Lab 3 – Wally's Watermelon Farm

Wally Wilkerson makes a living growing thousands of watermelons on his farm and selling them to grocers. The weight among the watermelons on his farm is a normally distributed random variable with a mean of $\mu = 20$ pounds and a standard deviation of $\sigma = 1.6$ pounds.

SAS and R

Answer the following two questions using your calculator and Z-table, then double-check your answers using SAS and R (see the SAS4 and R4 handouts).

What is the probability a randomly selected watermelon is heavier than 21.92 pounds? Round off to the fourth decimal place.

16.6% of the watermelons are lighter than k pounds. k = ? Round off to the third decimal place.

R

Take your own random sample of 1000 of Wally's watermelons.

```
> wweights=rnorm(1000,20,1.6)
```

Find the mean and standard deviation of your data and confirm that your sample mean and sample standard deviation are close to the corresponding population values.

$\overline{X} \approx \mu = 20$ and $s \approx \sigma = 1.6$

Based on your answer to the first question on this handout, approximately how many of your 1000 watermelons will be heavier than 21.92 pounds?

```
> sum(wweights>21.92)
```

How many actually were? ______________________

Based on your answer to the second question on this handout, approximately how many of your 1000 watermelons will be lighter than k pounds?

How many actually were? ______________________

Since we know the watermelon weights is normally distributed and the sample sze is very large, we are very confident the QQ plot will look like what? Confirm your answer.

```
> qqnorm(wweights)
```

****

SAS and R

Wally has the same set of tables as you. He has an interest in “extreme” observations from normal distributions. For example, the probability an observation will be 3 or more standard deviations greater than the mean is about 0.001349898 and would only occur about once in every 741 observations (on average).

R code:

```
> 1-pnorm(3,0,1)
[1] 0.001349898
> 1/(1-pnorm(3,0,1))
[1] 740.7967
```

SAS code:

```
data extreme;
answer1=1-cdf('normal',3,0,1);
answer2=1/(answer1);
proc print data=extreme noobs;
run;
```

An observation 4 or more standard deviations greater than the mean will occur about once in every how many observations? Round off to the nearest integer.

An observation 5 or more standard deviations greater than the mean will occur about once in every how many observations? Round off to the nearest integer.

An observation 6 or more standard deviations greater than the mean will occur about once in every how many observations? Round off to the nearest integer.

Using what you've learned on the SAS4 and R4 handouts, answer the following questions using SAS and R.

On a Z-curve, what z-value has an area to its left equal to $\frac{1}{1 \text{ million}}$? Round off to the third decimal place.

On a t-curve with 10 degrees of freedom, what t-value has an area to its left equal to 0.84? Round off to the third decimal place.

On a t-curve with 7 degrees of freedom, what t-value has an area to its right equal to 0.65? Round off to the third decimal place.

On a χ^2 curve with 60 degrees of freedom, what χ^2-value has an area to its left equal to 0.99? Round off to the third decimal place.

On a χ^2 curve with 32 degrees of freedom, what χ^2-value has an area to its right equal to 0.04? Round off to the third decimal place.

On a F-curve with 4 numerator and 3 denominator degrees of freedom, what F-value has an area to its left equal to 0.80? Round off to the third decimal place.

On a F-curve with 18 numerator and 35 denominator degrees of freedom, what F-value has an area to its right equal to 0.05? Round off to the third decimal place.

Wally took a random sample of 24 watermelons and counted the number of seeds in each. His data is located in a file named watermelons.csv. The first five rows of the file look like this:

	A
1	seeds
2	304
3	265
4	312
5	328

$$\overline{X} =$$

$$s \approx$$

Find Wally's 95% confidence interval for $\mu =$ the population mean number of seeds per watermelon. (Wally doesn't know σ, so should you compute a Z-interval or T-interval?)

It is necessary to look at QQ plot of the data? Why? What conclusion do you draw from the QQ plot?

****

After you have completed this handout, complete the Canvas quiz titled:
<u>Lab 03 – Wally's Watermelon Farm</u>