

*Faculty Development Program*

Clinical Epidemiology and Clinical Research

TOPIC: Biostatistics I: Descriptive

DATE: April 11 (1:30 PM – 5:00 PM)

LEADERS: Art Evans

OBJECTIVES:

1. Describe differences between nominal, ordinal, interval, and ratio data and between independent and dependent variables.
2. Describe the best ways of organizing and summarizing the different types of data (nominal, ordinal, interval/ratio), with respect to:
  - Center of the data; central tendency; typical values: means, medians, modes
  - Variability or dispersion of the data: range, min/max, quartiles, percentiles, standard deviation, variance
  - Shape of the data: stem-leaf, bar chart, histogram, pie chart, polygon/line graph.
3. Calculate a standard deviation and variance by hand.
4. Calculate a standard error of the mean if given data about the mean, standard deviation, and sample size.
5. Describe the difference between a standard deviation and standard error of the mean, and the different interpretations of  $\text{Mean} \pm 2 \times \text{SD}$  and  $\text{Mean} \pm 2 \times \text{SEM}$ .
6. Use a table that describes the area under a Normal curve to calculate the proportion of a normal population that have values more extreme than any given value.
7. Use SPSS to produce each of the descriptors above.
8. Use SPSS to compute a new variable.
9. Use SPSS to recode a variable into a new variable.

READING:

Norman and Streiner. *Biostatistics: The Bare Essentials*. Preface. Pages vii-ix.

Read enough to accomplish the objectives above. The topics are covered in the two books below. Pick the one that best fits your needs.

Norman and Streiner. *PDQ Statistics*. 2<sup>nd</sup> Ed. Pages 3-14.

Norman and Streiner. *Biostatistics: The Bare Essentials*. Chapters 1-4. Pages: 2-28.

## QUESTIONS:

1. A large group of college students first watched a video of an auto accident. Then they were randomized into two groups. One group was asked: “About how fast were the cars going when they contacted each other?” The other group was asked: “About how fast were the cars going when they collideded with each other?”

- Do you expect the average responses in each group to be nearly equal?
- If not, which group would report the highest average speed?
- What would you consider the smallest important difference in speeds (between groups)?

*(Teaching point: Never forget the possible role of bias. Biostatistics doesn't assess for bias.)*

2. A similar study was done that compared two different versions of the same question:

- 1) “Do you think the U.S. should forbid public speeches against democracy?”

(1=Agree; 2=Disagree)

- 2) “Do you think the U.S. should allow public speeches against democracy?”

(1=Agree; 2=Disagree)

- If you “reverse-code” the second question, would you expect identical responses to the two questions?
- If not, which group would have the greatest proportion answering “1”? Why?

3. In a study of attitudes toward ethnic groups, almost 30% of respondents had an opinion (ie, answered something other than “neutral” or “N/A”) about the fictional Wisians, rating their social standing *above* six other real groups, including “Mexicans”, “Vietnamese”, and “African Blacks”. (source: Crossen. Tainted Truth. Simon and Schuster. 1994. Page 24. Discussed in Utts. Seeing Through Statistics. Page 36.)

Do you think that any of the respondents in your study answered questions that they really didn't understand?

*(Teaching point: measurement bias must always be considered.)*

4. Advertisement in Great Britain: “TRIUMPH beats MERIT—an amazing 60% said Triumph tastes as good or better than Merit.”

What was not reported was that there were three response options:

- 1) T tastes better;
- 2) M tastes better;
- 3) T and M taste about the same.

Responses were: 1) 36%; 2) 40%; 3) 24%. What do you think Merit's advertisement said? Which brand wins?

*(Teaching point: Careful when collapsing categories. The message can change dramatically. To avoid bias, it's best if you make a priori decisions about how to collapse.)*

5. Consider these nine values. 1,2,4,5,7,9,15,25,40. Which is the best summary of the data, (a) or (b)?

a) Mean	12
Variance	166
SD	13
SEM	4
Lower 95% CL	2
Upper 95% CL	22

b) Five-number summary:

	Median		
	7		
Lower quartile	3	20	Upper quartile
Min	1	40	Max

6. The mean height for adult men in the U.S. is about 70 inches and the standard deviation is about 3 inches. The shape of the distribution is bell-shaped (normal).

- Within what range of heights would 68% of men's heights fall?
- Are 68% of all men greater than 5 ft. 8.6 in?
- Art Evans is 6 ft. 7 in. tall. What proportion of men are taller?

7. Calculate the mean and standard deviation for these two sets of five numbers.

	<u>Mean</u>	<u>Standard Deviation</u>
b) 100,100,100,100,100		
c) 90,90,100,110,110		