Masoud Nasari - mmnasari@math.carleton.ca
HP5250, 1-2pm Mondays.
5 x Assignments, 5% each
Midterm: 25% - good practice for final
Final 50% close book, cumulative

Assignments – computer component, software:MINITAB, written answer component.
Due Dates not yet determined

What is the definition of Statistics?
Branch of mathematics with applications in almost every facet of our lives.
Making educated decisions in the presence of uncertainty is what statistics does for us.

Variable: a characteristic that changes/varies over time and/or for different individuals or objects of consideration. Eg. Height of a person at this moment.

Experimental unit: individual or object on which the variable is measured

Data Value: (single measurement): value of the variable taken from a single measurement of one experimental unit.

Population – is perfect knowledge – the entire target group. Sets definitional boundaries – sets full area of study. It is seldom realistic to test or get data from the entire population. Set of all measurements of interest. Eg height of all Canadians is a population – each Canadian is an experimental unit. Each height is a Data Value. The entire set is the Population.

Sample is a (hopefully representative) subset of the population, something that can be compared to the population. Approximation/Best Guess…

Univariate data: results when a single variable is measured on an experimental unit. i.e. height.

Bivariate/Multivariate data: when two or more variables are measured over an experimental unit: I.e. measuring both height and gender or both weight and blood pressure. Often related or dependent.

Interval between a and b (a, b), includes all values BETWEEN (BUT NOT INCLUDING) a and b.

Quantitative variables : numbers – two types:

1. DISCRETE variable: assumes a finite or countable number of values:
    a. i.e. number of Canadians that will buy a car in 2014. Min = 0, max = 34×10^6
    b. i.e. number of times to flip a coin to get heads – infinite, but countable.
2. CONTINUOUS variable: assumes infinite (uncountable) # of values (interval, ratio scales)
    a. . i.e. a person's height or weight. $(0, \infty)$ - in general, DISTANCE-related or TIME-related variables are CONTINUOUS.

Descriptive stats – describes the data
Descriptive statistical tools:

Graphs:

1. Pie charts
2. Line charts
3. Stem and leaf charges
4. Histogram
5. Box plot

Numerical descriptions:

1. Mean
2. Median
3. Variance
4. Std deviation
5. Range
6. Stderr

inferential statistics (to be covered later) – used to make predictions or inferences from the data

Qualitative variables: nominal

What two types of information should a graph provide?

1. The measured values of variable of interest
2. How often the values occurred

Pie charts and Bar Charts (for qualitative/nominal variables) – use colour to distinguish different qualitative variables. (i.e. measure of party support in student population)

To produce pie or bar chart:

1. F: Frequency: # of times each value is observed
    a. F_s: Frequency of a given data point (i.e. number of times shoe size of 10 is measured)
2. RF: (Relative Frequency) F_s/total number of observations (n) (i.e. size 10 observed 6 times, total number of observations = 60, RF_s = 6/60 = 1/10
3. $\%_s = RF_s * 100$ = (i.e. 10%)
4. $Angle_s = RF_s * 360°$

Pie and bar charts for quantitative variables
Same idea as for qualitative
Eg.
F_s = Avg annual income for a category: $x.
RF_s = $x/$total of all annual incomes

Line charts for Time series data – allow you to discern (identify) a pattern or trend that will most likely continue to hold into the immediate future.
X axis → time
Y axis → data values

Stem and Leaf Plots

1. divide each measurement into two parts: part on left is stem, part on right is leaf
2. list the stems in a column from smallest to largest
3. record the leaves for each stem
4. order the leaves from lowest to highest

***include LEAF UNITS

Eg. The following 15 numbers represent the shoe size of 15 people:

34, 31, 30,
38, 30, 41,
42, 36, 43,
40, 37, 46,
45, 41, 39

3 and 4 become the stems (30 and 40):

3|0 0 1 4 6 7 8 9
4|0 1 1 2 3 5 6

Unit of leaf = 1
Unit of stem = 10

Frequency Histograms

1. Chose a number of bins between 5 and 12
    a. To calculate the number of bins to use, take the square root of the sample size
    b. ALWAYS ROUND UP. i.e. for a sample size of 37, the total number of bins should be 7.
2. Range of X axis = (largest measurement- smallest)
3. To determine the width of each bin: Range/# of bins - ALWAYS ROUND UP. i.e. if width calc = 1.65, round up to 1.7 or 2.0
4. If the measurements are discrete, and the sample size is small, then each distinct value can be taken as a bin (i.e. sample size of 12 or less)
5. Identify the boundaries of bins:
    a. first boundary of the first bin: the smallest measurement
    b. Second boundary of first bin = smallest measurement+ width of a bin
    c. First boundary of 2^{nd} bin = 2^{nd} boundary of first bin + width of a bin
    d. …
6. Determine the total number of samples that reside in each bin (frequency). Construct a statistical table, based on the frequency of each bin. Sum of frequencies should = total sample size (n)
7. Plot the RF for each BIN

Bin	Bin Boundary	Frequency	RF
1	Boundary 1 <Boundary 2	F_1	F_1/n
2	Boundary 2 <Boundary 3	F_2	F_2/n
3	Boundary 3 <Boundary 4	F_3	F_3/n
…			
Total		n	1

Eg. Number of Liters of milk purchased by 25 households. (Quantitative, Discrete)

Liters of milk:
0, 3, 5, 4, 3,
2, 1, 3, 1, 2,
1, 1, 2, 0, 1,
4, 3, 2, 2, 2,
2, 2, 2, 3, 4

Square root of 25 = 5. +1 = 6

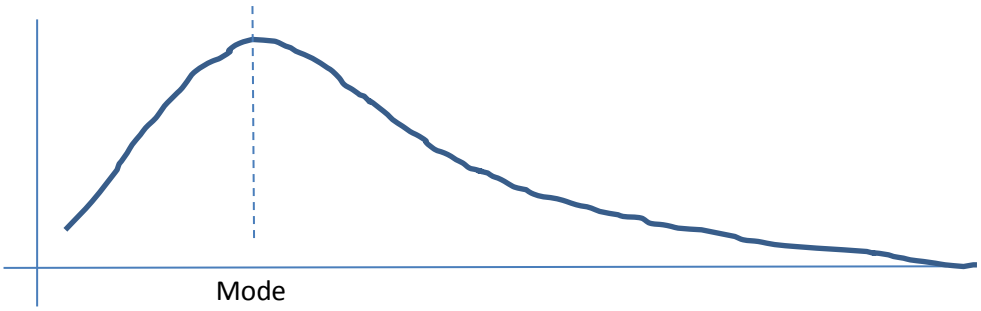
Bin	Bin Boundary	Frequency	R.F_0
0	No boundaries if each number represents a class	2	2/25
1		5	5/25
2		9	9/25
3		5	5/25
4		3	3/25
5		1	1/25
Total		25	25/25 = 1

What is LEFT INCLUSION in plotting histograms?
It means the left boundary of the bin is INCLUDED in the values plotted, but the RIGHT BOUNDARY IS NOT.

Assignment 1 to be posted this weekend – Assignments on Chapter 1 & 2

Symmetric distribution (graph): a distribution which forms a mirror image about the middle class/bin
Normal or bimodal

Right skew is a distribution or graph that the greater portion of measurements lie on the right hand side of the mode



Sample Variance = S^2

Compute Sample variance for the following data: (n=7)

	Xi	Xi^2	
	4	16	
	6	36	
	5	25	
	5	25	
	3	9	
	10	100	
	7	49	
Σ	40	260	

$s^2 = \frac{260 - \frac{40^2}{7}}{7-1} = 5.238$ --- recommended formula for SAMPLE VARIANCE

S^2 cannot be negative unless degenerate case (will not deal with this)

$S = \sqrt{s^2}$ = 2.28

$$\boldsymbol{s^2 = \frac{\Sigma(x^2) - \frac{(\Sigma x)^2}{n}}{n-1}}$$

Given the Frequency HISTORGRAM

(3,5) = 3

(5,7) = 3

(7,9) = 6

(9,11)= 5

(11,13) = 1

Σ = 18

PROPORTION of measurements between 3 and 7 (3,7) = 6/18 = 33%

PROPORTION of measurements up to and including 9 (-∞, 9) = 12/18 = 66%

Tchebyshev's THEOREM:

Tells us the proportion of measurements within an interval

$(\bar{x} - ks, \bar{x} + ks); k = 1,2,3$

k = number of STD DEVIATIONs from the MEAN

s = STD DEVIATION of the mean

$(\bar{x} - ks, \bar{x} + ks)$ = a given interval

$(\bar{x} - s, \bar{x} + s)$

$(\bar{x} - 2s, \bar{x} + 2s)$

$(\bar{x} - 3s, \bar{x} + 3s)$

Tchebyshev's theorem states that AT LEAST 1 -1/k^2 of the measurements lie inside the interval of interest.. this is applicable to ANY KIND OF DISTRIBUTION, regardless of shape

Tchebyshev's theorem: $\mathbf{1 - \frac{1}{k^2}}$ **samples will fall within *k* standard deviations of the mean**

k=1, AT LEAST **0% of the measurements will fall within -1 and +1 STDDEV**

k=2, AT LEAST **75% of the measurements will fall within -2 and +2 STDDEV**

k=3, AT LEAST ~= **89% of the measurements will fall within -3 and +3 STDDEV**

HOWEVER… EMPIRICAL RULE states that If the distribution is a normal distribution (bell shaped AND symmetric), then:

- ±1 STDDEV = 68%
- ±2 STDDEV ~= 95%
- ±3 STDDEV ~= 99.7%

If we don't know the shape of the distribution, we use Tchebyshev's theorem!

Relationship between Range and *s*: $s \approx \frac{Range}{4}$

Measures of relative stand:

z-score: a way to standardize all values into a measure of the number of standard deviations from the mean

$$z = \frac{x - \bar{x}}{s}$$

If |z| > 3 then it is an outlier

Percentiles

The p percentile is a number that is larger than p% of the measurements and smaller than 100%-p%

Eg. when you are in the 90th percentile, It means that 90% of the population are scoring below you.

Quartiles:

Q1 represents the upper bound of the quartile– lower quartile (25th percentile)
Q2 represents the upper bound of second quartile = 50th percentile – AT THE BOUNDARY (= median)
Q3 represents the upper bound of third quartile = 75th percentile

How to bin into quartiles?

1. Order data from lower to highest.
2. Position of Q1 = i = (n+1)*.25 – if this is an integer then Q1 = x_i sample
3. Position of Q2 = i = (n+1)*.5 – if this is an integer then = Q2 = x_i sample
4. Position of Q2 = i = (n+1)*.75 – if this is an integer then = Q3 = x_i sample

When you divide any number by 4, the remainder will be 0, 1, 2, 3
Remainder is 1 = .25
Remainder is 2 = .50
Remainder is 3 = .75

Eg. Multiply (n+1)*.25 = **4.25**
Find the just below measurement + (quartile)*(the just above-just below)

X_4 + 0.25(X_5-X_4) = Q1 ????

Eg, Find Q1 of the following set of measurements
11, 19, 0.1, -4. 9, 27, 2, 8, 30, 32, 27
ORDER them:

-4	0.1	2	8	9	11	19	27	30	32
X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}

n= 10
Use 0.25 to find Q1: 0.25(10+1) = 11/4 = 2.75
X_2 =0.1, X_3 =2.0
$Q1 = X_2 + 0.75(X_3-X_2) = 1.525$
= 0.1+0.75 (2-.01)
=.85*1.99
= 1.525

Use 0.75 to find Q3: 0.75(10+1) = 11/4 = 8.25
Q3 = 27+0.25 (30-27)
$X_8 + 0.75(X_9-X_8) = 27.75$

Boxplots
IQR = Q3-Q1
Lower Fence = Q1 – (1.5*IQR)
Upper Fence = Q3 + (1.5*IQR)
Minitab – measurements below lower fence, or above upper fence are considered outliers

The five-number summary is a descriptive statistic that provides information about a set of observations. It consists of the five most important sample percentiles:

1. the **sample minimum** (smallest observation)
2. the **lower quartile** or first quartile
3. the **median** (middle value)
4. the **upper quartile** or third quartile
5. the **sample maximum** (largest observation)

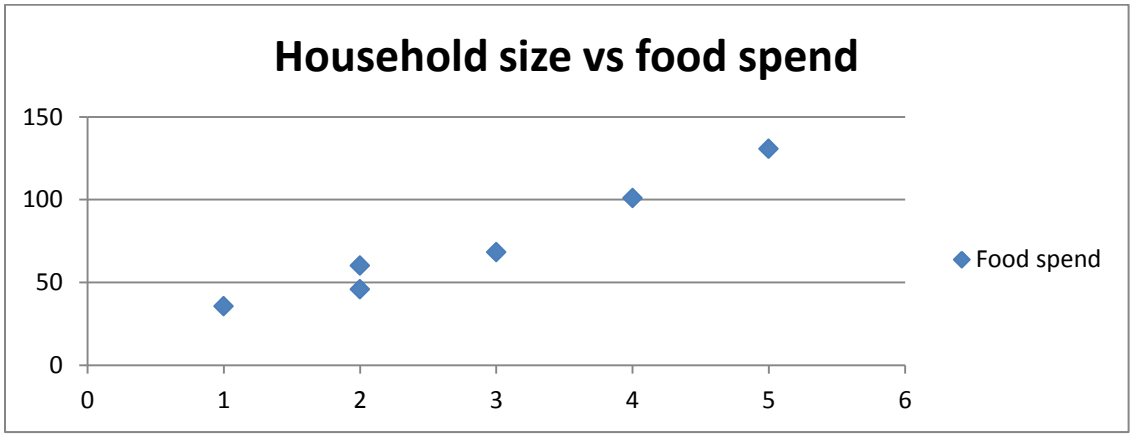
Chapter 3: Bivariate data

One way to represent independent data – two different independent variables: construct a pie chart for each quality, (i.e. weight and blood pressure.)

Another way: SCATTER PLOT – display QUANTITATIVE BIVARIATE measurements

Eg.
Let X be the number of household members
Let y be the amount of money each household spends on food

# of ppl	2	2	3	4	1	5
Food spend	45.75	60.19	68.33	100.92	35.56	130.62



In looking at a scatter plot determine the following:

1. Is there a pattern?
2. If there is a pattern, how strong is the pattern?
3. Are there outliers/unusual points?

Correlational coefficient: $\mathbf{r} = \frac{S_{xy}}{s_x s_y}$ important: = -1 <= $\mathbf{r}$ <= +1 (s = stddev)

If r >0 and <= 1 then increasing pattern
If r >= -1 and <0 then DECREASING PATTERN
If r = 0 : no correlation

Compute r for household example above:

	X_i	X_i^2	Y_i	Y_i^2	X_iY_i
	2	4	45.75	2093.06	91.5
	2	4	60.19	3622.83	120.38
	3	9	68.33	4737.56	204.99
	4	16	100.92	10189.84	403.68
	1	1	35.56	1.58283.01	35.56
	5	25	130.62	170061	653.1
SUM:	17	59	441.37	38565.23	1504.21

$\bar{y} = \frac{441.37}{6} = \ 73.52833 \qquad \bar{x} = \frac{17}{6} = \ 2.833$

Pearson's r coefficient $\qquad r_{XY} = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$

$\mathbf{r = \frac{1509.21}{\sqrt{59 * 38895.8}} = 0.99}$

$S_x = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x)^2}{n}}{n-1}} \qquad S_y = \sqrt{\frac{\sum y_i^2 - \frac{(\sum y)^2}{n}}{n-1}}$

$S_{xy} = \frac{1}{n-1}\left(\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}\right) \qquad r_{xy} = \frac{S_{xy}}{S_x S_y} \qquad b = r\,\frac{S_y}{S_x}$

$S_x = \sqrt{\frac{59 - \frac{(17)^2}{6}}{5}}$ **= 1.47196** $\qquad S_y = \sqrt{\frac{38895.8 - \frac{(441.37)^2}{6}}{5}}$ **= 35.85**

$S_{xy} = \frac{1}{5}\left(1509.21 - \frac{(17)(441.37)}{6}\right)$ **= 51.73**

$r_{xy} = \frac{51.73}{(1.47196)(35.83)} = .9809 \qquad b = .9809\ \frac{35.85}{1.47196} = 23.873$

$y = \bar{y} - \left(r\frac{S_y}{S_x}\right)\bar{x} + \left(r\frac{S_y}{S_x}\right)x$

$a = \ 73.52833 - (23.873)2.833$ **= 5.8961**

a= 5.7824, b=23.9103, r= 0.9802 Sx =1.47196 Sy=35.85

Bivariate data
Eg. Household – x = # of members, Y= amount of $ on food spent per week
r =.99 --- r = +1, then x & y exhibit an increasing linear relation
r = -1, then x & y exhibit a decreasing linear relation

LEAST SQUARE REGRESSION LINE - Patterns enable us to make near-future predictions
When X&Y exhibit a linear relation, then in this case we use a REGRESSION LINE to predict values of Y into the near future. (only where r is close to +1 or -1)

Parabolic relationship: $y = x^2$ CANNOT USE Least Squares Regression

To compute y, for a given value of x:
$\boldsymbol{y = a + bx}$

Where a = $\boldsymbol{\overline{y} - b\overline{x} \quad and \quad b = r\frac{s_y}{s_x} \quad so:}$ $\boldsymbol{y = \overline{y} - \left(r\frac{s_y}{s_x}\right)\overline{x} + \left(r\frac{s_y}{s_x}\right)x}$

For each X_i, the error is: $\boldsymbol{\sum_{i=1}^{n}(y_i - bx_i - a)^2}$

Y = 5.26 + 24.15x

CROSS-VALIDATION: Text your regression line against an actual value in your data set.
i.e. plug in x = 5, and you get 126.01 – the actual value is 130.62. very close – there is not much error

CHAPTER 4 probability

Probability and statistics are closely related.
In fact probability is used as a tool to evaluate the reliability of our inference about the population based on a sample

Experiment: is a process by which observations (measurements) are observed
The measurements could be numerical or non-numerical..
i.e. recording of a test grade or recording people's answers to the question, "Do you like math?"

SIMPLE EVENT: the result of repeating the event only one time. I.e. ONE toss of the dice
For example in rolling dice, all individual possibilities are simple events: E_1={1}, E_2={2}, E_3={3}, E_4={4}, E_5={5}, E_6={6}
EVENT: - any COLLECTION OF SIMPLE EVENTS

Eg. In rolling a dice experiment,
Let E be observing an even number, E = {2,4,6}
Let O be Observing an ODD number, O = {1,3,5}
Let A be a number between 2&4 inclusively, A = {2,3,4}

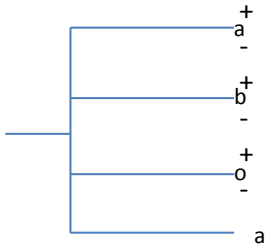
Set theory revisited
A = {1,2,3}
B ={1,2,6}

$A \cup B = \{1,2,3,6\}$
$A \cap B = \{1,2\}$

EVENT sets can be said to be MUTUALLY EXCLUSIVE if the INTERSECT is the EMPTY SET ($\varnothing$) - no common elements

SAMPLE SPACES: Collection of all sample events
Eg rolling a die: S ={1,2,3,4,5,6}
Eg Flipping a coin: S ={H,T}

Tree diagram for S{a+, a-, b+, b-, o+, o-, ab}



Probability of an event A – we say that A occurs when we at least one of its elements is observed. I.e. rolling a dice

$$P(A) = \lim_{n \to \infty} \left(\frac{f_a}{n}\right)$$

Method of computing quartiles….
To compute Upper quartile bound of Q1:
(n+1) * .25 = z.yy
$Q1 = X_z + yy(X_{z+1} - X_z)$

Can be used for 30th percentile, or 33rd percentile.

Every probability function should satisfy these two conditions:
(PROBABILITY FUNCTIONS)

1. For each event, A, P(A) is $\geqq 0$
2. If S is the sample space, P(S) = 1 (i.e. for flipping a coin, S={H,T}

For event A, if P(A) = 0, we say event A NEVER OCCURS
For a given event A, if P(A) = 1, then we say event A ALWAYS OCCURS

Evaluating the probability of an event A:
P(A) is the sum of the probabilities of its simple events.
The probability of an individual number (simple) event in rolling the dice is 1/6, because the sum of the sample space is always = 1
S={1,2,3,4,5,6} ∴ P(1)+P(2)+P(3)+p(4)+P(5)+P(6) = 1
The probability of rolling an even number: E={2,4,6} = P(2)+P(4)+P(6) = 1/6 +1/6 +1/6 = 3/6 = ½

Example, 5 simple events: S={E1,E2,E3,E4,E5}
Given the following probabilities of the simple events, we can calculate P(E1) because the sum of all simple events must = 1 (P(E1) = 2/10
P(E2) = 2/10
P(E3) = 1/10
P(E4) = 2/10
P(E5) = 3/10

EG… draw a card out of the standard deck (52 cards)
The probability that the chosen card is a king is 4/52
$A = \{K_s, K_h, K_d, K_c\}$
$P(A) = P(K_s) + P(K_h) + P(K_d) + P(K_c) = 4/52$

What is the probability of getting 5 cards of the same suit (FLUSH)?
4 suits, each suit has 13 cards

$$\frac{C_1^4 \, C_5^{13}}{C_5^{52}} = \frac{\frac{4!}{3!} * \frac{13!}{5!(8!)}}{52!/5!(47!)} = .00198$$

THE mn RULE

If the first part of the problem can be solved m ways and the 2nd part can be solved n ways, then the total number of solutions is m*n - i.e 2 dice, each die has 6 possible single throws, therefore the total number of combinations is 6*6 = 36. The probability of any single combination is 1/36

EG – what are all the possible ways that you can get from Montreal to Ottawa to Toronto
Bus, Fly, Car, Train to Ottawa (4 ways)
Train or Drive to Toronto (2 ways)
4*2 = 8 possible ways to get from montreal to Ottawa
If the final destination is Vegas, and there are 3 ways to get to Vegas from Toronto, then the total number of ways to get to Vegas from Montreal is 4*2*3 = 24

Flip a coin 10 times = how many different outcomes can you expect?
2*2*2…. (2^{10}) = 1024

Assume I have N people and I want to arrange them in a set of rooms (without replacement)
In the first room, I could put from 1 to N people in the rooms, so I have N options
In the second room I have N-1 options,
In the third room, I have N-2 options,
…. In the last room I have 1 option
(N)(N-1)(N-2)(N-3)….1 = N!
5! = 5x4x3x2x1 = 120

The EXTENDED mn RULE:

If an experiment is performed in k stages, with n_1 ways to accomplish the first stage, and n_2 ways to accomplish the second stage, …. And n_k ways to accomplish the kth stage, then the number of ways to accomplish the experiment is $n_1 n_2 n_3 \ldots n_k$

PERMUTATIONS and COMBINATIONS

Permutation: $P_r^n = \frac{n!}{(n-r)!}$, where r <= n $\quad special\ case: P_n^n = n!$

total number of ways that I can choose r out of n objects **when ORDERING MATTERS**

Combination: $C_r^n = \binom{n}{r} = \frac{n!}{r!(n-r!)}$, where r <= n

Total number of ways that I can choose r out of n objects when **ORDERING DOES NOT MATTER**

Eg. 50 people, n= 50
I have 3 positions to fill (CEO, ASSISTANT, SECRETARY), r = 3 (position matters)
How many ways can I fill the three positions?
1st position : 50 ways
2nd position: 49 ways
3rd position: 48 ways
Total probability = 50*49*48, but the easiest way to calculate this is to calculate 50!/(50-3)!

Eg Build team of 3 people out of 50 people
In 3! Ways I can reorganize the 3 people I selected, the team will not change (position does not matter)
Therefore the total number of ways that I can build a team is 50x49x48/3!
Therefore Total probability = 50!/3!(50-3)!

52 cards – choose 5, what is the probability of getting them all from the same suit? (flush)
Ordering does not matter
13/52, 12/52,11/52,10/52,9/52
52 cards in a deck, 13 cards in a suit, 4 suits

5 cards from 13 in a suit
5 cards from 52 in a deck
1 suit out of 4

$$\frac{\binom{4}{1}\binom{13}{5}}{\binom{52}{5}}$$

EVENTS and the RELATIONSHIPS between them
When A⋂B = ⌀ then we say that A&B are MUTUALLY EXCLUSIVE.
The COMPLEMENT of an event A, is the set of all elements that are NOT A: A′ (A PRIME)

We say that events A and B are INDEPENDENT if: P(A⋂B) = P(A)*P(B)
i.e. probability of a coin flip being heads twice in a row – each event is Independent – each event has an individual probability of 50%
i.e. PERSON chosen at random:

- Let A be the event that a person's favourite food is STEAK
- Let B be the event that the person's liberal party is LIBERAL

(these two are not dependent on each other – they do not provide any information about each other)
∴ A & B are independent
Let C be the event that the person is COLOURBLIND
Let D be the event that the person is MALE
C and D are INTERRELATED – they are NOT INDEPENDENT

If you try to predict the probability of one event based on the probability of the other, unrelated events tell you nothing in the prediction. DEPENDENT events have more impact on predicting outcome.

<u>THE ADDITION RULE of PROBABILITY:</u> P(A⋃B) = P(A) + P(B) – P(A⋂B)
A rule for evaluating the probability of a UNION of two events (A⋃B) = (probability of A OR B happening) – probability of an event that falls in the area covered by: all members of A and all members of B.

SPECIAL CASE: If A and B are MUTUALLY EXCLUSIVE: (no overlap): **(A⋃B) = P(A) + P(B)**
: $P(\varnothing) = 0$

<u>CONDITIONAL PROBABILITIES</u>
P(A|B) → probability of **A GIVEN B**
i.e. When B is given, what is the probability of A?

$P(A|B) = \frac{P(A \cap B)}{P(B)}$; when P(B) ≠ 0 - If the probabilities are MUTUALLY EXCLUSIVE, P(A|B) = 0

If A and B are INDEPENDENT EVENTS, then P(A|B) can be written as P(A) * P(B)/P(B), therefore
$P(A|B) = \frac{P(A)*\cancel{P(B)}}{\cancel{P(B)}}$ = P(A) --- knowing B doesn't help you at all to evaluate P(A)

In general P(A|B) is NOT THE SAME as P(B|A)!

MULTIPLICATION RULE:
P(A⋂B) = P(A|B) * P(B)
P(B⋂A) = P(B|A) * P(A)

Eg: roll dice and flip coin simultaneously in a FAIR toss
S(A)={1,2,3,4,5,6}
S(B)={H,T}
S (A∪B)={1T,2T,3T,…..6T,1H,2H….6H}
TOTAL probability of any given S = 1, so probability of any simple event in S(A∪B) = 1/12

Let P(A) be the probability of the dice coming up 1
Let P(B) be the probability of the coin coming up heads
P(A) = 1/6
P(B) = ½
P(A)*(PB) = 1/12

What is the probability that the DICE shows LESS THAN 3 **AND** Coin shows HEADS?
(**AND** means INTERSECT: (A⋂B))
Let P(A) be the probability of the dice coming up less than 3 = 2/6
Let P(B) be the probability of the coin coming up HEADS = 1/2
P(A) = 2/6
P(B) = ½
P(A)*(PB) = 2/12 = .167

What is the probability that the DICE shows LESS THAN 3 OR the COIN shows HEADS?
(OR means UNION: (A∪B)
P(A) + P(B) – P(A⋂B) = (2/6 + ½) – 2/12 = 4/12 + 6/12 – 2/12 = 8/12

If you know the coin showed HEADS, probability of dice being less than 3 (P(A|B)) is still 2/6 because A and B are independent.

BAYE'S RULE/THEOREM (using CONDITIONAL COEFFICIENTS):

Law of total probability

Eg. Partition sample space S for rolling dice, into 3 partitions (Cartesian)

S1 ={1,2}, S2={3}, S3={4,5,6}

Si ⋂ Sj = ⌀, for each i and j

$S_1 \cup S_2 \cup S_3 = S = 1$

If we select a subset of all the samples and call it A, then we create three intersects:

$A \cap S_1$

$A \cap S_2$

$A \cap S_3$

We can calculate the probability of A, then by looking at the probabilities of the intersects:

$P(A) = P(A \cap S_1) + P(A \cap S_2) + P(A \cap S_3) \ldots + P(A \cap S_K)$

$= P(A|S_1)*P(S_1) + P(A|S_2)*P(S_2) + P(A|S_3)*P(S_3) \ldots P(A|S_k)*P(S_k)$

$= \sum_{i=1}^{k} P(A|S_i) * P(S_I)$

Bayes theorem: Let $S_1, \ldots S_k$ be a partition of S. With PRIOR PROBABILITIES $P(S_1) \ldots P(S_K)$, UNION of all prior probabilities = S. Which means sum of the probabilities should be equal to SUM of probabilities of S, P(S) = 1

$$\sum_{i=1}^{k} P(S_i) = P(S) = 1$$

Then for any event A in S, $P(Sj \mid A) = \frac{P(A|S_j)P(S_j)}{\sum_{i=1}^{k}(P(A|S_i)*P(S_j))}$

Eg. IF 50% of the students are male, and 40% of male students are smokers, and 30% of female students are smokers

a) What is the probability that a student chosen at random is a smoker?

Sample set is partitioned into MALE/FEMALE

S_1 be the subset that is male (M) = event of observing a male = ½

S_2 be the subset that is female (F) = event of observing female = ½

S_m = subset that are smokers (an intersect of the S_1 and S_2) =

Probability of the person being a smoker if you know they are male: $P(S_m|S_1)$ = 40% or .4

Probability of the person being a smoker if you know they are female: $P(S_m|S_2)$ =30% or .333

Probability that a person chosen at random is a smoker:

$P(S_m) = P(S_m \cap S_1) + P(S_m \cap S_2)$

$= (P(S_m|S_1) * P(S_1)) + (P(S_m \mid S_2) * P(S_2))$

$= (.4*.5) + (.333*.5)$

If one person is chosen at random, and someone tells you the chosen person is a smoker, what is the probability that the person is male?

$P(S_1 | S_m)$ ---- have to use BAYES theorem to calculate it

$$\frac{P(S_m|(S_1)* P((S_1)}{P(S_m|(S_1)*P((S_1) + P(S_m | S_2)*(P(S_2)} = \frac{.4 * .5}{(.4*.5) + (.3*.5)} =$$

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

Cabinet problem:

Cabinet 1		Cabinet 2		Cabinet 3	
Gold coin	Gold coin	Gold Coin	Silver Coin	Silver Coin	Silver Coin

Open a drawer in a cabinet at random
You observe a gold coin.
Choose a drawer at random
What is the probability that the second drawer (in that cabinet) also contains a gold coin?

Probability of choosing a gold coin at random is 1/6, therefore P(A) = 1/6

If we kept the first coin, then there are only 5 drawers left to choose from and the probability of B is still independent, so P(B) = 2/3

DISCRETE RVs are QUANTITATIVE RVs which assume finite or countable number of values
CONTINUOUS RVs are QUANTITATIVE RVs which are infinite and uncountable… if a RV is not discrete, it's continuous.

Random Variable r.v. sets
A variable X, is a RANDOM VARIABLE if the value it assumes, corresponding to outcome of an experiment, are pure chance or a random event.
Examples of random variables:
i.e. ROLLING DICE experiment. Let X be the appearing number. $S_x=\{1,2,3,4,5,6\}$
I.e. choose a person at random and measure their height. Let X be the height of the chosen person. $S_x=(0,\infty)$
i.e. Lifetime of an LED TV . Let be the lifetime of a chosen LED TV, $S_x=[0,\infty)$
I.E.. # of Canadians who will buy a car next year. $S_x=\{0,1,\ldots.,33x10^6\}$ (Poisson RV}

The DISTRIBUTION FUNCTION of a DISCRETE RV, P(x), is a graph, table, or formula that gives the probability related to each possible value of the discrete RV of x.
2 defining Properties of a DISTRIBUTION FUNCTION P(x):

1. **For each value of x, P(x) ≧ 0**
2. **ΣP(x) = 1**

EG. Flip a coin (Bernoli random variable)
Let X be 1 if you observe Heads
Let X be 0 if you observe Tails
If you are given the probability of 2/3 for 0 and 1/3 for 1, What is the probability distribution?

Table definition:

x	0	1	
P(x)	2/3	1/3	= 1

Satisfies the definition of a **DISTRIBUTION function because each value of X is >=0 and sum of P(x) = 1**

Formula definition: $P(x) = (1/3)^x * (2/3)^{1-x}$; x=0,1
P(0) = 2/3
P(1) = 1/3

Eg. Rolling a dice

x	1	2	3	4	5	6
P(x)	1/12	1/6	2/6	1/12	2/12	?
	1/12	2/12	4/12	1/12	2/12	= 2/12

b) What is the probability given the above table, that random variable X assumes a value > 1?
P(x>1) = P(2) +P(3) + P(4) + P(5) + P(6) = 1 - P(1) = 11/12

c) What is the probability of observing a number >= 2 and <= 4?
P(2)+(P3)+(4) = 7/12

POPULATION (Expected value of X) in a PROBABILITY DISTRIBUTION:
$\mu = E(x) = \Sigma x * P(x)$

POPULATION VARIANCE:
$\sigma^2 = \Sigma (x-\mu)^2 P(x)$ >= 0 (cannot be negative)
EASIER ALTERNATIVE for POP VARIANCE: **$\Sigma(x^2 * P(x)) - \mu^2$**

Eg. 4.82 from text
X: = # of laptops sold in one day
Probability Distribution is given as

X	0	1	2	3	4	5
P(x)	.1	.4	.2	.15	.1	.05
P(x)x	0	.4	.4	.45	.4	.25
$P(x)x^2$	0	.4	.8	1.35	1.6	1.25

Sum of (P(x)x) = 1.9
Sum of $(P(x)x^2)$ = 5.4
How many laptops can they expect to sell in any given day?

$\mu = E(x) =$ sum of $(x*px) = 1.9$ laptops per day.
What is the Variance? $\sigma^2 = \text{Sum}(x^2P(x)) - \mu^2 = 5.4 - (1.9)^2 = 1.79$
What is the standard deviation of X? $\sigma = \text{sqrt}(1.79) = 1.34$

Is it likely or unlikely to sell 5 laptops per day?
Must evaluate if the number is within +- 2 σ
2 σ = 2x1.34 = 2.68
Is 5 greater within (1.79-2.68) to (1.79+2.68)? No. It is an outlier.

Chapter 1- 4 for Mid Term
Further info for Chapter 4 which is not included in the book.

DeMorgan Law

1) $(A \cap B)^C = A^C \cup B^C$
2) $(A \cup B)^C = A^C \cap B^C$

$P(A^C \cap B^C) = P(AUB)^C = 1-P(AUB)$
$P(A^C \cup B^C) = P(A \cap B)^C = 1-P(A \cap B)$

1 . $P(A' \cap B) = P(B) - P(A \cap B)$
$P(A \cap B') = P(A) - P(A \cap B)$

2. $P(A' \cup B) = 1- P(A) + P(A \cap B)$

3. $P(A \cup B') = 1- P(B) + P(A \cap B)$

4. $P(A' \cup B') = P[(A \cap B)'] = 1 - P(A \cap B)$

5. $P(A' \cap B') = P[(A \cup B)'] = 1 - P(A \cup B)$.

EG X : -1.5, 0, 1, 2 ← X is discrete random variable
P(X) : 0.1, 0.4, 0.2, 0.3 (sum of probabilities = 1)

$E(X) = \text{SUM}(X*P(X))$
$E(X^2) = \text{SUM}(X^2*P(X))$
$\sigma^2 = E(X^2) - \mu^2$

You can take any function g(X) and calculate expected value = SUM(g(X)*P(X))
I.E. E(2X+1) = (2(-1.5)+1)*0.1 + (2(0)+1)*.4 + (2(1)+1)*.2 + (2(2)+1)*.3

P(X <= 1) = P(1) + P(0) + P(-1.5) = .2 +. 4 +. 1 = .7
P(X < 1) = P(X <= 0) = P(0) + P(-1.5) = .4 +. 1 = .5
P(X=1) = .2 BUT THIS IS also equal to P(X <= 1) - P(X < 1)
If I am interested in X = A, and I am given P(X<=A) and P(X < A), then I can calculate P(X=A)

Chapter 5 - (for final exam)

BINOMIAL EXPERIMENT: (must satisfy these 5 conditions)

1. **Consists of N identical trials (i.e. flip the same coin 10 times**
2. **The result of each trial is either the Success (S) or Failure(F)**
3. **The probability of success (0<P<1) remains the same for each trial**
4. **Trials are independent**
5. **We are interested in the NUMBER OF SUCCESSES (S), which we represent by X{0,1,2, 3,,,n}**

Eg. Shoot a gun 10 times independently at a target. Each attempt may result in a HIT the target with a probability of success P=0.1. X= the number of times that the person hits the target. X is a binomial random variable

Eg. 100 bulbs, 20 of them are defective, a sample size of 5 from this box results in the following probabilities (20/100, 19/99, 18/98, etc…) – this is NOT a binomial experiment because the sampling is done WITHOUT REPLACEMENT. This is HYPER-GEOMETRY – p of success changes from trial to trial

Same light bulb experiment, WITH replacement, P(Defective) = 20/100 = 2/10 = 1/5
Now probabilities are the same for each trial. This is now a binomial experiment.

However, if the population you are taking the sample from is huge, and your sample size is small, then the probability of observing redundancy, becomes really low, and Without replacement, you can have **ALMOST binomial IF: Sample size/population size is < 0.05**

If Sample size/population size is >= 0.05 then it is no longer binomial

Probability distribution of a BIONOMIAL RANDOM VARIABLE

$$p(k) = \binom{n}{k} p^k q^{n-k} \qquad q = 1 - p$$

P(X=k)
X has a binomial distribution (n, p)
n = number of trials
p = probability of success
q = 1-p = probability of failure

EG: Flip coin 10 times. The probability of observing H is 1/5.
Let x = # of H in 10 times of flipping the coin

$$p(k) = \binom{10}{k} \left(\frac{1}{5}\right)^k \left(\frac{4}{5}\right)^{10-k}$$

What is the probability that exactly 3 heads will be observed?

$$p(x = 3) = \binom{10}{3} \left(\frac{1}{5}\right)^3 \left(\frac{4}{5}\right)^7$$

0!=1

MEAN and VARIANCE of a BINOMIAL DISTRIBUTION (n,p)
Mean of x = E(x) = np
Variance of X = σ^2 = npq
Std dev = $\sqrt{\sigma^2}$

CUMULATIVE DISTRIBUTION FUNCTION at k : P(X<=k)
Eg. P(x<=k) = $\sum_{x=0}^{k} P(x)$
= P(0) + P(1) +P(2) + P(3)…. + P(k)
This is true for any kind of distribution, including the binomial distribution

P(x<=2) = P(0)+P(1)+P(2)
P(x<2) = P(X<=1) = P(0) + P(1)
P(X=2) = P(X<=2) - P(X<=1)

COMPLEMENT OF (X <= y) is 1- (X < y)

Time consuming to compute cumulative distributions, so tables have been created for cumulative Distribution functions n=2 to n=25

I,e. Bin (3, 0.3) n = 3, p=.3
Probability in any given table cell is the probability **that X <= given k value -** to get an exact value, you have to subtract two values from the table.

If X has a binomial distribution of (8, 0.95), what is the probability that x<=7?
P(x<=7) = 0.337 (from table)
P(x>2) = 1-P(x<=2) = 1 - 0000 = 0.

If X has binomial distribution of (8, 0.4)
P(x>=3) = 1- P(X<=2) = 1 - .315 =
Cumulative DISTRIBUTION TABLES – gives cumulative probabilities for binomial distributions
P(x=k) = P(x≤k) – P(x≤k-1) also same as P(x=k) = P(x≤k) – P(x<k) aslo P(x=k) = P(x≤k) – P(x≤k$^-$)

Eg. X	-1.5	0.5	1
P(x)	.1	.6	.3

Shows that there is no guarantee that the X value will increment in units of 1.
Therefore P(x=0.5) = P(x ≤ 0.5) – P(x ≤ 1.5)

Poisson distribution

A random variable that is associated with a certain event in a specific location.
i.e. number of accidents at an intersection at a certain time of day
number of earthquakes in Ottawa in a year
number of cases coming into the emergency ward on a Saturday night

When X counts the number of events in a period of time and/or space so that the average, μ of these events are expected to happen, then X has a Poisson(μ)

Probability distribution of Poisson (μ)

PMF – Probability MASS Function $P(x) = \frac{e^{-\mu}\mu^k}{k!}$, $k = 0, 1, 2, 3, \ldots$

Poisson is a discrete random variable
e = 2.71828

eg. If the # of car accidents at the intersection of Carleton U and colonel by has poisson distribution with mean or average of 6 accidents per year.

a) What is the probability that next year, there will be exactly 5 accidents?
   $P(5) = \frac{e^{-6}6^5}{5!}$
b) What is the probability that there will be at most 2 accidents next year?

   $P(x \le 2) = P(0) + P(1) + P(2) = \frac{e^{-6}6^0}{0!} + \frac{e^{-6}6^1}{1!} + \frac{e^{-6}6^2}{2!}$

   (POISSON DISTRIBUTION TABLES IN BOOK provide these numbers)

c) What is the probability of having exactly 1 car accident in the next **6 MONTHS**?
   Divide **μ/2 = 6/2 = 3**

   P(x=1) $= \frac{e^{-3}3^1}{1!}$

Mean and variance of a V.V. X which has poisson (μ):
Mean = E(x) = μ = np
Variance = σ^2 = μ
Std dev = σ = $\sqrt{\mu}$

(Note that text has Poisson distribution tables – column is **μ,** rows are **x<=n**)

P(x = 2) = P(x <= 2) – P(x <= 1)
P(x >= 2) = 1- P(x <= 1)
P(2 <= x <= 5) = P(x<=5) – P(x<=1)
P(2 < x <= 5) = P(x <=5) – P(x <= 2)

Be comfortable with being able to do these kinds of permutations and calculations

Approximating a BINOMIAL distribution by Poisson:

Let X have a binomial distribution(n,p)

P(x<= k) - you can get either x or k or both from the Cumulative Binomial Probability tables (table stops at n=20 or 25, with jumps of 5 numbers

Use POISSON APPROXIMATION instead of binomial calculation, **IF n is large AND n*p = < 7 then**

P(x = k) is $\approx \frac{e^{-np}(np)^k}{k!}$ **And** $P(x \leq k) \approx P(Y \leq k)$ **when Y has poisson(np)**

Eg. If a person shoots a gun 500 times with a P = 0.01

a) Probability that the person will **hit the target 4 times**

X = # of times he hits the target

X has a binomial distribution X(500,0.01)

$P(X=4) = \binom{500}{4}(0.01)^4(0.99)^{496}$ note that P(x <= 4) means that you would have to compute probability of P(0) +P(1) +P(2) +P(3)+P(4)

np = 500*0.01 =5 (this is less than 7 so we can use poisson approximation)

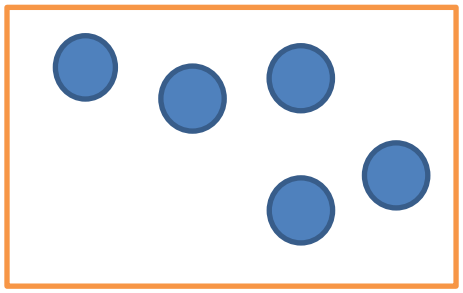
P(x = 4) $\approx \frac{e^{-5}(5)^4}{4!}$ **= 0.1755**

= P(Y<=4) - P(Y<=3) = 0.175

0.44 – 0.265 (from tables)

(For poisson approximation, μ = np, then use tables)

HYPERGEOMETRIC DISTRIBUTIONS:



Box with N balls. M are Red. N-M are blue
Take a sample of size n, WITHOUT replacement.
Let X be the number of red balls in the sample of n
What are the possible values of X?
X can be = 0, 1, 2, up to n
P(M) = M/N-M

Box of N Balls, M are red, N-M are blue
Take a sample of size n, let x = # of red balls in the sample of size n

P(k) = P(x=k) $= \dfrac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$ k = 0,1,2,… k is restricted to min(n, M)

(in how many ways can you choose n out of N? C $\binom{N}{n}$)

n= sample size
M = total SUCCESSES in POPULATION
N = total POPULATION size
k = # of successes in sample

Mean of HYPER DISTRIBUTION = $n\left(\frac{M}{N}\right)$
Variance of HYPER DISTRIBUTION = $n\left(\frac{M}{N}\right)\left(\frac{N-M}{N}\right)\left(\frac{N-n}{N-1}\right)$

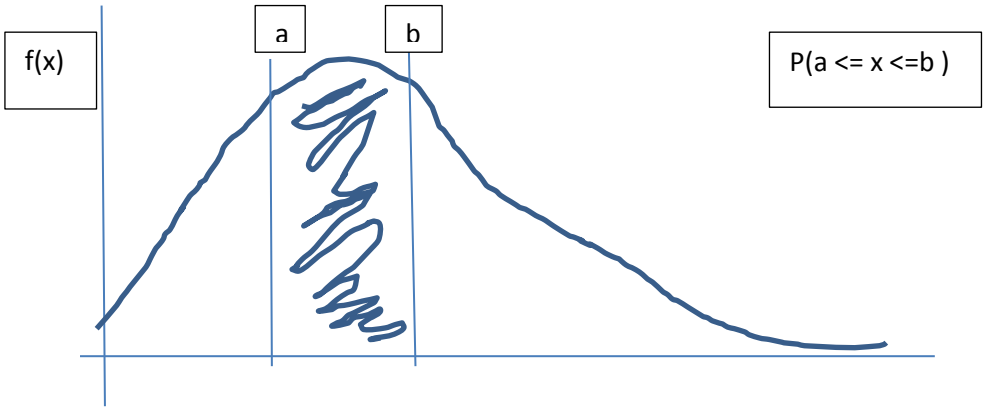
Hyper geometry from binomial:
Binomial probabilities don't change (with replacement)
In Hyper geometries the probability changes with each sample (without replacement)
Eg. A box of 25 lamps, with 6 defective. Take a sample of n=3, WITHOUT replacement.
What is the probability of having 2 defective lamps in the sample set of 3?

P(k) = P(x=2) $= \dfrac{\binom{6}{2}\binom{25-6}{3-2}}{\binom{25}{3}} = \dfrac{\binom{6}{2}\binom{19}{1}}{\binom{25}{3}}$ =

Male/female in large population can follow hypergeometric distribution.

CHAPTER 6

Probability DENSITY Functions: Pdf : works on interval
Probability MASS Function: Pmf : Works on discrete variables – you can assign probabilities to each value of x and sum the probabilities to equal 1



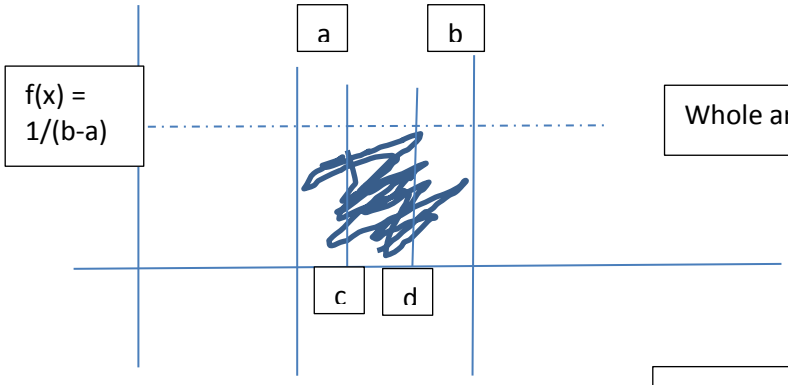
PROPERTIES of Probability Density function:

- For ALL x, F(x) >= 0
- The area under f(x) and x=(a,b) , the x-axis gives P(a ≤ x ≤ b) = P(a ≤ x < b) = P(a < x ≤ b), because P(x=integer) = 0

If x is a CONTINUOUS RANDOM VARIABLE, the propbability that x is equal to any SINGLE number is 0
P(x=a)=0 (i.e. P(x=5) = 0)

UNIFORM RANDOM VARIABLE: x has uniform(a,b)

F(x) $= \frac{1}{b-a}$ if a ≤ x ≤ b, and 0 otherwise



Whole area is = 1

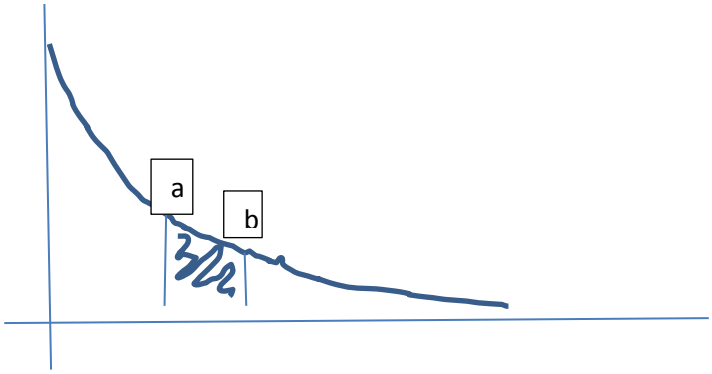
Look at textbook example

If X has uniform[0,2]
P(c <= x <= d) = (d-c)/b-a)
therefore
P(1 <= x <=1.5) = (1.5-1)/(2-0)

EXPONENTIAL RANDOM VARIABLE/DISTRIBUTION
Related to Poisson Random variable
Exp(μ) or Exponential(μ)

We say that x has an exponential distribution with mean **μ**, IF it has $f(x) = \begin{cases} \frac{1}{\mu} e^{-\frac{x}{\mu}} & ; x \geq 0 \\ 0 \ \ Otherwise \end{cases}$

P(a<=x<=b) (area = 1)



$P(x > a) = e^{-\frac{a}{\mu}}$

$P(a <= x <= b) = e^{-\frac{a}{\mu}} - e^{-\frac{b}{\mu}}$

Poisson distribution counts the number of events. The time you have to wait until the next event is an exponential distribution.
Amount of time it takes to do a task after learning can be modeled by an exponential random variable

Eg. The lifetime of a particular brand of LED TV is EXPONENTIAL with mean μ= 4 years
If you know that a certain LED TV of the same brand, has been working for at least 3 years, what is the probability that the TV's lifetime will be more than 8 years?
X= Lifetime of an LED TV

$P(X \geqq 8 \mid X \geqq 3) = \frac{p(x \geqq 8 \cap x \geqq 3)}{P(x \geq 3)} = \frac{p(x \geqq 8)}{P(x \geq 3)} = \frac{e^{-\frac{8}{4}}}{e^{-\frac{3}{4}}} = e^{-\frac{5}{4}} = P(x \geq 5)$

The intersection of two sets where one is inside the other, is always the probability of the smaller set.

In the above example, it says ***forget about the past***. It isn't relevant to the current probability

Which means, p(x ≧ a+b | x ≧ a) = p(x ≧ b) - this is **the memoryless property of exponential distributions.**
Read example 6.2 of the book.
NORMAL DISTRIBUTION: N(μ,σ²)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$ **, -infinity < x < + infinity**

Given $f(x) = \frac{1}{\sqrt{16\pi}}\, e^{-\frac{(x-2)^2}{16}}$ **, you can determine that σ² = 8, and μ = 2**
X has N(2,8)

The two parameters of a normal distribution are its MEAN and its VARIANCE

$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$ **, you can determine that σ² = 1, and μ = 0 -- STANDARD NORMAL DISTRIBUTION. X has N(0,1)**

$f(x) = \frac{1}{3\sqrt{2\pi}}\, e^{-\frac{(X+1.5)^2}{18}}$ **X has N(-1.5,9), MU = -1.5, variance = 9**

<u>Probability DENSITY FUNCTION (PDF)</u>

X has a NORMAL DISTRIBUTION N(μ,σ²)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

Can obtain μ from distribution if we have graph.

STANDARD NORMAL DISTRIBUTION: Z has standard normal (normal zero, one) if PDF N(0,1)

$$f(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{(Z)^2}{2}}$$

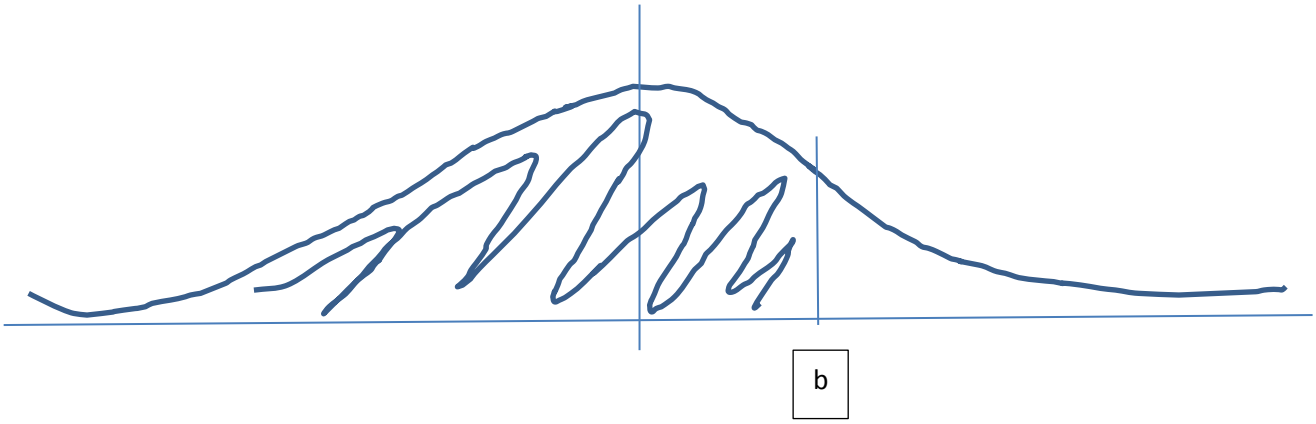
Expected Value (μ) = 0, and variance =1
A very particular case. Why is this case so important?
Because you can take ANY NORMAL distribution and transform it to a STANDARD NORMAL distribution
Let x be **N(μ,σ²), then Z =** $\frac{x-\mu}{\sigma}$ **and Z has N(0,1)**

Eg. X HAS N(-11, 16), μ=-11, σ²=16

$$Z = \frac{x-(-11)}{\sqrt{16}} = \frac{(x+11)}{4}$$

CUMULATIVE DISTRIBUTION FUNCTION: P(X≤k)
x has N(μ,σ²)
P(X ≤ k) eg. P(X ≤ 1.5)

How do we get probabilities of a cumulative distribution as a STANDARD NORMAL distribution.
P(Z ≤ b) - probability that you are going to observe a VALUE for Z that is less than or equal to B.



Rather than do integrations, use the STANDARD NORMAL distribution tables.
If Z standard normal is less than or equal to -1.53, how do you get it from the table?
First find -1.5 on the left margin, then move across the columns to 0.03, and the intersection is -1.53 and the result is 0.0630.

P (z <= 1.46) = 0.9279
P (z > 1.52) = 1- P(z <= 1.52) = 1 – 0.9357 =
P (1.55 < Z < 1.68) = P(z < 1.68) – P(z < 1.55) - we are dealing with a continuous random variable – it does not matter whether or not it is < or ≤ - the result is the same.

If P (Z ≤ **b) = 0.99, what is b? (should be able to get this from the table). b = 2.33**

P Z <= 0) = ½ (50%) → should be able to solve without the table! ☺

What if x has N(μ,σ²)? P(x ≤ a) = P(z ≤ (a-μ)/σ)
TRANSFORM it to STANDARD NORMAL and then you can use the table

Eg. X has N(100, 100); P(x >110) = P (z > (110-100)/sqrt(100)) = P (z > 1)
To get the probability, must calculate 1- P(z ≤ 1) = 1 - 0.8413

Eg x has N(-10,25)
What is P(x ≤ a) = 0.9251?
P(z ≤ (a-(-10))/sqrt(25)) = 0.9251 (.9251 is 1.44 in the tables), therefore: a+10/5 = 1.44 --- solve for a, a = -2.8

Normal has approximation to BINOMIAL (n,p)

Eg $P(x \leq k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i} = p(o) + (p1) + \cdots p(k)$

q =1-P

Can only use STANDARD NORMAL APPROXIMATION if the following TWO CONDITIONS ARE TRUE

1) **If np > 5 and nq > 5 (where q = 1-p)**
2) **p should not be close to 0 or 1 – best is when p is close to .5**

If so, you can approximate P(x ≤ k) ≈ $\boldsymbol{P\left(z \leq \frac{k+0.5-np}{\sqrt{npq}} \right)}$

Adding .5 of a unit is the CONTINUITY CORRECTION. It is absolutely vital. This is because X is a DISCRETE random variable and Z is a CONTINUOUS random variable. Adding .5 is necessary to adjust for this discrepancy.

Eg. n = 25, p = 0.6; z has Binomial(25,0.6)

P(x ≤ 17) = 0.846 (from binomial table)

Investigate the NORMAL approximation to see how good it is - FIRST – verify the two conditions!

$\left(z \leq \frac{17+0.5-(25*.6)}{\sqrt{25*.4*.6}} \right)$ = $\left(z \leq \frac{17.5-(15)}{\sqrt{25*.4*.6}} \right) = P(z \leq 1.02)$

Find probability of 1.02 in the Normal table = .8461 -- VERY close to actual value of 0.846 from binomial table.

P(x=17) = P(x <= 17) – P(x <=16) → USE NORMAL approximation for both and substitute in.

(end of chapter 6)

CHAPTER 7

CONCEPT OF PARAMETERS
i.e. why is it important to know the Average value of X?

eg.

Height of people:
Group1: 189,171,162
Group2: 180, 173, 176, 178

Which one of these two groups is taller, which is shorter? Only by taking averages… we cannot compare different groups without creating one representative number from each group and comparing them.
Can calculate MEANS…
Group 1: 172cm
Group 2: 176cm

Can calculate MAXIMUMS:
Group 1: 189cm
Group 2: 182cm

PARAMETER: ANY CHARACTERISTIC OF A POPULATION.
Eg. Population Mean, Variance, Median, Maximum, Minimum.
two groups (annual income of Canadians), (annual income of Americans) – data has to be compressed into two sets.
MEAN is the most natural to compare. Can compute MEDIAN. (If dist is symmetric, mean and median are the same.)

PROBLEM: In majority of cases, the parameter of a POPULATION is UNKNOWN.

1. **In order to make an inference on the unknown parameter, we obtain a sample of size *n* from that population**
2. **How you choose your sample is extremely important (not covered in this course…. But MUST BE RANDOM, and your sample size must be large enough, and it must be REPRESENTATIVE of the population as a whole)**

STATISTIC: Any function of the data in the sample, I.E. $\bar{x}$, s^2, $max(x_1, x_n)$
Each STATISTIC is a RANDOM VARIABLE. Why? Because it is computed from a set of RANDOM VARIABLES.
DISTRIBUTION of any statistic is called the SAMPLING DISTRIBUTION of that statistic.

3 ways to derive the sampling distribution of a statistic of interest:

1) We can compute it mathematically
2) We can use simulations to approximate it.
    a. i.e. for mean: get multiple samples from the population and compute mean for each sample. Construct a frequency histogram from the sample means - will be a good simulation of the sampling distribution of the mean
3) We use statistical theorems to approximate the sampling distribution of the statistic in hand

In the rolling of a die, roll twice and take the mean: (x1 + x2)/2
If we determine the probabilities of any given mean, we see a symmetrical distribution of the probabilities.

Probability of any given number rolled is 1/6, therefore the probabilities for the following means are:

$\bar{x}$	1	1.5	2	2.5	3	**3.5**	4	4.5	5	5.5	6
P($\bar{x}$)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Expected value of ($\bar{x}$) in this kind of distribution is always the MEAN of the MEANS
E($\bar{x}$) = 1*1/36 + 1.5*2/36….. + 6*1/36 = 3.5

PROPERTIES OF THE SAMPLE MEAN $\bar{x}$:

1) $\bar{x}$ is an UNBIASED ESTIMATOR for μ; E($\bar{x}$) = μ (regardless of sample size)
2) The standard deviation of $\bar{x}$ is $\frac{\sigma}{\sqrt{n}}$ where n is the sample size
3) The CENTRAL LIMIT THEOREM APPLIES: if the sample size n is LARGE, then the sampling distribution of $\bar{x}$ is approximately NORMAL with mean of μ and stddev of $\frac{\sigma}{\sqrt{n}}$

Sufficiently large samples of a RANDOM distribution converge to a NORMAL distribution.

if n >= 30 then $\bar{x}$ is APPROXIMATELY N(μ, $\frac{\sigma^2}{n}$)
If $\bar{x}$ is APPROXIMATELY Normal, THEN $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ is APPROXIMATELY STANDARD NORMAL, N(0,1)

Rules for determining if SAMPLE SIZE n is LARGE enough to use the CENTRAL LIMIT THEOREM:

1) If the original population is normally distributed, then I can use the CLT for any size of *n*, and the approximation becomes EXACT.
2) If the population distribution is fairly symmetrical, then for a relatively small sample size, you can use the CLT
3) If we the population distribution is skewed or unknown, we must use a sample size of at least 30 to use CLT.

Eg. Alzheimer's disease from onset to death range from 3 to 20 years, with mean **μ = 8 years** (population mean actually given in this example), and **σ = 4 years**.
Choose a sample of **n=30** patients and record their lifetimes (x_1,….x_{30})
Find the approximate probability that ($\bar{x}$ <= 7)
Given that we have the population mean and a sample size of 30, we can use CLT to assume
P($\bar{x} \leq 7$) ≈ P($z \leq \frac{(7-8)}{\frac{4}{\sqrt{30}}}$)

$\sigma^2 = 16$

$\bar{x}$ is approximately N(8,16/30) therefore Z = ($\bar{x} - \frac{8}{\frac{4}{\sqrt{4}}}$)) is approx.. N(0,1)

CLT P(Z ≤ -1.37) = 0.0853 (from tables)

$\hat{p}$ (sample proportion) is approximately N(P, (P(1-p))/n)

IF np and n(p-1) > 5 then $\boldsymbol{z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}}$

Eg. p is the proportion of Canadians who think that sport is equally important for both genders.
If the actual population proportion is p=0.55, and we get a sample of size n=500

Then the SAMPLE proportion, $\hat{p} = \frac{\sum_{i=1}^{500} y_i}{500}$

What is the probability that the sample proportion will show a number >= 60% ?

P($\hat{p} \geq 0.6$)

First, make sure that np, and n(1-p) are > 5:

n(1-p) = .45*500 (yes, > 5)

np = .55*500 (yes, > 5)

Can use CLT

$P(\hat{p} \geq 0.6) \approx P(z \geq \frac{0.6-0.55}{\sqrt{\frac{0.55(1-0.55)}{500}}} = P(z \geq 2.25) = 1 - P(z \leq 2.25) = 1 - 0.9878 = 0.0122$

(from tables)

Not very likely to occur.

Chapter 8 – LARGE SAMPLE ESTIMATION (n ≧ 30)

Let Θ represent the statistic of interest in the population
Obtain a sample of size 30 or more
How do we use the sample to estimate the population mean?

If we use ONE single value to estimate Θ, it is a POINT ESTIMATION.

Or, we can use 2 values from the sample to estimate Θ, it is an INTERVAL ESTIMATION
Eg, average height of Canadians is 172cm – H_0, the null hypothesis
We believe the number is higher than 172cm –H_a, the alternative hypothesis

Inferential statistics:
Methods of inference

1) **Estimation** – estimating or predicting the value of the unknown parameter. Can be POINT or INTERVAL estimation.
2) **Hypothesis testing** – process of making a decision about a parameter based on some preconceived idea about it.
   Read examples 8.1 and 8.2 in the book

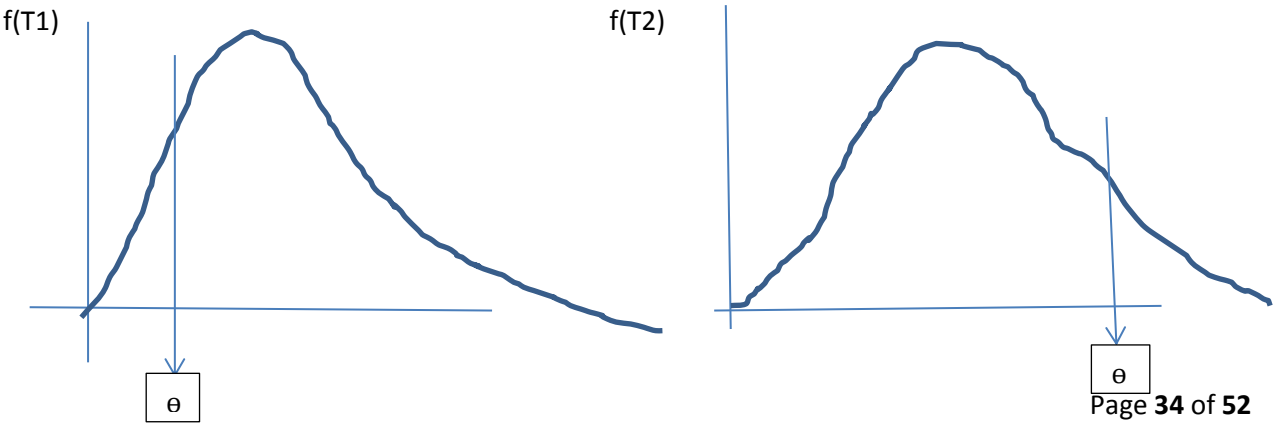
Types of estimators:
An ESTIMATOR is a FORMULA that tells us how to calculate an estimation for an unknown parameter

1) POINT ESTIMATOR – Based on a sample, one single value is computed as an estimator for an unknown parameter
2) INTERVAL ESTIMATOR – based on a sample, TWO values are computed/calculated to form an interval that will contain the unknown parameter with high probability

POINT ESTIMATION:
Assume parameter of interest in the population is μ
One way is to calculate $T1(x_1,\ldots. x_n)$ and multiply them all to get one number
Another way $T2= \max(x_1, x_n)$
Another way $T3 = \text{Sum}(x_i , i=1 \text{ to } n)/n$ (sample mean)

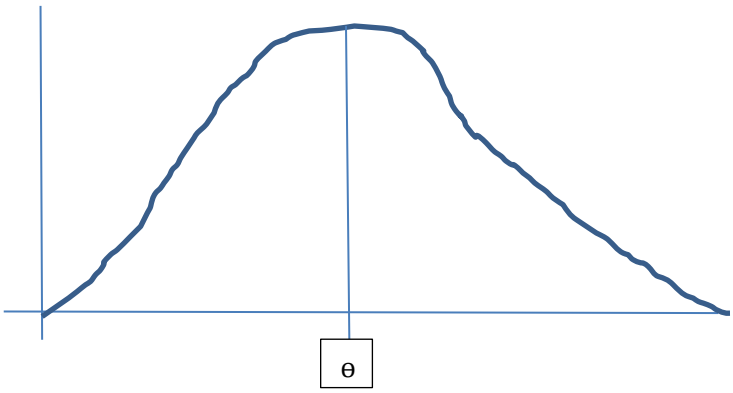
Some estimators are better than others. How to choose a good estimator?
Choose estimators which satisfy the following conditions:

1) **Pick the ones who are UNBIASED - the sampling distribution is CENTERED over the parameter of interest**



T1 over estimates ɵ in the majority of cases.
T2 underestimates ɵ in the majority of cases
T3 is better because 50% of the time theta is overestimated, 50% it's underestimated

f(T3)



Unbiased ESTIMATOR for the parameter of interest (ɵ)
We say the estimator $T(x_1,\ldots,x_n)$ is UNBIASED for ɵ if $E(T(x_1,\ldots,x_n)) = ɵ$

2) Choose (among the UNBIASED estimators) that have the SMALLEST Variance (or SD)

$\bar{x}$ is an unbiased estimator for μ
I can calculate $\bar{x}1$ as (x1+x2)/2. Or I can calculate $\bar{x}2$ as (x1+x2+…x10)/10.
Which to choose?
Variance of $\bar{x}2$ will have a SMALLER variance, as a result you should choose it.

Use sample mean to estimate population mean – **error is $|\bar{x} - \mu|$**
Use sample proportion to estimate population proportion.

Proportion is $\hat{p} = \frac{\sum_{i=1}^{n} y_i}{n}$ error is $|\hat{p} - p|$

Error of ANY estimation is $|\hat{\theta} - \theta|$

MARGIN OF ERROR (will use 95% margin)
For $\mu = 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$ but we may not have population STDDEV, so use sample stddev
$= 1.96\left(\frac{s}{\sqrt{n}}\right)$
very large sample results in s being divided by very large number, reduces it to effectively zero
therefore increasing sample size REDUCES margin of error

For p $= 1.96\sqrt{\frac{p(1-p)}{n}}$ - but use sample $= 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ IF $n\hat{p} > 5$ & $n(1-\hat{p}) > 5$

Eg. – example 8.4 from book
Investigation possibility of watching TV over internet.
n=50, sample mean = 11.5 hrs/day, s=3.5 hrs/day
Estimation is that unknown μ is 11.5
95% margin of error = 1.96(3.5/sqrt(50)) = 0.97 ~= 1
Means that with high probability (95%), our sample mean is ~ +/- 1 hour from actual mean.

INTERVAL ESTIMATION
Come up with two numbers which will capture the statistic of interest within them.
$a \leq \Theta \leq b$ you want to ensure that $P(a \leq \theta \leq b)$ is high

Interval estimation is a rule for calculating two numbers, between which the unknown parameter of interest will lie with HIGH PROBABILITY (given in terms of CONFIDENCE LEVEL)
Confidence level is 1-α, where α is a number between 0 and 1
Or as C.I. (100 * (1-α))% (Represents probability that interval will contain the parameter of interest)
Common confidence levels: 90%, 95%, 99%

How to construct a 100(1- α))% C.I. for the population mean μ, based on a sample size of n ≧ 30
How can I get **a** and **b** from sample mean and sample stddev? – use CLT.
Since n>= 30, sample mean is approximately $\mathbf{N(\mu, \frac{\sigma^2}{n})}$ then $\boldsymbol{Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}}$ - this is why Standard Normal confidence intervals are often called Z-intervals

$$P(-z_{\frac{\alpha}{2}} \leq \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\frac{\alpha}{2}}) \approx 1-\alpha$$

$$\bar{x} - \mu = z_{\frac{\alpha}{2}}(\frac{\sigma}{\sqrt{n}})$$

Lower bound becomes $\bar{x} - z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$, upper bound becomes $\bar{x} + z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$

a b

Large sample (n>= 30) what is the 100(1-α)% C.I. for μ?
Case 1: σ is KNOWN: $[\bar{x} - z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right), \ \bar{x} + z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)] \quad \leftrightarrow \quad \bar{x} \pm z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$
Case 2: σ is UNKNOWN: $[\bar{x} - z_{\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right), \ \bar{x} + z_{\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right)] \quad \leftrightarrow \quad \bar{x} \pm z_{\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right)$

"Student Statistic" –

Eg. Example 8.6 from the text
To study a chemical contaminant in the food, a sample size of n=50, $\bar{x} = 756\ g/day$, s=35 g/day.
Construct a 95% confidence interval for μ
α/2 = 0.025
Need to find value from Standard Normal table where Z = 1- 0.025 probability
If you want to use the table, you need to use a Z=.975, which = 1.96 in the table
Calculating confidence interval:
$756 \pm 1.96\left(\frac{35}{\sqrt{50}}\right)$ → = [746.3, 765.7] - we are 95% confident that μ will be within this interval
Length (range) of the interval is upper bound minus lower bound 19.4

Construct a 90% confidence interval for μ
α/2 = 0.05, use Z = .95 from the table. Equivalent table value is 1.65 (or 1.645)
(8.16)
$756 \pm 1.65\left(\frac{35}{\sqrt{50}}\right)$ → = [747.84, 764.16] - we are 90% confident that μ will be within this interval
length (range) of the interval is upper bound minus lower bound 16.32
In general THE HIGHER the CONFIDENCE, the WIDER the INTERVAL
Increase confidence comes at a price – it increases the RANGE of values

FINAL exam will ASK question about CONFIDENCE interval ranges.

INTERPRETATION OF 100(1-α)% C.I.
What if someone tells you that μ is actually 750, the 95% confidence interval captures it.
If μ is actually 767, we missed the interval.
If we generate 100 Intervals of level 100(1-α)%, then 100(1-α) of them should capture the unknown parameter
It means we have a 100*α chance of getting it wrong.

If we have a binomial distribution of Bin(10,0.95), the average of binomial distribution is np, therefore 9.5 of the samples will capture the mean.

Properties of sample proportion $\hat{\mathbf{p}}$

1. $(\mathbf{1}-\hat{\mathbf{p}})$ **is an unbiased estimator for p; E($\hat{\mathbf{p}}$) = p**
2. S.E. of $\hat{\boldsymbol{p}}$ **is** $\sqrt{\frac{pq}{n}}$; if np >= 5 AND nq >= 5 then the S.E. is estimated by $\sqrt{\frac{\hat{p}\hat{q}}{n}}$
3. The CLT for $\hat{\mathbf{p}}$: **If n is large so that $n\hat{p}$ and n $n\hat{q}$ >= 5 then $\hat{\mathbf{p}}$ is approximately N(P, $\frac{pq}{n}$)**

Z = $\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}\hat{q}}{n}}}$ **is approximately N(0,1)**

3. **100(1-α)% Confidence Interval for p, when $n\hat{p}$ and n $n\hat{q}$ >= 5 is:**

$$p \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} = \left[\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}, \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right]$$

$$\mathbf{P}\left(-z_{\frac{\alpha}{2}} \leq \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}\hat{q}}{n}}} \leq z_{\frac{\alpha}{2}}\right) \approx \mathbf{1-\alpha} = \mathbf{P}\left(\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} \leq \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) \approx \mathbf{1-\alpha}$$

Example 1

Population Proportion for Canadians who are going to vote with Confidence level of 90%
CI of 90% means α/2 (therefore z=.05) = z= 1.65
Sample size of n=985.
592 said they would vote.
Therefore $\hat{p} = \frac{592}{985} = .601$

.601 +- 1.65($\sqrt{\frac{(.601)(0.399)}{985}}$ **--- will lose ½ marks in exam if you do NOT check CONDITIONS!!! 985*(592/985) = 592 >= 5; 985*(392/985) = 393 >= 5**

Example 2

Mean lifetime of rabbits who may or may not get a drug treatment

Population1 = rabbits in CONTROL GROUP =μ1 sample n1 >= 30
Population2 = rabbits in TREATMENT GROUP = μ2 sample n2 >= 30
n1 and n2 don't have to be the same size – as long as both are greater than 30
Of interest is to compare the mean difference of the two populations μ1 - μ2 (the order is important)
If μ1 - μ2 = 0 then then the drug has NO effect
If μ1 - μ2 > 0 then then the drug has a NEGATIVE effect
If μ1 - μ2 < 0 then then the drug has a POSITIVE effect

Use $\bar{x}_1 - \bar{x}_2$ to estimate μ1 - μ2.

1. $\bar{x}_1 - \bar{x}_2$ is an unbiased estimator for μ1 - μ2
2. S.E. OF $\bar{x}_1 - \bar{x}_2$ IS $\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$, estimated by $\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$
3. CLT for $\bar{x}_1 - \bar{x}_2$ is approximately N(μ1 -μ2, $\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$) - but we don't have pop stddev, so

Z = $\frac{\bar{x}_1-\bar{x}_2-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$ is approximately N(0,1)

4. 100(1-α)% CI for μ1-μ2 is $(\bar{x}-\bar{x}_2) \pm \boldsymbol{z_{\frac{\alpha}{2}}}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$

P(-$z_{\frac{\alpha}{2}}$ $\leq$ $\frac{\bar{x}_1-\bar{x}_2-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$ $\leq$ $z_{\frac{\alpha}{2}}$) ≈ 1-α

= P($(\overline{x}_1-\bar{x}_2) - z_{\frac{\alpha}{2}}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} \leq \mu_1-\mu_2 \leq (\bar{x}_1-\bar{x}_2) + z_{\frac{\alpha}{2}}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$)≈ 1-α

If the value 0 shows up inside the confidence interval $(\bar{x}-\bar{x}_2) \pm z_{\frac{\alpha}{2}}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$ then you can conlude that it is likely that μ1 -μ2 = 0

(if zero is on the boundary, then it's less assured)

Read examples 8.9 and 8.10 in the book

Ex. 8.45 –
Pop1: n1 = 64 sample mean = 2.9 sample variance = 0.83
Pop2: n2 = 64 sample mean = 5.1 sample variance = 1.67

a) Construct a 90% CI for μ1-μ2
$\alpha/2 = 0.05$, $Z_{0.95}$=1.65

$2.9-5.1 \pm \sqrt{\frac{.83}{64}+\frac{1.67}{64}}$ = $[-5, .45, 1.05]$ **- because 0 is in this interval, we can conclude at the level of 90% that μ1=μ2**

b) If we construct a 95% CI for these samples, can we conclude without calculations that **μ1=μ2?**
Yes, because the 90% CI is included IN the 95% CI, we can conclude that 0 is in the 95% CI
c) If we construct a 99% CI, with $z_{0.005}$=2.57 we get [-7.26, 2.86] → We can conclude that **μ1=μ2**

Ex. What if the parameter of interest is the population proportion?
The two samples MUST be INDEPENDENT – extremely important
I.e. proportion of Canadians making 500K/year, vs proportion of Americans making more than 500K/yr
$\hat{p}_1$ = canadians
$\hat{p}_2$ = americans

Use $\hat{p}_1 - \hat{p}_2$ to estimate $p_1 - p_2$ (unbiased estimator)
If SE for the population cannot be determined, then use SE of $\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}$, **only *IF***
$n_1\hat{p}_1, n_2\hat{p}_2, n_1\hat{q}_1$ & $n_2\hat{q}_2 \geq 5$
Then $\hat{p}_1 - \hat{p}_2$ is approximately N($p_1 - p_2, \frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}$)

$$Z = \frac{\hat{P}_1 - \hat{P}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}}$$ is approximately N(0,1)

100(1-α)% CI for p1- p2= $\hat{p}_1 - \hat{p}_2 \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$

One-sided confidence limits

$\mu \quad \left[(\bar{x} - z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}), (\bar{x} + z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}})\right]$

Lower Confidence Level (LCL) = Point Estimator – Zα * SE of the point estimator

Upper Confidence Level = Point Estimator + Zα * SE of the point estimator

Construct 100(1-α)% LCL for μ: $[\bar{x} - z_{\alpha} \frac{s}{\sqrt{n}}, +\infty]$

Construct 100(1-α)% UCL for μ: $[-\infty, \bar{x} + z_{\alpha} \frac{s}{\sqrt{n}},]$

Construct 100(1-α)% LCL for P: $[\hat{p} - z_{\alpha}\sqrt{\frac{pq}{n}}, +\infty]$

Construct 100(1-α)% UCL for P: $\left[-\infty, \hat{p} + z_{\alpha}\sqrt{\frac{pq}{n}}\right]$

EG>
Choosing sample size (n), from population with unknown mean μ – what size of sample should I select to ensure a 95% CI for $\bar{x}$? (100(1-α)% margin of error)
To construct Confidence interval we replaced σ with S, because we didn't know σ…. THIS Implies a MARGIN OF ERROR because we used the SAMPLE standard deviation.
The desired margin of error for capturing the population mean using $\bar{x}$ should be $z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq B$
(Where B is the DESIRED margin of error)
So… $n > \left(\frac{z_{\frac{\alpha}{2}}}{B}\right)^2 \sigma^2$
There are two ways to calculate n for a confidence interval, if you don't have access to σ:
1) you can use S, if you trust the data: $n > \left(\frac{z_{\frac{\alpha}{2}}}{B}\right)^2 s^2$
2) you can use the $\left(\frac{RANGE}{4}\right)^2$ **:** $n > \left(\frac{z_{\frac{\alpha}{2}}}{B}\right)^2 \left(\frac{RANGE}{4}\right)^2$

ALWAYS ROUND UP!

EG.

Construct a 95% CI for μ with a desired margin of error of 4 (I.e. no more than 4 units from the mean) A previous study for the same population gave S=21, what is n?
(desired margin of error) B = 4
1-α = .05, so α/2 = .025
$n > \left(\frac{.025}{4}\right)^2 21^2$ = 105.8 → therefore we need n of at least 106.EG.

Choosing a SAMPLE SIZE n to study a POPULATION PROPORTION, P.

To construct a 100(1-α)% CI for P, we would use $\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}$

But we don't have P, so use

$z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}} < B$ **therefore** $n > \left(\frac{z_{\frac{\alpha}{2}}}{B}\right)^2 pq$

$f(p) = P(1-P), where\ \ 0 < P < 1$
IF P = ½ MAX $(f(p)) = 1/4$

Eg. Construct a 90% CI for P. What should n be if the desired margin of error is 0.04?
(1-α)/2 = .05 - from the table, Z=1.65
B = .04
$n > \left(\frac{1.65}{.04}\right)^2 1/4$ **= 422.7, therefore n=423**

Read example 8.14 and table 8.7 – IMPORTANT

Chapter 9 - Large Sample (N>= 30) Hypothesis Testing

Measured μ of Canadian height = 170. This is the NULL HYPOTHESIS, $H_\emptyset$
μ is hypothesized to be > 170cm (ALTERNATE HYPOTHESIS, one sided), H_a
μ is hypothesized to be ≠ 170cm (ALTERNATE HYPOTHESIS)

To test a hypothesis on a large sample you need:

1. Null and alternative hypotheses
2. Test Statistic,; computed based on the data
3. Critical point(s) are identified by the so-called level of the test α – (and α should be something small, i.e. 0.01, 0.05, 0.1), or by p-value
    a. P-value is the minimum value of α that results in rejection of the null hypothesis. (obtained from the data)
4. Conclusion: Either REJECT $H_\emptyset$ in favor of H_a, OR DO NOT reject $H_\emptyset$

	REJECT $H_\emptyset$	Fail to REJECT $H_\emptyset$
$H_\emptyset$ is TRUE	Type I error	CORRECT decision
$H_\emptyset$ is NOT TRUE	CORRECT decision	Type II error

Cannot control both types of errors at the same time so we generally try to control Type I error.
α = P(Type I Error)

Eg. Case 1
$H_\emptyset : \mu = \mu_0$
$H_a : \mu > \mu_0$

The test statistic is $\mathcal{L} = \frac{(\bar{x} - \mu_0)}{s/\sqrt{n}}$

$\mathcal{L}$ is the maximum alpha value you can accept
Critical value at level α is Z_α
p-value = P(Z > largest value of alpha we can accept)
Where Z is Z standard normal
Conclusion:
Critical value approach says that we can reject H0 if $\mathcal{L} > Z_\alpha$
P-value approach says REJECT H_0 IF P-value is <= α

In Case 2, if $H_a : \mu < \mu_0$, then the P-value remains the same, but the Critical value at level α is $-Z_\alpha$

Case 3
$H_\emptyset : \mu = \mu_0$
$H_a : \mu \neq \mu_0$

Test statistic $\mathcal{L} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
Critical points at level α are $Z_{\alpha/2}$ and $-Z_{\alpha/2}$
P-Value = P(Z < - |$\mathcal{L}$|) + P(Z > |$\mathcal{L}$|)
= 2P(Z > |$\mathcal{L}$|)

Z_α is a critical point. It's under the standard normal distribution. It divides the real line into the ACCEPTANCE region and the REJECTION region. **Anything LESS than the critical point will fall into the acceptance region – you will ACCEPT the NULL HYPOTHESIS**.
You can REJECT the NULL hypothesis, only if your sample statistic is LARGER than Z_α

MARGIN OF ERROR

$|\hat{p} - p| \leq z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}$ → **100%(1-α) MARGIN of error for P**

$|\bar{x} - \mu| \leq z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ → **100%(1-α) MARGIN of error for μ**

Directionality is crucial – pay attention to the direction of the inequality

"CRITICAL VALUE APPROACH" – ONE –sided test.
Eg. The average weekly earing of women in prof. positions is $670. To study the average weekly earning of men for the same positions, we take a sample size of n=50 men, $\bar{x}$ = $725, with s of $102
H_0 : μ = $670
H_a : μ > $670
Are men making more to the confidence level of α = 0.01
Critical value for $Z_{0.01}$ = 2.33

$\mathcal{L} = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{725-670}{102/\sqrt{50}} = 3.41$

Conclusion: 3.41 > 2.33 ∴ REJECT H_0

To study the average number of minutes of commercials are run before a movie starts.
H_0 : μ = 3 minutes
Take sample of 50 movies and determine $\bar{x} = 3.25\ minutes,\ S = 0.5min$
From this, you develop an alternative hypothesis:
H_a : μ > 3 minutes
Set α = 0.01

$\mathcal{L} = \frac{3.25-3}{\frac{.05}{\sqrt{50}}} = 3.57$ ← if this is NEGATIVE then we got the DIRECTIONALITY of the effect wrong in our alternative hypothesis

Critical Value for $Z_{0.01}$ = 2.33
Conclusion: 3.57 > 2.33, therefore REJECT H_0

Eg (9.9 from text)
H_0 : μ = 3300mg/day sodium
H_a : μ > 3300 mg/day

N=100 (>30)

$\bar{x} = 3400, S = 1100$

$\mathcal{L} = \frac{3400-3300}{\frac{1100}{\sqrt{100}}} = 0.91$

P-Value = $P(Z_\alpha > 0.91) = .1814$ ← minimum alpha to reject H_0
If we decide that we want an alpha level of 0.05, then .1814 !< 0.05, so CANNOT reject H_0

α = P(Type I error)
= P(rejecting H_0 when H_0 is true)

POWER OF TEST = 1-β = P(Reject H_0 when H_a is true)
= P (Truly rejecting H_0)

Pop 1, μ_1 n >= 30
Pop 2, μ_2 n >= 30 (treatment group)

$\bar{x}$ = Measure # of days to recover with/without treatment

H_0 : $\mu_1 - \mu_2 = 0$
H_a : $\mu_1 - \mu_2 > 0$ -- # of days to recover WITHOUT treat. should be larger than # of days to recover WITH treat.

Practice Exam to be posted on CULearn – not representative of difficulty or type of questions.
Review session on Dec 9th….

When working with two SAMPLES from two populations, they must be INDEPENDENT

One sided:
H_0 : $\mu_1 - \mu_2 = D_0$
H_a : $\mu_1 - \mu_2 > D_0$
Test Statistic $\mathcal{L} = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
At the level alpha, the critical value of Z_α P-value = $P(Z > \mathcal{L})$
Conclusion: Critical Value approach: Reject H_0 IF $\mathcal{L} > Z_\alpha$
P-value approach: Reject H_0 IF P-value $\leq \alpha$

If the direction of the evaluation for H_a reverses : $\mu_1 - \mu_2 < D_0$ then
Conclusion: Critical Value approach: Reject H_0 IF $\mathcal{L} < -Z_\alpha$
P-value approach: Reject H_0 IF P-value $\leq \alpha$

Two SIDED:
H_0 : $\mu_1 - \mu_2 = D_0$
H_a : $\mu_1 - \mu_2 \neq D_0$
Calculate Z. look up probability for Z from tables, and multiply by 2 , as this is the probability
P-value = $P(Z > |\mathcal{L}|) + P(Z < -|\mathcal{L}|)$ = **$2P(Z > |\mathcal{L}|) = 2(1-P_{(\mathcal{L})})$**

Critical value approach: Reject H_0 IF $\mathcal{L} < -Z_{\alpha/2}$ OR $\mathcal{L} > Z_{\alpha/2}$

P-Value approach: Reject H_0 IF P-value $\leq \alpha$

Example: Study the effect of owning a car on a university student's performance.
Conduct a TWO-SIDED test.
Take a sample of size = n_{nocar}=100, average GPA $\bar{x}_{nocar} = 2.7$; $s^2{}_{nocar} = 0.36$
Take a sample of size = n_{Car}=100, average GPA $\bar{x}_{Car} = 2.54$; $s^2{}_{Car} = 0.4$
H_0 : $\mu_1 - \mu_2 = 0$
H_a : $\mu_1 - \mu_2 \neq 0$

$$\mathcal{L} = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{2.7 - 2.54 - 0}{\sqrt{\frac{.36}{100} + \frac{.4}{100}}} = 1.84,$$ P-value = 2P(Z > 1.84) = 2(1 - .96712) = 0.068

from the normal tables, when Z=1.84, P=0.96712 therefore need 2* (1 - .96712) <- This is the minimum alpha you need to reject H_0

if we set alpha at 0.07 then we can reject H_0
if we set alpha at 0.05, then we fail to reject H_0

Relationship between CIs and two-sided hypothesis testing
100(1-α)% CI for any parameter is the ACCEPTANCE REGION of a two sided hypothesis test for that parameter at level α

How to construct a HYPOTHESIS TEST using CONFIDENCE INTERVALS

$(\bar{x}_1 - \bar{x}_2) \pm z_{\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ **-> 100(1-alpha)% CI for $\mu_1 - \mu_2$**

H_0 : $\mu_1 - \mu_2 = 0$
H_a : $\mu_1 - \mu_2 \neq 0$ at level α

Eg… calculate a 95% CI for $\mu_1 - \mu_2$…. i.e. [-1.5,2.3] at alpha = 0.05….
IF ($\mu_1 - \mu_2$) is within the acceptance region of the confidence interval (i.e. =0) , you must accept H_0
Otherwise REJECT H_0

Eg. car/no car example used earlier.
Construct a 95% confidence interval based on prior data:

$(2.7 - 2.54) \pm 1.96\sqrt{\frac{.36}{100} + \frac{.4}{100}}$ **= 95%CI[-.01, 0.33] → because ZERO is inside the acceptance region, as a result you accept H_0 (conclusion is the same as the direct P-value approach)**

Hypothesis testing for Population Proportion

One SIDED:

$H_0 : P = P_0$
$H_a: P > P_0$

Test Statistic $\mathcal{L} = \dfrac{\hat{p} - p_o}{\sqrt{\dfrac{p_o q_o}{n}}}$

CRITICAL VALUE at level alpha is Z
p-VALUE = $p(Z > \mathcal{L})$

Two sided approach:

CRITICAL VALUE at level alpha is $Z_{\alpha/2}$

p-VALUE = $P(Z > |\mathcal{L}|) + P(Z < -|\mathcal{L}|) = 2P(Z > |\mathcal{L}|)$

Critical value approach: Reject H_0 IF $\mathcal{L} < -Z_{\alpha/2}$ OR $\mathcal{L} > Z_{\alpha/2}$
P-Value approach: Reject H_0 IF P-value $\leq \alpha$

Eg. Expected proportion of the population making 80K more:
P = 0.2
$H_0 : P = 0.2$
$H_a: P < 0.2$

Take a sample of n=100 canadians and record income.
15 of them are making $80,000 or more. $\hat{p} = \frac{15}{100}$
Compute P-value
First compute Z

$$\frac{.15 - .2}{\sqrt{\frac{0.2x0.8}{100}}} = -1.25$$

P-value = P(Z < -1.25) = 0.105 ← this is the minimum alpha you need to reject H_0
The larger the P-value, the harder it is to reject H_0

Hypothesis testing for
DIRECTIONAL DIFFERENCE between TWO POPULATION PROPORTIONS
Eg. cold treatment, does a drug decrease cold symptoms
Control group
Treatment group
$P_1 = 0.2$
$H_0 : \hat{p}_1 - \hat{p}_2 = 0$
$H_a: \hat{p}_1 - \hat{p}_2 > 0$

POOLED PROPORTION: $\hat{p} = \frac{\hat{p}_1 n_1 + \hat{p}_2 n_2}{n_1 + n_2}$; $\hat{q} = 1 - \hat{p}$

TEST STATISTIC $\mathcal{L} = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}\hat{q}\frac{1}{n_1} + \frac{1}{n_2}}}$

MUST CHECK CONDITIONS!: $n_1\hat{p}_1$, $n_1\hat{q}_1$, $n_2\hat{p}_2$, $n_2\hat{q}_2$ >= 5

Reject if test statistic is LESS THAN Z_α

Conclusion for two sided uses $Z_{\alpha/2}$

EG…

52 men in 1000 have heart disease: $\hat{p}_1 = \frac{52}{1000} = .052$

23 women in 1000 have heart disease: $\hat{p}_2 = \frac{23}{1000} = .023$

At a significance level of 0.05, is there significant evidence indicating that the proportion of men with heart conditions is greater than the proportion of women with heart conditions?
(n1 and n2 can be different values – but here they are the same)
State the hypothesis:
$H_0 : \hat{p}_1 - \hat{p}_2 = 0$
$H_a: \hat{p}_1 - \hat{p}_2 > 0$
$n_1 = 1000$
$n_2 = 1000$

$\hat{p} = POOLED\ PROPORTION = \frac{52 + 23}{1000 + 10000} = 0.0375$

$\hat{q} = 1 - \hat{p} = 0.9625$

$\mathcal{L} = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}\hat{q}\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{0.052 - 0.023}{\sqrt{(0.0375)(0.9625)\frac{2}{1000}}} = 3.41$

Critical value is Z of 0.05% $(Z_\alpha) = Z_{0.95} = 1.645$
Reject H_0 because $\mathcal{L} > Z_{0.95}$
CONCLUSION STATEMENT: The proportion of men of with heart disease is higher than that of women to a significance level of α=0.05
p-value = P(Z > 3.41) = 1- P(Z ≤ 3.41)

CHAPTER 10: Inference from SMALL SAMPLES (n <30)

When n is < 30, we cannot call the test statistic ***Z***, we call it ***t*** because it is no longer normally distributed – it will have a Student t-distribution, with **n-1 degrees of freedom**

$t = \frac{(\bar{x} - \mu)}{s/\sqrt{n}}$ Can be described as "has t-distribution with n-1 df" or "has t(n-1)"

PROPERTIES OF THE t-Distribution with *r* degrees of freedom (*df*) (where *r ≥ 1*)

1. It is BELL SHAPED, with heavier tails than a standard normal distribution
   As the degrees of freedom, r, increases, the t-distribution approaches N(0,1). It becomes normal when r ≥ 30
2. The t-distribution is identified by its *df*, *r*
3. *t-distribution ONLY APPLIES if the ORIGINAL POPULATION is NORMALLY DISTRIBUTED*

Using the table, given the df, and t value, find the probabilities:
P(t(9) > 1.383) = 0.1
P(t9) > 2.821) = 0.01
P(t(14) > 1.761) = 0.05

Using the table, given the df and probability, find the t value:
P(t(14)> ?) = 0.01 = 2.624

To have a df of 30, you should have n of 31, so that n-1 = 30

Due to symmetry of the bell curve, the two tails are the same, so P(t(14) < -1.761) is exactly the same as P(t(14) > 1.761)

P(t(14) < -1.761) P(t(14) < -1.761)

Small sample INFERENCE is about the population mean, μ
When the population is NORMAL (μ,σ^2)

ONE SIDED:
H_0 : $\mu = \mu_0$
H_a : $\mu > \mu_0$ **Test statistic** $\mathcal{T} = \frac{(\bar{x}-\mu_0)}{s/\sqrt{n}}$ **critical value at level α, $t_\alpha(n-1)$**
P-value = P(t(n-1) > $\mathcal{T}$)
Conclusion:
critical value approach: Reject h0 if $\mathcal{T} > t(n-1)$
P-value: Reject H_0 IF p-VALUE ≤ α

Two-sided:
H_0 : $\mu = \mu_0$
H_a : $\mu > \mu_0$ **Test statistic** $\mathcal{T} = \frac{(\bar{x}-\mu_0)}{s/\sqrt{n}}$ **critical value at level α, - $t_{\alpha/2}(n-1)$ and $t_{\alpha/2}(n-1)$**
P-value = P(t(n-1) < $|\mathcal{T}|$) + P (t(n-1) > $|\mathcal{T}|$) = 2 P(t(n-1) >$|\mathcal{T}|$)

Conclusion:
critical value approach: Reject h0 if $\mathcal{T} < -t_{\alpha/2}(n-1)$ **OR** $\mathcal{T} > t_{\alpha/2}(n-1)$
P-value: Reject H_0 IF P-VALUE ≤ α

Confidence intervals based on n< 30 : $\bar{x} \pm t_{\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}}$

Eg. Oxygen dissolved in a river's water system in july (low water)
Years 1 thru 6 in ppm: 4.9, 5.1, 4.9, 5.0, 5.0, 4.7
n = 6; mean = 4.93; $s^2 = 0.0186$
Assume POPULATION is normally distributed.
It is said that the actual mean level of dissolved oxygen in the water is μ=5ppm (Based on looking at the data assume a DIRECTIONAL alternative hypothesis):
H_0 : μ = 5ppm
H_a : μ < 5ppm

Test at level α = 0.05 $\mathcal{T} = \frac{(\bar{x}-\mu_0)}{s/\sqrt{n}} = \frac{(4.93-5)}{0.1366/\sqrt{6}} = -1.195$
Critical value = $-T_{0.05}(5)$ (*FROM TABLE*) = -2.015
Cannot reject H_0 because -1.195 ***is NOT*** < -2.015

Case 1: Directional hypothesis
H_0 : $\mu_1 - \mu_2 = D_0$
H_a : $\mu_1 - \mu_2 > D_0$
Test statistic $\mathcal{T} = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ $\quad s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$

Critical value at level alpha $t_\alpha(n1+n2-2)$
P-value = $p(t_\alpha(n1+n2-2) > \mathcal{T})$
Conclusion:
Critical value approach: Reject H_0 IF $\mathcal{T} > t_\alpha(n1+n2-2)$
P-value approach: Reject H_0 IF P-value <= α

Case 2: Opposite direction:
H_0 : $\mu_1 - \mu_2 = D_0$
H_a : $\mu_1 - \mu_2 < D_0$
Test statistic $\mathcal{T} = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ $\quad s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$

Critical value at level α: $-t_\alpha(n1+n2-2)$
P-value = $p(t_\alpha(n1+n2-2) < \mathcal{T})$
Conclusion:
Critical value approach: Reject H_0 IF $\mathcal{T} < -t_\alpha(n1+n2-2)$
P-value approach: Reject H_0 IF P-value <= α

Case 3: TWO SIDED

H_0 : $\mu_1 - \mu_2 = D_0$

H_a : $\mu_1 - \mu_2 \neq D_0$

Test statistic $\mathcal{T} = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ $\quad s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$

Critical value at level α: $-t_{\alpha/2}(n1+n2-2)$, $+t_{\alpha/2}(n1+n2-2)$

P-value = $P(t_{\alpha/2}(n1+n2-2) < -|\mathcal{T}| + t_{\alpha/2}(n1+n2-2) > |\mathcal{T}|)$ **$= 2P(t_{\alpha/2}(n1+n2-2) > |\mathcal{T}|)$**

Conclusion:

Critical value approach: Reject H_0 IF $\mathcal{T} < -t_{\alpha/2}(n1+n2-2)$ OR $\mathcal{T} > t_{\alpha/2}(n1+n2-2)$

P-value approach: Reject H_0 IF P-value <= α

CONFIDENCE INTERVALS when n1 and n2 are small….

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}(n_1+n_2-2)}\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

Read example 10.1 in the book?

Chapter -10, continued – small sample hypothesis testing

CI for n > 30: $\bar{x} + z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$

CI for n < 30 : $\bar{x} + t_{\frac{\alpha}{2}(n-1)}\left(\frac{s}{\sqrt{n}}\right)$

Small sample inference for the difference between two population means

- **TWO SAMPLES must be INDEPENDENT**
- **Populations must be NORMALLY DISTRIBUTED**
- **If even one sample is less than 30, you must use the t-test**
- **The VARIANCES of the two populations MUST coincide $\sigma_1^2 = \sigma_2^2$**
    - **Test this by dividing LARGER S^2/SMALLER S^2 <= 3 means we can conclude $\sigma_1^2 = \sigma_2^2$**

Eg.
Sample of size n_1 < 30, n_2 <30, compute sample mean and variance from each populations
Case 1

Example:
Tony's Tire Company, Will's Tire company
Tony claims that their tires last MORE than 1500 miles longer than Will's tires do.
Take 10 tires from each company, n1=10, n2=10
Tony's; n1=10; $\bar{x}_1$ = 16,700 miles; $s_1 = 1{,}700$
Will's; n2=10; $\bar{x}_2$ = 15,100 miles; $s_2 = 1{,}350$
At α =0.1 is there significant evidence to suggest that tony's claim is correct?

H_0 : $\mu_1 - \mu_2 = 1500$
H_a : $\mu_1 - \mu_2 > 1500$

Is $\frac{s_1^2}{s_1^2} \leq 3$ **?** 1700/1350 = 1.26 ; yes

$$s_p^2 = \frac{(10-1)1700^2 + (10-1)1350^2}{10 + 10-2} = 2{,}356\ 250$$

$$\mathcal{T} = \frac{16700 - 15100 - 1500}{\sqrt{2{,}356\ 250\left(\frac{1}{10}+\frac{1}{10}\right)}} = 0.14$$

Look up α=0.1 in t-table for 18 *df* = *1.330*

0.14 IS not > 1.330 therefore cannot reject H_0 – therefore TONY'S claim is false?
(something like this on the final exam – compare 2 small samples)

Situation where TWO SAMPLES are NOT independent: PAIRED DIFFERENCE INFERENCE

Tires A, and Tires B, and you are interested in average lifetime of each.
Of interest is $\mu_A - \mu_B$ – Driving each tire on different cars may not have same conditions, need to create the same conditions for tires in order to make an inference about the respective average lifetimes. Mount Tire A_x, & B_x on the back of the same car. This makes the two samples dependent.

Tire A, $n_1 = 5$
Tire B, $n_2 = 5$
Drive each car for 4000 miles, and measure wear.
Wear test Results:

Tire A	Tire B	differences
10.6mm	10.2mm	0.4
9.8mm	9.4mm	0.4
12.3mm	11.8mm	0.5
9.7mm	9.1mm	0.6
8.8mm	8.3mm	0.5
		$\bar{d} = \quad 0.48$
		$s_d = \quad 0.0837$

Take differences – treat as a sample of n=5 coming from ONE population:
H_0 : $\mu_A - \mu_B = 0$ - replace with H_0 : $\mu_d = 0$
H_a : $\mu_A - \mu_B \neq 0$ - replace with H_a : $\mu_d \neq 0$

With $\alpha = 0.05$, $\alpha/2 = 0.025$

Test statistic: $\mathcal{T} = \frac{(\bar{d} - 0)}{s_d/\sqrt{n}} = \frac{(0.48 - 0)}{.0837/\sqrt{5}} = 12.87$

Critical value at level α: $t_{\alpha/2(n-1)} : 2.776$

P-Value = $2P(t_{\alpha/2(n-1)} > |\mathcal{T}|)$: (SUCH a large $\mathcal{T}$ implies a very very small probability – near zero – too small to be in the tables)

Conclusion:
Critical Value Approach: Reject H_0 if $\mathcal{T} > t_{\alpha/2(n-1)}$ or $\mathcal{T} < -t_{\alpha/2(n-1)}$
P-value approach: Reject H_0 if P-value $\leq \alpha$