Bar of Soap

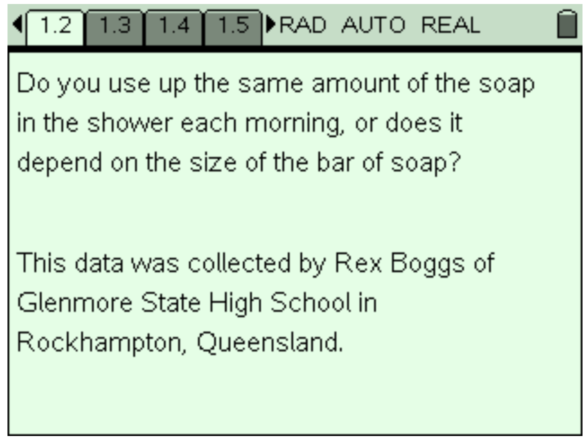
ID: XXXX

Name ______________________

Class ______________________

In this activity, you will investigate the decay over a three-week period of a bar of soap. While a quadratic model was hypothesised initially, the data over most of the period was surprisingly linear. Students explore, analyse and interpret the data provided using a variety of statistical tools and methods.

Open the file *StatActXX_Bar_of_Soap_EN.tns* on your handheld and follow along with your teacher to work through the activity. Use this document as a reference and to record your answers.



The Problem

"I had a hypothesis that the daily weight of my bar of soap in my shower wasn't a linear function, the reason being that the tiny little bar of soap at the end of its life seemed to hang around for just about ever. I wanted to throw it out, but I felt I shouldn't do so until it became unusable. And that seemed to take weeks.

"Also I had recently bought some digital kitchen scales and felt I needed to use them to justify the cost. I hypothesised that the daily weight of a bar of soap might be dependent upon surface area, and hence would be a quadratic function.

"I kept records for three weeks (the life of the bar), and was amazed to find that the data was linear with a very high r^2 value, until the last few days of its life.

"The data ends at day 22. On day 23 the soap broke into two pieces and one piece went down the plughole ..."

Date	Day	Weight
"30Aug99	0	124
"31Aug99"	1	121
"3Sept99"	4	103
"4Sept99"	5	96
"5Sept99"	6	90
"6Sept99"	7	84
"7Sept99"	8	78
"8Sept99"	9	71
"10Sept99"	11	58
"11Sept99"	12	50
"16Sept99"	17	27
"18Sept99"	19	16
"19Sept99"	20	12
"20Sept99"	21	8
"21Sept99"	22	6

EXERCISES

1. Would you agree that the data appears to be linear? Use a movable line to find the best fit you can and give the equation of this line of best fit and the sum of squares.
2. Which regression model would you predict should give a better result for this data: **linear** (such as **a*x+b**) or **median-median**? Why?
3. Use the data values on page 1.11 to compute a linear regression, and then plot the residual values against "day" on the D&S page following. What do these residual values tell you about this regression method?
4. Can you find a better model for this data? Justify your result.
5. How would describe the way that the weight of a bar of soap decreases over time with reference to the data provided.

SUGGESTED ANSWERS

1. If we remove the first few and the last few days, then the linear fit is a good one. The equation weight = -5.7*day+124 appears a good fit with sum of squares 121.3.
2. The median model is generally less affected by outlier values than the other methods, but there are no extreme outlier values involved here. Since there is variance both above and below a each end there may be an averaging effect and we might expect the linear model to give a slightly better result.
3. The residual plot displays the extent to which our data differs from the predicted model. The residual plot suggests that a linear model may not be the best fit for our data since a clear pattern is displayed although the correlation coefficient is very high.
4. The quadratic hypothesis appears reasonable and worth testing. By plotting the day data against the square root of the weight, the sum of squares for a linear regression is extremely small and the range of the residual values is much smaller than for the straight linear model. Both suggest that the quadratic model is a very good fit.
5. The linear model has a very high r^2 value, whereas the quadratic hypothesis offers a very low sum of squares, so both are reasonable. By changing only one value (the first day) this quadratic model can be made even closer to a perfect fit, whereas the last few days would need to be changed for a linear fit - hence the quadratic hypothesis would appear to be the better one.

