APPENDIX N

Summary Statistics:

The "Big 5" Statistical Tools for School Counselors

This appendix describes five basic statistical tools school counselors may use in conducting results based evaluation. Skill in using these five concepts will make any presentation on results based evaluation a meaningful one.

The "Big Five" Statistical Tools for School Counselors

The Question:
How can school counselors find meaningful patterns in existing school data?

The Answer:
By mastering five statistical concepts (Mean, Standard deviation, Percentages, Correlation, and T-Test).

1. Mean – a value that tells us something very important about a distribution of scores. It is the average that balances the variability or distances between scores; a measure of central tendency that tells us something important about the group of scores as a whole (like the median and mode).

Example: Number of absences for 10 at-risk sophomores for November

Students	# of Absences		Deviation from the mean
1	2		-3
2	7		2
3	8		3
4	6		1
5	3		-2
6	6		1
7	2		-3
8	3		-2
9	8		3
10	5		0
Sum =	50	Deviation sum =	0

Mean = Sum/the total number of scores = 50/10 = 5

Scores (absences) vary around the mean. A mean of 5 tells you something important about the variability in this distribution of scores. It is the center of these deviations.

2. Standard deviation – tells us how large the variability of the scores around the mean is.

Example: Number of referrals for misbehaving in class

Students	# of Referrals	Deviation from the mean Deviation (squared)
6	1	1
8	3	9
2	-3	9
4	-1	1
4	-1	1
3	-2	4
7	2	4
6	1	1
Sum = 40	Sum of Dev = 0	Squared Deviation Sum = 30

Standard deviation = the square root of the (Squared Deviation Sum), divided by the total number of subjects.

Standard deviation = 30/8 = 3.75 = 1.94 Mean = 40/8 = 5

Standard deviation tells us something important about how far scores vary from the mean. Look at your normal curve graph, about 68% of all the scores vary within a range of 1 standard deviation above the mean and 1 standard deviation below the mean. In our example, this range would be from 3.06 to 6.94 office referrals.

Z scores tell us how far a score is from the mean in Standard Deviation units. It is like a common denominator that we can use to put different standard deviations on the same scale.

Z = someone's score minus the mean, divided by the standard deviation

For example, for student 3 it would be z = 2-5/1.94 = -1.55

Where does -1.55 z-score units fall on your normal curve graph?

3. Percentages – Scores have meaning only in reference to how everyone else did on a test. Percentages are a way to tell us what the location of someone's score is in relation to all the other scores on the test. It lets us know what percentage of all the people who took the test scored at or below a given score. It makes the meaning of an individual score more meaningful and interpretable.

Look at your graph of the normal curve. What % of the people are expected to score at or below 2 standard deviations above the mean?

What percentage of the people are expected to score at or below -1 standard deviations below the mean?

4. Correlations – if one set of scores varies around the mean, 2 sets of scores can covary with each other. Do scores on different measures covary (do they go up and down together, in sync, in rhythm with each other)?

Positive correlation: two scores go up and down together, correlation goes from zero to +1 (e.g., height and weight)

Negative correlation: two scores go in opposite directions, one goes up and the other goes down, correlation goes from zero to -1 (e.g., perfectionism and tolerance for making mistakes)

Non-significant correlation: two sets of scores have nothing to do with each other, they go up and down irrespective of what the other score does.

Students	Score	Dev	z for X	Score	Dev	z for Y X times Y
14	1	0.4	24	3	1.00	0.400
16	3	1.2	23	2	0.67	0.804
13	0	0	21	0	0.00	0.000
13	0	0	20	-1	-0.33	0.000
14	1	0.4	23	2	0.67	0.268
14	1	0.4	23	2	0.67	0.268
10	-3	-1.2	18	-3	-1.00	1.200
9	-4	-1.6	16	-5	-1.67	2.004
17	4	1.6	25	4	1.33	2.128
10	-3	-1.2	16	-5	-1.67	2.004

Example: What is the relationship between a student's self-confidence in their ability to do mathematics and whether or not they experience feelings of anxiety about doing mathematics?

X = Self-confidence
(higher scores mean you are more self-confident)

Y = Anxiety
(higher scores mean you are less anxious about mathematics)

Mean of X = 13 Mean of Y = 21 Sum of XY = 9.2

SD of X – 2.5 SD of Y = 3.0

Correlation r = Sum of the XY cross products divided by the number of pairs of scores

r = 9.2/10 = .92

Interpretation:

1. **What does this high positive correlation mean?**
   Answer: Higher self-confidence ratings are associated with less self-reported anxiety about mathematics.
2. **Is this .92 correlation statistically significant?**
   Answer: Yes, the p value is <.01 (look it up in the table)
3. **What does statistical significance mean here?**
   Answer: A significance level is the likelihood (the probability) that the result we get is due to sampling error. When this value gets small enough, we decide that the correlation we got was not due to sampling error.

   p <.05 This means that our correlation is likely to occur by chance less than 5% of the time. So, we decide to reject the idea that our correlation is due to sampling error and accept that there is a significant relationship between math self-confidence and math anxiety. While it is possible that our finding is due to chance, we expect this to occur less than 5% of the time.

   p <.01 Now, you interpret this probability value in relation to our .92 correlation. How often does a .92 correlation happen by chance at the .01 level? How about at the .001 level?

   It is a way to make an inference about the amount of relationship between these two variables in the population. It could be that our .92 correlation happened by chance, maybe we picked 10 students not in any way representative of the population we are interested in.

5. <u>T-test</u> – when we want to compare the differences in the performance of two different groups, we can test to see if the means of the two groups are statistically different from each other. A t-Test lets us assess whether or not observed differences in the performance of two different groups are statistically significant or due to sampling error.

t = the mean for group 1 minus the mean for group2, divided by the standard error of the difference

$$t = \frac{\text{group 1 mean} - \text{group 2 mean}}{\text{standard error of the difference}}$$

Example: A small rural school is concerned that this year's sophomore class is missing substantially more days of school than previous sophomore classes. To test this, they want to compare the mean number of absences the 2004 sophomore class has had to the mean number of absences the 2003 sophomore class had.

Mean number of absences for 2004 = 7.6, standard deviation = 3.26
Mean number of absences for 2003 = 7.0, standard deviation = 2.12
Standard error of the difference = 1.62
N1 (sample size for group 1) is 5
N2 (sample size for group 2) is 8

Degrees of freedom = (N1 – 1) + (N2 – 1) = 4 + 7 = 11

$$t = \frac{7.6 - 7.0}{1.62} = \underline{0.6} = .37$$

Now: Look up the value that the T-test has to reach to be significant with 11 degrees of freedom. At the $p < .05$ level it has to be ≥ 2.201.

Is there a statistically significant increase in the number of absences from this year's sophomore class compared to last year's class? How would you report that probability value?
(Answer) $p > .05$