Your Name: ______________________

Section: ______________________

36-201 INTRODUCTION TO STATISTICAL REASONING
Computer Lab #14 [there was no lab #13]
Time Series Data and Time Series Transformations

Objectives:

1. To learn about the main features of a *time series* (a series of measurements on the same variable taken at repeated intervals in time), including *trend*, *seasonal variation*, and *cycle*.

2. To review exploratory methods such as the *boxplot*.

3. To explore the effect of location and scale transformations ($y = ax + b$) on time series data.

Getting Started

For today's lab, you will need some data from the file server. Before starting MINITAB, follow these steps to get the files `usemploy.mtp` and `temp.dat` for today's lab:

1. Open the **Student** folder. To do this, double click on the "Cluster HD" icon. Then double click the "documents" folder. You will eventually move the data into this folder.
2. In the following order: From the Apple menu in the upper left hand corner of the titlebar, choose **Chooser**. In the dialog box that appears *click* **AppleShare**. In the bottom left box called "AppleTalk Zones" *select* **BH**. In the top right box where it says "Select a file server" select **HSSHELIOS**. *Click* **Ok**.
3. In the dialog box that appears login using "36201" as your Name and "36201" as your password. Click **Ok**. In the dialog box that appears highlight "Class." Click **Ok**.
4. A "Class" icon will appear to the right. Double click on the icon. Then double click on 36201. You will see many files in a folder.
5. Copy all of the files that you will need to the **Student** folder. Do this by dragging the files into the Student folder. For today, you need the files `usemploy.mtp` and `temp.dat`.
6. Close all windows.
7. **IMPORTANT:** Drag the *class* icon into the trash. This is necessary to let others in the lab get access to the data.

Background: Time Series Data

Repeated measurements of the same variable over a long period of time is called a *time series*. Examples include, the measurement of a person's body temperature every hour for 24 hours; the number of U.S. exports per month for the last 60 months; and, the number of traffic fatalities each year for the past 50 years. A feature of time series data is that we are measuring the same variable on an individual or on a particular process (e.g., number of exports) regularly at different points in time. This is in contrast to measuring, say body temperature, just once on each of 24 people. To use exploratory data analytic methods to describe a time series, we will use a plot called a *time series plot* which is simply a scatterplot of the variable at each point in time at which it was measured. Often a time series is erratic with a lot of fluctuations. Our objective is to try to see an overall pattern in the time series by describing features of the time series. These features include

- the *Trend* (how the level of the variable changes over the long term),
- *Cycles* (short-term, regular fluctuations with units of time such as years, months, weeks, or days),
- *Seasonal Variation* (regular fluctuations that repeat from year to year, especially in economic and social data), and any
- *Small Scale Fluctuations* or *Changing Variability*, either of which may indicate that something interesting is going on that is not captured by the overall trend, cycle or seasonal variation in the data.

Example I

A simple measure of the growth of the US economy is the number of workers employed each month. The files `usemploy.mtp` and `temp.dat` contains the number of workers employed in the US economy (excluding farm workers, who are counted separately), every month from January 1986 to January 1996 (source: US Bureau of Labor Statistics WWW site `http://stats.bls.gov/`). The data itself is reproduced in Table 1, page 3, for reference.

1. In `MINITAB` pull down the **File** menu and select **Open Worksheet**. On the right side of the dialogue box choose **Desktop** and then select the **Student Folder**. At the bottom of the dialogue box set "List Files of Type" to **ALL**.

2. Choose `usemploy.mtp` and click **open**

3. Name the columns **Obs-No**, **Year**, **Month** and **Workers**.

Obs. No.	Year	Month	Workers	Obs. No.	Year	Month	Workers	Obs. No.	Year	Month	Workers
1	1986	1	97247	41	1989	5	108127	81	1992	9	109249
2	1986	2	97438	42	1989	6	108796	82	1992	10	109827
3	1986	3	97987	43	1989	7	107725	83	1992	11	110064
4	1986	4	98822	44	1989	8	107803	84	1992	12	110042
5	1986	5	99513	45	1989	9	108599	85	1993	1	107875
6	1986	6	99819	46	1989	10	109083	86	1993	2	108496
7	1986	7	99064	47	1989	11	109514	87	1993	3	108935
8	1986	8	99236	48	1989	12	109495	88	1993	4	109936
9	1986	9	100160	49	1990	1	107460	89	1993	5	110901
10	1986	10	100726	50	1990	2	107952	90	1993	6	111499
11	1986	11	100983	51	1990	3	108606	91	1993	7	110636
12	1986	12	101138	52	1990	4	109257	92	1993	8	110749
13	1987	1	99173	53	1990	5	110192	93	1993	9	111719
14	1987	2	99494	54	1990	6	110864	94	1993	10	112411
15	1987	3	100202	55	1990	7	109499	95	1993	11	112737
16	1987	4	101213	56	1990	8	109346	96	1993	12	112869
17	1987	5	101980	57	1990	9	109917	97	1994	1	110592
18	1987	6	102584	58	1990	10	110113	98	1994	2	111168
19	1987	7	101831	59	1990	11	110059	99	1994	3	112141
20	1987	8	102061	60	1990	12	109761	100	1994	4	113172
21	1987	9	102830	61	1991	1	107263	101	1994	5	114179
22	1987	10	103736	62	1991	2	107227	102	1994	6	114961
23	1987	11	104074	63	1991	3	107507	103	1994	7	114004
24	1987	12	104313	64	1991	4	107913	104	1994	8	114266
25	1988	1	102209	65	1991	5	108608	105	1994	9	115269
26	1988	2	102785	66	1991	6	109110	106	1994	10	115829
27	1988	3	103535	67	1991	7	107867	107	1994	11	116414
28	1988	4	104433	68	1991	8	107959	108	1994	12	116415
29	1988	5	105186	69	1991	9	108666	109	1995	1	114014
30	1988	6	106059	70	1991	10	109027	110	1995	2	114645
31	1988	7	105169	71	1991	11	109036	111	1995	3	115307
32	1988	8	105325	72	1991	12	108888	112	1995	4	116133
33	1988	9	106180	73	1992	1	106520	113	1995	5	116858
34	1988	10	106841	74	1992	2	106775	114	1995	6	117615
35	1988	11	107306	75	1992	3	107300	115	1995	7	116423
36	1988	12	107496	76	1992	4	108143	116	1995	8	116588
37	1989	1	105569	77	1992	5	108959	117	1995	9	117430
38	1989	2	105976	78	1992	6	109456	118	1995	10	117960
39	1989	3	106624	79	1992	7	108430	119	1995	11	118158
40	1989	4	107433	80	1992	8	108479	120	1995	12	118176
								121	1996	1	115378

Table 1: 1000's of Non-farm workers employed in the US, January 1986 to January 1996.

Question #1: How many non-farm workers were there in the US economy in March 1986? How many in March 1987?

To investigate the time dependence of these data, use MINITAB to make a time series plot of US employment levels displayed in the table:

> Pull down the **Graph** menu, and select **Time Series Plot**. In the dialog box that appers, type C4 in the first row of the **Graph Variables** table (leave the other rows blank). Select **Calendar** to the right of the **Graph Variables** table, click on the word next to **Calendar**, and select **Month Year** from the list. At the bottom of the dialog box select **Options**, and a new dialog box will appear. In first row of the **Start Time** table type 1 for **Month** and 1986 for **Year**. Finally, uncheck the first **Show** box on the far right, and check the second **Show** box. Now click on **Ok** in the **Options** dialog box and **Ok** in the **Time Series Plot** dialog box, to see the time series plotted.

After the time series plot appears, make it as large as possible by clicking on the resize box in the upper right corner of the plot, before answering the next several questions.

Question #2: You can see a regular *cycle* in the time series; how long (how many months) is it?

♣ **Question #3: Find the points on the graph corresponding to March 1986 and March 1987. Identify any peaks and valleys you see in the period March 1986 to March 1987. Does this pattern of peaks and valleys repeat each year (i.e., does it represent a regular *seasonal variation*)? Give a plausible explanation for the pattern of peaks and valleys you have identified.**

Question #4: Looking at the times series as a whole from 1986–1996, do you see a general *trend*? That is, is the number of workers in the US economy tending to increase, decrease, or stay about the same? (Imagine a straight line through these data. Is it increasing, decreasing or flat? This is the trend.)

To explore the trend in the data and also the variability more closely we "smooth away" some of the smaller fluctuations in the data by plotting boxplots for each year's employment figures on a single graph.

> Pull down the **Graph** menu and select **Boxplot**. In the dialog box that appears, type C4 (i.e. Workers) for **Y** and C2 (i.e. Year) for **X** on the first line of the **Graph Variables** table; leave the other lines of the table blank. Click on the ▽ next to **Frame** near the bottom of the screen, select **Multiple Graphs** from the menu that appears, and select **Overlay graphs on same page** in the dialog box you get. Click **Ok** in this dialog box and then click **Ok** in the **Boxplot** dialog box to see the result.

You may wish to maximize the size of this plot also once it is made.

Question #5: The median in each boxplot represent the median number of workers employed in that year. Imagine a smooth curve drawn through all the median lines; this better represents the *trend* than a straight line. Sketch that curve in the space below.

Question #6: Describe the 10 boxplots you obtained. Does the *variability change* over the years (look at the IQR's)? Are all the boxplots symmetric? Is there skewing? Are there outliers? Do any of these features change over the years?

Question #7: Review the time series plot you made and the boxplot plot. From 1990–1992 describe what happens to the number of workers.

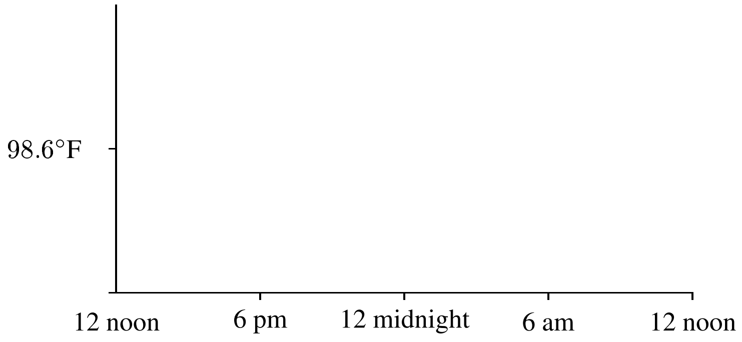
♣ **Question #8: When Bill Clinton was running for president in 1992 against incumbent George Bush (SR.), his principal political consultant James Carville had the slogan "It's the economy, stupid!" posted on his office wall. Explain what Carville meant,** *using evidence from your* MINITAB *analysis*.

We are done with this example. Under the **File** menu select **Restart Minitab**. Do not save any changes.

Example II

My Mom always told me that normal body temperature is 98.6°F (37°C). Suppose that you were able to take measurements of your body temperature every 20 minutes for 24 hours.

<u>Question #9:</u> On the graph below draw what you would guess to be the trend and cycle, if any, for the time series of your body temperature.



The data file `temp.dat` contains core body temperature readings for a 28 year old healthy female obtained every 20 minutes for approximately 5 days. The temperature readings are in °C. There are a total of 352 observations (3 readings/hr x 24 hours/day x 5 days = 360 readings). Since there are 3 readings of temperature/hr every 72 measurements (3x24=72) corresponds to 1 day. These data were obtained as part of a study of biological rhythms at the University of Pittsburgh Medical School.

In `MINITAB`, pull down the **File** menu, and select **Open Worksheet**. Choose `temp.dat` and click **open**

Name column C1 "CTEMP". Make a time series plot of CTEMP as follows:

> Pull down the **Graph** menu, and select **Time Series Plot**. In the dialog box that appers, type C1 in the first row of the **Graph Variables** table (leave the other rows blank). Select **Clock** to the right of the **Graph Variables** table, click on the word next to **Clock**, and select **Day Hour Minute** from the list. At the bottom of the dialog box select **Options**, and a new dialog box will appear. In the first row of the **Start Time** table type 1 for **Day**, 0 for **Hour** and 0 for **Minute**. Under the heading **Assignment of Time to Data**:
>
> - The long box to the right of **Day** should be empty, and the **Show** box to the far right should be checked.
> - The long box to the right of **Hour** should be empty, and the **Show** box next to it should be <u>*un*</u>checked.
> - In the long box to the right of **Minute**, type 0:59/20 (so that MINITAB knows that there is a measurement every 20 minutes of each hour), and make sure the **Show** box is <u>*un*</u>checked also.
>
> Now click on **Ok** in the **Options** dialog box and **Ok** in the **Time Series Plot** dialog box, to see the time series plotted.

The plot begins at Noon on the first day.

Question #10: Draw a rough sketch of the time series of body temperature.

♣ **Question #11: Describe the features of the time series plot of body temperature. (Is there a *trend*? What is the *cycle*? Are there *small-scale fluctuations* or *changing variability*?) Could you see from the plot approximately where the 5 days are, if the labels on the x-axis were missing?**

Question #12: Reading from the plot, what are the approximate maximum and minimum values of temperature in °C. What is the range of temperature values?

Question #13: What are the mean and standard deviation of the series?

Question #14: Make a boxplot of the distribution of temperature. Describe the distribution. Can you detect from the boxplot the important features of the time series that you described in Question #11? Why or why not?

The values of "CTEMP" are given in °C. It might be useful in order to help us interpret these data more easily to transform (or recode) the data to °F. The mathematical relationship between °C and °F is as follows. If CTEMP is a temperature value in °C then the corresponding temperature in °F, call it FTEMP, is

$$\text{FTEMP} = 1.8 * \text{CTEMP} + 32$$

> To transform the values of "CTEMP" from °C to °F, let's create this new variable. To do this, pull down the **Calc** menu, and select **Mathematical Expressions**. Type C2 in the **Variable** box at the top, and type `1.8*c1 + 32` in the large **Expressions** box. Then click on **Ok**. Name the new variable in C2 "FTEMP".

♣ **<u>Question #15:</u> Make a time series plot of the temperature data in $^\circ$F using the variable FTEMP, just as you did for CTEMP for Question #11. Are the features of this time series plot different than the features you described in Question #11 when the data were in $^\circ$C?**

Question #16: What are the mean and standard deviation of the FTEMP series?

Question #17: Given what you expected to be the time series of temperature that you described in Question #10, are you surprised by the actual time series observed? Comment.

We are done with this example. Under **File**, *choose* **Close Datafile**. Don't save changes.