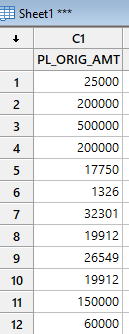
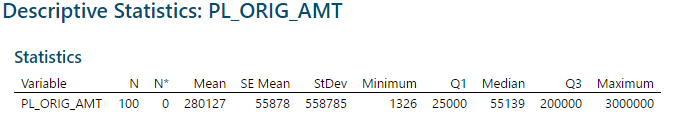
This is an example of running minitab to recode data into high low fields for Project 1.

I populated column 1 in minitab with some data that I had for the amount of purchases. There are 100 data fields in the column. If I wanted to split the column into high low values, I first run the descriptive statistics on the column to determine the median and number of values.

This data continues beyond this, but I’m only showing part to make this shorter.

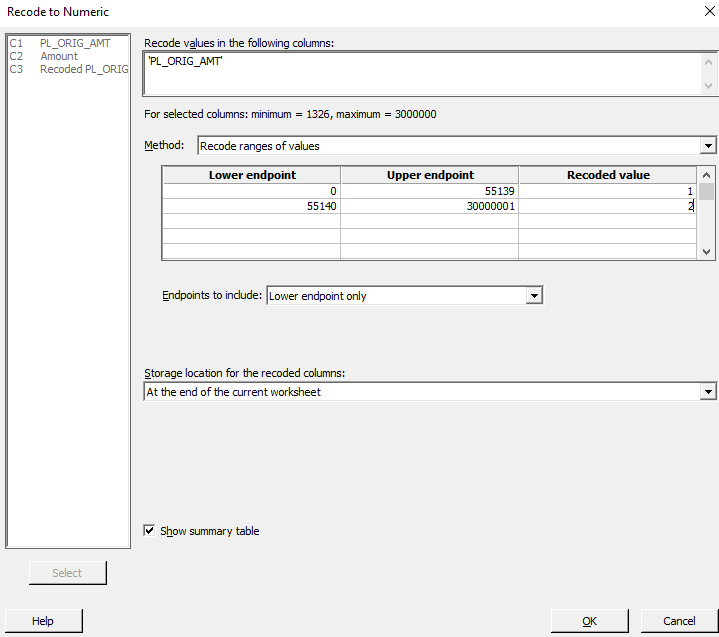
When I run descriptive statistics on this column of data, I get these:



From these statistics we see the median is 55,139 and there are 100 values. So, we want to recode all the values equal to or below the median as “Low” or “1” (whatever category you want to use) and then recode the values above the median as “High” or “2” (whatever category you want to use). We Click on Data, Recode, Recode as numbers or Recode as text (depending upon what new category you want). I used “1” and “2” so I chose, “recode to numeric”.

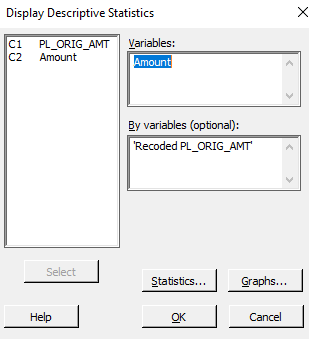
Pick Method “Range of Values”. This will ask you for the ranges of the numbers in the original column to split the data into the new categories. You can see I created two new values “1” for all the numbers up to and including the median, and “2” for all numbers over the median to our highest value.

I also chose to store the recoded values “at the end of the current worksheet” (this will create a new column after your other columns in the sheet – to the far right). The recode screen is shown below:

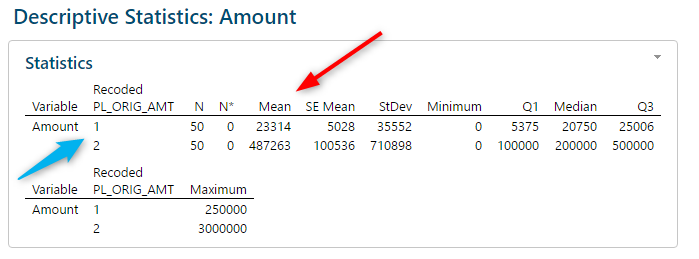


So now when I look at my data sheet, at the bottom of the minitab screen, I’ll see a new column with the “high” “low” values in it. Column C3-T shows “Recoded PL\_Orig\_Amt”, or my original data from column 1 recoded as categorical values representing the lowest ½ of the values from column 1 (1s) and the highest values (2s).

Say the data in Column 2 represents Obesity data. Now we can create the average mean of each of the “low” and “high” values of column 1 (in the project column 1 could have been the smoking data column). We can do this by running descriptive statistics, but splitting the values by the new categorical variable. To do this we can pick, Stat, descriptive statistics, display descriptive statistics and on the menu that appears, chose column 2 (Amount) as the “variables” to create statistics on (think of this as the obesity values) and in the “By variables” window, chose column 3 or the Recoded Pl\_Orig\_Amt (our recoded high low variable). Click on OK. (see screen shot below)



When you run this, you will now get descriptive statistics, including the mean for your “high” and “low” groups. Think of this for the mean obesity for the high and low group for the project. You can use these to run your hypothesis test with. For the 4-way split, you can recode the data to the quartile breaks of the original file and use the means to run ANOVA.



You can get creative and run the hypothesis test right from the data file here too.

Hopefully, this gives you a good path for creating your “high” and “low” breaks, your 4 way break for exercise, and how to get the stats for generating your hypothesis tests.