GRADE 12 APPLIED
UNIT D – STATISTICS WORKBOOK

Name: _____

Round decimal answers to nearest 0.01 where appropriate. **SHOW YOUR WORK**

Finding the Central Tendency of a set of samples or collected numbers

1. Manually (without an App) find the sample size, n, the sum of the data, $\sum x$, and the mean, $\bar{x}$, of the following given data sets.

a. {6 6 8 2}

n = 4 $\sum x$= 22

$\bar{x}$= 22/4 = 5.50

b. {3 4 9 7 2 7}

n = $\sum x$=

$\bar{x}$=

c. {7 7 5 6 3 1}

n = $\sum x$ =

$\bar{x}$ =

d. {2 0 9 5 9 1}

n = $\sum x$ =

$\bar{x}$ =

e. {2.5, 0.2, 9.0, 4.9, 8.8, 0.7}

n = $\sum x$=

$\bar{x}$=

f. {3.2, 5.1, 0.3, 6.2, 1.4, 6.8}

n = $\sum x$=

$\bar{x}$=

g. {4.9, 8.8, 3.9, 8.4, 1.7, 8.5, 7.0, 7.6, 4.0}

n = $\sum x$=

$\bar{x}$=

h. { 4, 1, $^{-}$8, $^{-}$9, 5, 5, $^{-}$9, $^{-}$4, 2, 1, 2, $^{-}$6, 1 }

n = $\sum x$=

$\bar{x}$=

Do you get the sense this ⇑ could be burdensome and error prone?
Check your answers using an App or maybe your calculator does Stats too
PS: I have never figured out how to do it on a basic calculator!
PPS: Have you noticed the mean has to be in between the smallest and the largest values of *x*?

i. {4, 4, 5, 5, 5, 5, 5, 6, 6, 7, 7, 7, 8, 8, 9}

n = $\sum x$=

$\overline{x}$=

j. {7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7}

n = $\sum x$=

$\overline{x}$=

k. {2, 2, 2, 2, 2, 2, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5,6, 7, 7, 7, 7, 7, 7, 7, 7, 7, 8, 8 }

n = $\sum x$=
$\overline{x}$=

What if we '*trimmed'* off the highest five numbers and the lowest five numbers?

n = $\sum x$=

$\overline{x}$=

m. {2.1, 2.1, 2.1, 2.1, 2.1, 2.1, 2.2, 2.2, 2.2, 2.2, 2.2, 2.2, 2.2, 2.3, 2.4, 2.4, 2.4, 2.4, 28.5}

n = $\sum x$=
$\overline{x}$=

That 28.5 looks out of place, what if we ignored it? Recalculate if you trim it out.
n = $\sum x$=

$\overline{x}$=

By how much did the mean change?

2. Can you think of an easier way to do these types of boring calculations if you have lots of the same number? Flip to the end section called **Frequency Distribution Table**, you might find that easier.

3. Does your calculator have statistical buttons on it? (Bet it does! Even $2 calculators have statistical buttons, unfortunately I have never figured out how to use them). Or maybe you found a good app for your smartphone. (which you are not allowed to use anyway if you have to show your work!)

4. Manually, without an App, find the sample size, n, the sum of the data, $\sum x$, the mean, $\bar{x}$, and the median, $\tilde{x}$, and mode(s) of the following sets of numbers, x. If there are two (or more modes) state them all.

a. {6 6 9 2 5}

n = 5 $\sum x$ = 28

$\bar{x}$ = 28/5 = 5.60

$\tilde{x}$ = {2, 5, (6) 6, 9}
$\tilde{x}$ = 6

mode(s) =

b. {3 4 9 7 2 7 4}

n = $\sum x$ =

$\bar{x}$ = $\tilde{x}$ =

mode(s) =

c. {7 7 5 6 3 1}

n = $\sum x$ =

$\bar{x}$ = $\tilde{x}$ =

mode(s) =

d. {2 0 9 5 9 1}

n = $\sum x$ =

$\bar{x}$ = $\tilde{x}$ =

mode(s) =

e. {2.5, 0.2, 9.0, 4.9, 8.8, 0.7}

n = $\sum x$ =

$\bar{x}$ = $\tilde{x}$ =

mode(s) =

f. The temperatures in Winnipeg in early March:
{6, 4, 4, 5, 4, −5, −5, 6, 4, −7, −7, 7, 8, −7, −9}

n = $\sum x$ =

$\bar{x}$ = $\tilde{x}$ =

mode(s) =

g. {7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7}

n = $\sum x =$

$\bar{x} =$ $\tilde{x} =$

mode(s) =

h. {2.3, 2.5, 2.1, 2.1, 2.1, 2.4, 2.2, 2.2, 2.2, 2.0, 2.0, 2.1, 2.2, 2.3, 2.4, 2.4, 2.4, 2.4, 58.5}

n = $\sum x =$

$\bar{x} =$ mode(s) =

That 58.5 looks out of place, what if we ignored it?

n = $\sum x =$

$\bar{x} =$ mode(s) =

By how much does knocking off that 58.5 change:
The median? _____ The Mode? ____
The Mean?_____

Now check the above ↑ with an App of some sort. The TI-83 one preferably. DESMOS does it too, but not as well yet.

WORD PROBLEMS

5. Your daughter's nursery school is looking for funding. The government needs to know some statistics about the school. You are in the parent teacher association and asked to take the ages of your daughter's class (in whole years). Determine the statistics manually and then check with an App.

Nancy 3	Kelvin 4	Diane 4	Michelle 5	Betty 3
Karen 3	Debbie 4	Charlene 5	Charnelle 2	Steph 2
Carla 2	David 4	Brandon 6	Brian 4	Kyle 4

Calculate the following for the application form:

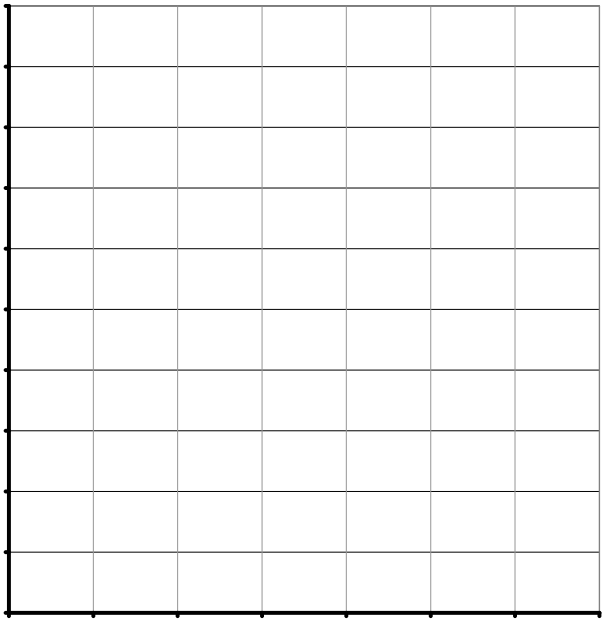
a. number of ages collected; n = ____

b. sum of all the ages, $\sum x =$ ____

c. mean age, $\bar{x} =$ ____ d. median age, $\tilde{x} =$ ____

e. age mode(s) = _____

6. Of course a picture of your Nursery School data is always important to see how data is distributed or if there is any 'lopsided-ness' (also known as '*skewed'* data) that favours one end or the other of the range of data and that might throw off the central values or make some irregular 'clumping' of data.
Graph a histogram here and indicate where the statistics would be displayed on the histogram.



Now do a histogram on a Statistics App or in a spreadsheet or in DESMOS or on the Texas Instrument Graphing Calculator and compare.

7. You have started to record the time that the bus arrives at your bus stop. The variable you measure is how many whole minutes after 8 am. Here is the data you collected after a month of school. Do this manually without an App.

8	8	9	10	7	7	8	9
10	8	4	3	9	8	9	6
10	2	5	45				

Calculate:

a. number of times sampled; n = ____

b. sum of all the sampled arrival times (after 8 am), $\sum x =$ ____ minutes

c. mean arrival time, $\bar{x} =$ _____ minutes

d. median arrival time, $\tilde{x} =$ _____ minutes

e. arrival time mode(s) = _____ minutes

There was a very late bus the last day of your samples (collected times). Which central tendency statistic is most affected by the late arriving bus?? _____

Does the data make more sense if you graph it?

Sketch a histogram here→

By **sketch** we mean ' back of a cigarette pack' type sketch. Not exact, not to scale, just a rough diagram.

Now compare the stats and the histogram with whatever technology you prefer.

8. An archaeologist discovered an ancient indigenous meeting ground. He found a hundred arrowheads and measured each one to the nearest 0.1 cm. Here is what he got:

4.6	4.2	4.4	4.8	4.6	4.6	4.5	4.6	4.8	4.7
4.1	4.4	4.1	4.1	4.8	4.1	4.4	4.5	4.4	4.7
4.3	4.2	4.3	4.2	4.5	4.6	4.2	4.8	4.4	4.2
4.8	4.0	4.5	4.8	4.4	4.6	4.7	4.7	4.8	4.4
4.8	4.1	4.1	4.4	4.4	4.2	4.7	4.8	4.1	4.4
4.3	4.2	4.1	4.4	4.6	4.0	4.0	4.3	4.8	4.1
4.2	4.1	4.1	4.3	4.5	4.3	4.4	4.4	4.7	4.4
4.4	4.4	4.1	4.2	4.4	~~44.4~~	4.8	4.7	4.7	4.7
4.8	4.1	4.2	4.8	4.4	4.3	4.1	4.7	4.3	4.5
4.6	4.4	4.2	4.4	4.3	4.4	4.5	4.2	4.6	4.5

That 44.4 looks rather rogue, an '*outlier*', his assistant probably just wrote it down wrong so the scientist discards it!

Calculate:

a. number of arrowheads sampled; n = ____

b. sum of all the sample lengths, $\sum x$ = ____

c. mean length, $\bar{x}$ = ____ median length, $\tilde{x}$ = ____

d. length mode(s) = _____

Why should you be cautious throwing out (discarding) some data?

Try this using the FREQUENCY TABLE METHOD at the end of this workbook, bet it is easier.

Now use technology to calculate the arrowhead statistics and to do a histogram. Discard the rogue 44.4 though. Compare the results. Sketch the histogram below

TRIMMED MEANS

9. Calculate a 10% trimmed mean (10% off each end) for the following:

a. {1, 5, 6, 7, 8, 3, 5, 2, 10, 5}

b. {3, 3, 6, 7, 8, 9, 1, 3, 6, 5, 2, 3, 1, 4, 5, 7, 7, 2, 1, 7}

c. {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}

Calculate a variable (an unknown amount)

10. Jason wants a pass mark of 75% for his course so he will not have to do remedial math when he gets to post-secondary. Presently he has the following marks: 70%, 80%, 60% and 70%. What mark will he need to get on his final test? (all marks are equally weighted)

11. Cameron wants to jog an average of 6 km per day for the entire week. The first six days she jogs: 5 km, 3.5 km, 7 km, 0 km, 10 km, and 4 km. How far does she have to jog on day 7 to get an average of 6 km per day.

12. Rick loves curling. Normally he makes 75% of his shots; a success rate of 75%. However, this morning he is having a bad morning! (maybe because he skipped breakfast?). In his first 10 rocks he threw he made 6 successful shots. How many successful shots must he make in his last 6 rocks to maintain a decent 75% success rate? (a player normally throws 16 rocks a game).

WEIGHTED MEANS

13. Quiz Marks of 45%, 65%, 75%, and the 4^{th} mark of 90% was a test worth **twice as much** as any of the quiz marks. Calculate the overall mark as a weighted mean.

14. Quiz Marks of 75%, 65%, 75%, and the 4^{th} mark of 40% was a test worth three times as much as any of the quiz marks. Calculate the overall mark as a weighted mean.

15. Jason wants a good overall mark to go to university to take bio-chemistry. When the institution is selecting candidates for admission to the course they use a **weighted mean**. English and History are not as important to a scientist as Math and Science marks. So the institution applies the following **weight factors** to marks from school transcripts. English **10%**; History **10%**; Biology **40%**; and Mathematics **40%** weight factors. Jason's marks for those key courses were respectively: **65%**; **50%; 88%**, and **92%**.

a. What will be the **weighted mean mark** that the institution uses when considering Jason's admission to the Biology program?

b. the institution requires that for admission to the select Bio-Chemistry that Jason's weighted mean is 85%. Will he be accepted into the program?

16. The Education Director of our college wants to know the overall mean mark of all students in English and Math combined. Courtney has 30 students and says the mean mark for her classes is 75%. Rick has 50 students and says the mean mark of all his students is 60%. What is the average (mean) mark of the Math and English marks for both classes combined?

Find the variable value in a weighted mean

17. Bruce wants to be admitted to a Business Course at university. He needs an overall mark from his school of 80% in mathematics to be accepted. So far math his marks work out to be an average of 76%. He still has the final exam to write, it is worth 25% of his final mark. What mark does Bruce need on his final exam to get and overall mathematics average of 80%?

18. **Find the value of x**:

a. $80 = 0.75*76+0.25*x$

b. $60 = 0.8*80+0.2*x$

c. $50=\frac{45+75+60+2x}{5}$

$250 = 180 + 2x$

$-180 \quad -180$

$70 = 2x$

$x = 35$

d. $80=\frac{45+75+60+3x}{6}$

PERCENTILE RANKING

19. Roberta is in a class of 40 students. Her overall mark for the class is 65%. She wonders how she is doing compared to her classmates. (her progress report would show her if she looked at that). The teacher posted all the marks on–line, but without names. She noticed that she and three other students have the same mark (65%) and that 15 students have a mark lower than hers. What percentile rank is Roberta?

20. Jennifer is preparing for a career in the military, she is getting in shape with lots of sit-ups at a local gym. She knows she can do 45 sit-ups in 2 minutes and thinks she is doing pretty well. She records her performance on a chart on the wall with other members of the gym and notices that 15 people do fewer than her, she is the only one who does 45 in two minutes, and 38 people do better than her. She wonders how she is doing compared to the others.

a. What is Jennifer's percentile rank?
b. In which decile is she?
c. In which quartile is she?

FREQUENCY DISTRIBUTION TABLES

Instead of adding a bunch of the same number together lots of times we use multiplying! (Duh!) Use the nursery school ages in previous pages.

variable value, x **age**	**frequency the value of x happens, f_x**	**product of the variable amount time how many time times** **$x * f_x =$**
age 2	**3 times**	**2 * 3 =** **6**
3	3	9
4		
5		
6		
	n =	$\sum(x * f_x) =$ $\bar{x} = \sum(x * f_x) / n =$

Find the middle data/datum in the table and find the median, $\tilde{x}$?

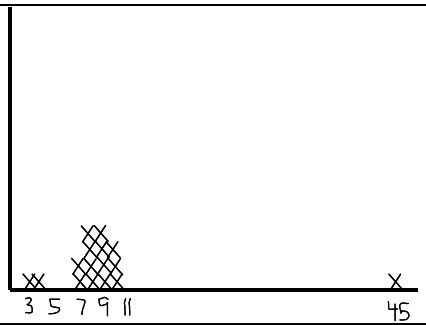
$\frac{(n+1)}{2}$ th position

The mode is much more obvious now! The mode(s) is: _____

A frequency data table to record large samples.

Frequency Data Table (to calculate statistics of large samples)					
x Value of variable being measured	Tally ticks (if necessary)	**f** frequency each value happens [count]	**Cumulative Frequency**	**f*x** *f times x*	
					Mode; most frequent x: _____
					Mean: $\frac{\Sigma f * x}{n} =$
					Median Halfway up the data; in between two values if n is even. _____
					$(n+1)/2^{th}$
		sum: n = _____		sum $\Sigma all\ the\ f * x's$ _____	

ANSWERS

1a. n = 4; Σx = 22; mean = 5.5	1b. n = 6; Σx = 32; mean = 5.33
1c. n = 6; Σx = 29; mean = 2.19	1d. n = 6; Σx = 26; mean = 4.33
1e. n = 6; Σx = 26.1; mean = 4.35	1f. n = 6; Σx = 23; mean = 3.83
1g. n = 9; Σx = 54.8; mean = 6.09	1h. n = 13; Σx = ⁻15; mean = ⁻1.15
1i. n = 15; Σx = 91; mean = 6.07	1j. n = 13; Σx = 91; mean = 7
1k. n = 30; Σx = 143; mean = 4.77 Trimmed: n = 20; Σx = 96; mean = 4.8$\underline{0}$	1m. n = 19; Σx = 68.4; mean = 3.6$\underline{0}$ Trim off outlier: n = 18; Σx = 39.9; mean = 2.22 Mean changed by 1.38 lower just by knocking off that one rogue(?) outlier
2a. n = 5; Σx = 28; $\bar{x}$ = 5.60 , $\tilde{x}$ = 6 mode: 6	2b. n = 7; Σx = 36; $\bar{x}$ = 5.14 , $\tilde{x}$ = 4 mode: 4 and 7 (*bi-modal*)
2c. n = 6; Σx = 29; $\bar{x}$ = 4.83 , $\tilde{x}$ = 5.5 mode: 7	2d. n = 6; Σx = 26; $\bar{x}$ = 4.33 , $\tilde{x}$ = 3.5 mode: 9
2e. n = 6; Σx = 26.1; $\bar{x}$ = 4.35 , $\tilde{x}$ = 3.7 mode: nil	2f. n = 15; Σx = 8; $\bar{x}$ = 0.53 , $\tilde{x}$ = 3.7 mode: 4
2g. n = 13 ; Σx = 91 ; $\bar{x}$ = 7 , $\tilde{x}$ = 7 mode: 7	2h. n = 19 ; Σx = 98.8 ; $\bar{x}$ = 5.20 , $\tilde{x}$ = 2.20 mode: 2.1 and 2.2 (bi-modal) Knocking off the outlier 58.5: n = 18 ; Σx = 40.3 ; $\bar{x}$ = 2.23 , $\tilde{x}$ = 2.20 mode: 2.1 and 2.2 (bi-modal)
4. n = 15 ; Σx = 55 ; $\bar{x}$ = 3.67 , $\tilde{x}$ = 4; mode: 4	
5.	
6. n = 20 ; Σx = 185 ; $\bar{x}$ = 9.25 mins after 8am , $\tilde{x}$ = 8 minutes after 8am; mode: 8 minutes after 8	

7. n = 99 ; Σx = 435 ; $\bar{x}$ = 4.39 , $\tilde{x}$ = 4.4; mode: 4.4
Caution throwing out outliers, maybe that one extra long arrow head was from a different tribe and was a clue that there had been a confrontation! And you just threw it away!

Easier to do this: 4.0*3, 4.1*14, 4.2*12 , etc to calculate the sum; use the *frequency* of each repeated data value.

8. n = 99; Σx = 4.0*3+4.1*14+4.2*12+4.3*9+4.4*21+4.5*8+4.6*9+4.7*10+4.8*13
mean = 4.42 cm; median = 4.2 cm; mode: 4.1cm
Always careful throwing out data! That long arrowhead could have been evidence of an enemy that had attacked or something! You would certainly want to be on the look out for those again. Or maybe just go back to your understudy and see if he/she recalls there being such an arrowhead, to see if perhaps he had just recorded it wrong

9a. n = 8; Σx = 41; mean = 5.13
9b. n = 16; Σx = 71; mean = 4.44
9c. n=8; Σx = 44; mean = 5.50