Solutions for all questions are in blue

1. (35 points) The following Excel spreadsheet contains the maximum wind speed in miles/hour of 8 recent hurricanes (data from the 2003 Atlantic Hurricane Summary of the National Hurricane Center).

	Max wind (mph)			
Ana	40	Mean	53.75	
Two	35	Standard deviation	21.00	
Bill	60	Variance		
Claudette	90	Median	50	
Danny	75	Q1		
Six	35	Q3		
Seven	35	Min		
Erika	60	Max		
		IQR		

(a) Write the Excel command that was used in cell D3.
= stdev(b2:b9)

(b) Make a stemplot of this data.

3	|	5	5	5
4	|	0		
5	|			
6	|	0	0	
7	|	5		
8	|			
9	|	0		

(c) Find the missing information in cells D6 to D10.

$$\begin{aligned} Q1 &= \frac{35+35}{2} = 35 \\ Q3 &= \frac{60+75}{2} = 67.5 \\ IQR &= Q3 - Q1 = 32.5 \\ Min &= 35 \\ Max &= 90 \end{aligned}$$

(d) Are there any outliers in this data?

$$\begin{aligned} 1.5 * IQR &= 48.75 \\ Q1 - 1.5 * IQR &= -13.75 \text{ and} \\ Q3 + 1.5 * IQR &= 116.25 \end{aligned}$$

Since all observations are higher than -13.75 and lower than 116.25 then there are no outliers.

(e) Find the sample mean and sample variance of the data in km/hour (1 km = 0.62 miles).
This is a linear transformation $x_{\text{km/hour}} = \frac{1}{0.62} x_{\text{miles/hour}}$

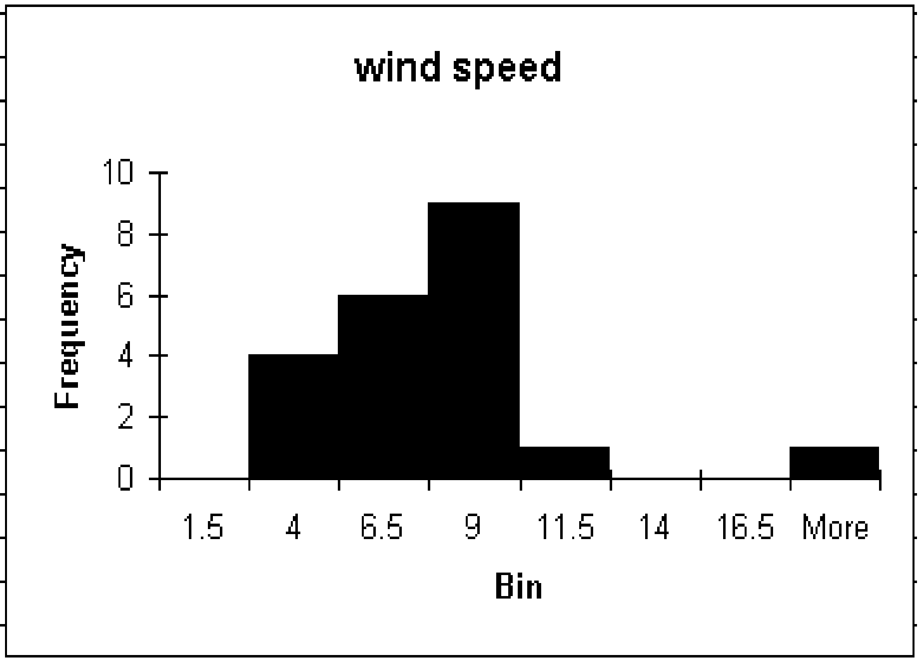
$$\begin{aligned} \bar{x}_{\text{km/hour}} = \frac{1}{0.62} * \bar{x}_{\text{miles/hour}} &= 86.69 \text{ km/hour} \\ s_{\text{km/hour}} = \frac{1}{0.62} * s_{\text{miles/hour}} &= 33.87 \text{ km/hour.} \\ \text{Variance } = s^2_{\text{km/hour}} = (s_{\text{km/hour}})^2 &\approx 1147.24 \text{ km/hour.} \end{aligned}$$

(f) Write an Excel command that you would write in field E2 that would give the standardized version of the data point in field B2, in a form that could be "dragged downward" to give the standardized versions of all numbers in the list.
One possible correct answer is = ($B2-$D$2)/($D$3).

(g) Find the standardized median.
Standardized median is
$\frac{50-53.75}{21} = -0.18$.

2. (15 points) The histogram below represents the distribution of day-average wind speed (miles/hour) for 21 days in September 2003 recorded in Raleigh/Durham International Airport in North Carolina.

Bin	Frequency
1.5	0
4	4
6.5	6
9	9
11.5	1
14	0
16.5	0
More	1



(a) Describe the overall pattern of this histogram.

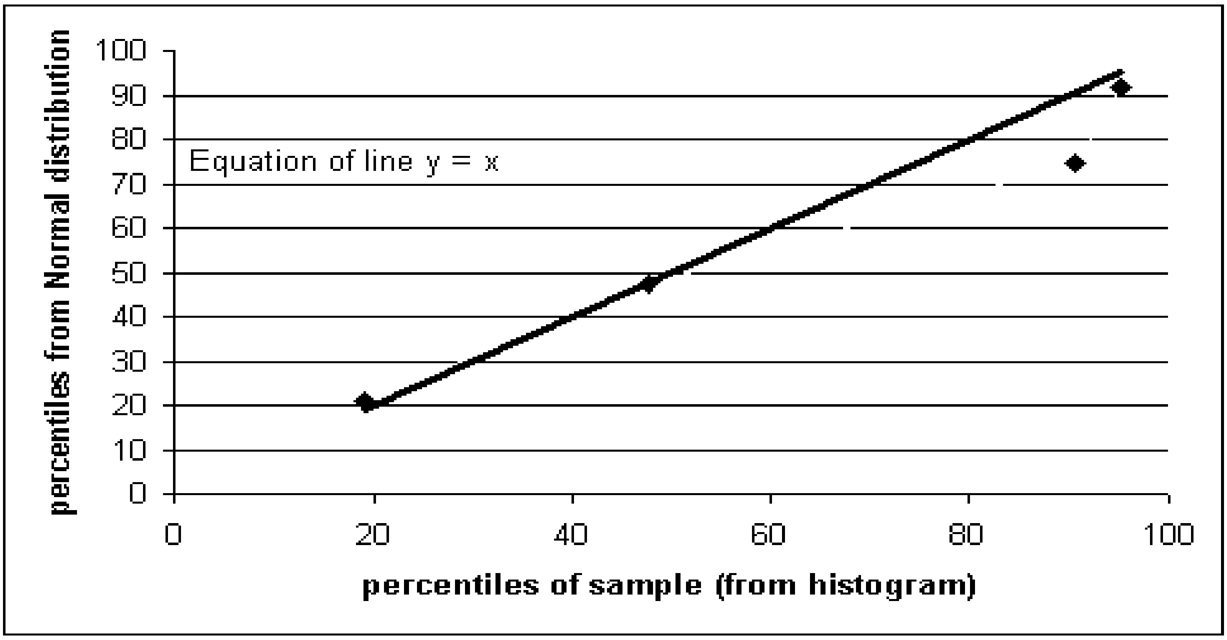
The distribution of wind speed is unimodal, slightly skewed to the left. The range of the data is approximately [1.5-19], with a center or midpoint in the range [6.5-9]. There is one possible outlier in the range [16.5-19].

(b) What percent of days have an average wind speed of less than 6.5 miles per hour?
There are 10 observations out of 21 that are less than 6.5 miles per hour, so the relative frequency of these observation is $\frac{10}{21} = 47.62\%$.

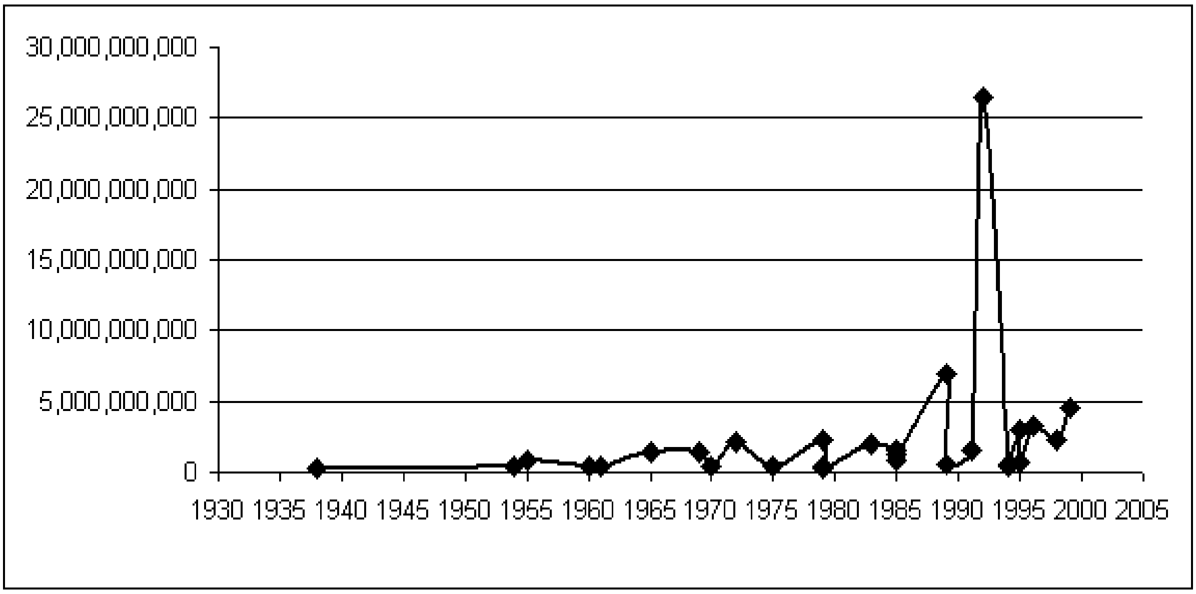
(c) Would the normal distribution $N(6.74, 3.39)$ be a good approximation for this data? Compare the shape of this histogram to the shape of a normal distribution, and compare the values of 3 different percentiles of this data to the same 3 percentiles of the normal distribution.
Comparison of shapes: the shape of the histogram is unimodal as is the normal distribution but the histogram is not symmetric and doesn't have a bell shape. Comparison of percentiles: From the histogram, we see that 4 miles/hour is the 19.04 percentile, 6.5 miles/hour is the 47.61 percentile and 9 is the 90.48 percentile. We have the following table:

Wind speed	percentile(from histogram)	standardized wind speed	percentile from N(0,1)
4	19.04	$\frac{(4-6.74)}{3.39} \approx -0.81$	20.90
6.5	47.62	$\frac{(6.5-6.74)}{3.39} \approx -0.07$	47.21
9	90.48	$\frac{(9-6.74)}{3.39} \approx 0.67$	74.86
11.5	95.23	$\frac{(11.5-6.74)}{3.39} \approx 1.40$	91.92

We see from the table and corresponding figure below that because of the skewness of the data, the largest percentiles of the sample disagree with the largest percentiles of the normal distribution. So, the normal distribution is not a very good approximation of the distribution of wind-speed.



3. (20 points) The following time series represents the cost (in US dollars) of 28 tropical cyclones the year they stroke the U.S. mainland for years 1938 to 1999.



(a) Give an approximate range of the costs and describe the main features of this time plot.
The approximate range of costs is [0-27,500,000,000]. There is no apparent increasing trend between 1938 and 1995 of the costs of tropical cyclones, although there seem to be an increasing trend after 1995. Between 1970 and 1995, there seems to be a 5-year seasonal variation. We notice from the time plot two unusually large cost cyclone, the first hit between 1985 and 1990 and the second hit between 1990 and 1995. We see from the time plot that from 1937 to 1960, the cost of only 4 cyclones was recorded. After 1960 the cost of at least 3 tropical cyclones every period of five years were recorded and after 1985, the cost of at least 4 tropical cyclones every period of five years was recorded.(Remark: It is not clear from the description of the data if this is due to increase of incidence of tropical cyclones on the US mainland or if the costs per cyclone increased).

(b) Using the axis scale of the time plot, in which range do you expect the median cost to be?

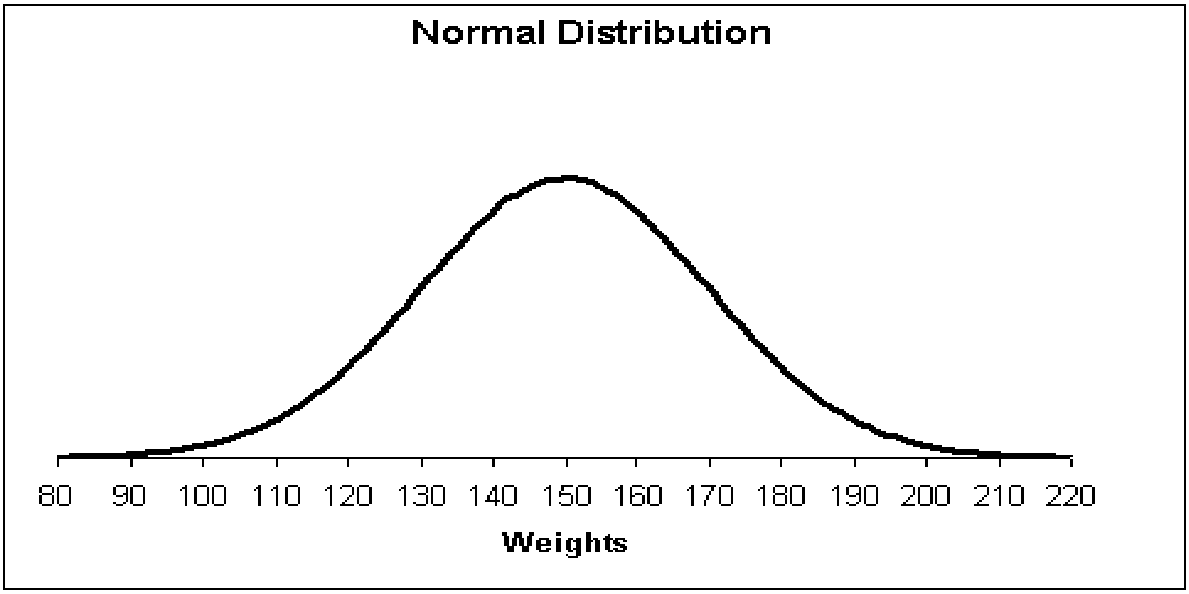
Since only two observations out of 28 exceed 5,000,000,000 dollars, then the median cost is in the range [0-5,000,000,000].

(c) Would you expect the sample mean cost to be lower or higher than the median cost? Explain briefly why.
Because of the two large cost observation, the mean cost will be higher than the median cost

(d) The cost of each cyclone was measured the year it stroke the U.S mainland. Are the 28 costs comparable to each other over time? Why or Why not?
No, the cost of the 28th cyclone are not comparable. The data should be corrected by inflation.

4. (15 points) The weight distribution of male players in hockey teams is approximately normal with mean $\mu = 150$ pounds and standard deviation $\sigma = 20$ pounds.

(a) Draw the distribution of weights and numerically label the horizontal axis.



(b) What is the relative frequency of hockey players heavier than 162.5 pounds.

$$x = 162.5$$
$$z = \frac{162.5 - 150}{20} = 0.625$$

The relative frequency of observations larger than 0.625 is given by the table as : $1 - \frac{(0.7324+0.7357)}{2} \approx 26.59\%$.

(c) A Hurricane hockey player weighs 162.5 pounds. At least how much weight should he gain to be in the top 20% of the distribution?

From the standard normal table, the top 20% corresponds to the z value 0.84. To be in the top 20% of the distribution, a player needs to weigh at least 166.8 pounds i.e $166.8 = (0.84) * 20 + 150$. So the player needs to gain at least 4.3 pounds.

5. (15 points) Center and Spread:

(a) Give a list of 5 numbers with mean 1 and standard deviation 0. Are there more possible choices?

1, 1, 1, 1, 1. There is no other choice because if the standard deviation is 0 than all numbers have to be equal and since the mean is 1 then all numbers have to be equal to 1.

(b) Give a list of 5 numbers with mean 0 and standard deviation 1.

-1 -1 0 1 1.

(c) Draw a density curve with median 0 and IQR 1. (Label numerically both the vertical and horizontal axis, or give the mean and standard deviation).

One possibility is the uniform distribution (shape: rectangle) with values on vertical axis from (-1) to (+1) and on horizontal axis from (0) to (0.5). One other possibility is a Normal distribution with mean 0 and standard deviation ≈ 0.74.

6. (1 point) Bonus question

   (a) From the data in question 2, in which day of the month do you think the highest value of wind speed occurred ?

   It was September 18th, the day Hurricane Isabel reached the North Carolina triangle area.