**Although students are responsible for this material on exams, they do not need to turn in the worksheet answers.**

Data: http://seeds.okstate.edu/SeedsPPP/CN,1/GulfWheat/GulfWheatPrices.xlsx

**Motivation:** In real applications the data you are given are not clean and will need considerable transformation. This means throwing out certain variables and sorting the data. This assignment presents you with very messy data and shows you how to use the *IF* and *AND* functions to pick out the data you want, and the *DATA* , *SORT* routine to sort data. Most people who work extensively with Excel will tell you they are forced to use *IF* statements and *DATA , SORT* routines all the time to get data to a manageable format. *On Exam 1 you will be given a very messy dataset and will be asked to transform it in ways similar to this worksheet.*

**Warning:** This project may be very time-consuming and often confusing—and frustrating.

**(A)  Histogram for Wheat Prices at the Gulf of Mexico Ports**

The aforementioned link contains data on wheat prices from 1974 until 2007. Make a clear and attractive histogram of these prices which occur only on (or immediately after) <u>December 10</u>. The data are messy and numerous, and I have not made the bins for you, so you will have to do the following to construct this graph.

- Insert a new worksheet into the Excel Workbook.
- Create three new columns called *Year, Month,* and *Day* , using the functions *Year(), Month(),* and *Day().*[Learn how to copy and paste quickly, without scrolling down, by holding down *ctrl* + *shift* then pressing one of the arrow keys]
- Then, you want to create new variables that help you identify the price for each year on or closest to December 10, but not before December 10
- . I do this by creating the following two variables, the first of which calculates the day minus 10, but only in December. This is a number that turns from negative to zero or positive once December 10 is reached or breached.

| | D | E | F | G | |
|---|---|---|---|---|---|
| | year | month | day | | indicate |
| 0 | 1974 | 12 | 2 | -8 | |
| 0 | 1974 | 12 | 3 | -7 | |
| 5 | 1974 | 12 | 4 | -6 | |
| 0 | 1974 | 12 | 5 | =IF(E242<>12,-200,(F242-10)) | |
| 0 | 1974 | 12 | 6 | -4 | |
| 0 | 1974 | 12 | 9 | -1 | |

- The next variable equals one whenever the previous data was before 12/10 but the next data is on or after 12/10.

| | D | E | F | G | H | I |
|---|---|---|---|---|---|---|
| | year | month | day | | indicate | indicate2 |
| 0 | 1974 | 12 | 2 | -8 | 0 | |
| 0 | 1974 | 12 | 3 | -7 | 0 | |
| 5 | 1974 | 12 | 4 | -6 | 0 | |
| 0 | 1974 | 12 | 5 | -5 | =IF(AND(G241<0,G242>=0)=TRUE,1,0) | |
| 0 | 1974 | 12 | 6 | -4 | 0 | |
| 0 | 1974 | 12 | 9 | -1 | 0 | |

- The variable *indicate* will always equal one for each year where the date is equal to or immediately after December 10.
- Then I copy and paste the values (*Paste Special , Values*) into a new column I call *indicate2*. This way the values will not change when I sort the data.
- Then I can select all the data except row 1 and use the sort function (*Data , Sort*) and sort the variables from largest to smallest on variable *indicate2*.
- The first rows will then have the rows corresponding to each year, on or immediately after December 10. I copy and paste these rows into a new worksheet and make a histogram from that.
- On that new page with December 10 prices, select all the December data and name it *decprice*.
- Use the *MAX(decprice)* and *MIN(decprice)* functions in Excel to determine the highest and lowest wheat prices in the data. Use that information to set reasonable bins, using your judgment. Don't over-think it.
- Use the *Data , Data Analysis* tool to construct a histogram of these wheat prices. Follow the histogram tutorial in Video 3 in the article *(AD.A.4) Histograms and VaR*. The video link is http://youtu.be/2x3O0cN7XTM.
- Note that if the *Analysis* toolpak doesn't shown under the *Data* tab, thus not allowing you access to the histogram tool, this tutorial (http://www.youtube.com/watch?v=cq-o4chUHrQ) will show you how to add it.
- I want this to be a near-perfect histogram, with appropriate axis values and labels. It should be clear what the histogram is about and what everything means. It should be attractive. Follow all the suggestions at the histogram tutorial and you will be fine.
- The histogram should resemble the following. It doesn't have to be identical, but it should have the same features.

**Histogram of December Wheat Prices in U.S. Gulf (1974-2007)**



prices in $ per bushel

- % prices in interval (left y-axis)
- % price in interval or less (right y-axis)