Class 8

Daniel B. Rowe, Ph.D.

Department of Mathematical and Statistical Sciences

Agenda:

Review Chapters 1 – 5
(Exam 1 Chapters)

Just the highlights!

1. Summation Notation

$$\sum_{i=1}^{n} f(x_i) = f(x_1) + f(x_2) + \ldots + f(x_n)$$

2. Factorials

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$$

3. Computations

x=20, y=14, s=16, w=-2, m=15, n=10

Compute $x + y \cdot \frac{\sqrt{s}}{n} = 25.6$

4. Simple Linear Equations

$2 - 2x = 3x + 3 \qquad x = -1/5$

1: Statistics

1.1 Americans Here's Looking at you

Statistics is all around us!

How much time between Internet usage?

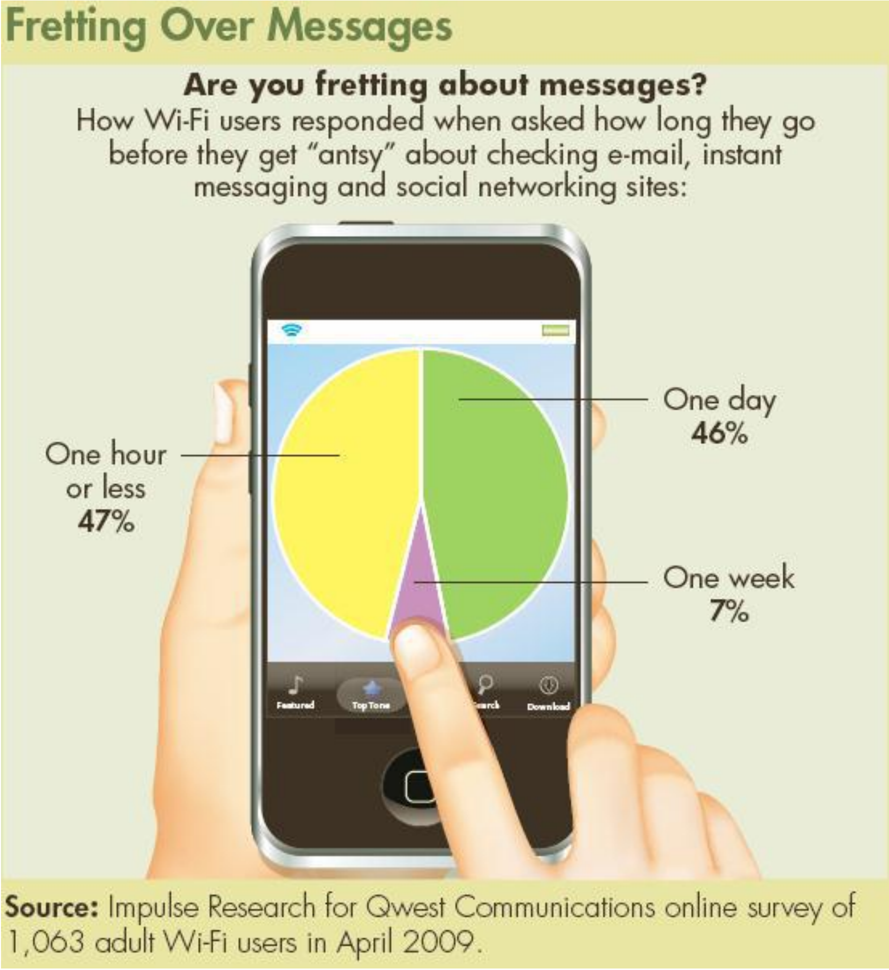


Figure from Johnson & Kuby, 2012.

1: Statistics

1.1 What is Statistics?

Population: A collection, or set, of individuals, objects, or events whose properties are to be analyzed.

Sample: Subset of the population.

Variable: A characteristic of interest about each individual element of a population or sample.

Data value: The value of the variable associated with one element of a population or sample.

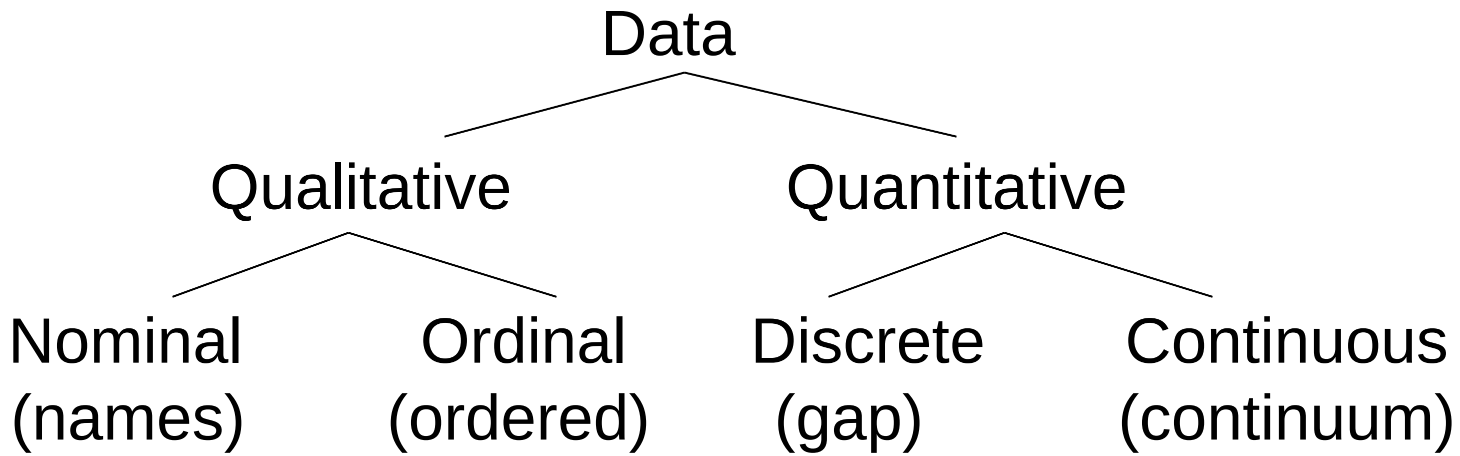
Parameter: A numerical value summarizing all the data of an entire population.

Statistic: A numerical value summarizing the sample data.

1: Statistics

1.1 What is Statistics?

Data: The set of values collected from the variable from each of the elements that belong to the sample.



1: Statistics

1.1 What is Statistics?

Qualitative variable: A variable that describes or categorizes an element of a population.

Nominal variable: A qualitative variable that characterizes an element of a population. No ordering. No arithmetic.

Ordinal variable: A qualitative variable that incorporates an ordered position, or ranking.

Quantitative variable: A variable that quantifies an element of a population.

Discrete variable: A quantitative variable that can assume a countable number of values. Gap between successive values.

Continuous variable: A quantitative variable that can assume an uncountable number of values. Continuum of values.

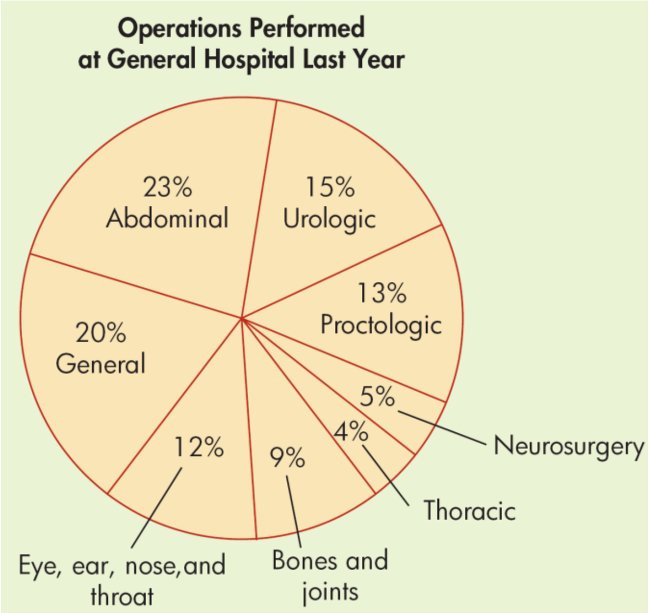
2: Descriptive Analysis and Single Variable Data

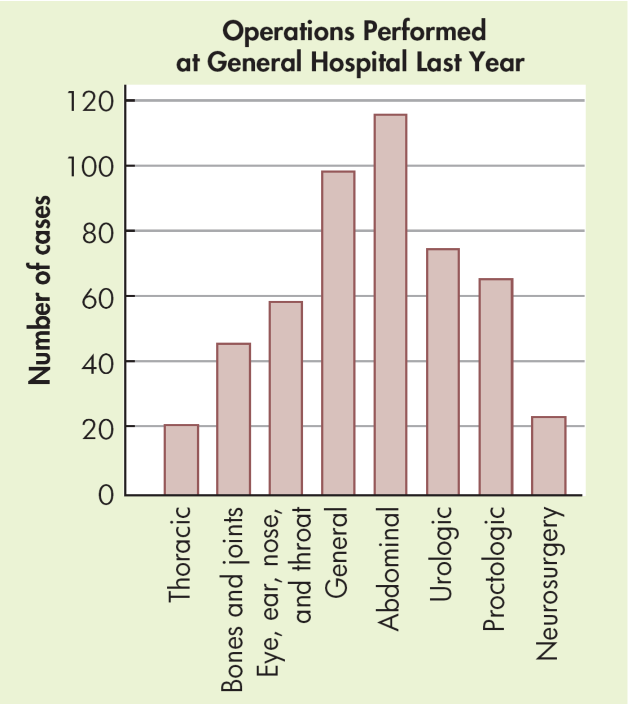
2.1 Graphs - Qualitative Data

Circle (pie) graphs and bar graphs:

Circle is parts to whole as angle.

Bar graph is amount in each category as rectangular areas.





Figures from Johnson & Kuby, 2012.

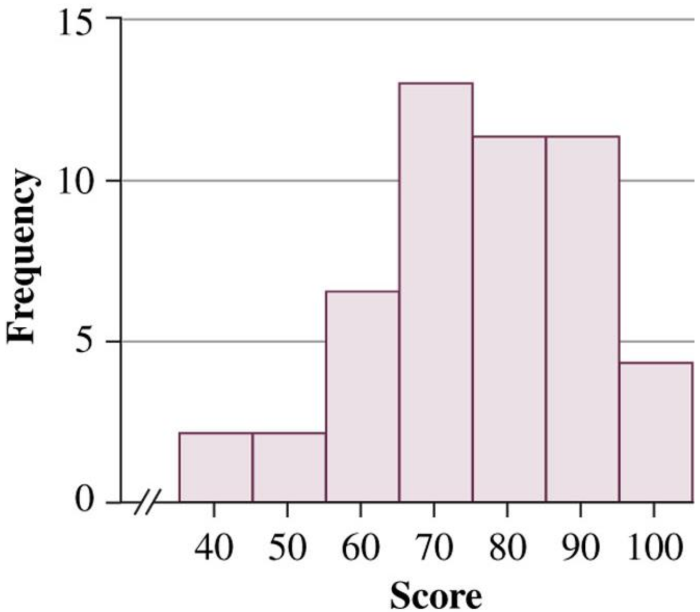
2: Descriptive Analysis and Single Variable Data

2.2 Frequency Distributions and Histograms

Statistics Exam Scores [TA02-06]

60	47	82	95	88	72	67	66	68	98	90	77	86
58	64	95	74	72	88	74	77	39	90	63	68	97
70	64	70	70	58	78	89	44	55	85	82	83	
72	77	72	86	50	94	92	80	91	75	76	78	

Boundaries	Frequency
$35 \leq x < 45$	2
$45 \leq x < 55$	2
$55 \leq x < 65$	7
$65 \leq x < 75$	13
$75 \leq x < 85$	11
$85 \leq x < 95$	11
$95 \leq x \leq 105$	4
	50



Figures from Johnson & Kuby, 2012.

2: Descriptive Analysis and Single Variable Data

2.3 Measures of Central Tendency

Sample Mean: Usual average, p. 63 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Sample Median: Middle value, p. 64 n odd, $\tilde{x} = \frac{n+1}{2}$ value

n even, avg $\frac{n}{2}$ & $\frac{n}{2}+1$ values

Sample Mode: Most often, p. 66 $\hat{x} = \text{most often}$

Measures of central tendency characterize center of distribution.

Measures of dispersion characterize the variability in the data.

2: Descriptive Analysis and Single Variable Data

2.4 Measures of Dispersion

Range: $H - L$, p. 74

Deviation from mean: value minus sample mean, p. 74

$$i^{th} \ deviation \ from \ mean = x_i - \bar{x}$$

Sample Variance: avg squared dev using n-1 in den, p. 76

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} x_i^2 - \left[\left(\sum_{i=1}^{n} x_i\right)^2 / n\right]\right\}$$

Sample Standard Deviation: $s = \sqrt{s^2}$

2: Descriptive Analysis and Single Variable Data

2.3, 2.4 Measures of Central Tendency and Dispersion

Example: Data values: 1,2,2,3,4

$\bar{x} = 2.4$ $\hat{x} = 2$ $\tilde{x} = 2$

$s^2 = 1.3$ $s = 1.1$

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

$\hat{x} =$ most often value

$\tilde{x} =$ middle value

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

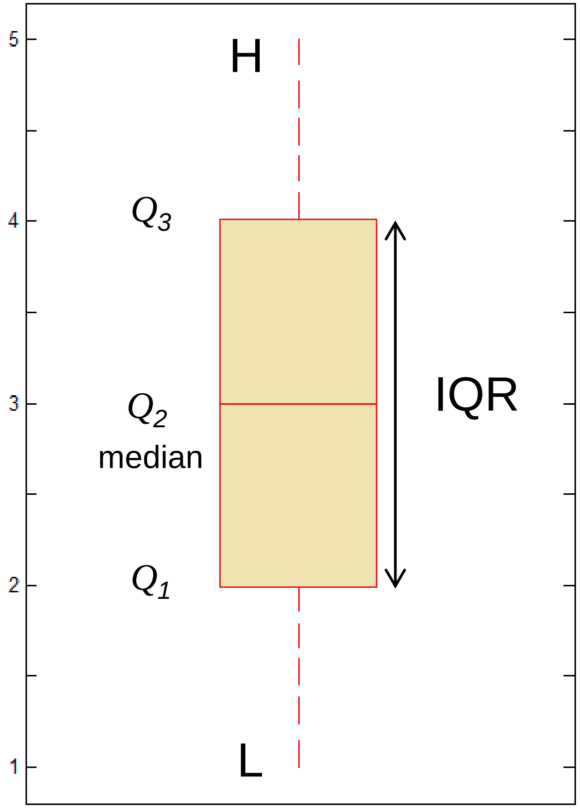
$$s = \sqrt{s^2}$$

2: Descriptive Analysis and Single Variable Data

2.5 Measures of Position

Measures of Position: Quartiles - ranked data into quarters

L = lowest value
H = highest value
Q_2 = median
Q_1 = 25% smaller
Q_3 = 75% smaller
IQR = Q_3 - Q_1



2: Descriptive Analysis and Single Variable Data

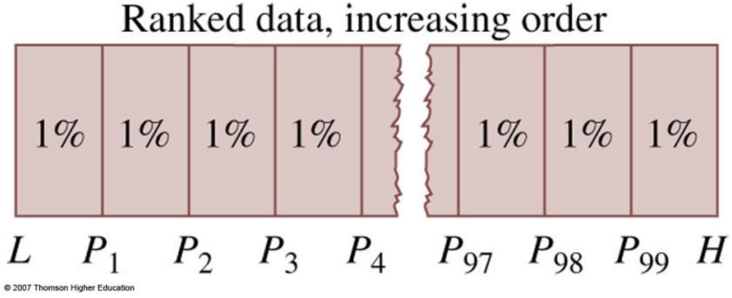
2.5 Measures of Position

Measures of Position: percentiles - rank data into 100^{ths}

L = lowest value

H = highest value

P_k = value where k% are smaller



rank data $\frac{nk}{100}$

integer, A → p_k halfway between value and next one average of A^{th} and $(A+1)^{th}$ values

not Integer, B → p_k is value in next largest position, B+1 value

Figure from Johnson & Kuby, 2012.

2: Descriptive Analysis and Single Variable Data

2.5 Measures of Position

Standard score, or z-score: The position a particular value of x has relative to the mean, measured in standard deviations.

$$z_i = \frac{i^{\text{th}} \text{ value - mean}}{\text{std. dev.}} = \frac{x_i - \bar{x}}{s}$$

There can be n of these because we have $x_1, x_2, \ldots, x_n$.

3: Descriptive Analysis and Bivariate Data

3.1 Bivariate Data: two qualitative

Cross-tabulation tables or **contingency tables**

Example:
Construct a 2×3 table.
Know different %ages.

Name	Gender	Major	Name	Gender	Major	Name	Gender	Major
Adams	M	LA	Feeney	M	T	McGowan	M	BA
Argento	F	BA	Flanigan	M	LA	Mowers	F	BA
Baker	M	LA	Hodge	F	LA	Ornt	M	T
Bennett	F	LA	Holmes	M	T	Palmer	F	LA
Brand	M	T	Jopson	F	T	Pullen	M	T
Brock	M	BA	Kee	M	BA	Rattan	M	BA
Chun	F	LA	Kleeberg	M	LA	Sherman	F	LA
Crain	M	T	Light	M	BA	Small	F	T
Cross	F	BA	Linton	F	LA	Tate	M	BA
Ellis	F	BA	Lopez	M	T	Yamamoto	M	LA

	Major			
Gender	LA	BA	T	Row Total
M	5	6	7	18
F	6	4	2	12
Col. Total	11	10	9	30

M = male
F = female
LA = liberal arts
BA = business admin
T = technology

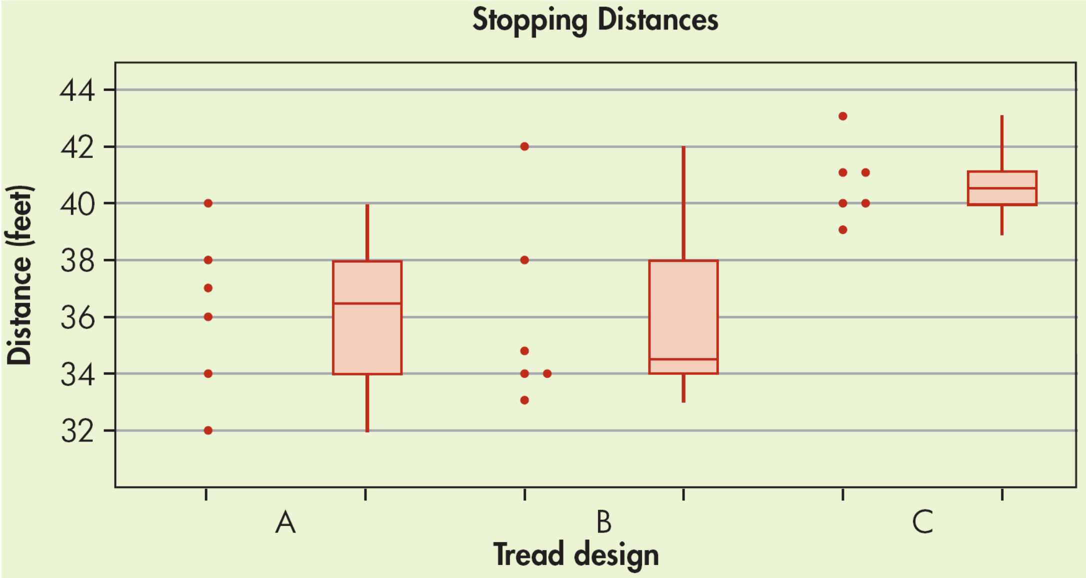
Figures from Johnson & Kuby, 2012.

3: Descriptive Analysis and Bivariate Data

3.1 Bivariate Data: one qualitative and one quantitative

Example:

Design A (n = 6)			Design B (n = 6)			Design C (n = 6)		
37	36	38	33	35	38	40	39	40
34	40	32	34	42	34	41	41	43



Vertical box-and-whiskers

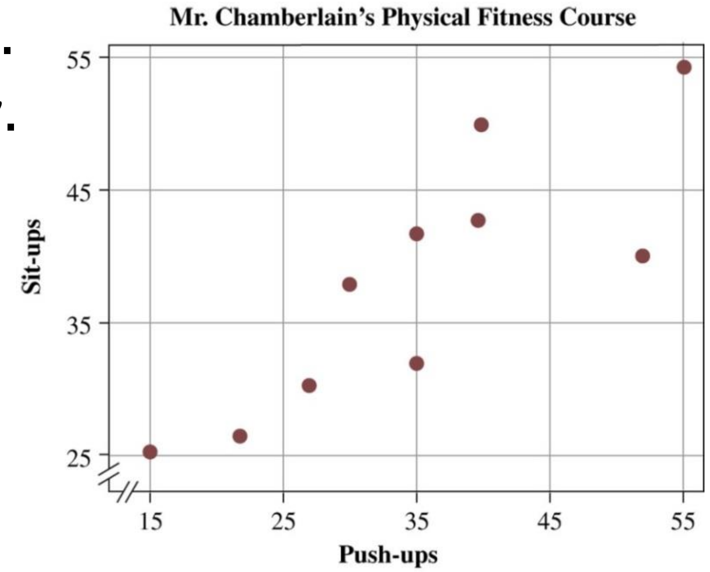
Figures from Johnson & Kuby, 2012.

3: Descriptive Analysis and Bivariate Data
3.1 Bivariate Data: two quantitative, Scatter Diagram

Example: Push-ups

Student	1	2	3	4	5	6	7	8	9	10
Push-ups, x	27	22	15	35	30	52	35	55	40	40
Sit-ups, y	30	26	25	42	38	40	32	54	50	43

Input variable: independent variable, x.
Output variable: dependent variable, y.



Scatter Diagram: A plot of all the ordered pairs of bivariate data on a coordinate axis system.

(x,y) ordered pairs.

Figures from Johnson & Kuby, 2012.

3: Descriptive Analysis and Bivariate Data
3.2 Linear Correlation

Linear Correlation, $\boldsymbol{r}$, is a measure of the strength of a linear relationship between two variables x and y.

$$-1 \leq r \leq 1$$

Will discuss its computation in a minute.

$r \approx 0$ $\quad$ $r \approx 0.5$ $\quad$ $r \approx 0.8$ $\quad$ $r \approx -0.5$ $\quad$ $r \approx -0.8$

Figure from Johnson & Kuby, 2012.

3: Descriptive Analysis and Bivariate Data

3.2 Linear Correlation

Example:

Push-ups, x	27	22	15	35	30	52	35	55	40	40
Sit-ups, y	30	26	25	42	38	40	32	54	50	43

Student	Push-ups, x	x^2	Sit-ups, y	y^2	xy
1	27	729	30	900	810
2	22	484	26	676	572
3	15	225	25	625	375
4	35	1,225	42	1,764	1,470
5	30	900	38	1,444	1,140
6	52	2,704	40	1,600	2,080
7	35	1,225	32	1,024	1,120
8	55	3,025	54	2,916	2,970
9	40	1,600	50	2,500	2,000
10	40	1,600	43	1,849	1,720
	$\sum x = 351$ sum of x	$\sum x^2 = 13{,}717$ sum of x^2	$\sum y = 380$ sum of y	$\sum y^2 = 15{,}298$ sum of y^2	$\sum xy = 14{,}257$ sum of xy

Figure from Johnson & Kuby, 2012.

3: Descriptive Analysis and Bivariate Data

3.2 Linear Correlation

Example:

Push-ups, x	27	22	15	35	30	52	35	55	40	40
Sit-ups, y	30	26	25	42	38	40	32	54	50	43

$$\sum_{i=1}^{n} x_i = 351$$

$$\sum_{i=1}^{n} x_i^2 = 13717$$

$$\sum_{i=1}^{n} y_i = 380$$

$$\sum_{i=1}^{n} y_i^2 = 15298$$

$$\sum_{i=1}^{n} x_i y_i = 14257$$

$$SS(x) = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 = 13717 - \frac{(351)^2}{10} = 1396.9$$

$$SS(y) = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 = 15298 - \frac{(380)^2}{10} = 858.0$$

$$SS(xy) = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) = 14257 - \frac{(351)(380)}{10} = 919.0$$

$$r = \frac{SS(xy)}{\sqrt{SS(x)SS(y)}} = \frac{919.0}{\sqrt{(1396.9)(858.0)}} = 0.84$$

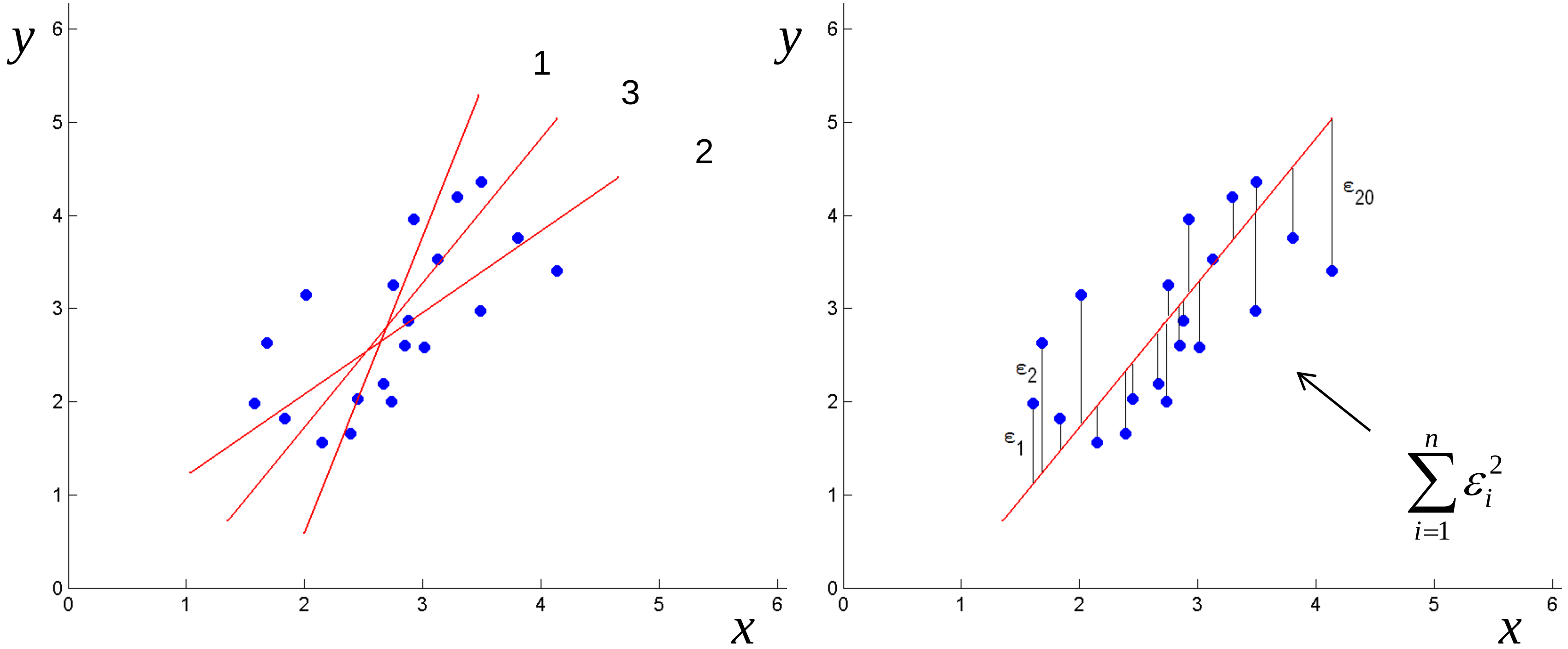
Figure from Johnson & Kuby, 2012.

3: Descriptive Analysis and Bivariate Data

3.3 Linear Regression

b_0 is estimated y-intercept
b_1 is estimated slope.

We try different lines until we find the "best" one, $\hat{y} = b_0 + b_1 x$.



Move line until sum of the squared residuals is a minimum.

3: Descriptive Analysis and Bivariate Data

3.3 Linear Regression

(x,y) pairs: (1,1),(3,2),(2,3),(4,4)

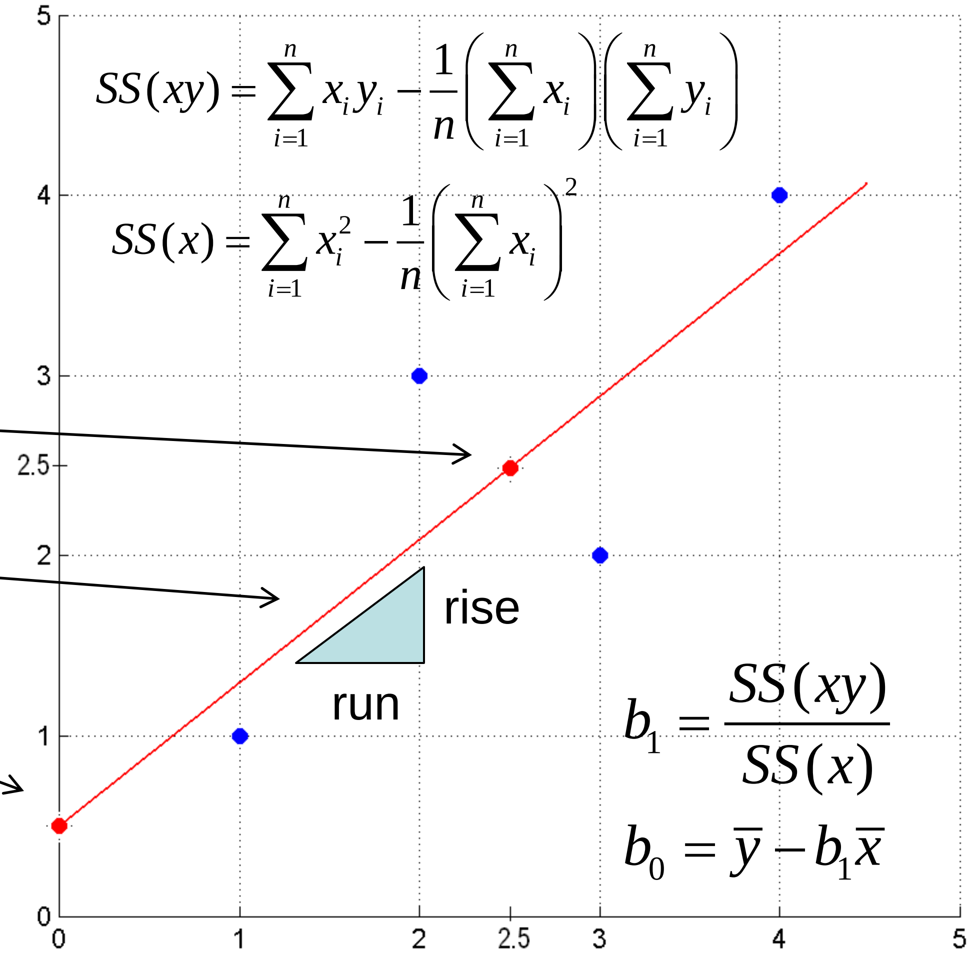
Plotted points.

The line goes through $(\bar{x},\bar{y})$.

The slope is b_1=0.8.

The y - intercept b_0=0.5 .

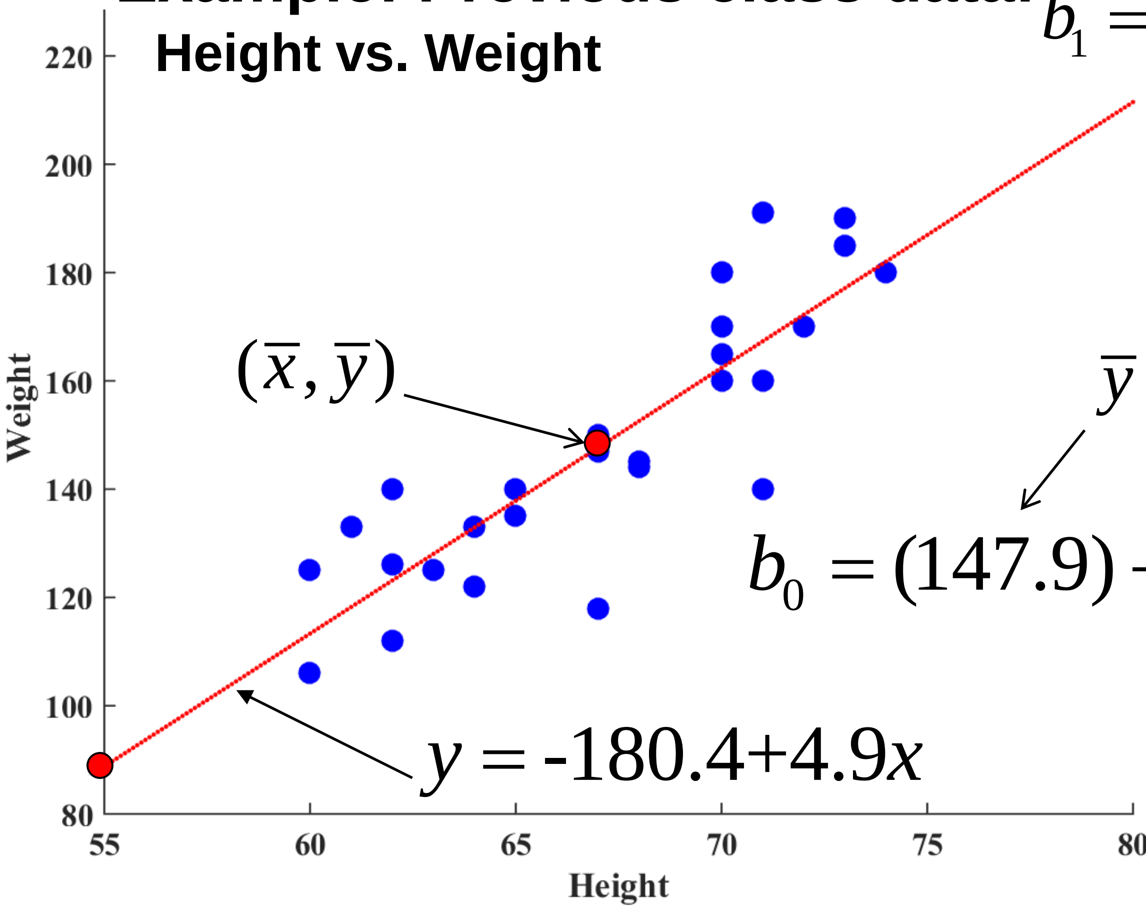
Two points (2.5,2.5) and (0,.5).



3: Descriptive Analysis and Bivariate Data

3.3 Linear Regression

Example: Previous class data!



$$b_1 = \frac{SS(xy)}{SS(x)} = \frac{2370.1}{483.0} = 4.9$$

units of lbs/in

$\bar{y}$ b_1 $\bar{x}$

$$b_0 = (147.9) - (4.9)(67.0) = -180.4$$

point-slope formula

4: Probability

4.1 Probability of Events

An **experiment** is a process by which a measurement is taken or observations is made. i.e. flip coin or roll die

An **outcome** is the result of an experiment. i.e. Heads, or 3

Sample space is a listing of possible outcomes. i.e. $S=\{H,T\}$

An **event** is an outcome or a combination of outcomes.
i.e. even number when rolling a die

4: Probability

4.1 Probability of Events

A_i are events

Property 1: $0 \le P(A_i) \le 1$

Property 2: $\sum_{i=1}^{n} P(O_i) = 1$

$i = 1, \ldots, n$

O_i are outcomes

Approaches to probability.

(1) Empirical (AKA experimental)

$$\text{empirical probability of } A = \frac{\text{number of times } A \text{ occured}}{\text{number of trials}}$$

(2) Theoretical (AKA classical or equally likely)

$$\text{theoretical probability of } A = \frac{\text{number of times } A \text{ occus in sample space}}{\text{number of elements in the sample space}}$$

4: Probability - Empirical

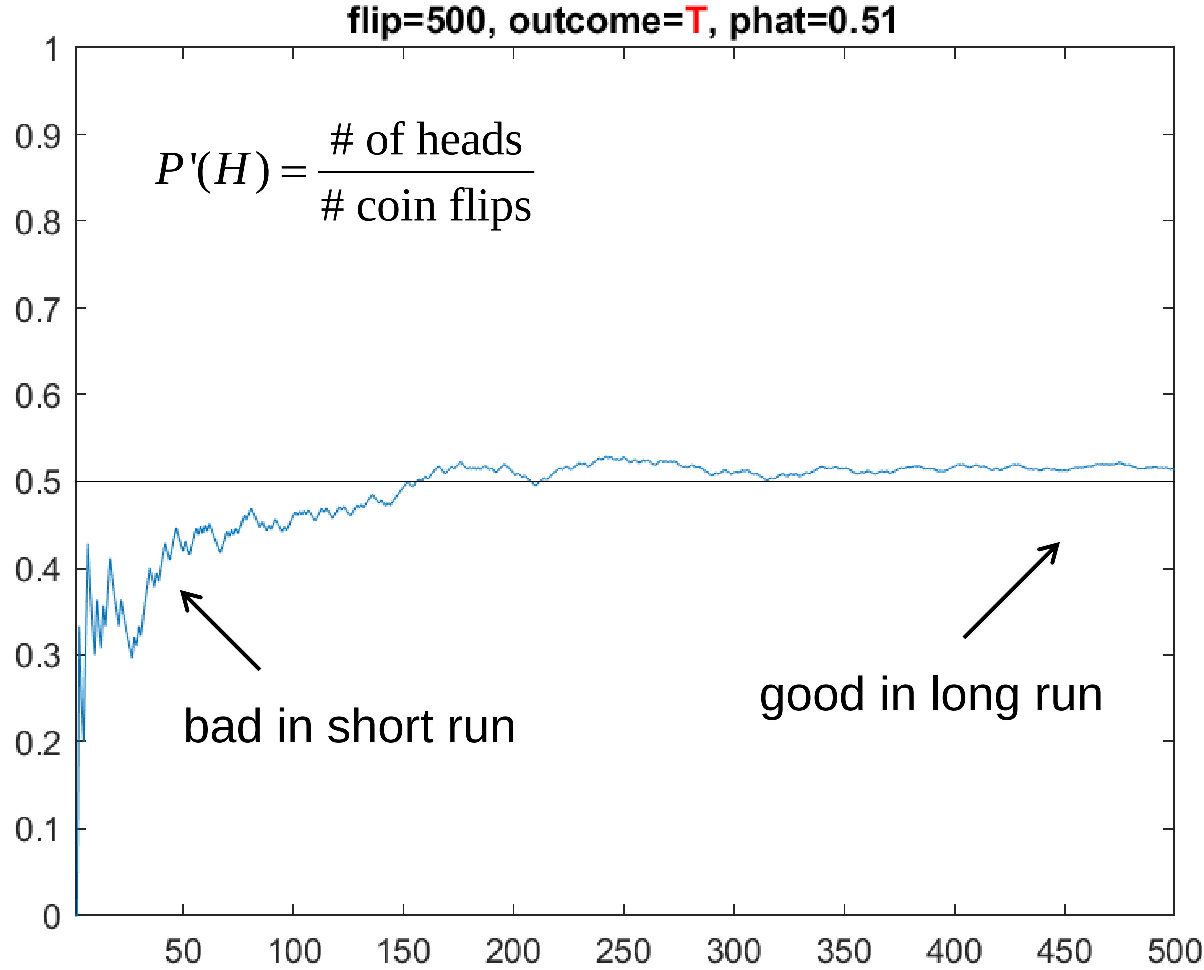
4.1 Probability of Events – Law of large numbers

Had computer flip a single coin 1000 times.

Flip # on x axis

$P'(H)$ on y axis.

This shows convergence to true value of 1/2.



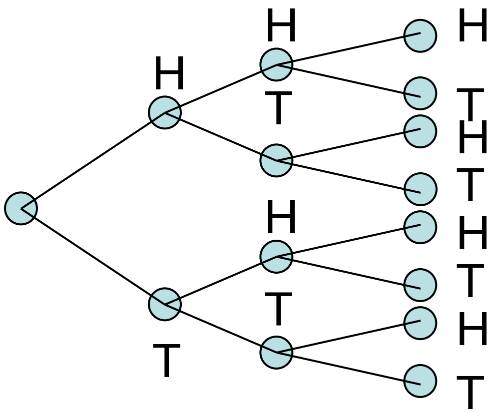
4: Probability

4.1 Probability of Events

So let's flip a coin three times.

Sample space:
listing of outcomes
for 3 flips

Can flip three times.



$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

$$P(HHH) = \frac{\#\text{ times } HHH \text{ occurs in } S}{\#\text{ elements in } S}$$

4: Probability

4.2 Rules of Probability – "*A* or *B*"

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

Example: Pick Card, A=Heart, B=Ace

$P(\text{Heart or Ace})$

$P(\text{Heart}) = 13/52$

$P(\text{Ace}) = 4/52$

$P(\text{Heart and Ace}) = 1/52$

double counted

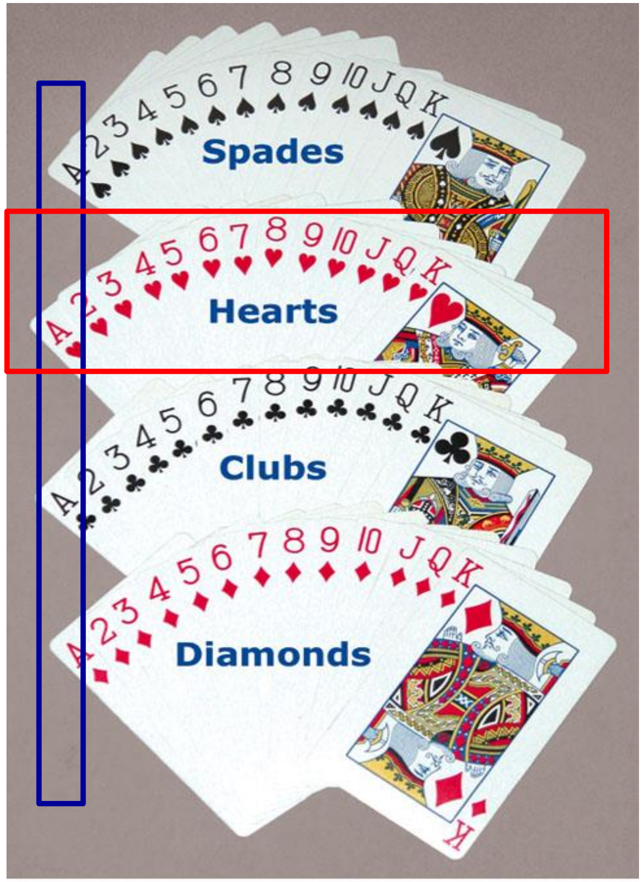


Figure from Johnson & Kuby, 2012.

$$P(\text{Heart or Ace}) = P(\text{Heart}) + P(\text{Ace}) - P(\text{Heart and Ace})$$

$$P(\text{Heart or Ace}) = 13/52 + 4/52 - 1/52 = 16/52$$

4: Probability
4.3 Rules of Probability

Example: Pick Card, A=Heart, B=Ace

$$P(\bar{A}) = 1 - P(A)$$

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

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

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

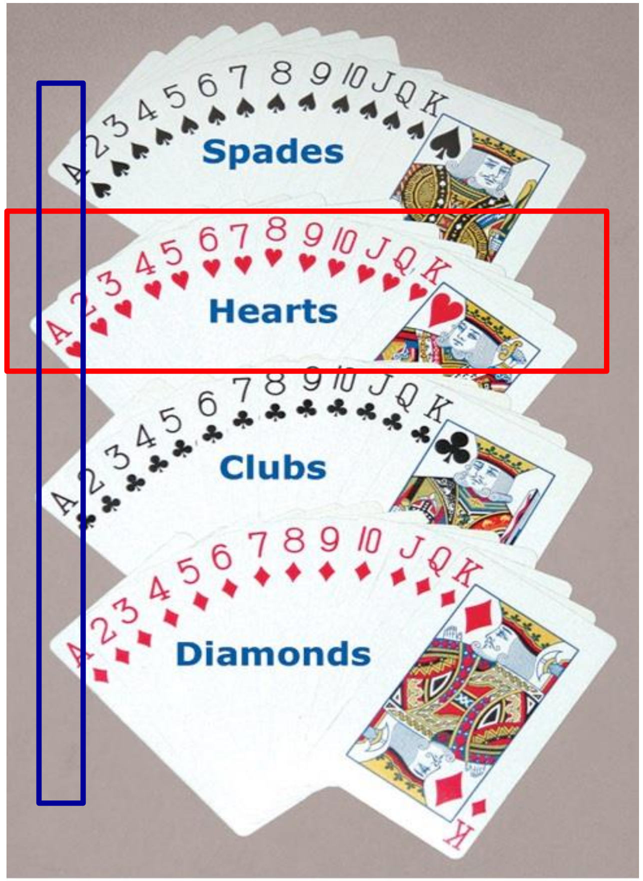


Figure from Johnson & Kuby, 2012.

4: Probability

4.5 Independent Events

Independent events: … the occurrence or nonoccurence of one gives no information about … occurrence for the other.

$$P(A) = P(A \mid B) = P(A \mid \text{not } B)$$

Dependent events: … occurrence of one event does have an effect on the probability of occurrence of the other event.

$$P(A) \neq P(A \mid B)$$

Special multiplication rule:
Let A and B … independent events … in a sample space S.

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

5: Probability Distributions (Discrete Variables)

5.1 Random Variables

Random Variables: A variable that assumes a unique numerical value for each of the outcomes in the sample space of a probability experiment.

Example:

Let x = the number of heads when we flip a coin twice.

$x=\{0,1,2\}$ $\longleftarrow$ numerical values

for

$\{TT,TH,HT,HH\}$ $\longleftarrow$ outcomes in sample space

5: Probability Distributions (Discrete Variables)

5.1 Random Variables

Discrete Random Variables: A quantitative random variable that can assume a countable number of values.

Continuous Random Variable: A quantitative random variable that can assume an uncountable number of values.

Continuum of values.

Examples:

Discrete: Number of heads when we flip a coin ten times.

Continuous: Distance from earth center to sun center.

5: Probability Distributions (Discrete Variables)

5.2 Probability Distributions of a Discrete Random Variable

Random Variables: … assumes a unique … value for each of the outcomes in the sample space … .

Probability Function: A rule $P(x)$ that assigns probabilities to the values of the random variable x.

Example:

Let x = # of heads when we flip a coin twice.

$x=\{0,1,2\}$

$$P(x)=\frac{2!}{x!(2-x)!}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{2-x}$$

x	$P(x)$
0	$\frac{1}{4}$
1	$\frac{1}{2}$
2	$\frac{1}{4}$

5: Probability Distributions (Discrete Variables)

5.2 Mean and Variance of a Discrete Random Variable

$$\mu = \sum_{i=1}^{n} [x_i P(x_i)] = x_1 P(x_1) + x_2 P(x_2) + \ldots + x_n P(x_n)$$

For the # of H when we flip a coin twice discrete distribution:

$$\mu = (x_1) \cdot P(x_1) + (x_2) \cdot P(x_2) + (x_3) \cdot P(x_3)$$

$$\mu = (0) \cdot P(0) + (1) \cdot P(1) + (2) \cdot P(2)$$

$$\mu = (0) \cdot (1/4) + (1) \cdot (1/2) + (2) \cdot (1/4)$$

$$\mu = 0 + 1/2 + 1/2$$

$$\mu = 1$$

	x	$P(x)$	
x_1	0	$\frac{1}{4}$	$P(x_1)$
x_2	1	$\frac{1}{2}$	$P(x_2)$
x_3	2	$\frac{1}{4}$	$P(x_3)$

5: Probability Distributions (Discrete Variables)

5.2 Mean and Variance of a Discrete Random Variable

$$\sigma^2 = \sum_{i=1}^{n}[(x_i-\mu)^2 P(x_i)] = (x_1-\mu)^2 P(x_1) + (x_2-\mu)^2 P(x_2) + \ldots + (x_n-\mu)^2 P(x_n)$$

For the # of H when we flip a coin twice discrete distribution:

$\mu = 1$

x	$P(x)$
0 (x_1)	$\frac{1}{4}$ ($P(x_1)$)
1 (x_2)	$\frac{1}{2}$ ($P(x_2)$)
2 (x_3)	$\frac{1}{4}$ ($P(x_3)$)

$\sigma^2 = (x_1-\mu)^2 \cdot P(x_1) + (x_2-\mu)^2 \cdot P(x_2) + (x_3-\mu)^2 \cdot P(x_3)$

$\sigma^2 = (0-1)^2 \cdot P(0) + (1-1)^2 \cdot P(1) + (2-1)^2 \cdot P(2)$

$\sigma^2 = (-1)^2 \cdot (1/4) + (0)^2 \cdot (1/2) + (1)^2 \cdot (1/4)$

$\sigma^2 = 1/4 + 0 + 1/4$

$\sigma^2 = 1/2 \quad \rightarrow \quad \sigma \approx 0.7071$

5: Probability Distributions (Discrete Variables)

5.2 Mean and Variance of a Discrete Random Variable

$$\sigma^2 = \sum_{i=1}^{n}[x_i^{\,2}P(x_i)] - \mu^2 = [x_1^{\,2}P(x_1) + x_2^{\,2}P(x_2) + \ldots + x_n P(x_n)] - \mu^2$$

Alternate Formula

For the # of H when we flip a coin twice discrete distribution:

$$\sigma^2 = x_1^{\,2} \cdot P(x_1) + x_2^{\,2} \cdot P(x_2) + x_3^{\,2} \cdot P(x_3) - \mu^2$$

$$\sigma^2 = [0^2 \cdot P(0) + 1^2 \cdot P(1) + 2^2 \cdot P(2)] - 1^2$$

$$\sigma^2 = [0 \cdot (1/4) + 1^2 \cdot (1/2) + 2^2 \cdot (1/4)] - 1$$

$$\sigma^2 = 0 + 1/2 + 4/4 - 1$$

$$\sigma^2 = 1/2 \quad \rightarrow \quad \sigma \approx 0.7071$$

$\mu = 1$

	x	$P(x)$	
x_1	0	$\frac{1}{4}$	$P(x_1)$
x_2	1	$\frac{1}{2}$	$P(x_2)$
x_3	2	$\frac{1}{4}$	$P(x_3)$

5: Probability Distributions (Discrete Variables)

5.3 The Binomial Probability Distribution

Bi means two like bicycle

An experiment with only two outcomes is called a Binomial exp.
Call one outcome *Success* and the other *Failure*.
Each performance of expt. is called a trial and are independent.

$$P(x) = \underbrace{\frac{n!}{x!(n-x)!}}_{\text{num}(x \text{ successes})} \underbrace{p^x (1-p)^{n-x}}_{P(x \text{ successes and } n-x \text{ failures})} \qquad x = 0,\ldots,n \qquad (5.5)$$

Prob of exactly x successes

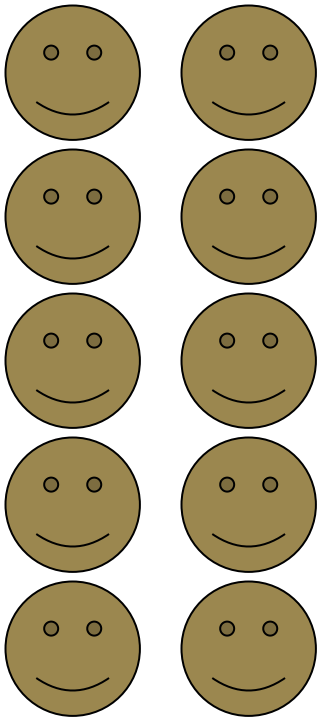
n = number of trials or times we repeat the experiment.
x = the number of successes out of n trials.
p = the probability of success on an individual trial.

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

5: Probability Distributions (Discrete Variables)

5.3 The Binomial Probability Distribution

Flip coin ten times.



$$P(x)=\frac{n!}{x!(n-x)!}p^{x}(1-p)^{n-x}$$

n = number of trials
x = the number of successes
p = the probability of success

n=10, x=7, p=.5

$$P(7)=\frac{10!}{7!(10-7)!}\left(\frac{1}{2}\right)^{7}\left(1-\frac{1}{2}\right)^{10-7}$$

$$P(7)=\frac{10\cdot 9\cdot 8\cdot \cancel{7!}}{\cancel{7!}3!}\left(\frac{1}{2}\right)^{7}\left(\frac{1}{2}\right)^{10-7}$$

$$P(7)=\frac{10\cdot 3\cdot \cancel{3}\cdot \cancel{2}\cdot 4}{\cancel{3}\cdot \cancel{2}}\left(\frac{1}{2}\right)^{10} \longrightarrow P(7)=\frac{120}{1024}$$

5: Probability Distributions (Discrete Variables)

5.3 The Binomial Probability Distribution

TABLE 2 **page 713**

Binomial Probabilities $\left[\binom{n}{x}\cdot p^x \cdot q^{n-x}\right]$ (continued)

n=10, x=7, p=1/2

$$P(x)=\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$

Figure from Johnson & Kuby, 2012.

n	x	p 0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99	x
10	0	.904	.599	.349	.107	.028	.006	.001	0+	0+	0+	0+	0+	0+	0
	1	.091	.315	.387	.268	.121	.040	.010	.002	0+	0+	0+	0+	0+	1
	2	.004	.075	.194	.302	.233	.121	.044	.011	.001	0+	0+	0+	0+	2
	3	0+	.010	.057	.201	.267	.215	.117	.042	.009	.001	0+	0+	0+	3
	4	0+	.001	.011	.088	.200	.251	.205	.111	.037	.006	0+	0+	0+	4
	5	0+	0+	.001	.026	.103	.201	.246	.201	.103	.026	.001	0+	0+	5
	6	0+	0+	0+	.006	.037	.111	.205	.251	.200	.088	.011	.001	0+	6
	7	0+	0+	0+	.001	.009	.042	.117	.215	.267	.201	.057	.010	0+	7
	8	0+	0+	0+	0+	.001	.011	.044	.121	.233	.302	.194	.075	.004	8
	9	0+	0+	0+	0+	0+	.002	.010	.040	.121	.268	.387	.315	.091	9
	10	0+	0+	0+	0+	0+	0+	.001	.006	.028	.107	.349	.599	.904	10

x	0	1	2	3	4	5	6	7	8	9	10
$P(x)$	$\frac{1}{1024}$	$\frac{10}{1024}$	$\frac{45}{1024}$	$\frac{120}{1024}$	$\frac{210}{1024}$	$\frac{252}{1024}$	$\frac{210}{1024}$	$\frac{120}{1024}$	$\frac{45}{1024}$	$\frac{10}{1024}$	$\frac{1}{1024}$

Questions?

Exam 1 on Tuesday!

Bring pencil, calculator, scratch paper.

(And caffeinated beverage!)

Will be given exam, formula sheet, binomial tables.