

R TUTORIAL, #4: **MEASURES OF SPREAD**

The (>) symbol indicates something that you will type in.

A bullet (•) indicates what the R program should output (and other comments).

TEST SCORES DATA SET

> Type: scores <- c(80, 76, 100, 83, 100)

> Type: scores

- Check to make sure you entered the values in correctly.

RANGE

- **WARNING:** The ‘range’ command gives the min and max values of the data set.
- We will use a different definition for “range”: max – min.

- Find the min and max values.

> Type: range(scores)

- Find [what we call] the range of the scores.

> Type: diff(range(scores))

- The ‘diff’ command, when applied to a list of two values, takes “second value – first value.”

SAMPLE VARIANCE

- Find the sample variance of the scores.

> Type: var(scores)

SAMPLE STANDARD DEVIATION; ROUNDING

- Find the sample standard deviation of the scores.

> Type: sd(scores)

- Round it off to one decimal place.

> Type: round(sd(scores), digits=1)

POPULATION VARIANCE

- To find the population variance of 5 data values, take the sample variance and multiply it by 4/5. If we have 6 data values, we multiply by 5/6, and so forth.

> Type: PopVar <- var(scores)*4/5

- The asterisk (*) denotes multiplication.

> Type: PopVar

POPULATION STANDARD DEVIATION

- The population standard deviation is the square root of the population variance.

> Type: sqrt(PopVar)

CHEBYSHEV'S THEOREM

> Type: k <- seq(1:5)

> Type: Cheb <- sapply(k, function(k) 1-1/k^2)

- This applies the function f , where $f(k) = 1 - \frac{1}{k^2}$, to each element of the 'k' list.

- The carat (^) denotes exponentiation.

> Type: data.frame(k, Cheb)

- You will see, for example, that at least 96% of a data set must lie within 5 standard deviations of the mean.