STAT*2040 Statistics I

G. Umphrey **Question Set #1** September, 2012

Note: Work out full solutions and save these. An online assessment will ask you questions based on your solutions. Some extra questions for practice are recommended; remember solutions to odd-numbered text questions are in the Student Solutions Manual. Assistance with R will be provided in the labs. Solutions for all questions (other than those answered in the Student Solutions Manual) will be posted.

1-1. Text, Exercises 1.2 (p.8), 1.34 (p. 16)

1-2. For each of the following sample data sets, use your calculator to efficiently obtain the (i) mean, (ii) standard deviation, and (iii) variance (keep minimum three decimal place accuracy for all statistics):

(a) 6.8, 4.3, 4.8, 5.2, 9.1

(b) −6.8, −4.3, −4.8, −5.2, −9.1

(c) 6.8, 4.3, 4.8, −5.2, −9.1

(d) 26.8, 24.3, 24.8, 25.2, 29.1

(e) 19.8, 19.8, 19.8, 19.8

(f) −19.8, −19.8, −19.8, −19.8

(g) −19.8, −19.8, 19.8, 19.8

(h) −19.8, 19.8, 19.8, 19.8

1-3. For the data set in 1-2 (c), add 10 to each observation. This eliminates all negative values. How does this affect the mean, standard deviation, and variance?

1-4. For the data set in 1-2 (a), triple each observation. How does this affect the mean, standard deviation, and variance?

1-5. Which data set(s) in 1-3 has (have) the largest range, and what is this value?

1-6. (a) The sample variance can be calculated in various ways, since the following expressions for the Sum of Squares are mathematically equivalent:

$$\Sigma(x_i - \bar{x})^2 = \Sigma x_i^2 - (\Sigma x_i)^2/n = \Sigma x_i^2 - n\,\bar{x}^2.$$

Prove the equivalence of the three expressions above. (This is not on the online assessment but make sure you try it!)

(b) If a sample of size n = 26 has a sample mean of 18.6 and $\Sigma x_i^2 = 9402$, what are the values of the variance and standard deviation? [Note the expressions for the Sum of Squares in (a) need to be divided by n − 1 to get the variance.]

1-7. Consider a data set $x_1, x_2, ..., x_n$. A linear transformation $y = a + bx$ is made on each of the x_i values to create a new data set $y_1, y_2, ..., y_n$.

(a) Why is $y = a + bx$ called a "linear transformation"?
(b) How are the values of the sample statistics between the y and x data sets related for each of the following statistics: (i) mean, (ii) median, (ii) mode, (iv) range, (v) variance, (vi) standard deviation?

(c) The z-score for an observation in any sample data set is obtained by subtracting the sample mean and then dividing the difference by the sample standard deviation. Use the relationships you found in (b) to show that the z-scores for any sample data set have a mean of 0 and standard deviation of 1.

1-8. Two sections of the same course wrote a common exam. The mean for the entire class of 540 students was 65.4. If the mean for one of the sections with 290 students was 67.7, what was the mean of the other section?

1-9. Suppose 1000 shares of a particular company were bought, at an average (mean) price of $46.50 per share. If another 2000 shares were purchased, what would the average price of these shares have to be so that the average purchase price of all 3000 shares was $48.00?

1-10. The mean and variance for a sample of body temperatures measured in degrees Fahrenheit were 98.8 and 3.23, respectively. What would be corresponding mean and variance if temperature was measured in degrees Celsius? What if temperature was measured in degrees Kelvin? [Recall $°C = (°F - 32) \times 5/9$ and $°K = °C + 273$.]

Note: Assistance will be available in the labs to help you solve the next question using R statistical software.

1-11. The following data comes from "The Handbook of Small Data Sets" by Hand et al. (1993). This data set will also be made available on our course website in both "unstacked" and "stacked" formats.

308. Presidents, popes and monarchs

Lunn, A.D. and McNeil, D.R. (1991) *Computer-Interactive Data Analysis,* Chichester: John Wiley & Sons, 86.

Listed below are survival times in years from inauguration, election or coronation to death of US Presidents, Roman Catholic Popes and British Monarchs from 1690 to the present.

One question of interest is whether the survival times for the three groups differ in any marked way.

Presidents		Popes		Kings and Queens	
Washington	10	Alex VIII	2	James II	17
J. Adams	29	Innoc XII	9	Mary II	6
Jefferson	26	Clem XI	21	William III	13
Madison	28	Innoc XIII	3	Anne	12
Monroe	15	Ben XIII	6	George I	13
J.Q. Adams	23	Clem XII	10	George II	33
Jackson	17	Ben XIV	18	George III	59
Van Buren	25	Clem XIII	11	George IV	10
Harrison	0	Clem XIV	6	William IV	7
Tyler	20	Pius VI	25	Victoria	63
Polk	4	Pius VII	23	Edward VII	9
Taylor	1	Leo XII	6	George V	25
Filmore	24	Pius VIII	2	Edward VIII	36
Pierce	16	Greg XVI	15	George VI	15
Buchanan	12	Pius IX	32		
Lincoln	4	Leo XIII	25		
A. Johnson	10	Pius X	11		
Grant	17	Ben XV	8		
Hayes	16	Pius XI	17		
Garfield	0	Pius XII	19		
Arthur	7	John XXIII	5		
Cleveland	24	Paul VI	15		
Harrison	12	John Paul	0		
McKinley	4				
T. Roosevelt	18				
Taft	21				
Wilson	11				
Harding	2				
Coolidge	9				
Hoover	36				
F. Roosevelt	12				
Truman	28				
Kennedy	3				
Eisenhower	16				
L. Johnson	9				

(a) Use R to obtain key summary statistics (including the mean, standard deviation and variance) for the combined "Leaders" data as well as for each of the three types of leaders.
(b) Show how the sample means for the three types of leaders can be used to calculate the sample mean of the combined "Leaders" data.
(c) Use R to obtain four frequency histograms, one for the combined "Leaders" and one for each leader group. Keep a common x-axis scale for all graphs. Use a class width of 5, with the first class midpoint being 2.0. (This will give 13 classes for the combined "Leaders" data.)
(d) Describe the shape of each histogram.