
Min	-6	-6.686256511
Max	4.2	7.426868456
Range	10.2	14.11312497
Median	1	2.277733847
Mode	-2	#N/A
Variance	11.46047035	13.24393094
Std Deviation	3.385331646	3.6392212
Quartile 1	-2	-0.161714953
Quartile 2	1	2.277733847
Quartile 3	3.14159	3.45906713

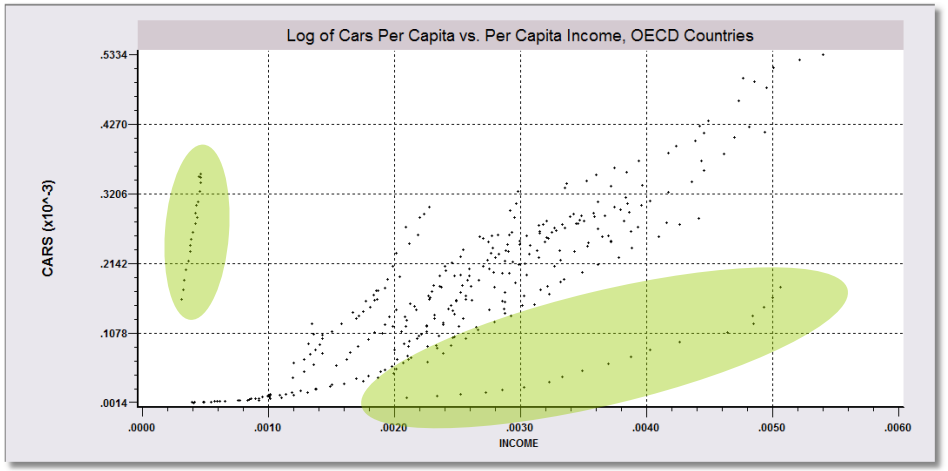
COVARIANCES	
11.46047035	10.58259554
10.58259554	13.24393094

Positive Covariation: The distribution of one variable depends on another variable.

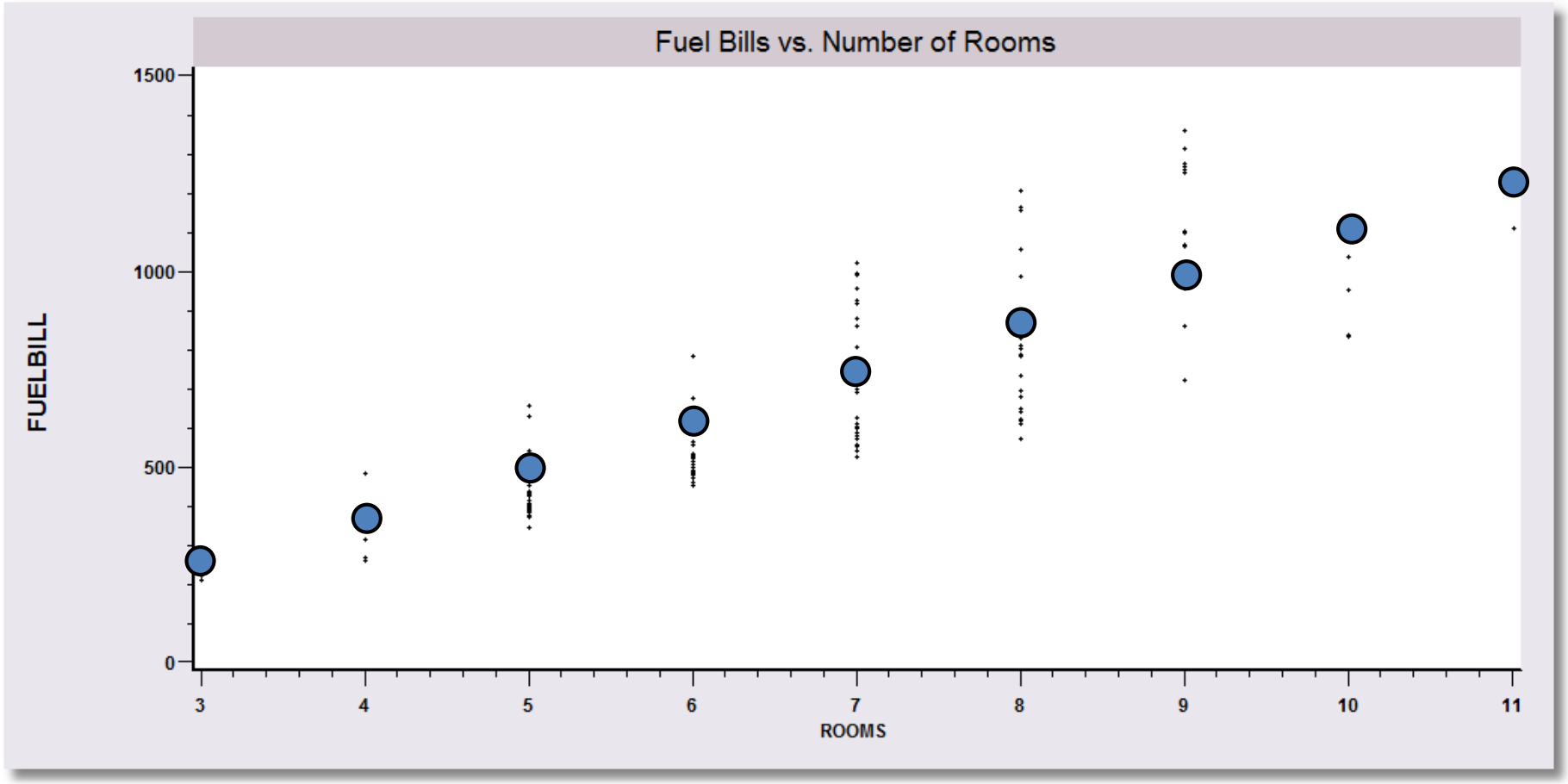


Distribution of fuel bills changes (moves upward) as the number of rooms changes (increases).

The per capita number of cars varies (positively) with per capita income. The relationship varies by country as well.



(Linear) Regression of Bills on Rooms



Measuring How Variables Move Together: Covariance

$$\text{Cov}(X,Y) = \sum_{\text{values of X}} \sum_{\text{values of Y}} P(X=x, Y=y)(x-\mu_X)(y-\mu_Y)$$

Covariance can be positive or negative

The measure will be positive if it is likely that Y is above its mean when X is above its mean.

It is usually denoted σ_{XY}.

Correlation is Units Free

$$Correlation \text{ Coefficient}$$

$$\rho_{XY} = \frac{\text{Covariance(X,Y)}}{\text{Standard deviation(X) Standard deviation(Y)}}$$

$$-1.00 \leq \rho_{XY} \leq +1.00.$$

Correlation

	Real Estate			
Finance	0	1	2	Total
0	.15	.10	.05	.30
1	.30	.20	.20	.70
Total	.45	.30	.25	1.00

$\mu_R = .8 \quad \mu_F = .7$

$\text{Var}(F) = 0^2(.3)+1^2(.7) - .7^2 = .21$

Standard deviation = ..46

$\text{Var}(R) = 0^2(.45)+1^2(.30)+2^2(.25) - .8^2$
$= .66$

Standard deviation = 0.81

Covariance = +0.04

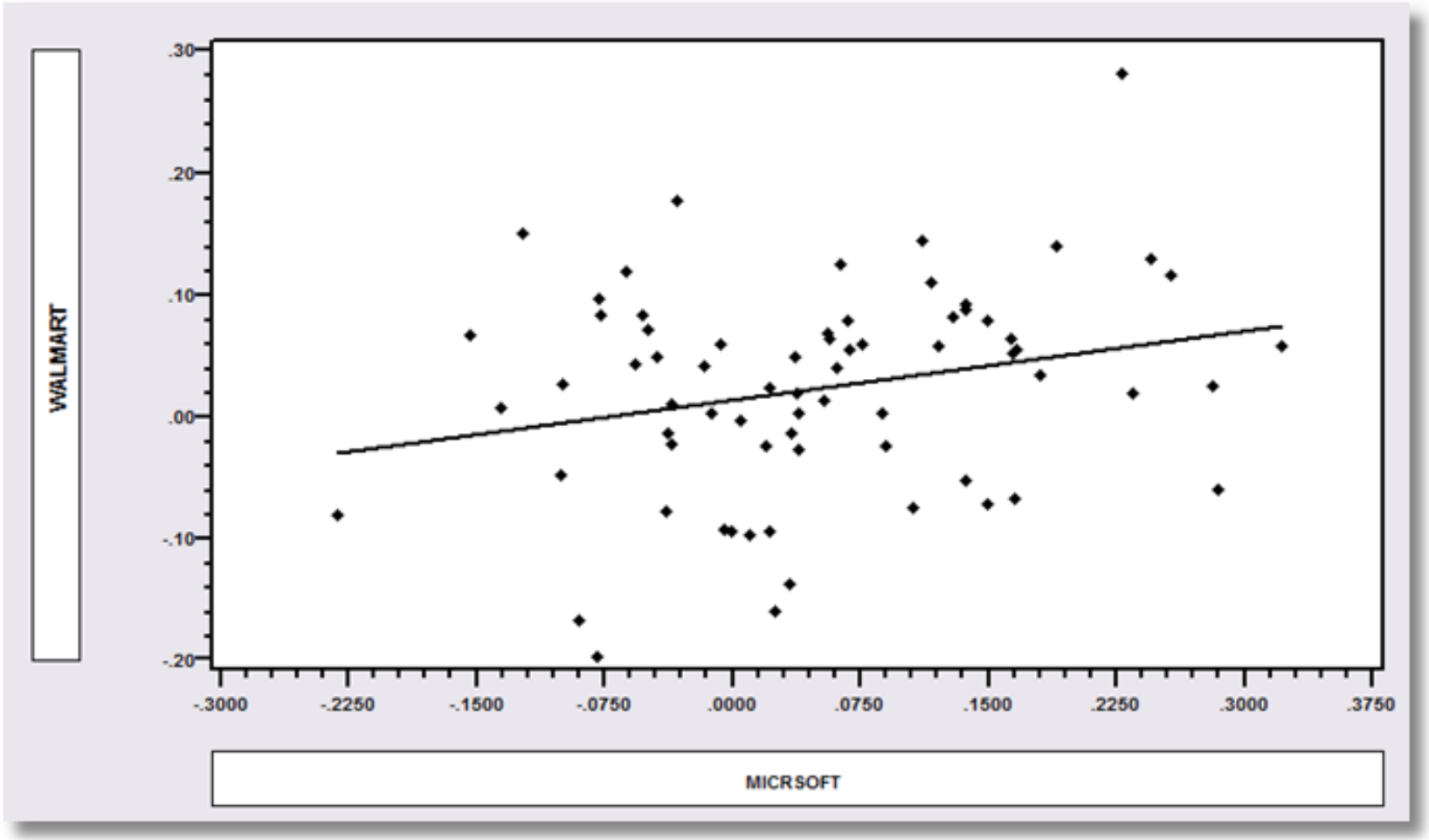
$$\text{Correlation} = \frac{0.107}{.46 \times .81} = \frac{.04}{.46 \times .81}$$

Correlated Variables: Returns on Two Stocks*

Line	Observation	MICROSOFT	WALMART
1	1995.Ja	.06105	.03825
2	1995.Fb	.12897	.08112
3	1995.Mr	.14939	-.07317
4	1995.Ap	.03592	.04737
5	1995.Ma	.06717	.07750
6	1995.Jn	-.07927	-.19758
7	1995.Jl	.02197	-.09463
8	1995.Au	.16347	.05133
9	1995.Sp	.14906	.07767
10	1995.Oc	.07591	.05774
11	1995.Nv	-.12341	.14858
12	1995.Dc	-.00512	-.09360
13	1996.Ja	.03478	-.01491
14	1996.Fb	-.00668	.05844
15	1996.Mr	.02507	-.16006
16	1996.Ap	.03730	.01830
17	1996.Ma	-.05005	.07009
18	1996.Jn	-.08999	-.16833
19	1996.Jl	.02205	.02185
20	1996.Au	.03833	-.02855
21	1996.Sp	-.23185	-.08104
22	1996.Oc	.13686	.08677
23	1996.Nv	-.03846	-.01522
24	1996.Dc	.13650	.09063
25	1997.Ja	.06838	.05314
26	1997.Fb	.25603	.11412
27	1997.Mr	.28069	.02381
28	1997.Ap	.05665	.06262
29	1997.Ma	-.00059	-.09532
30	1997.Jn	.00507	-.00469
31	1997.Jl	.22767	.28018
32	1997.Au	.23415	.01870
33	1997.Sp	-.03533	.00943
34	1997.Oc	.11070	.14296
35	1997.Nv	.05510	.06759
36	1997.Dc	.10587	-.07535
37	1998.Ja	-.05267	.08254
38	1998.Fb	.24470	.12798
39	1998.Mr	-.03216	.17625
40	1998.Ap	-.03886	-.07843
41	1998.Ma	-.07853	.09587
42	1998.Jn	.18007	.03312
43	1998.Jl	.06274	.12368
44	1998.Au	.32138	.05658
45	1998.Sp	.13594	-.05313
46	1998.Oc	.12014	.05717
47	1998.Nv	-.01213	.00188
48	1998.Dc	-.06274	.11772
49	1999.Ja	.28434	-.06084
50	1999.Fb	.11580	.10886
51	1999.Mr	-.15430	.06616
52	1999.Ap	.01907	-.02551
53	1999.Ma	-.13560	.00580
54	1999.Jn	-.01654	.03969
55	1999.Jl	.18923	.13860
56	1999.Au	-.07742	.08195
57	1999.Sp	.08946	-.02582
58	1999.Oc	-.10049	-.04920
59	1999.Nv	.08774	.00121
60	1999.Dc	.03362	-.13887
61	2000.Ja	.16265	.06255
62	2000.Fb	.03871	.00144
63	2000.Mr	-.05742	.04246
64	2000.Ap	.00938	-.09826
65	2000.Ma	.16490	-.06795
66	2000.Jn	.16640	.05390
67	2000.Jl	.05384	.01202
68	2000.Au	-.03610	-.02408
69	2000.Sp	-.09943	.02524
70	2000.Oc	-.04410	.04711

*** Averaged yearly return**

The two returns are positively correlated.



```
Descriptive Statistics for   2 variables
--------+-------------------------------------------------------------------
Variable|       Mean       Std.Dev.     Minimum      Maximum     Cases Missing
--------+-------------------------------------------------------------------
MICRSOFT|      .050071      .114264     -.231850      .321380        70       0
 WALMART|      .021906      .086035     -.197580      .280180        70       0
--------+-------------------------------------------------------------------
Correlation =   .2486345
```

Descriptive Statistics: The Relationship between Two Variables

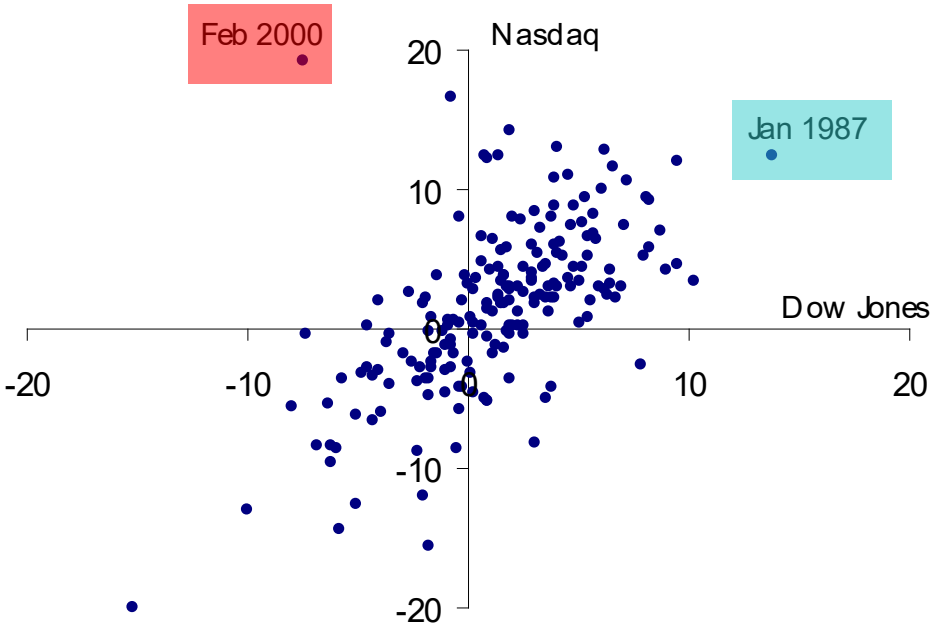
Scatter Diagram: Visual Illustration of How Two Variables Are Related

Monthly Percentage Growth Rate of Dow Jones Industrial Average

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1985	6.21	-0.22	-1.34	-0.69	4.55	1.53	0.90	-1.00	-0.40	3.44	7.12	5.07
1986	1.57	8.79	6.41	-1.90	5.20	0.85	-6.20	6.93	-6.89	6.23	1.94	-0.95
1987	13.82	3.06	3.63	-0.79	0.23	5.54	6.35	3.53	-2.50	-23.22	-8.02	5.74
2000	-4.84	-7.42	7.84	-1.72	-1.97	-0.71	0.71	6.59	-5.03	3.01	-5.07	3.59

Monthly Percentage Growth Rate of NASDAQ Composite Average

Year	Jan	Feb	Mar	Apr	May
1985	12.79	1.97	-1.76	0.50	3.64
1986	3.35	7.06	4.23	2.27	4.44
1987	12.41	8.39	1.20	-2.86	-0.31
2000	-3.17	19.19	-2.64	-15.57	-11.91



Each point on the scatter diagram represents the growth rate of the Dow and the growth rate of the Nasdaq for one specific month.

From Prof. Frank Westhoff
Amherst University

Correlation Coefficient

$$\text{CorrCoef}[x, y] = \frac{\text{Cov}[x, y]}{\sqrt{\text{Var}[x]}\sqrt{\text{Var}[y]}}$$

The denominator is positive.

Numerator: Up by a factor of 2.54

Denominator: Up by a factor of 2.54

How are the covariance and correlation coefficient similar?

The sign of covariance and the sign of the correlation coefficient are the same.

How do the covariance and correlation coefficient differ? Two important ways:

Correlation Coefficient Is Unaffected by the Choice of Units

Question: Again, what if we were to measure precipitation, the "x variable," in centimeters rather than inches?

$$\text{Var}[x] = \frac{\sum_{t=1}^{T}(x_t - \bar{x})^2}{T} \qquad \text{Cov}[x, y] = \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{T}$$

$(x_t - \bar{x})$ up by a factor of 2.54

$(x_t - \bar{x})^2$ up by a factor of 2.54^2

What happens to $\text{Var}[x]$?

$\text{Var}[x]$ up by a factor of 2.54^2

$\sqrt{\text{Var}[x]}$ up by a factor of 2.54

$\text{Cov}[x, y]$,up by a factor of 2.54.

What happens to $\text{Cov}[x, y]$?

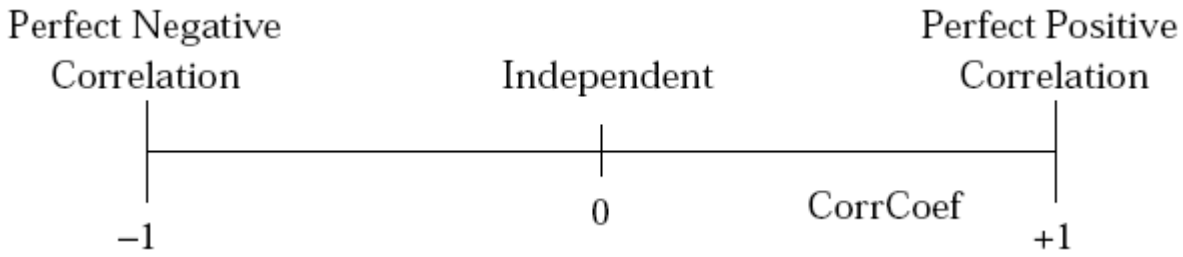
Both the numerator and denominator of CorrCoef[x, y] increase by a factor of 2.54

$\text{CorrCoef}[x, y]$ remains constant.

From Prof. Frank Westhoff Amherst University

How do the covariance and correlation coefficient differ – continued?

Correlation Coefficient Has a Limited Range: –1 to +1.



Consequently, the correlation coefficient reflects the magnitude of the correlation.

To illustrate that the correlation coefficient has a limited range, we shall consider the two polar cases:

Perfect positive correlation.

Perfect negative correlation.

From Prof. Frank Westhoff
Amherst University

To show that the correlation coefficient lies between –1 and +1, we consider two polar cases.

An Example of Perfect Positive Correlation: The two variables have identical values:

$$y_t = x_t \quad \text{for each } t = 1, 2, \ldots, T$$

What does Var[$\boldsymbol{y}$] equal?

$$\bar{y} = \bar{x}$$

$$y_t - \bar{y} = x_t - \bar{x}$$

$$(y_t - \bar{y})^2 = (x_t - \bar{x})^2$$

$$\frac{(y_t - \bar{y})^2}{T} = \frac{(x_t - \bar{x})^2}{T}$$

$$\text{Var}[\boldsymbol{y}] = \text{Var}[\boldsymbol{x}]$$

What does Cov[$\boldsymbol{x}, \boldsymbol{y}$] equal?

$$(x_t - \bar{x})(y_t - \bar{y}) = (x_t - \bar{x})(x_t - \bar{x})$$

$$= (x_t - \bar{x})^2$$

$$\frac{(x_t - \bar{x})(y_t - \bar{y})}{T} = \frac{(x_t - \bar{x})^2}{T}$$

$$\text{Cov}[\boldsymbol{x}, \boldsymbol{y}] = \text{Var}[\boldsymbol{x}]$$

$$\text{Var}[\boldsymbol{x}] = \frac{\sum_{t=1}^{T}(x_t - \bar{x})^2}{T} \qquad \text{Var}[\boldsymbol{y}] = \frac{\sum_{t=1}^{T}(y_t - \bar{y})^2}{T} \qquad \text{Cov}[\boldsymbol{x}, \boldsymbol{y}] = \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{T}$$

$$\text{CorrCoef}[\boldsymbol{x}, \boldsymbol{y}] = \frac{\text{Cov}[\boldsymbol{x}, \boldsymbol{y}]}{\sqrt{\text{Var}[\boldsymbol{x}]}\sqrt{\text{Var}[\boldsymbol{y}]}} = \frac{\text{Var}[\boldsymbol{x}]}{\sqrt{\text{Var}[\boldsymbol{x}]}\sqrt{\text{Var}[\boldsymbol{x}]}} = \frac{\text{Var}[\boldsymbol{x}]}{\text{Var}[\boldsymbol{x}]} = 1$$

From Prof. Frank Westhoff Amherst University

An Example of Perfect Negative Correlation: One variable is the negative of the other:

$$y_t = -x_t \qquad \text{for each } t = 1, 2, \ldots, T$$

What does Var[$\boldsymbol{y}$] equal?

What does Cov[$\boldsymbol{x}, \boldsymbol{y}$] equal?

$$\bar{y} = -\bar{x}$$

$$y_t - \bar{y} = -(x_t - \bar{x})$$

$$(y_t - \bar{y})^2 = (x_t - \bar{x})^2$$

$$\frac{(y_t - \bar{y})^2}{T} = \frac{(x_t - \bar{x})^2}{T}$$

$$\text{Var}[\boldsymbol{y}] = \text{Var}[\boldsymbol{x}]$$

$$(x_t - \bar{x})(y_t - \bar{y}) = -(x_t - \bar{x})(x_t - \bar{x})$$

$$= -(x_t - \bar{x})^2$$

$$\frac{(x_t - \bar{x})(y_t - \bar{y})}{T} = -\frac{(x_t - \bar{x})^2}{T}$$

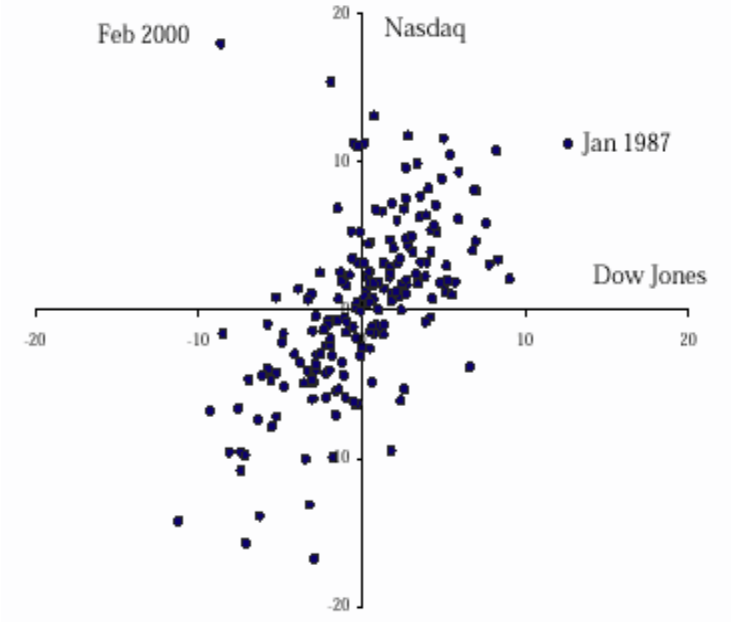
$$\text{Cov}[\boldsymbol{x}, \boldsymbol{y}] = -\text{Var}[\boldsymbol{x}]$$

$$\text{Var}[\boldsymbol{x}] = \frac{\sum_{t=1}^{T}(x_t - \bar{x})^2}{T} \qquad \text{Var}[\boldsymbol{y}] = \frac{\sum_{t=1}^{T}(y_t - \bar{y})^2}{T} \qquad \text{Cov}[\boldsymbol{x}, \boldsymbol{y}] = \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{T}$$

$$\text{CorrCoef}[\boldsymbol{x}, \boldsymbol{y}] = \frac{\text{Cov}[\boldsymbol{x}, \boldsymbol{y}]}{\sqrt{\text{Var}[\boldsymbol{x}]}\sqrt{\text{Var}[\boldsymbol{y}]}} = \frac{-\text{Var}[\boldsymbol{x}]}{\sqrt{\text{Var}[\boldsymbol{x}]}\sqrt{\text{Var}[\boldsymbol{x}]}} = \frac{\text{Var}[\boldsymbol{x}]}{\text{Var}[\boldsymbol{x}]} = -1$$

From Prof. Frank Westhoff Amherst University

Summary Scatter Diagrams, of Deviations From Means, Covariance, and Correlation Coefficient

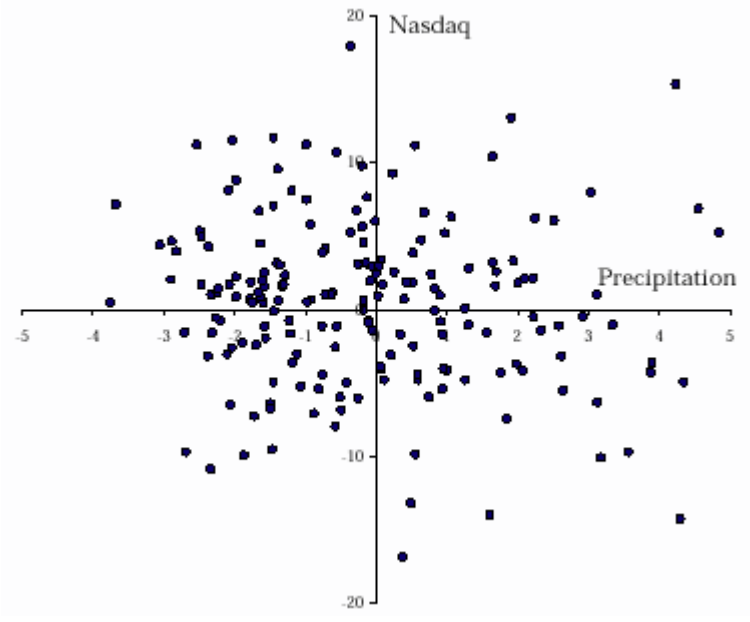


Not independent.

Positively correlated: Knowing one variable helps us predict the other.

Cov = 19.61

CorrCoef = .67

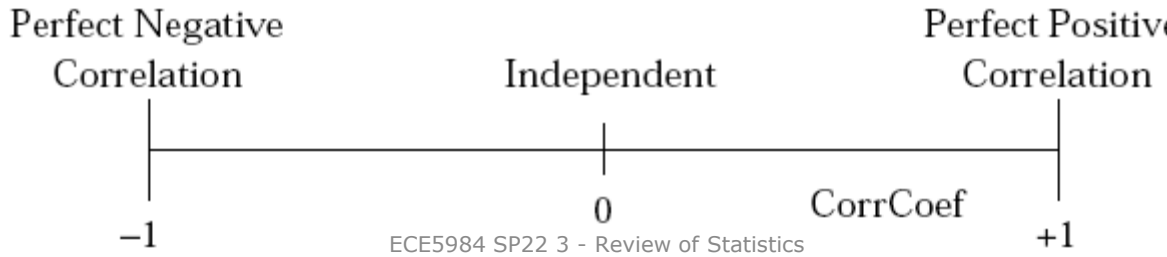


Independent.

Uncorrelated: Knowing one variable does not help us predict the other.

Cov = –.91 ≈ 0

CorrCoef = –.07 ≈ 0



From Prof. Frank Westhoff Amherst University

Let's correct a well-known fallacy: *correlation does not imply causation*

- Just because two variables are well-correlated does not mean that there is a causal link, either way
  - They might both be caused by a hidden third variable
  - It might be a coincidence

- Proving causality is **really hard**
  - Strong evidence (but not proof) of causation can come from randomized controlled experiments
    - Randomly split a group of people who suffer from a disease
    - Give one group a proposed cure, give the other group a *placebo*
    - Compare the outcome of the two groups
    - Key element: groups are statistically alike in all other ways
  - Often cannot do this

We often use the *covariance matrix* to examine the relationships between many variables (for example, the predictive features in an ADS)

- $\mathrm{K}_{X_i X_j} = \mathrm{cov}[X_i, X_j] = \sigma_{X_i X_j} = \mathrm{E}[(X_i - \mathrm{E}[X_i])(X_j - \mathrm{E}[X_j])]$
- $K_{XX} = \begin{bmatrix} \mathrm{E}[(X_1 - \mathrm{E}[X_1])(X_1 - \mathrm{E}[X_1])] & \mathrm{E}[(X_1 - \mathrm{E}[X_1])(X_2 - \mathrm{E}[X_2])] & \cdots & \mathrm{E}[(X_1 - \mathrm{E}[X_1])(X_n - \mathrm{E}[X_n])] \\ \mathrm{E}[(X_2 - \mathrm{E}[X_2])(X_1 - \mathrm{E}[X_1])] & \mathrm{E}[(X_2 - \mathrm{E}[X_2])(X_2 - \mathrm{E}[X_2])] & \cdots & \mathrm{E}[(X_2 - \mathrm{E}[X_2])(X_n - \mathrm{E}[X_n])] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}[(X_n - \mathrm{E}[X_n])(X_1 - \mathrm{E}[X_1])] & \mathrm{E}[(X_n - \mathrm{E}[X_n])(X_2 - \mathrm{E}[X_2])] & \ldots & \mathrm{E}[(X_n - \mathrm{E}[X_n])(X_n - \mathrm{E}[X_n])] \end{bmatrix}$
- $K_{XX} = \begin{bmatrix} \sigma^2{}_{X_1} & \sigma_{X_1 X_2} & \cdots & \sigma_{X_1 X_n} \\ \sigma_{X_2 X_1} & \sigma^2{}_{X_2} & \cdots & \sigma_{X_2 X_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_n X_1} & \sigma_{X_n X_2} & \ldots & \sigma^2{}_{X_n} \end{bmatrix}$, where $\sigma^2{}_{X_i} = \sigma_{X_i X_i} = \mathrm{E}[(X_i - \mathrm{E}[X_i])(X_i - \mathrm{E}[X_i])]$

The covariance matrix for a set of variables carries a lot of information about their relationship

- Non-zero off-axis elements indicate some correlation between the variables
  - Any independent (uncorrelated) variables will have zeroes off-axis
- The magnitude of the off-axis elements is related to the correlated (or joint) variance
  - In relation to the on-axis variances, we can determine degree of correlation

- For any two variables, the correlation coefficient can be calculated from the covariance matrix
- $$\rho_{ij} = \frac{\mathrm{E}\left[(X_i - \mu_i)(X_j - \mu_j)\right]}{s_i s_j} = \frac{\sigma_{X_i X_j}}{\sigma_{X_i}\sigma_{X_j}} = \frac{\sigma_{X_i X_j}}{\sqrt{\sigma^2{}_{X_i}}\sqrt{\sigma^2{}_{X_j}}} = \frac{K_{ij}}{\sqrt{K_{ii}}\sqrt{K_{jj}}}$$

G	AB	APP	H	2B	3B	HR	HRBASES	RBI	JUNK
1	4	1	0	0	0	0	0	0	188
25	118	25	32	6	0	0	0	13	96
29	137	29	40	4	5	0	0	19	10
27	133	27	44	10	2	2	8	27	188
25	120	25	39	11	3	0	0	16	17
12	49	12	11	2	1	0	0	5	2
1	4	1	1	0	0	0	0	2	183
31	157	31	63	10	9	0	0	34	166
1	5	1	1	1	0	0	0	1	89
18	86	18	13	2	1	0	0	11	101
22	89	22	27	1	10	3	12	18	123
1	3	1	0	0	0	0	0	0	138
10	36	10	7	0	0	0	0	1	190
3	15	3	6	0	0	0	0	5	44
20	94	20	33	9	1	1	4	21	72
29	128	29	32	3	3	0	0	23	97
1	4	1	0	0	0	0	0	0	4
1	4	1	1	0	0	0	0	0	163
17	73	17	17	1	1	0	0	8	155
1	2	1	0	0	0	0	0	0	5
25	106	25	28	2	5	0	0	13	180
29	152	29	46	3	3	0	0	24	157
30	134	30	30	4	0	0	0	21	170
3	14	3	1	0	0	0	0	0	79
12	63	12	15	2	3	1	4	14	181
19	87	19	20	2	0	0	0	10	147
29	127	29	32	8	1	0	0	18	157
19	77	19	20	3	1	0	0	16	191
7	33	7	7	0	0	0	0	2	149
27	118	27	38	8	1	0	0	26	99
28	150	28	37	7	5	3	12	30	183
6	22	6	4	0	1	0	0	2	62

COV	G	AB	APP	H	2B	3B	HR	HRBASES	RBI	JUNK
G	2210.474	8168.775	2210.474	2277.836	398.0319	79.29819	202.1473	808.5891	1077.756	5.702245
AB	8168.775	33961.27	8168.775	9537.537	1660.304	342.2181	826.0891	3304.357	4493.652	42.63293
APP	2210.474	8168.775	2210.474	2277.836	398.0319	79.29819	202.1473	808.5891	1077.756	5.702245
H	2277.836	9537.537	2277.836	2745.703	479.7856	100.5687	237.0437	948.1746	1297.509	13.42671
2B	398.0319	1660.304	398.0319	479.7856	93.46701	16.45938	45.13218	180.5287	234.5061	1.981836
3B	79.29819	342.2181	79.29819	100.5687	16.45938	6.802997	5.763516	23.05406	45.49688	0.332044
HR	202.1473	826.0891	202.1473	237.0437	45.13218	5.763516	40.91035	163.6414	141.6601	0.453692
HRBASES	808.5891	3304.357	808.5891	948.1746	180.5287	23.05406	163.6414	654.5656	566.6403	1.814767
RBI	1077.756	4493.652	1077.756	1297.509	234.5061	45.49688	141.6601	566.6403	693.6194	5.187408
JUNK	5.702245	42.63293	5.702245	13.42671	1.981836	0.332044	0.453692	1.814767	5.187408	3364.827

CORREL	G	AB	APP	H	2B	3B	HR	HRBASES	RBI	JUNK
G	1	0.942806	1	0.924598	0.875681	0.646652	0.672216	0.672216	0.869378	0.002091
AB	0.942806	1	0.942806	0.987683	0.931894	0.711969	0.700839	0.700839	0.924371	0.003988
APP	1	0.942806	1	0.924598	0.875681	0.646652	0.672216	0.672216	0.869378	0.002091
H	0.924598	0.987683	0.924598	1	0.94709	0.735844	0.70727	0.70727	0.93854	0.004417
2B	0.875681	0.931894	0.875681	0.94709	1	0.652731	0.729862	0.729862	0.919224	0.003534
3B	0.646652	0.711969	0.646652	0.735844	0.652731	1	0.345478	0.345478	0.663192	0.002195
HR	0.672216	0.700839	0.672216	0.70727	0.729862	0.345478	1	1	0.838508	0.001223
HRBASES	0.672216	0.700839	0.672216	0.70727	0.729862	0.345478	1	1	0.838508	0.001223
RBI	0.869378	0.924371	0.869378	0.93854	0.919224	0.663192	0.838508	0.838508	1	0.003395
JUNK	0.002091	0.003988	0.002091	0.004417	0.003534	0.002195	0.001223	0.001223	0.003395	1

Today's Objectives

- Probability
- Conditional Probability
- Dependence
- Descriptive Statistics
- Covariance
- Correlation
- Covariance Matrix