6/07/20

Agenda for Math 5710 ♬ Meeting #20 ☻☺ 7/20/20 (8:00 a.m. – 9:10 a.m.)

1. Hello:

Brigham City:	Adam Blakeslee	Ryan Johnson	Tyson Mortensen	
Logan:	David Allen	Natalie Anderson	Kameron Baird	Stephen Brezinski
	Zachary Ellis	Adam Flanders	Brock Francom	Xiang Gao
	Ryan Goodman	Janette Goodridge	Hadley Hamar	Phillip Leifer
	Brittney Miller	Jonathan Mousley	Erika Mueller	Shelby Simpson
	Steven Summers	Matthew White	Zhang Xiaomeng	

2. Note the syllabus' activity list for today:

20 M/7/20	1. Construct the following concept, comprehend associated communication structures, and employ associated algorithms: odds, uniform discrete distribution functions, population statistical parameters, expected values of discrete random variables, and variance of discrete random values 2. Take advantage of Quiz 20

3. Briefly raise issues and questions prompted by the following homework assignment:

A. Study our notes from Meeting #19.

B. Comprehend Jim's sample response to Quiz 19.

C. Comprehend Entry #038 & #39A–B of our *Glossary*.

D*. Please solve the following problems; display the computations, and upload the resulting pdf document on the appropriate Canvas assignment link:

> The diameter of flat metal disk manufactured by a factory is a random number between 4 and 4.5. What is the probability that the area of such a flat disk chosen at random is at least 4.41π ?

E. From the Video Page of *Canvas*, view with comprehension the video names "intro to discrete random variables and discrete probability distributions."

F. Comprehend Jim's sample responses to the homework prompts that are posted on *Canvas*.

4. Muse a bit about the following definition we comprehended for homework:

039B. Definition for discrete *uniform probability distributions* :

Given $m \in \{$ discrete distribution functions for X on Ω $\} \wedge |\, \Omega\, | = n \ni n \in \mathbb{N}$,
($m \in \{$ discrete uniform probability distribution functions for X on Ω $\} \Leftrightarrow$
$m(x) = \frac{1}{n} \; \forall x \in X$

5. Ask ourselves what we mean when we refer to odds of an event happening:

A. Based on the estimate that 3.5% of the world population are infected with COVID-19, what are the odds that a randomly selected person will be infected.?

B. Associate our concept of *odds the occurrence of an event* and the following definition:

040. The relationship between probabilities and odds:

Definition for the odds of an event occurring: Given E is an event of Ω ,

(the odds in favor of E occurring = $r : s$ where $r, s \in [\, 0, \infty\,) \Leftrightarrow \frac{r}{s} = \frac{p(E)}{1-p(E)}$)

6. Revisit the world of statistical population parameters before we return to Math 5710 topics; do so as we run through Items #7–10 of this agenda.

7. Consider the question: How large is 27? Is it small or is it large?

A. Analyze the following problem:

i. Theresa responds to the prompts on a test resulting in a raw score of 27. Interpret that measurement result (i.e., 27).

ii. Does it help to know that for that same test, the score associated with Eva's responses is 21?

iii. Does it help to see a string of scores based on the responses of 37 other students?

Table I: Norm-Group Scores for the Test Theresa Took

i	Student	x_i	i	Student	x_i
1	A.J.	41	20	Jose	39
2	Arlene	10	21	Joy	31
3	Bill	42	22	Lucinta	43
4	Bonita	30	23	Luke	34
5	Buddy	17	24	Michael	18
6	Buhla	38	25	Melinda	12
7	Cameron	10	26	Nadine	37
8	Candice	34	27	Nettie	40
9	Celeste	16	28	Ory	23
10	Charles	37	29	Phyllis	28
11	Eloise	35	30	Quinn	33
12	Elvira	36	31	Raul	18
13	Evelyn	25	32	Ricardo	39
14	Garret	34	33	Rufus	39
15	Gilda	39	34	Sidney	14
16	Hans	32	35	Singh	21
17	Hilda	29	36	Tomaria	37
18	Ira	40	37	Zelda	40
19	Jacob	35			

iv. Is it convenient to view that string of scores in descending order?

Table II: Table I's Scores in Descending Order

i	Rank	x_i	i	Rank	x_i
22	1	43	23	19	34
3	2	42	30	21	33
1	3	41	16	22	32
18	5	40	21	23	31
27	5	40	4	24	30
37	5	40	17	25	29
15	8.5	39	29	26	28
20	8.5	39	13	27	25
32	8.5	39	28	28	23
33	8.5	39	35	29	21
6	11	38	24	30.5	18
10	13	37	31	30.5	18
26	13	37	5	32	17
36	13	37	9	33	16
12	15	36	34	34	14
11	16.5	35	25	35	12
19	l6.5	35	2	36.5	10
8	19	34	7	36.5	10
14	19	34			

v. What if we had a second string of norm scores to interpret 27?

Table III: A Second String of Norm-Group Scores to Which Theresa's Score Is Compared

i	Rank	w_i	*i*	Rank	w_i
9	1	45	6	16	19
19	2	37	10	18	16
1	3	34	20	18	16
13	4.5	33	14	18	16
30	4.5	33	8	20	15
12	6	30	2	21	13
29	7	29	27	22	12
7	8	28	17	23.5	10
23	9.5	27	15	23.5	10
25	9.5	27	16	25.5	9
18	12	24	26	25.5	9
24	12	24	21	27	8
4	12	24	5	28	3
3	14.5	20	22	29.5	0
28	14.5	20			

v. Let's briefly turn our attention away from interpreting Theresa's score of 27 and toward the question of interpreting a population's score.

8. Given a data string $D \ni D = (x_1, x_2, x_3, ..., x_{n-1}, x_N)$, wrap our minds around three conventional measures of central tendency and the rational for each:

A. The median of *D*

B. The mode(s) of *D*

Sidebar Musing:

While designing algorithms for quantifying an entire data string (e.g., X to Y as rewritten below, why not compare the two sums?

X = (33, 32, 32, 30, 29, 29, 28, 27, 27, 25, 25, 25, 25, 24, 24, 22, 21, 20, 18, 18) $\wedge$

Y = (70, 63, 58, 55, 49, 42, 42, 36, 31, 29, 24, 22, 18, 14, 10, 10, 4, 1, 0, 0)

In other words, how reasonable would it be to suggest that the scores from X are less than the scores from Y because of the following:

$$33+32+32+30+29+29+28+27+27+25+25+25+25+24+24+22+21+20+18+18 = 515 \ \wedge$$

$$70+63+58+55+49+42+42+36+31+29+24+22+18+14+10+10+4+1+0+0 = 578$$

Letting $a_1 = 33$, $a_2 = 32$, $a_3 = 32$, $a_4 = 30$, ..., $a_{19} = 18$, $a_{20} = 18$ $\wedge$

$b_1 = 70$, $b_2 = 63$, $b_3 = 58$, $b_4 = 55$, ..., $b_{19} = 0$, $b_{20} = 0$,

we could restate the above comparison using some shorthand:

$$\sum_{i=1}^{20} a_i = 515 \ \wedge \ \sum_{i=1}^{20} b_i = 578$$

Now let's use our summing algorithm to comparing X to a third data string W where W = (88, 63, 63, 44, 17).

So with $w_1 = 88$, $w_2 = 63$, $w_3 = 63$, $w_4 = 44$, and $w_5 = 17$, we have $\sum_{i=1}^{5} w_i = 275$

Are we ready to conclude that the scores from X are greater than those from W since 275 < 515 ?

So maybe our summing algorithm needs some adjustments to account for difference in the number of scores or numbers.

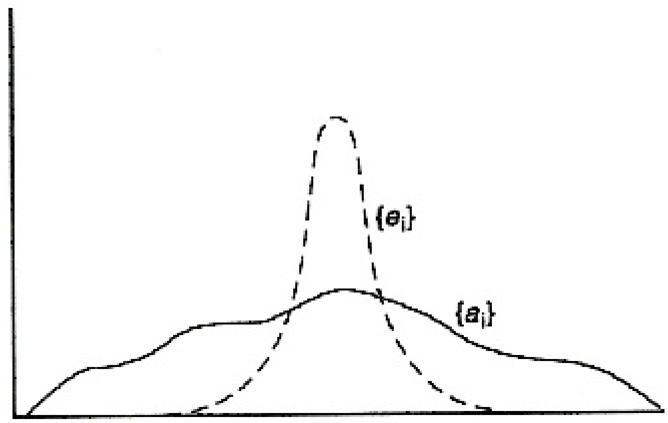
C. The arithmetic mean of X (i.e., μ_X) defined as follows:

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

> Sidebar note:
>
> Note the distinction between the parameter statistic μ_X and the sample statistic $\bar{X}$ as well as the conventional distinction between N and n.

9. Focus our attention on developing a norm-referenced units for comparison – measures of variability:

    A. Which of the two graphs reflects the data string with greater variability (i.e., heterogeneity), { e_i } or { a_i }?



    B. So let's develop a reasonably valid, but necessarily unusable, algorithm for metric that reflects variability:

b_i
46
40
38
30
24
20
16
16
10
0

46–40=6								
46–38=8	40–38=2							
46–30=16	40–30=10	38–30=8						
46–24=22	40–24=16	38–24=16	30–24=6					
46–20=26	40–20=20	38–20=18	30–20=10	24–20=4				
46–16=30	40–16=24	38–16=22	30–16=14	24–16=8	20–16=4			
46–16=30	40–16=24	38–16=22	30–16=14	24–16–8	20–16=4	16–16=0		
46–10=36	40–10=30	38–10=28	30–10=20	24–10=14	20–10=10	16–10=6	16–10=6	
46–0=46	40–0=40	38–0=38	30–0=30	24–0=24	20–0=20	16–0=16	16–0=16	10–0=10

So the number of pairwise differences for 10 scores is 45 (i.e., $\binom{10}{2} = 45$). The median of the pairwise difference is 16; the modes are 6, 8, 10, 14, 16, and 30. And the mean of those pairwise differences is approximately 17.33.

What do we think? Valid? Usable?

C. So let move to something more usable but nearly as valid. Ask ourselves: Instead of comparing each score to each other score, what if we identified a metric that quantifies the whole string of scores as one number? If we did that then we could reduce the number of differences we need to compute from $\binom{N}{2}$ to *N*.

D. Develop mean deviation (i.e., MD_x)

E. Try on this one: $MD_x = \frac{1}{N}\sum_{i}^{N}|x_i - \mu_X|$

F. Let's yield to convention and develop standard deviation (i.e., σ_X).

> Sidebar note:
> Note the distinction between the parameter statistic σ_X and the sample statistic *s*.

G. Try this on for size: $\sigma_X = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \mu_X)^2}$

10. Focus our attention on *z*-score.

A. Note that the number of standard deviations a number in a data string (e.g., a raw score on a test) is greater than the arithmetic mean of the comparison group is a *z*-score.

B. Suppose *G* is a data string ∋ ($\mu_G = 40 \ \wedge\ \sigma_G = 12$), compute:

i. z_{40}

ii. z_{28}

iii. z_{34}

iv. z_{48}

v. z_{43}

C. Adopt the following definition iff it is consistent with our concept of z-score:

Given data string $X \ni,\ X = (\, x_1,\ x_2,\ x_3, \ldots,\ x_N \,)$,

$$z_{xi} = \frac{x_i - \mu_x}{\sigma_x}$$

11. Upon our return to Math 5710 topics, examine the concept of expected value of a discrete random variable:

A. Begin by revisiting Staggerlee's problem with an eye toward the mean of the random variable:

Staggerlee wants to conduct a coin-flipping experiment for the purpose of determining the probabilities of randomly obtaining various events when a fair coin is flipped exactly three times in succession. He plans to use the resulting probability distributions to hedge his bets in a variety of games of chance. Please design the experiment for him so that it yields probability values for the each of the following events: X_j is the event in which exactly j tails turn up for $j \in \{\, 0, 1, 2, 3 \,\}$. Describe the experiment – identifying the sample space and discrete probability distribution.

Sample description:

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

Let X_j = the event that there exactly j tails. Thus,

$| X_0 | = | \{$ HHH $\} | = 1$

$| X_1 | = | \{$ HHT, HTH, THH $\} | = 3$

$| X_2 | = | \{$ TTH, THT, HTT $\} | = 3$

$| X_3 | = | \{$ TTT $\} | = 1$

Let p be our random probability function. Since $|\Omega| = 8$, we have the following probability values:

$$p(X_0) = \frac{1}{8} \wedge p(X_1) = \frac{3}{8} \wedge p(X_2) = \frac{3}{8} \wedge p(X_3) = \frac{1}{8}$$

B. Comprehend the following definition and relate it to μ_X :

041A. Definition for *expected value* for a discrete random variable X:

Given $X \in$ { discrete random variables of Ω } $\wedge$
($m \in$ { discrete probability distribution for X },
($E(X)$ is the *expected value* of $X \Leftrightarrow E(X) =$
$\sum_{x \in \Omega} x m(x)$ provided that the series converges absolutely).

Note: If the series does not converge absolutely, then X does not have an expected value.

12. Comprehend the following definitions and relate them to σ_X^2 and σ_X :

41B. Definition for *variance* for a discrete random variable X:

Given $X \in$ { discrete random variables of Ω } $\wedge$ the expected value of X is $E(X)$,
($V(X) =$ *variance of* $X \Leftrightarrow V(X) = E((X - E(X))^2))$

41C. Definition for *standard deviation* for a discrete random variable X:

Given $X \in$ { discrete random variables of Ω } $\wedge$ the variance of X is $V(X)$,
($D(X) =$ *standard deviation of* $X \Leftrightarrow D(X) = \sqrt{V(\mathrm{X})}$

13. Take advantage of Quiz 20.

14. Complete the following assignment prior to Meeting #21:

A. Study our notes from Meeting #20.

B. Comprehend Entry #041A–C of our *Glossary*.

C. From the Video Page of *Canvas*, view with comprehension the video names "expected value and variance of discrete random variables (better)."

D. Re-analyze our old friend Sybil's problem and ponder the following questions:

i. What is the random variable X ?

ii. How does μ_X and σ_X^2 compare to $E(X)$ and $V(X)$?

And for the sake of convenience, here is a copy of Sybil's problem:

For the purpose of formulating the rules of a game of chance in which a pair of fair dice (a red die and a yellow die) are rolled one time, Sybil wants to determine the likelihood of each of the possible *events* determined by the sum of the number of dots that appear on the top face of the yellow die and on the top face of the red die.

> Sybil thinks, "Each die has six faces – a face with one dot, a face with two dots, a face with three dots, a face with four dots, a face with five dots, and a face with six dots. So there are 36 possible *outcomes* since 36 is the cardinality of the following set:
>
> { (r, y) : r = the number of dots on the red die's top face $\wedge$
> y = the number of dots on the red die's top face } =
> { (1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6),
> (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6),
> (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5),
> (6, 6) }
>
> And since I'm interested in the *random probability* of each of the possible *events*, the *events* of interest are the sum of numbers associated with the two top faces. So each event is associated with one element in the following set: { 2, 3, 4, 5, ..., 12 }. I'll count the number outcomes are associated with each of the 12 events:
>
> Let X_i = the event that the sum is i. Thus,
> $|X_2|$ = \|{ (1, 1) }\| = 1
> $|X_3|$ = \|{ (1, 2), (2, 1) }\| = 2
> $|X_4|$ = \|{ (1, 3), (2, 2), (3, 1) }\| = 3
> $|X_5|$ = \|{ (1, 4), (2, 3), (3, 2), (4, 1) }\| = 4
> $|X_6|$ = \|{ (1, 5), (2, 4), (3, 3), (4, 2), (5, 1) }\| = 5
> $|X_7|$ = \|{ (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1) }\| = 6
> $|X_8|$ = \|{ (2, 6), (3, 5), (4, 4), (5, 3), (6, 2) }\| = 5
> $|X_9|$ = \|{ (3, 6), (4, 5), (5, 4), (6, 3) }\| = 4
> $|X_{10}|$ = \|{ (4, 6), (5, 5), (6, 4) }\| = 3
> $|X_{11}|$ = \|{ (5, 6), (6, 5) }\| = 2
> $|X_{12}|$ = \|{ (6, 6) }\| = 1
>
> So the most likely event is X_7 and the two events that are less likely than any of the others are X_2 and X_{11}. And my chart above shows the scale for each events. Then I'll use the chart to compute the exact theoretical probabilities by reporting the probabilities as the ratio of each of those events and 36 (i.e., $p(X_i) = X_i \div 36$."

---

15. And from Charles Babbage (1791–1871): *Errors using inadequate data are much less than those using no data at all.*