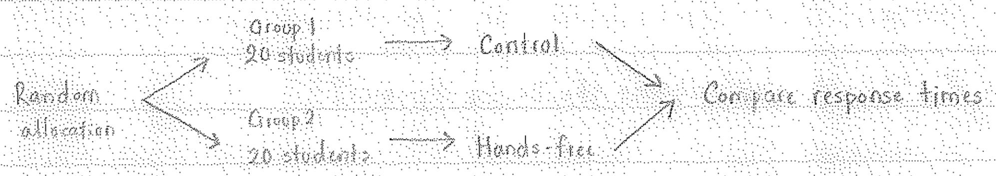
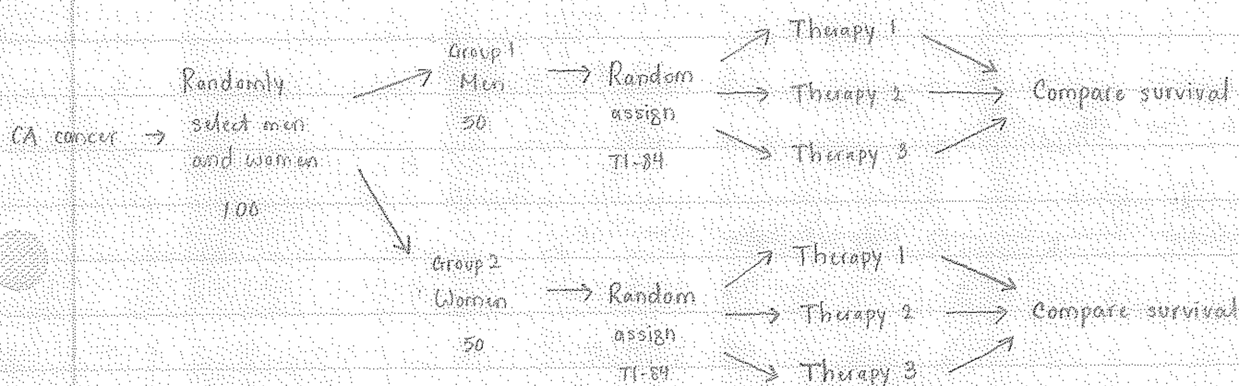
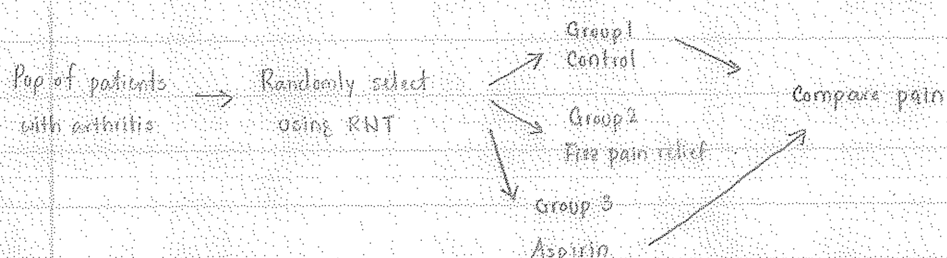
Do Now (09/23/11)

Randomly select 40 students from a population of all undergraduates. Randomly assign 20 to a control and 20 to the experimental group. Give the experimental group the hands-free device and nothing to the control group. Compare their response times.







① "Raise your hand if you believe the Beijing Olympics were better than Athens?"

- Verbal / peer influence, suggestive, x vs. x·4

② Write your opinion of "Do you enjoy watching the Beijing Olympics?"

- Way written suggests a Yes answer, leading

③ Please rate the 2008 Olympics

low satisfaction ⟶ high satisfaction

1 2 3 4

- Even number of choices, no unclear terms, evenly spaced

④ Log hours ____.

Strive for clear, concise questions

Make respondent feel comfortable and give truthful answers

- Fears
- Peers
- Response bias < squirrelly / deceiving

- Information can be presented quickly, efficiently, attractively
- Keep it simple and understandable
  - 7 secs to viewer
- Starts with a title

Bar Graph
- can see differences in data easily
  - good for comparing
- title
- label horizontal and vertical axes
- Keep steps the same and bars the same width
- can have double bar graphs
- zoom in for sharper comparison or deceptively exaggerate?
- option: put exact number inside bar
- Pareto charts vertical bar graphs arranged tall to small

Pie Graph
- good for comparing parts of a whole
- begins with disc divided into sectors
- put number or percentage inside sector
- pieces of pie cannot overlap

Line Graph
- good for showing change and trends
- title
- label axes
- keep steps (strides) congruent
- put dot where the data should go
- connect dots from left to right with straight edge
- can have dotted, dashed, multiple lines
- Time plots are special line graphs involving time (x-axis)

Pictographs

- rare but cute
- uses pictures / symbols to show info
- full symbol = amount set in key
  - partial symbol < amount in key
- maintain same symbol and same spacing
- keep it simple

Histograms

- Bar graph in which bars touch
- Width of bar has meaning (time, age, etc)
- Data grouped into classes
  - Useful with large amounts of data
- Overall shape assists in recognizing patterns
- Deviations from the norm
- Exact / specific examples are hidden
- How to choose # of classes? (5-10 is good)
  - Would info be hidden / lost
  - Will nothing stand out
- How wide?
  - $\frac{\text{largest} - \text{smallest}}{\#\text{ of bars}}$
  - If you get decimal, round up
- Begin with smallest value, then step by width
- Normal (symmetric)

  Rectangular

  Bimodal (peak, valley, peak)

  Skewed Right (peak on left)

  Skewed Left (peak on right)

Frequency Table

- Columns to organize raw data before graphing
- Has classes / groups ; tally marks ; frequency number ; class midpoints
- First class begins with smallest number (next class add width)

Class	Tally	Frequency	Relative Frequency
133 - 144	III	3	$3/12$
145 - 154	~~IIII~~	5	$5/12$
155 - 164	IIII	4	$4/12$

Analysis

- Center
- Shape
- Spread
- Outliers

0. _ _ _ _
1. _ _ _ _
2. _ _

Mean

- usually meant by "average"
- $\frac{\Sigma x}{n}$
- based on numeric values
- can be easily affected by outliers (pulled toward outlier)

$\bar{x}$ = sample average

μ = population average

Median

- middle value when numbers are ordered from small to large
- position is emphasized, not numeric values
- resistant to outliers

Mode

- most frequently occurring value
- overlooked, but often more appropriate
- bimodal or no mode
- outliers do not affect mode

Trimmed Mean

- mean that resists extremes
- eliminates pull of extremely low or high values in data set
- 10% trimmed mean → remove highest 10% and lowest 10%

PG. 88-98

(2) One number may not represent an entire set of #s well, so a cross reference is the measure of dispersion / variation / fluctuation / spread

Name 3 measures of variation

(3) (1) Range

(2) Standard deviation

(3) Variance

(4) Advantages and disadvantages of range

- Advantage ⇒ easily computed
- Disadvantage ⇒ does not tell how much values vary from one another or from mean

(5) Mesurement that helps us see how data is different from the mean is standard deviation.

(6) Why divide by n-1 sometimes and by N other times

n-1 is used to calculate sample standard deviation (s)

N is used to calculate population standard deviation (σ)

(7) Obtuse Ollie believes he discovered an easier method for ex 5 $\Sigma(8-5.5)^{2/4}$. Comment on his new formula

He used median (5.5) instead of mean.

(8) Compare and contrast sample and population formulas for the mean and standard deviation

$\sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$ = sample standard deviation

$\sqrt{\frac{\Sigma(x-\mu)^2}{N}}$ = population standard deviation

2 , 4 , 6 , 8 , 10

standard deviation

x	$x - \bar{x}$	$(x - \bar{x})^2$
2	-4	16
4	-2	4
6	0	0
8	2	4
10	4	16

$\bar{x} = \frac{30}{5} = 6$

$\Sigma = 40 / n-1 = \frac{40}{4} = 10$ square root : $\sqrt{10} = \boxed{3.16}$

x	$x - \mu$	$(x - \mu)^2$
19.8	-20.5	420.3
43.8	3.5	12.3
36.1	-4.2	17.6
52.4	12.1	146.4
63.1	22.8	519.8
20.7	19.6	384.2
46.3	6.0	36.0

$\mu = \frac{282.2}{7} = 40.3$

$\sigma = 1536.6 / 7 = \sqrt{219.5} = \boxed{14.81}$

Coefficient of Variation

- allows comparisons between populations

$$CV = \frac{s}{\bar{x}} \times 100 \quad \text{or} \quad \frac{\sigma}{\mu} \times 100$$

$$\frac{\sigma}{\mu} = \frac{14.82}{40.31} = 0.3677 \times 100 = 36.77\%$$

0 - 10%	Tight
13% - 30%	Low
33% - 65%	Medium
67% - 95%	High
100% +	Wild

Quartiles

1st Quartile → Q_1 = 25th percentile

2nd Quartile → Q_2 = 50th percentile = median

3rd Quartile → Q_3 = 75th percentile

① Arrange # small to large

② Median (Q_2)

③ Median of groups left and right of median (Q_1 and Q_3)

10 11 12 13 16) (19 20 21 22 23

Median: 17.5, $Q_1 = 12$, $Q_3 = 21$ Interquartile Range: 21-12 = 9, Range = 13

= spread

Box and Whisker

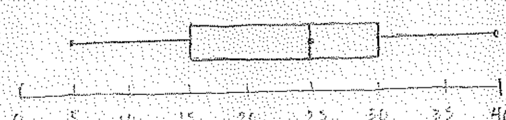
① # line

② Median

③ Lower Quartile, Upper Quartile

④ Small, Large

⑤ Draw box and line segment



Correlation ⇒ study variables simultaneously to find degree of relation

Regression Problems ⇒ variables used to predict how one variable affects another

Scatter Plot ⇒ display data in situations of 2 numbers for each item
- Bivariate / paired data

- +/- Association
- Low / Med / High Correlation
- Influential Observations (↔), Outliers (↕)
- Trends
  - As the class' wingspans increase, so does its heights

- Florida Boat Registration v. Manatee Deaths

~~85~~	~~92~~
~~98~~	~~85~~
~~78~~	~~97~~
~~82~~	~~91~~
~~88~~	~~93~~
~~91~~	~~86~~
~~94~~	~~92~~
~~95~~	~~94~~
~~97~~	95
~~91~~	~~93~~
~~96~~	~~92~~
~~80~~	
99	
~~92~~	

Is there a relationship between x and y?
How strong is the relationship?

Least Square Line: $y_p = a + bx$

$$b = \frac{SSxy}{SSx} \qquad a = \bar{y} - b\bar{x}$$

$$SSxy = \Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n}$$

$$SSx = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$$

① Organized table

$$\begin{array}{cccc} x & x^2 & y & xy \\ \downarrow & \downarrow & \downarrow & \downarrow \\ \Sigma x & \Sigma x^2 & \Sigma y & \Sigma xy \end{array}$$

② Find SSx
③ Find $SSxy$
④ Find $\bar{x}$
⑤ Find $\bar{y}$
⑥ Find b
⑦ Find a
⑧ $\hat{y} = a + bx$
⑨ Graph it

x	y
$\bar{x}$	$\bar{y}$
0	a
x	y_p

$\bar{x} = 15.02$

$\bar{y} = 8.46$

$\Sigma xy = 618.32$ $\qquad \Sigma x = 75.10$

$\Sigma x^2 = 1224.83$ $\qquad \Sigma y = 42.30$

$(\Sigma x)^2 = 5640.01$

$b = -0.1759$

$a = 11.10$

If residual plot has a PATTERN, then the least square regression line is NOT a good representation of relationship between x and y.

If residual plot does NOT PATTERN, then x and y ARE linear.

Karl Pearson (England)(1857-1936)

Correlation coefficient $(r) = \frac{SSxy}{\sqrt{SSx \cdot SSy}}$ $\quad -1 \le r \le 1$

r	
1.0	positive association and perfect correlation
-1.0	negative association and perfect correlation
0	no association and no correlation
0 - 0.25	low correlation
0.3 - 0.65	moderate correlation
0.67 - 0.99	high correlation

Coefficient of Determination $= r^2$ $\quad 0 \le r^2 \le 1$

if $r^2 = 0.25$, then 25% of y is explained by the x

$r^2 = 0.4489$ 45% of variation in GPA is explained by variation in IQ

55% of variation in GPA is left unexplained

→ what else affects GPA?

(6) $\bar{x} = 6.39$

$\bar{y} = -2.49$

$SS_{xy} = 3.93$

$SS_x = 30.41$ $\qquad r = \frac{SS_{xy}}{\sqrt{SS_x \cdot SS_y}} = \frac{3.93}{\sqrt{30.41 \times 72.85}} = 0.0835$

$SS_y = 72.85$ $\qquad r^2 = 0.0069$

$b = 0.1292$

$a = -3.3156$

out $(6.2, 3.6)$

low correlation

0% of change in y is explained by variation in x

Probability $\Rightarrow$ a measure of chance

$P(A)$ = probability of event A "P of A"

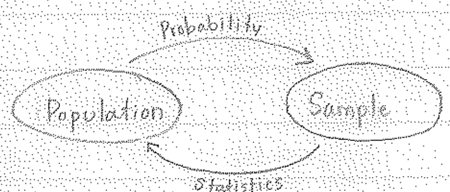
$$\text{Probability} = \frac{\#\text{ of desired results}}{\text{total }\#\text{ of results}} = \frac{f}{n} \qquad 0 \le p \le 1$$

impossible $\le p \le$ certain

$P(\text{not } A)$ = probability of not getting event A "Complement of an event"

$P(A') = 1 - P(A)$

Population is known and ask about sample



Independent Events

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

$P(5 \text{ and } 5) = P(5,5) = P(5) \cdot P(5) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} = 0.02\overline{7}$

Dependent Events

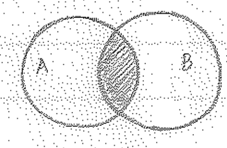
$P(A \text{ and } B) = P(A) \cdot P(B/A) = P(A) \cdot P(B)^*$ no replacement

$P(\text{ace and ace}) = P(\text{ace}) \cdot P(\text{ace})^* = \frac{4}{52} \cdot \frac{3}{51} = \frac{12}{2652} = 0.004525$

Compound Events

AND

"both"



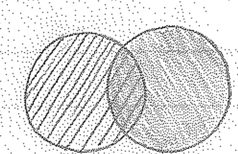
$P(A \text{ and } B) = P(A \cap B)$

Independent: $P(A) \cdot P(B)$

Dependent: $P(A) \cdot P(B|A)$

OR

"either"



$P(A \text{ or } B) = P(A \cup B)$

Disjoint: $P(A) + P(B)$

Non-disjoint: $P(A) + P(B) - P(A \text{ and } B)$

Probability of matching

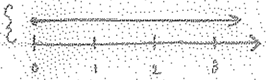
P(Bl, Bl) or P(Blu, Blu) or P(Br, Br)

Black = 10   Brown = 7   Blue = 8   Purple = 2   Green = 2   Red = 4   Stripe = 1

6.1

Discrete Random Variable $\Rightarrow$ quantitative observations that are "countable" $\{-2, -1, 0, 1, 2, \dots\}$

Ex. number of students who voted; students who earned "A's"

Continuous Random Variable $\Rightarrow$ "countless" observations $\{$ 

Ex. temperature of room; time (depends)

Probability distribution $\Rightarrow$ all probabilities add up to 1

- Has a mean, standard deviation

$\mu = \Sigma x P(x)$ $\sigma = \sqrt{\Sigma (x-\mu)^2 P(x)}$

(expected value)

b $P(1^+) = 0.396 + 0.264 + 0.088 + 0.015 + 0.001 = 0.763$

d $\mu = \Sigma x P(x) = 1.253$

c $\sigma = \sqrt{\Sigma (x-\mu)^2 P(x)} = \sqrt{0.9406} = 0.9698$

x	$P(x)$	$xP(x)$	$x-\mu$	$(x-\mu)^2$	$(x-\mu)^2 P(x)$
0	0.237	0	-1.253	1.570	0.3721
1	0.396	0.396	-0.253	0.064	0.0253
2	0.264	0.528	0.747	0.558	0.1473
3	0.088	0.264	1.747	3.052	0.2684
4	0.015	0.060	2.747	7.546	0.1132
5	0.001	0.005	3.747	14.040	0.0140
					0.9406

$CV = \frac{\sigma}{\mu} = \frac{0.9698}{1.253} = 0.7740$

6.2

① Who was Jocob Bernoulli?

Swiss mathematician from the 1600s who studied binomial experiments extensively.

② Problems that have exactly 2 possible outcomes is called binomial / Bernoulli experiments.

③ Describe the central problem of a binomial experiment

Find the probability of r successes out of n trials.

④ T / F Bernoulli experiments work only with dependent situations.

Only work with independent trials.

⑤ Each faculty member at Pepperdine has asked about recommending which new car Mr. Micek should purchase

ⓐ # of trials n

ⓑ # of outcomes possible As many outcomes as possible

NOT binomial

$P(r) = C_{n,r}\, p^r q^{n-r}$

2 success out of 3 trial

$P(2) = C_{3,2}\, p^2 q^1$

$= 3(0.25)^2(0.75)$

$= 0.1406$

n= 73 p= 0.80

r= 62 q= 0.20

s = within 3 minutes

F = more than 3 minutes

Assume independent YES binomial

n = 51 p= no constant probability

r = 43 q=

s= enjoy

F= did not enjoy

NOT binomial

n= 20 p= 0.71

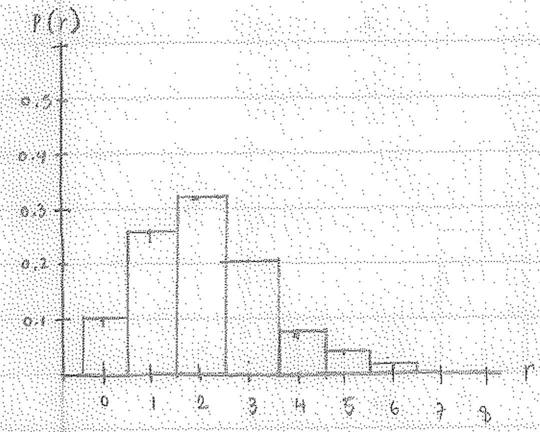
r = 18, 19, 20 q = 0.29

s = man would welcome woman

F = man would not welcome woman

Assume asked independently YES Binomial

6.3



Cntr:

Shp: skewed right

Sprd:

Out:

(5)

r	p(r)
0	0.100
1	0.267
2	0.311
3	0.208
4	0.087
5	0.023
6	0.004
7	0.000
8	0.000

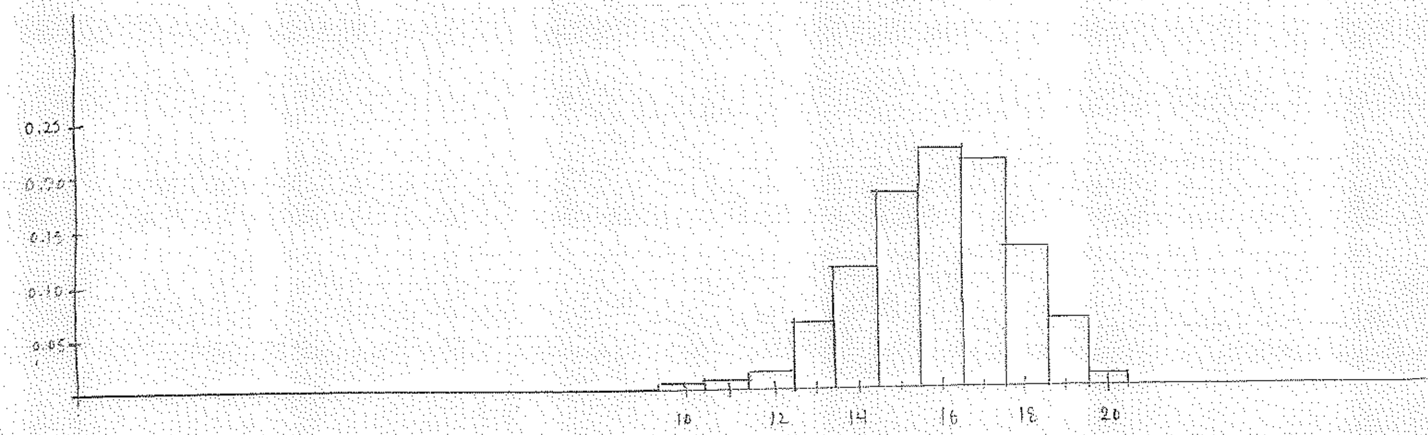
$\mu = np = 8(0.25) = 2$

$\sigma = \sqrt{\mu q} = \sqrt{2(0.75)} = 1.22$

$CV = \frac{1.22}{2} = 0.61$

20 students in class. 80% of students have passed. Pass or fail class.

1	0	9	0	17	0.205
2	0	10	0.002	18	0.137
3	0	11	0.007	19	0.058
4	0	12	0.022	20	0.012
5	0	13	0.055		
6	0	14	0.109		
7	0	15	0.175		
8	0.	16	0.218		



$\mu = np = 20(0.80) = 16$

$\sigma = \sqrt{npq} = \sqrt{16(0.2)} = 1.79$

$CV = \frac{1.79}{16} = 0.1118 = 11.18\%$