Statistics Ch 3 notes

Ch 3.1 Definitions and Terms:

Sample space (S): Total possible outcomes of a procedure. The outcomes are by chance and equally possible.
Event: outcome or results of a procedure. A, B, C ..
P(A) : probability of event A occurring.
Possible values for Probabilities
$0 \le P(A) \le 1$ (between 0 and 1, inclusive)
$P(A) \le 0.05$, A is unlikely.
P(A) = 1, A is certain, P(A) = 0, A is impossible
P(A) = 0.5 : A has a 50-50 chance.

Three Approaches to find probability of an event A.
Approach 1: **Theoretical probability**

$$P(A) = \frac{number\ of\ ways\ A\ occurs}{number\ of\ ways\ in\ sample\ space}$$

Example:
Ex1.: Select one card from a standard deck.
P(Heart) = 13/52 = ¼ = 0.25 = 25%
Ex 2: In a batch of 6500 light bulbs, 80 are defective.
Select one light bulb from the batch
P(defective) = 80/6500 = 0.0123
P(good) =(6500-80)/6500 = 0.988
Ex3: Find P(2 boys in three children family) = $\frac{3}{8}$
because sample space= {bbb, bbg, bgb, bgg,
gbb, gbg, ggb, ggg}

Approach 2: **Relative frequency approximation**

$$P(A) = \frac{number\ of\ times\ A\ occured}{number\ of\ times\ the\ procedure\ were\ repeated}$$

The probability is an estimate of chance.

Ex1: In year 2020, 86% of people use social media at least once per day. If one person is randomly selected, find the probability that the person uses social media at least once per day.
P (uses social media) = 0.86 (the relative frequency)

Ex2: A sample of 568 students shows 320 students are full-time student. If one student is selected from the whole population, find the probability that the student selected is a full-time student.
P(full-time) = 320/568=0.563

Ex3: A batch of seed results in 78 white flowers and 65 pink flowers. Find probability of getting a white flower from planting the same type of seed.
P (white flower) = 78/(78+65) = 0.545

Ex4. Play Monty Hall game, what strategy gives a higher chance of winning?
http://www.shodor.org/interactivate/activities/SimpleMontyHall/

Law of Large number: As the procedures are repeated more and more, the approximation will get close to the real probability.

Approach 3: **Subjective approach**
Use knowledge of the relevant circumstance to estimate the probability. May not be accurate.
Example: P(stuck in an elevator) = ??
This will probably be unlikely to occur, so Probability will likely be lower than 0.05.

Rounding and probability format:
Round to 3 significant digits unless fraction is a simple fraction of a/b where a, b are less than 10.
Use percentage only when communicating result to be the general public. Most software and professional journal use decimal notation.

Complement of Event A:
$\bar{A}$: event A does not occur, complement of A.

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

Ex: There is 20% chance of rain today. What is the probability of not rain today?
P(not rain) = 1 – P(rain) = 1 – 0.2 = 0.8

"OR" of two simple events.
An outcome is in A or B if the outcome is in A or in B or both.
P(A or B) = $\frac{sum\ of\ numbers\ of\ ways\ A,B,and\ both\ A\ and\ B}{Total\ number\ of\ ways}$

"AND" of two simple events.
An outcome is in A and B if the outcome is in both A and B.
P(A and B) = $\frac{numbers\ of\ ways\ both\ A\ and\ B\ occur}{Total\ number\ of\ ways}$

Conditional probability
An event written as A given B is a conditional probability that A will occur given that B has already occurred.
P(A|B) = $\frac{numbers\ of\ ways\ both\ A\ and\ B\ occur}{Total\ number\ of\ ways\ that\ B\ occurs}$
P(A Given B) = P(A|B) = P(A and B)/ P(B)

Ex1. Toss a 6-face die once, find the probability that the outcome is
a) a "four"
b) a "four" or "five"
c) a "four" and "five"
d) a "four" given that the outcome is an "even" number.
e) a "prime number"

a) 1/6, b) 2/6, c) 0, d) 1/3, e) ½

Ex2. A marble jar has 5 red, 3 blue and 7 white marbles. If one marble is randomly selected, find
a) P(red)
b) P(not red)
c) P(red or blue)
d) P(red and blue)
e) P(red given blue)

a) 5/15 = 1/3 = 0.333
b) 10/15 = 0.667
c) 8/15 = 0.533
d) 0 e) 0

Ex3. One card is drawn from a standard deck, find the following probability:
a) P(black) = 26/52 = 0.5
b) P(black and A) = 2/52 = 0.0385
c) P(four) = 4/52 = 0.0769
d) P(black or A) = (26+2)/52 = 0.5385
e) P(king given black card) = 2/26 = 0.0769
f) P(A | diamond) = 1/13 = 0.0769
g) P(not face card) = 40/52 = 0.7692

Contingency table: (two-way table)
A table used to summarize two categorical variables of a set of data.

P(A) $= \frac{\text{total counts in a row or column}}{\text{Grand total}}$

P(A and B) = $\frac{\text{one entry intersection of } A,B.}{\text{Grand total}}$

P(A or B) = $\frac{\text{Sum of counts for } A, B}{\text{Grand total}}$ (do not double count)

P(A given B) = $\frac{\text{Count of one entry intersection of } A,B}{\text{sum of colum } B \text{ or row } B}$

Ex1. Given the contingency table below, select one.

gender	Question
F	Y
F	N
F	N
M	Y
M	Y
M	Y
M	N
M	N
M	N
M	N

gender/ question	Yes	No
F	1	2
M	3	4

Grand Total = 10
P(F) = (1+2)/10 = 0.3
P(No) = (2+4)/10 = 0.6
P(M and No) = 4/10 = 0.6
P(F and Y) = 1/10 =0.1

P(F or Y) = (1+2+3)/10 = 6/10 = 0.6
p(M or Y) = (3+4+1)/10 = 8/10 = 0.8
= (3+4)/10 + (1+3)/4 - 3 /4 = 0.8
P(Yes GIVEN subject is male) = P(Y|M) = 3/7

Drug or clinical Diagnostic Test

	Positive	Negative
Uses Drugs	correct	false neg
Does not use drugs	false pos	correct

False positive: subject does not use drug but get a positive result.
False negative: subject uses drug but test does not detect it.

	Positive	Negative
Has disease	correct	false neg
not have disease	false pos	correct

False positive: subject is not sick but get a positive result.
False negative: subject is sick but test does not detect it.
Ex1. Use the contingency table below, select one.

	Positive	Negative
Uses Drugs	10	8
Does not use drugs	4	400

GT =422

Find P(positive) = (10+4)/422 = 0.033
Find P(positive given that subject uses drugs)
= 10/(10+8) = 0.556
note: If the subject uses drug, there is a higher chance of getting positive result.

Ch 3.2 Mutually exclusive events and Independent events

Mutually exclusive events:

Two events are mutually exclusive (disjoint) if they will not occur at the same time. P(A and B) = 0

Ex1: Toss a 6-face die, determine if the following are mutually exclusive:
a) Getting a "four" and "even"
No, four and even can occur at the same time
b) Getting a "four" and "five"
Yes, four and five cannot occur at the same time so four and five are mutually exclusive.
Ex2:

	Yes	no	no opinion
M	1	4	2
F	5	2	0

P(F and no opinion) = 0
F and no opinion are mutually exclusive.

Independent Events:

Two events are independent when any of the following is true:

P(A and B) = P(A) ·P(B)

P(A) = P(A | B)

P(B) = P(B | A)

Ex1. Given a 2-way table below:

	iPhone	Samsung	others
male	18	2	1
female	24	4	2

GT = 51

Is iPhone uses independent on gender?

Check if P(iPhone | female) = P(iPhone)

a) Find P(iPhone) = (18+24)/51 = 0.824

b) Find P(iPhone | female) = 24/30 = 0.8

P(iPhone) and P(iPhone Given female) are not exactly equal, so iPhone use is not independent of gender. We can conclude that iPhone use is not independent on gender. Gender affect the choice of iPhone.

Ex2. Given a 2-way table below

	Right-handed	Left-handed
Male	43	9
Female	44	4

Grand total = 100

Is Right-handedness independent on gender?

Check if P(M and R) = P(M) · P (R)

P(M) = 52/100 = 0.52

P (R) = (43+44)/100 = 87/100 = 0.87

P(M and R) = 43/100 = 0.43

But P(M) · P(R) = 0.52 (0.87) = 0.45

So they are not exactly equal.

We can conclude they are not independent.

Note: we will visit this again in Ch 11 to take into consideration of sampling variation.

Note: mutually exclusive events are not necessarily independent events. They are two different concepts.

Ch 3.3 Addition and Multiplication Rule

Addition Rule:

Addition Rule are used to find "OR" in a procedure.

P(A or B) = P(A) + P(B) – P(A and B)

If A and B are mutually exclusive: P(A and B) = 0

P(A or B) = P(A)+ P (B) when A, B are mutually exclusive.

Ex1. Toss a 6-face die once, use addition rule method to find P(one or odd).

P(one or odd) = P(one) + P(odd) – P(one and odd)

= 1/6 + 3/6 – 1/6 = 3/6 =0.5

Ex2. Toss a 6-face die once, use addition rule method to find P(one or even)

P (one or even) = P(one) + P(even)

= 1/6 + 3/6 = 0.667

Ex3. Use the contingency table below:

	iPhone	Samsung	others
male	18	2	1
female	24	4	2

GT = 51

Use addition rule to find P(male or iphone).

P(male or iphone) = P(male) + P(iphone) – P(male and iphone) = 21/51 + 42/51 – 18/51 = (21+41-18)/51 =45/51 = 0.8823

Multiplication Rule:

Multiplication Rule is used to find probability of two events: A and B.

$P(A \; and \; B) = P(A) \cdot P(B|A)$

If A and B are independent, P(B|A) = P(B) so

$P(A \; and \; B) = P(A) \cdot P(B)$ when A, B are independent.

A result of the multiplication rule gives the formula for conditional probability as:

$P(A|B) = \frac{P(A \; and \; B)}{P(B)}$

Ex1: Given the two-way table below:

	iPhone	Samsung	others
male	18	2	1
female	24	4	2

Find P(male |iphone) =

P(male and iphone)/P(iphone) =

$\frac{18/51}{42/51} = 18/42$ = 0.4286

Ch 3.4 Sampling w/wo replacement

Sampling with replacement – selected subjects are put back into the population before another subject are sampled. Subject can possibly be selected more than once.

Sampling without replacement – Selected subjects will not be in the "pool" for selection. All selected subjects are unique. This is the default assumption for statistical sampling.

Compound events involving multiple trials/steps

When events involve multiple steps, they are called compound events A, and then B, the compound event is also called A and B.

But P (A and then B) is not P(A and B) where A and B outcome of one step with 2 categories.

Multiplication rule for two events in two steps:

A and B : Event A occurs in one trial and Event B occurs in another trial.

P(A and B) = P(A) × P(B after A has occurred)

P(A and B) = P(A) × P(B |A)

Independent/Dependent events

1**) Dependent:** occurrence of one event affect the next event.

P(A and B) = P(A) × P(B| A)

2) **Independent:** occurrence of one event does not affect the other.

P(A and B) = P(A) x P(B)

In general : P(A and B and C …) = P(A) x P(B) x P(C) …

Ex1. A fair coin is tossed 5 times. Find P(first 3 tosses are heads) . F(first 4 are heads), P(at least one head in 4 tosses).

Coin tosses are <u>independent</u> events.

P(First 3 tosses are heads) =

P(First 4 tosses are heads) =

P(at least one head in 4 tosses) =

1 – P(all heads are tails)

Ex2. In a group of 300 all adults, 272 are right-handed, 3 adults are selected with replacement.

Find P(all 3 are right- handed) and P(all 3 left-handed).

P(one right-handed) = 272/300 =0.907

P(one left-handed) = (300-272)/300 = 28/300 = 0.093

P(3 right-handed) = (272/300) (272/300)(272/300)

= $(272/300)^3$ = 0.745

P(3 left-handed) = (28/300)(28/300)(28/300)

= $(28/300)^3$ = 0.0008

P(At least one right-handed) = 1 – P(All left-handed)

Ex3. In a jar with 5 red, 6 blue and 2 white marbles. Two marbles are selected, find the probability that both are red if:

a) If two marbles are selected with replacement.

b) If two marbles are selected without replacement.

Ans:

a) If marbles are replaced, the events are independent.

P(both are red) = 5/13 * 5/13 = 0.1479

b) If two marbles are selected without replacement, the events are dependent,

P(both are red) = 5/13 * 4/12 = 1282

Sampling and independent event

Sampling with replacement – independent events

Sampling without replacement – dependent events

Treating Sampling without replacement as independent if one of the following are satisfied:

a) Assume a very big population when population size is not given. Only P(A) is given.

b) **Use 5% guideline** for cumbersome calculations: When sampling without replacement and the sample size is no more than 5% of the size of population, treat sampling as independent. (Even though they are actually dependent.)

Ex1. Assume that 10% of adults in the United states are left handed. Find the probability that three selected adults all are left handed.

Since the population size is not given only P(L) = 0.1 is given, assume sampling are independent.

P(L and L and L) = P(L) × P(L) × P(L) = 0.1 × 0.1 × 0.1

= 0.001

Ex2. In batch of 6400 light bulbs, 80 are defective. If 12 light bulbs are selected from the batch without replacements, find probability that all are good.

Ans:

sample size = 12, population size = 6400

Since sampling proportion = 12/6400 = 0.00188 < 0.05 we can treat the sampling as independent by apply 5% guideline cumbersome calculation.

P (one good) = (6400-80)/6400 = 0.9875

P (one defective) = 80/6400 = 0.0125

P(all 12 are good) = 0.9875^{12} = 0.860

P (at least one defective) = 1 – P(all 12 are good)

Rare Event Rule

If, under a given assumption, the probability of a particular observed event is very small and the observed event occurs significantly less than or

greater than what we typically expect with that assumption, we conclude the assumption is not correct.
P(A) ≤ 0.05, then A is unlikely to occur by chance.

Use Probability to form Conclusion:
If P(A) > 0.05, A can occurs by chance, so there is <u>not sufficient evidence</u> to conclude that "the change" is effective.
If P(A) ≤ 0.05, A **cannot** have occurred by chance, so there is <u>sufficient evidence</u> to conclude that "the change" is effective.

Ex1. A study is done to test if vitamin C intake will reduce common cold. 6 out of 56 subject taking vitamin C catch the common cold compare to 8 out of 56 subjects not taking vitamin C. If vitamin C has no effect, there is a 0.32 chance of getting such sample result. What can we conclude?

Ans: Since probability 0.32 is not unlikely, the result could have occurred by chance rather than due to vitamin C treatment. There is not sufficient evidence to conclude vitamin C is effective.