Exam II Review Exercises (in-class version)

1. **Standard Normal Curve**: Sketch and compute the specified area under the normal curve. Write down the R command that you use.

a) $z < 2$

b) $z > 0$

c) $1 < z < 2$

d) What z is greater than 99.9% of all z scores?

2. **Heights**: Suppose that for a particular population subjects' heights average out to 64 inches with an sd of 6 inches and have a normal distribution.

a) What percent of the heights were below 72 inches?

b) What percent were above 60 inches?

c) What percent were between 60 and 72 inches?

d) What height was greater than 99.9% of all the heights?

3. **Linear Regression**: In a land animals data set, body weights average out to 4278 kg with an sd of 16,480 kg. Brain weights average out to 0.5 kg with an sd of 1.3 kg. The correlation is -0.005

a) Write down the equation for the regression line for predicting brain weight from animal weight.

b) Sketch the regression line.

c) Estimate the average brain size of an animals that weigh 1 kg (a guinea pig), 465 kg (a cow) and 87,000 kg (brachiosaurus). The true values are 0.0055 kg, 0.423 kg and 0.1545 kg. Why does the regression line do such a poor job?

d) Estimate the weight of an animal with a brain size of 0.157 kg (jaguar). The true value is 100 kg.

4. **Concept of Correlation**:

a) For a representative sample of people age 25 and older, would the correlation between age and systolic blood pressure be positive or negative? Explain.

b) The correlation between blood pressure and the number of cars that a person has ever owned turns out to be positive. How do you account for this association?

c) Suppose that your internet bill is always 30% of your phone bill. What can you say about the correlation?

d) Suppose that your power bill is always twice as much as your phone bill. What can you say about the correlation?

5. **Regression and the Normal Curve**: In the airquality data set, we find that

average temperature $\approx 78°$F,	sd $\approx 9°$F	
average solar radiation ≈ 186 lang,	sd ≈ 90 lang	$r \approx 0.28$

a) Predict the wind speed on a day in which the high temperature is 90°F.

b) Among 90 degree days, about 95% had radiation values in what range?

6. **NHANESIII Data Set**:

a) In the exam data set, what information does the variable `PEP13E5A` give? What kind of variable is it?

b) In the lab data set, what information does the variable `PLPSI` give? What kind of variable is it? What are its units?

c) In the lab data set, what information does the variable `HGP` give? Find the average of all the non-blank values (as you did for other variables in Report 2).

```
lab =read.csv("/Users/ralphwojtowicz/Desktop/math207/data/lab.csv", header=TRUE)
e1 = lab[lab$HGP != 88888, ]
mean(e1$HGP, na.rm=TRUE)
```

7. **Computing the Correlation and RMS Error**. Here is some data:

x	y	z_x	z_y	$z_x\, z_y$
3	9			
5	5			
7	1			

a) Complete the table then calculate the correlation.

b) Someone uses the line $y = 2 - 3\,x$ to approximate the data. Compute the RMS error for this line.

c) What is the RMS error for the regression line?