Research Orientation in Nursing XIV

NOW THAT I COLLECTED ALL THIS DATA WHAT DO I DO WITH IT?—PART I

Dr. (Mrs.) Marie Farrell and Dr. (Miss) Aparna Bhaduri

Dear Reader,

SO far you have completed two thirds of your work. In this letter we will discuss the analysis of your data, and in a later letter we will talk of interpretation, and then discussion with findings. Finally, some help on writing the report and preparing it for distribution, publication, or "sharing" will follow.

"Sounds good. Where do we begin."

Well, if you remember, in our last letter we suggested that certain kinds of data demand certain kinds of "treatments." And, we noted that you must determine the kind of treatment before you proceed, being sure, of course, to include the statistician in this important step.

"Right, And I understood the need to identify the levels of data, but can you review again what statistics go with what level."

Sure. You remember that the idea was to show the reason for taking care or being sensitive to the kinds of data collections techniques you use because.....

"Because the kind of data you have determined the treatment or use of statistics available to you."

That's right. And the analysis demands that you take the data and reduce it to some numerical measures before you can measure analyze and interpret them. The data must be quantified, and statistics computed on them in order to describe the characteristics of the group you are studying. These statistics will tell the reader the results of your study, and allow for the comparison of your study with other studies.

In general, we usually categorize statistics into two broad types. These are descriptive and inferential.

Descriptive Statistics

"Do descriptive statistics **describe** data?"

Yes, that's exactly right. **Descriptive statistics** are computed to determine the characteristics of the data at hand. These can be numbers, percentages, averages, indications of how "spread out" the data are (variability) or the relationships among data.

However, sometimes we wish to find out something about thousands of people. We can't possibly question every person, so we ask a sample of people, assuming that what the sample reports will be somewhat similar to what the thousands would report if we ask them. That is, we try to infer or to draw conclusions about the whole population on the basis of information we have about the sample. And really, this is the real purpose of science. We want to be able to predict future events, with a high degree of accuracy, by using inferences from past experience. Statistics enable us to judge the **probability** that our inference or estimates are close to the truth.[2]

In this letter, we are going to concentrate on descriptive statistics. We want to answer the questions: how typical are the data, how spread out are the data, and what are the relationships among or between sets of data.

"How do we look at 'typical-ness'—if there is such a thing?"

To answer that question, we look at three measures, called the measures of **central tendency.** These include: the **mean,** the **median** and the **mode,** which we will explain shortly. To measure 'spread-out-ness' or spread, we use the **range,** the **standard deviation,** and the **variance.** To measure relationships, we use **correlation co-efficient** sometimes referred to as "r". Measures of relationships or association answer the question: "within a given population, is there a relationship between one variable (or set of variables) and another variable (or set of variables)?"[3]

"Now what do these descriptive statistics have to do with our earlier discussion of the four levels of data?"

This will become clearer as we go along. As you learn what kinds of data are involved you will see hat, say, the mode is suited for nominal level data, and that certain kinds of correlation is

suitable for ordinal level data, etc. You can't add up a person's level in nursing and get an "average" because there is no number to add. So data having to do with degrees of "averageness" must be interval level.

Descriptive Measures

The Range

In this section we will describe how you prepare your data for overall description, and we will focus our attention on the qualities of accuracy, completeness, and objectivity.

"Alright, what do I do first?"

When you have many subjects on whom you have some measurements, it is cumbersome to look at the scores individually, we look at the **highest score** and the **lowest score** to find out **range** of observations.

For example, suppose we conducted a study where rated five villages (Villages A-E) on their level of sanitation, and, we derived the following **raw scores***.

Village	Score on Level of Sanitation	Systamized
A	65	98
B	43	65
C	27	43
D	98	27
E	24	24

Table 1—Selected Villages Scores, Levels of Sanitation

On a scale of 0 to 99, if the lowest score is 24 and the highest is 98, the range would be ——?

"74".

That's right. The highest integral minus the lowest integral equals the range, or

H — L = Range

This is simple enough; calculating the range **does** give you some quick information about the data.

The advantage of the range is that it is very easy to find. But, it is often subject to distortion. Take this series of numbers:

95 95 95 94 92 91 91 90 15

*Raw Scores are the only kind we have dealt with so far; others will be introduced later.

What do you see?

"All of the scores are in the nineties, except one which is very low, and tha numbers make the range quite large."

Right, so extremes of scores can affect the range; and this makes the range a poor measure of spread-out-ness or dispersion.

Use of Frequency Distribution to Compute the Mean

Sometimes, however, you have many scores, for in behavioral science research we work with human behaviour and humans have a high degree of variability. So you may end up with extensive sets of data which are so cumbersome that general trends cannot be discerned. Some techniques for condensing sets of data into more concise form are necessary so that your research questions can be answered. A **frequency distribution** is one technique used in research to bring order out of massed data with little or no loss of information.[4]

The frequency distribution of scores from 18 and 91 would be made like this:

The steps are quite simple:

1. Find the range: the higher integral—the lower integral = 91-18 = 73.
2. Divide the range by 15 (ideally for this number of scores, it is best to have between 10 o 15 class inervals if possible). 74 + 15 = 4.9 (so use 5 as class interval).
3. Find the lower integral unit (L.I.L.) which must be a multiple of 5, the class interval. Here it is 15.
4. Find the mid-point, the number that falls in the middle of each class interval. In our example, the mid-points are 97, 92, 87, 82, 77, etc.
5. Look at your raw data, and tally the number of, say, babies whose scores fell within the class intervals, and tabulate them in the frequency column. In our example, one baby fell in the 90-94 range, five in the 45-49 range, and none in the 20-24 range.

"So, I need to know the scores received by each infant, then. I systematize them, highest to lowest, determine the class interval and mid-points, and note the frequencies."

Rural Infants Development Scores	Systematized	Interval 5	Mid-point	Frequency
75	91	95-99	97	0
48	87	90-94	92	1
66	87	85-89	87	2
52	83	80-84	82	1
54	75	75-79	77	1
70	71	70-74	72	2
91	70	65-69	67	1
49	66	60-64	62	2
48	62	55-59	57	3
44	60	50-54	52	3
46	58	45-49	47	5
34	58	40-44	42	1
84	55	35-39	37	0
60	54	30-34	32	1
58	52	25-29	27	1
58	51	20-24	22	0
51	49	L.I.L. 115-119	17	1
55	49			n=25
43	48			
48	46	$\bar{X} = \frac{\Sigma fx^1}{n} = \frac{1440}{25}$		
27	45			
83	43	$\therefore \bar{X} = 57.6$		
87	34			
71	27			
62	18			

Table 2: Frequency Distribution Based on Infant Development Data for Rural Infants

Very good!

"I can see the need to do this, but why do I need the mid-point column?"

This is used to compute a very important measure, **the mean,** which we will discuss below. But first, several points must be made. When you create your intervals, they must not be so large that they hide some important qualities of your data. For example, if you are looking at infant growth, you would not want to lump babies from birth to one year in one pile because of the important monthly differences involved in such a sample. But also you do not want the frequency distribution to be so long that it isn't worth doing. Thirteen to 17 intervals is a workable number. In our example, we have 17.

"The next step is to determine "fx^1", which is the mid-point times the frequency.

"Alright. So in this example, $92 \times 1 = 92$, $87 \times 2 = 174$. Then these are added?"

That's right. This column is labelled "Σfz^1", or sum of the mid-point times the frequency.* And the sum is 1440. Then we divided the Σfx^1 by 'n" or numbers of infants, and we drive the mean, which is 57.6.

$$\bar{X} = \frac{f.X}{n}$$

$$\bar{X} = \frac{1440}{25}$$

$$\bar{X} = 57.6$$

Now, whenever the frequency distribution is fairly symmetrical and plenty of computation time is available, the mean is the statistic of choice.

Applications of the Mean

The mean is important to you because it is the most stable measure of central tendency and it lends itself to the computation of other stable measures.[5]

When we say it is "most stable," we mean that it "varies least among repeated random samples from the same population."[6]

Let us suppose you were involved in a research project where you were evaluating the effectiveness of a health training programme you conducted for school age children. You wanted to find the mean score on a test or **post-test** measure of their learning, but there were several hundred children involved in the programme. Well, you could test every school age child in the programme, sum their scores, and divide by n. In fact, that is simple procedure for calculating "$\bar{X}$" with small numbers; the simple formula is $\bar{X} = \frac{\Sigma X}{n}$ Again as we noted earlier, if you make a **frequency distribution** than to try to see all the information spread out about you. This helps us to get a better picture of the data, and to see how the measurements are distributed.

(Continued on page 23)

*Sum is denoted by the symbol "Σ", and is read, "sum of."

WHEREAS the graduate nurses with B.Sc. and M.Sc. Nursing programmes are adequately prepared to provide primary health care,

WHEREAS there are limited posts in the community health services for registered nurse midwives (R.N., R.M.), and

WHEREAS the avenues for registered nurse midwives to be involved in primary health care at various levels are scarce.

BE IT RESOLVED to urge the Central and State Governments to create positions for registered nurses at the block level and above.

7. **WHEREAS** W.H.O. Expert Committee in Community Health Nursing 1974, has stated that "A well organized nursing system responsible for education and practice of nursing personnel is needed in every country to cope with the problems of training and supervision of the health workers", and

WHEREAS only one post for nursing at the State Directorate of Health Services is existing to organise various professional activities,

BE IT RESOLVED to urge the State Governments to create posts at the State Directorate of Health Services, as is the pattern in West Bengal, i.e. to strengthen primary health care at the State level, posts of Deputy Director of Health Services (Nursing) with an Assistant Director of Health Services (Nursing) for community health nursing services, nursing education and nursing service respectively.

(Continued from page 4)

The Median

The second **measure of central tendency** is the median. This refers to the point above which has 50% of the total frequency, and below which is the other 50% of the total frequency. The median divides the distribution in half. In our example in Table 2, the median is 55, above and below which half the cases fall.

Use of the Median

If you want to know in which half of the distribution a particular case lies, then you find the median. When you have extreme cases, or when the distribution is not symmetrical, the median is useful.

The Mode

The mode refers to the point of greatest frequencies. Thus in a series such as: 9 1 4 2 6 9 8 7 9 3 5 9 would be the mode. It is used for very rough quick estimates, for identifying the most common score or the "typical case."

If we were to compute any other measure of central tendency the distribution formed would be greater or wider than sample means, thus the mean is more stable.

As we stated earlier, the other important aspect of the mean is that it is basic to other statistics. These include the standard deviation, product moment co-efficient and standard error.

In our next letter, we will explore some most essential measures—the measure of variability. As most interesting and understandable topic.

REFERENCES

1. Laboritz, S. Hagedorn, R. **Introduction to Social Research,** N.Y.: McGraw-Hill Book Co., 1976.
2. Stahl, S. Hermes, J. **Reading and Understanding Applied Statistics.** St. Louis: C.V. Mosby Co., 1975.
3. **Ibid.,** p. 233.
4. James, P., Klare, G. **Elementary Statistics: Data Analysis for the Behavioral Sciences.** N.Y.: McGraw-Hill Book Co., 1967.
5. Phillips, J.L., **Statistical Thinking.** San Francisco: W. H. Freeman and Co., 1973.
6. **Ibid.,** p. 15.