Sample Midterm #1 Exam Questions.

INSTRUCTIONS:

The maximum score you can earn on this exam is 100 points. Each question is worth up to 10 points. You may answer any question you wish and as many as you wish within the timeframe of the exam (2hrs.) You will hand in both your question booklet and your answer booklet. We mark the answers to questions you chose in the answer booklet until you either reach 100 points or we run out of questions to mark.

The exams in this course are all Short Answer Format. A short answer is paragraph or two (at the most) of factually correct, concise, grammatically correct English. Where calculations are required please show your intermediate steps. You may be awarded points for getting part way even if your final answer is incorrect.

Pocket scientific calculators are permitted in the exam room; however cell phones are not permitted at your exam desk.

If you have a question raise your hand and a proctor will attend you.

Sample Questions:

1. Jacob studied hard and scored in the 84^{th} percentile on his PSYC2002 final examination. The mean of his section was 71 with a standard deviation of 6. What was Jacob's raw score on this exam? What was his z-score? What would his equivalent score be on an exam with a mean of 100 and a standard deviation of 30?
ASSUMING THE RAW SCORES ARE NORMALLY DISTRIBUTED we know that a z score equivalent to the 84^{th} percentile is z=+1. Therefore Jacob's raw score is 71+(1)6=77. His z score is +1 so his equivalent raw score on an exam with a mean of 100 and standard deviation of 30 would be 100+ (1)30 = 130.

2. Alyssa wrote an exam in her neuroscience course and scored 23/30. Beth wrote an exam in her social science course and scored 58/75. If Alyssa's class average was 25/30 with a standard deviation of 2 and Beth's class average was 54/75 with a standard deviation of 4 who had the higher percentile rank? What would the student with the lower percentile rank's raw score have to increase to in order to match the percentile rank of the higher student?

Although both Alyssa and Beth got 77% on their exam, Alyssa's z-score is (23-25)/2 = -1, whereas Beth's z-score is (58-54)/4=+1. So, Beth had the higher percentile rank. Alyssa would have to increase her z-score from -1 to +1 which is 2 standard deviations or 2 z units. Since standard deviation = 2 for Alyssa's exam, Alyssa would have to increase her raw score by 2*2=4 raw score units to an exam mark of 27/30 to be in the same percentile rank as Beth

.

3. James took a proprietary standardized aptitude test for the company he works for and scored 65 on it. The mean of this test is 50 and the test standard deviation is 10. His friend Parham works for another company that also has its own aptitude test (with a mean of 300 and a standard deviation of 100). What would Parham's test score have to be if he scores at the same percentile rank on his company's test as James scored on his company test? James z score is +1.5 (50+ (1.5)10=65. For Parham to have the same percentile rank he'd need to have the same z score. On the other exam scale this corresponds to 300+1.5(100)=450.

4. If you want to accentuate the effect that extreme scores have on a frequency distribution measure of centre, which descriptive statistic measure of centre would you use? Why?
You'd use the arithmetic mean because it is most affected by extreme scores. The median and mode are not.

5. A bunch of students walked into the bookstore and spent a total of $1800. If the average amount each spent was $150 how many students were in the bunch? Solving the equation for the arithmetic mean by plugging in the 'knowns' and solving for the unknown, which is n. Rearranging the terms to put N on one side of the equation we get n = 1800/150 = 12 students.

6. What is the purpose for which the value in the denominator of the formula for the sample variance is 1 less than the value in the denominator of the formula for the population variance?
The formula for the sample variance with n in the denominator underestimates the population variance by a little bit for small samples. By using n-1 in the denominator the formula ratio becomes increasingly larger (especially for samples < 30) and therefore a better estimator of the population variance.

7. Explain why the **shape** of the Frequency Distribution of the z-scores calculated from a set of raw scores looks exactly the same as the FD for the raw scores themselves.

Subtracting a constant from a raw score and dividing the difference by a constant are both linear transformations on the raw score. Neither of these operations affect the Shape of the FD and so the FD shape of the equivalent z scores is exactly the same as for the raw scores.

8. How large is a population? How large is a sample?
A population is typically a very large entity that is, practically, both of indeterminate size, location in space and in time and therefore unmeasurable (except in special circumstances such as when the entire population can be defined – for example when it is 'captive' in public schools…).In contrast the size of a sample is known precisely because it can be identified in space and time and measurements can be taken on its characteristics.

9. You are a research psychologist who works for the Ontario Ministry of Education. You've been asked to estimate the cost of delivering an English language 'head start' program for 'at risk' grade 3 children enrolled in the elementary schools Ontario. There are currently 120,000 children in enrolled in grade 3 in Ontario. At the end of September all grade 3 children in the province were administered the Weschler Intelligence Scale for Children (the WISC-IV) from which their Full Scale IQ (FSIQ) is derived. (The Weschler FSIQ has a population mean of 100 and a standard deviation of 15 and is normally distributed).The Minister of Education wishes to be able to deliver this program to every student in grade 3 in the province whose FSIQ percentile rank is less than 16.

a) What is the cutoff FSIQ that defines a student who would be eligible for this program? The cutoff will be at the raw score equivalent for the 16^{th} percentile, which is z=-1. Thus, the FSIQ would be 100 – (1)15 = 85.
b) If the cost for this program is $5/day per student, what will the program cost the province for each year of operation? (Assume there are 200 days in the school year). The 16^{th} percentile means that 16% of the population has an FSIQ of 85 or less. Given there are 120,000 students in Grade 3 in Ontario, this means 0.16*120000=~19000 students. At $1000/yr (200days/yr*$5/day) the cost to the province will be ~$19,000,000.

10. The FD of a large sample is unimodal and positively skewed. Write the inequality that expresses the relative number of positive and negative deviation scores in this sample.

A positively skewed distribution has the long tail to the right. For positively skewed FDs the arithmetic mean is shifted to the right of the mode (the peak) and the median. Recall that a deviation score is the difference between a raw score and the sample mean. Recall also that the sum of positive deviation scores equals the sum of negative deviation scores (because the sum of all deviations scores about a sample mean = 0). If you look at the shape of the +skewed FD you can see that there are a small number of extremely large +deviation scores. To the left of the mean there are a larger number of much smaller – deviation scores. Because the sum of the+ deviation scores equals the sum of the – deviation scores in any sample, this means that there must be more – deviation scores than + deviation scores. That's the inequality.

11. **Estimate the proportion of the area under the Standard Normal Frequency Distribution (SNFD) to the LEFT of the following values of z**:

- z=0 50% or 0.5
- z=+1 84% or 0.84
- z=-1 16% or 0.16
- z=+2 ~98% or 0.98
- z=-2 ~ 2% or 0.02
- z=+1.65 ~95% or 0.95
- z=-1.65 ~ 5% or 0.05
- z=+3 ~ 100% or all of it.

12. **Estimate the 'ballpark' area under the SNFD between the following two values of z**:

- Between z = -1 and z=+1 68% or 0.68
- Between z= 0 and z=+2 47.5% or 0.475
- Between z=-1.96 and z=+1.96 95% or 0.95
- Between z= - 2.33 and z=+2.33 99% or 0.99
- Between z = -3 and z = +3 ~100% or all of it

13.**Estimate the area under the SNFD between the following:**

- Between the 16^{th} and the 68^{th} percentile 52%
- Between the mean and the 95^{th} percentile 45%
- Between z = - 3 and the 95^{th} percentile 95%
- Between the median and the mean 0%
- Between the mean and one standard deviation to the right of the mean 34%
- Between minus one and plus one standard deviation 68%
- Between minus one half a standard deviation and plus one half a standard deviation 53%

14. **In a normally distributed sample with a mean of 100 and standard deviation of 20:**

- What is the z score equivalent of a raw score of 40? $z = -3$
- What is the z score equivalent of a raw score equal to the sample mean? $z = 0$
- What are the z score equivalents of a raw score equal to the sample standard deviation? $z = -1$ and $z = +1$
- What is the raw score equivalent of a z score of +1.5? 130
- What is the raw score at a percentile rank of 95? $100+(1.65)*20 = 133$
- What is the raw score at a percentile rank of 5? $100 – 1.65(20) = 67$

15. **For a raw score of 60 from a sample mean of 50 and sample standard deviation of 5: $z = +2$**

- What is the equivalent raw score in a sample with a mean of 100 and a standard deviation of 12? $100+2(12)=124$
- What is the equivalent raw score in a sample whose mean is 1000 and whose standard deviation is 500? $1000+(2)500=1100$
- What is the equivalent raw score in a sample whose mean is 5 and whose standard deviation is 0.4? $5+2(0.4)=5.8$

16. Briefly explain the reasons for using a table over a graph when presenting your results, then explain the reasons for using a graph over a table. Tables are preferred over graphs when you have lots of data or when you need to report with high precision. Graphs are preferred when you want to show a functional relationship or a “take-home message”. One must be careful to avoid creating graphphitti – that is visual clutter that interferes with the transmission of the take-home message.

17. You have measured the pre-treatment weights of a sample of 100 rats as part of a ‘before and after’ study on the effects of chronic exposure to nicotine on appetite. You want to make a frequency distribution of these pre-treatment weights. The weight range is 350 grams about a mean of 545 grams. How many bins will you create to make your FD and what are the real limits of each of these bins? Upper real limit will be 750, lower real limit will be 350. Using 50 unit steps I’d use 9 bins with real limits as 349.5-399.5,399.6-449.5,449.6-499.5,

499.6-549.5,549.6-599.5,599.6-649.5,649.6-699.5,699.6-749.5

18. To the right of each, write out the Excel formula for calculating the following descriptive statistics on a sample of raw scores, located starting in cell C5 and ending in cell C500.

Descriptive Statistic:	Excel formula:
Maximum sample raw score	=max(c5:c500)
Minimum sample raw score	=min(c5:c500)
Sample Mean	=average(c5:c500)
Sample Median	=median(c5:c500)
Sample Standard deviation	=stdev(c5:c500)
Z score for cell c100	=(c100-average(c5:c500))/stdev(c5:c500)
Sample 38th percentile	=percentile(c5:c500,0.38)
Sample range	=max(c5:c500)-min(c5:c500)
Sample size (n)	=count(c5:c500)