Math 263, Section 5 **Name**_____

First Test Spring 2014 **Pin (4 figure number, *not* starting with 0)**_____

30 minutes; total 50 points. There are 5 problems.

1. (**5 points**) To determine the health care costs of its employees, a company interviewed a sample of 25 employees.[1] Their medical expenses for the previous year were recorded. The highest expense was accidentally recorded as 10 times its actual value. However, after correcting the error, the corrected amount was still greater or equal to any other expense in the sample. Which of the following must have remained the same after the correction was made?
(*Check those that must have remained the same; put a line through the others*):

~~___Mean~~
__✓_Median
__✓_IQR
~~___Range~~
~~___Standard deviation~~

2. (**6 points**) Ana's standardized score (z-score) for her systolic blood pressure, as compared to the blood pressure for other women her age, is 1.50. Which of the following is a correct interpretation of the z-score?
(*Check those that are correct; put a line through the others*).

~~___ Ana's systolic blood pressure is 150.~~
~~___Ana's systolic blood pressure is 1.5 above the mean systolic blood pressure of women her age.~~
~~___Ana's systolic blood pressure is 1.5 times the mean systolic blood pressure of women her age.~~
__✓_Ana's systolic blood pressure is 1.5 standard deviations above the mean systolic blood pressure of women her age.
~~___Only 1.5% of the women of Ana's age have systolic blood pressure higher than hers.~~
~~___50% of the women of Ana's age have systolic blood pressure higher than hers.~~

[1] Problems based on CB, 1997

3. (**8 points**) The Federal Highway Administration collects data on the number of vehicles in various countries.[2] The following table shows the data for one year in the US and Mexico, in millions of vehicles.

	US	Mexico
Cars	130	9
Other vehicles	80	5

(a) How many vehicles are there in the two countries combined? (*Give your answer in millions*.)

$$\text{Total} = 130 + 80 + 9 + 5 = 224 \text{ million vehicles.}$$

(b) Create the conditional distribution for each country by filling in the six cells outlined in black. Give your answers with 3 digits after the decimal point.

	US	Mexico
Cars	$130/(130+80) = 0.619$	0.643
Other vehicles	$1 - 0.619 = 0.381$	0.357
	1	1

Note:
A conditional distribution for each country is asking for the proportion of each type of vehicle in that country.

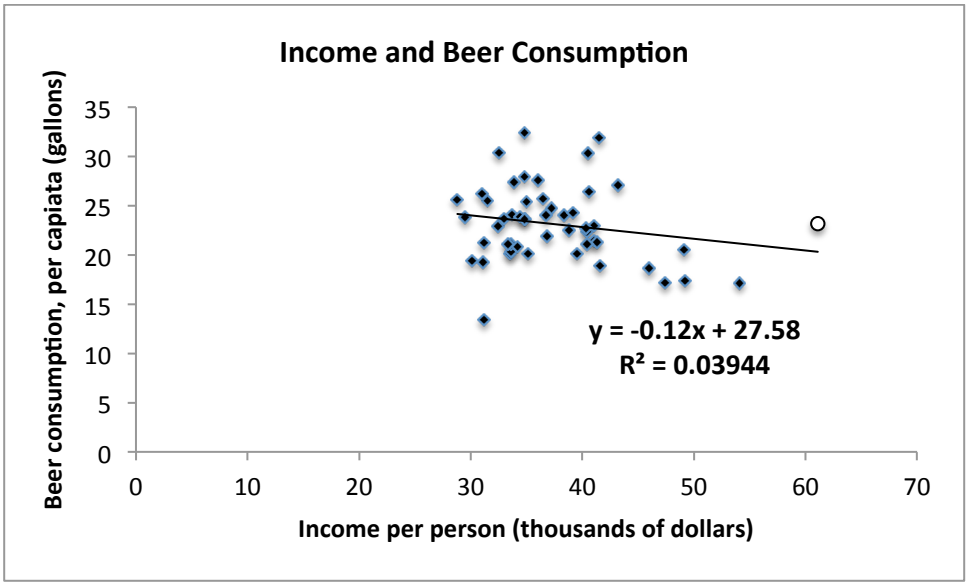
The next question is asking whether the distribution (cars vs. trucks) is affected by which country you are in. It is — but only slightly.
If you said the distributions are different, then there ***is*** an association between country and vehicle distribution. If you said the distributions are approximately the same, then there is no association.

(c) Is there an association between country and vehicle distribution? Yes __X__ No ____ (check one)
Reason for your answer:
The conditional distributions are not the same for each country: There is a larger proportion of cars in Mexico.

OR
Yes __ __ No ___ X _ (check one)
Reason for your answer:
The conditional distributions are almost the same for each country: There are 60-65% cars in each country.

[2] From *Highway Statistics*, reported by Weiss in *Introductory Statistics*, 9th edn (Addison Wesley, 2012).

4. (**12 points**) Do people with higher incomes spend more on beer? The scatter plot shows annual beer consumption (in gallons) plotted against personal income (in thousands of dollars per year) for the each of the US states and DC. A regression line has been fitted to the data.



(a) What is the correlation coefficient? (*Give three decimal places*.)
Since the slope is negative, the correlation coefficient is $r = -\sqrt{0.039} = -0.197$.

(b) What is the slope of the line?
The slope is -0.12.

(c) Interpret the slope in terms of beer. Give units.
The slope tells us that for every additional thousand dollars in personal income, the number of gallons of been consumed per capita is predicted to drop by 0.12 gallons.

(d) What is the intercept of the line?
The intercept is 27.58.

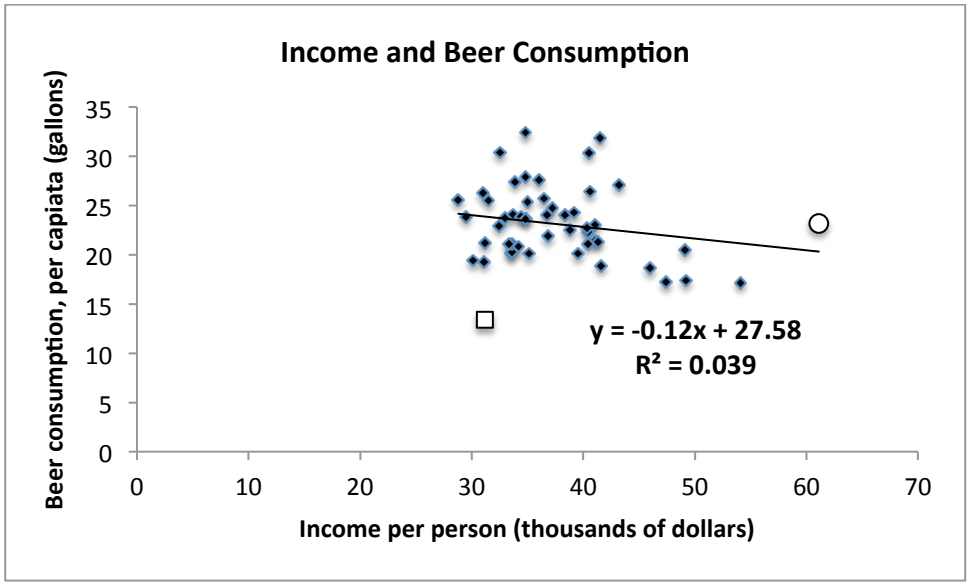
(e) Interpret the intercept in terms of beer. Give units.

The intercept tells us that a state with a personal income of zero is predicted to consume 27.58 gallons of beer. (An unreliable prediction since it involves extrapolating far from the given data.)

(f) If the point representing DC (marked by a hollow circle) were removed and a new regression line were drawn, would the slope of the new regression line be (*check one*):

Smaller (more negative) than the original slope __X__
Larger (less negative) than the original slope ____
Same as the original slope ____

(g) On the graph, circle and label the state (Utah) whose beer consumption is most substantially less than that predicted by the line.



(h) What percent of the variation about the mean beer consumption is predicted by per capita income?
Since $R^2 = 0.039$, we know 3.9% of the variation from the mean is explained by income. (Not a lot!)

5. (**19 points**) In Area A of Tucson, house prices are normally distributed, with mean \$350 thousand, median \$350 thousand, and standard deviation \$72 thousand.

(a) Find the proportion of houses in Area A that are worth more than half a million dollars. (*Show work.*)

Since half a million is \$500 thousand, we want the proportion with z-value greater than

$$z = \frac{500 - 350}{72} = 2.08.$$

From the table, the proportion is $1 - 0.9812 = 0.0188 - 1.88\%$.

(b) Find the proportion of houses in Area A that are worth between a quarter and half a million dollars. (*Show work.*)

For a quarter million dollars,

$$z = \frac{250 - 350}{72} = -1.39.$$

From the table, the proportion lower than a quarter million is 0.0823, so the proportion we need is $0.9812 - 0.0823 = 0.896 = 89.6\%$.

(c) Find the house price at the bottom of the top tenth percentile in Area A. (*Show work.*)

There are 90% of the houses below the top 10th percentile. Thus the z-value is $z = 1.28$, so the house price we are looking for satisfies

$$1.28 = \frac{x - 350}{72},$$

so $x = 350 + 1.28 \cdot 72 = 442.16$, or \$442,160 dollars.

(d) In Area B of Tucson, house prices have a mean of \$280 thousand, a median of \$250 thousand and a standard deviation of \$175 thousand. Mark the following statements as **T**(true) or **F** (false):

__F_ There is a smaller proportion of houses with prices below \$280 thousand in the Area B than in Area A.

Brief reason:
In Area B, more than 50% of the houses are below \$280 thousand because this is above the median of \$250 thousand. In Area A, fewer than 50% of the houses are below \$280 thousand because this is below the median of \$350 thousand.

__T_ The proportion of houses in the Area A with prices below \$350 thousand is the same as the proportion of houses in Area B with prices below \$250.

Brief reason:
Both are the medians, so both proportions are 50%.

__F_ Prices are more consistent (less variable) in Area B than in Area A.

Brief reason:
Prices are more variable in Area B because the SD is larger.

__T_ Prices in Area B are skewed right.

Brief reason:
Mean is larger (\$280 thousand) than the median (\$250 thousand).

__F_ Prices in Area B could be normally distributed.

Brief reason:
Mean and median are not equal. Alternatively, the SD in Area B is large enough that prices 3 SD below the mean would be negative ($280 - 3 \cdot 175 = -245$).