

Take-Home Test on Statistics

Please show all answers on this sheet. Work should not be shown, since it is assumed that you will use your calculator and/or Excel to do all the work for you. This worksheet will be graded based on *accuracy, completeness, and neatness*. If you need to start over, simply print out a blank copy and try again.

Before you start: Indicate whether you are using Excel or your TI-83/84 calculator to produce your results. Circle one: **Excel** **calculator**

The reason you have to circle the tool you are using is that the Q1 and Q3 values will differ slightly, depending on whether you use a calculator or a spreadsheet. If you use Excel for some things and the calculator for other things, that is also OK, but if so, please describe which tool you are using for which problems:

Excel: _____

calculator: _____

1. For all problems, you are to assume that a researcher on puppy obedience training is trying to determine whether there is a linear relationship between leash length (in cm) and the puppy's obedience score (on a scale from 0 to 10,000). Your data set has been e-mailed to you. (If there are any problems with reading your data, please contact Mr. Hansen immediately so that the issues can be resolved.)

In statistics, we use n to denote the sample size. In this case, n equals the number of data points, i.e., the number of puppies. What is the value of n ? Write an equation: $n = \text{whatever}$.

Answer: _____ = _____

2. For the x column (leash length in cm) and y column (obedience score), find the sample mean, sample standard deviation, and 5-number summary. Use correct notation: \bar{x} for the sample mean for x , \bar{y} for the sample mean for y , s_x and s_y for the sample standard deviations, and other labels just as you did in the Excelcise. An equation is required for each entry below.

	x	y
mean	_____ = _____	_____ = _____
s.d.	_____ = _____	_____ = _____
min	_____ = _____	_____ = _____
Q1	_____ = _____	_____ = _____
median	_____ = _____	_____ = _____
Q3	_____ = _____	_____ = _____
max	_____ = _____	_____ = _____

3. The *interquartile range* (IQR) is a measure of how spread out a collection of data values is. Compute the IQR (namely, $Q3 - Q1$) for each column:

IQR _____ _____

4. Use your 5-number summary information above to make a modified boxplot for each column of data. Remember that outliers are to be shown as dots. (An *outlier* is defined to be any data value that is more than $1.5(\text{IQR})$ less than Q_1 , or more than $1.5(\text{IQR})$ more than Q_3 .) The STAT PLOT feature of your calculator will make a modified boxplot for you, and all you have to do is to make a reasonable sketch of the calculator plot. Or, if you prefer, making modified boxplots by hand is certainly easy enough. Your sketch should be to scale. Show a scale, or if you prefer, simply label the 5 values of the 5-number summary on your boxplots.

Modified boxplot for x column: _____

Modified boxplot for y column: _____

5. Compute the linear correlation coefficient (r) that describes the strength and the direction of the linear relationship between the x column and the y column. Your calculator provides this information whenever you perform a linear regression (STAT CALC 8), but only if you have previously used the MODE button to turn the “stat diagnostics” feature on. If you have an older calculator that does not have that feature under the MODE menu, simply Google “DiagnosticOn” (all one word) and follow the first Google link that you get.

Excel computes the r value much more easily. Simply use the `=CORREL(range1,range2)` formula, as in the Excelcise.

Answer: $r =$ _____

6. Compute the equation of the line of best fit: $\hat{y} =$ _____
(Like all the other exercises on this worksheet, you can use either Excel or your calculator. Give the slope correct to at least 3 decimal places after the decimal point, and give the intercept correct to at least the nearest integer.)
7. Interpret the value of the slope in context.

8. Below, make a reasonably accurate scatterplot with the line of best fit shown. Each axis must be labeled with its real-world name (i.e., leash length or obedience score) and units, if applicable. Again, you can use either Excel or your calculator for this purpose. A neatly pasted Excel printout, with axes labeled properly, is acceptable.