Fall 2013

Northwestern University

STAT 202-0 Section 23

Introduction to Statistics

Midterm Exam

Name: _____

October 28, 2013

1. Please do not leave blank for any question.
2. There are 10 questions, each question is 10 points.
3. You may need a normal table to solve question 5 and 6. If you do not bring one, let your instructor know.
4. You have 50 minutes for this exam.
5. Explain briefly = Explain in one sentence or several phrases.

Formulas

Mean (average): $\mu, \overline{X}$
Standard deviation (SD): σ, s

$\sum_{i=1}^{n} X_i = X_1 + X_2 + \cdots + X_n$

$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$

$Z = \frac{X-\mu}{\sigma}$

$X = \mu + \sigma X$

$s_X = \sqrt{\frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n-1}}$

$r = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{s_X}\right) \left(\frac{Y_i - \overline{Y}}{s_Y}\right)$

$r = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{(n-1) s_X s_Y}$

$\widehat{Y} = a + bX$

$b = r \frac{s_Y}{s_X}$

$a = \overline{Y} - b\overline{X}$

1. Plots for Categorical Variables

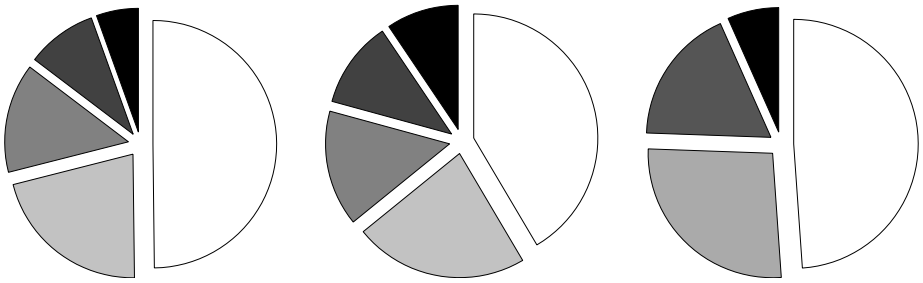
   The U.S. Auto sales in September 2013 are summarized in Table 1 (data source: the Wall Street Journal). The total is 1,139,050.

   Please find out the correct pie chart in Figure 1. The labels for the slices are all missing.

   Explain your choice briefly.

Table 1: Auto Sales, the United States Market, September 2013

	car	pickup	crossover	minivan	SUV
Segment totals	567,409	162,804	241,680	62,647	104,510



(a) (b) (c)

Figure 1: Auto Sales, Pie Charts

2. Plots for Quantitative Variables

   The stem-and-leaf display shows the number of home runs bit by Mark McGwire druing the 1986-2001 seasons.

Table 2: Home Runs (6|5 means 65)

stem	leaf
0	399
1	
2	29
3	22399
4	29
5	28
6	5
7	0

   Find out (a) the minimum, (b) the maximum, and (c) the median.

Figure 2: McGwire hitting a home run in St. Louis against the Tigers on July 14, 2001.

3. Histogram and Box-and-whisker Plot

   Three Physics classes all took the same test. Histograms and boxplots of the scores for each class are shown in Figure 3.

   Match each class with the corresponding boxplot, and explain your matching briefly.

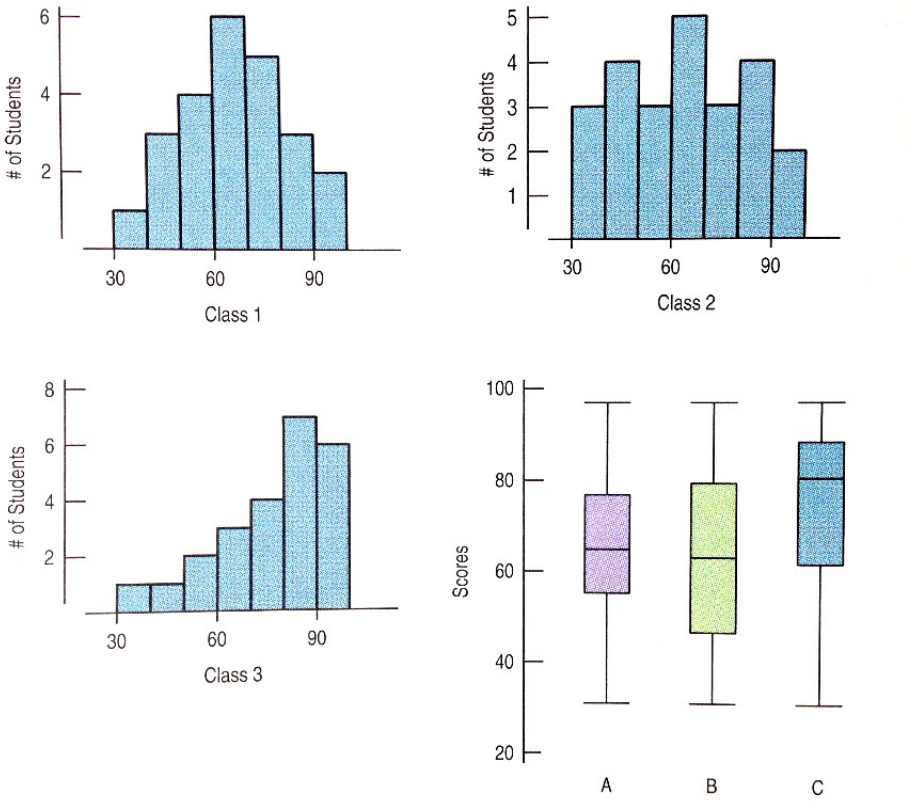


Figure 3: Histograms and boxplots of the scores for each Physics class

4. Describing Distributions with Numbers

The NBA All-Star Game is an exhibition game hosted annually by the National Basketball Association (NBA), matching the league's star players from the Eastern Conference against their counterparts from the Western Conference.

Table 3 shows the numbers of points scored by both teams in the recent 10 All-Star Games.

(a) Find the mean of the total number of points scored by both teams during the recent 10 games.

(b) Find the Standard Deviation (SD) of the total number of points scored by both teams during the recent 10 games.

Table 3: All-Star Game results (2004-2013)

Year	West	East	Total	Deviation	Deviation-squared
2004	136	132	268	-3.5	12.25
2005	115	125	240	-31.5	992.25
2006	120	122	242	-29.5	870.25
2007	153	132	285	13.5	182.25
2008	128	134	262	-9.5	90.25
2009	146	119	265	-6.5	42.25
2010	139	141	280	8.5	72.25
2011	148	143	291	19.5	380.25
2012	152	149	301	29.5	870.25
2013	143	138	281	9.5	90.25
Sum	1380	1335	2715	0.0	3602.50

Figure 4: All-Star Game 2013 at Houston.

5. Normal Distribution

   (You may need a Normal Table for this question.)

   Washington Police officers recorded the speeds of cars driving on a busy street by a school for a one-month period, where the speed limit read 20 mph. The mean of readings was 24.90 mph, with a standard deviation 7.52 mph.

   (data source: www.wtsc.wa.gov)

   What percent of the vehicles were exceeding the posted speed limit in school zones (20 mph)?

Figure 5: Speed limit 20 MPH.

6. Normal Distribution

   (You may need a Normal Table for this question.)

   In Fuel Economy Guide (Model Year 2013), Environmental Protection Agency (EPA) fuel economy estimates for automobile models tested predicted a mean of 23.8 mpg (miles per gallon) and a standard deviation of 6.2 mpg. Assume that a Normal model can be applied.

   (data source: www.fueleconomy.gov)

   An auto dealer introduced you a fuel-efficient car. He told you that this car's gas mileage is higher than 95% of vehicles.

   Find the gas mileage of this car.

Figure 6: Gas Mileage

7. Correlations

Total Fat versus Calories for 4 items on the BK menu are shown in Table 4. (data source: www.bk.com)

Table 4: Nutrition Facts

	Fat (g)	Calories
Whopper Sandwich	35	630
BK Quad Stacker	51	760
BK Veggie Burger	16	410
Alaskan Fish Sandwich	31	590

Table 5: Product of the deviations

	X_i	Y_i	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)$
Whopper Sandwich	35	630	1.75	32.5	56.875
BK Quad Stacker	51	760	17.75	162.5	2884.375
BK VEGGIE Burger	16	410	-17.25	-187.5	3234.375
Alaskan Fish Sandwich	31	590	-2.25	-7.5	16.875
Sum	133	2390	0	0	6192.5

In Table 5, the sum of product of the deviations is given, where X_i's denote total fat and Y_i's denote calories.

We know that the standard deviation of Total Fat content is 14.4 ($s_X = 14.4$), and the standard deviation of Calories is 144.5 ($s_Y = 144.5$).

Find the correlation between Total Fat content and Calories.

Are total fat content and calories related to each other? Explain briefly.

Figure 7: Burger

8. Simple Linear Regression

The data in Table 6 are the geographic latitude and the average August temperatures (Fahrenheit) for 4 cities in the United States.

Table 6: Geographic Latitude and Mean August Temperature

	Latitude	August Temperature
Miami FL	26	83
New York NY	41	76
Chicago IL	42	74
Portland OR	46	69

Table 7: Summary statistics for Geographic Latitude and Mean August Temperature

	Mean	SD	Correlation
Latitude	38.75	8.77	-0.960
August Temperature	75.5	5.80	

The correlation between Geographic Latitude and Mean August Temperature is -0.960. Summary statistics for the two variables are shown in Table 7.

Find the linear regression equation for predicting Mean August Temperature from Geographic Latitude.

Figure 8: Chicago

9. Stratified Sampling

   There are 6 female students and 11 male students in our class. Suppose that I randomly asked two female students and two male students about their heights, the sampling results are summarized in Table 8.

Table 8: Students' Heights

strata	Sample (in)	Sample Average (in)
Female Students	64, 66	65
Male Students	71, 74	72.5

   Estimate the average height of all students in our class based this sample (two female students and two male students).

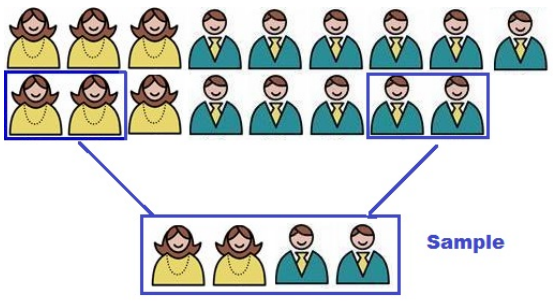


Figure 9: Stratified Sampling

10. Two-way Table and Simpson's paradox

(Effect of race on death-penalty sentences in Florida murder cases)

In 1991, Radelet and Pierce studied the effect of race on death-penalty sentences in murder convictions in the state of Florida.

From Radelet and Pierce's study, we see Caucasian defendants received the death penalty more often than African-American defendants.

But if we consider the very same data, except that we stratify according to the race of the victim of the murder, we see that when considering the cases involving Caucasian victims separately from the cases involving African-American victims, that the African-American defendants are more likely than Caucasian ones to receive the death penalty in both instances.

The following table (two numbers are missing) was obtained tabulating the death-penalty sentences (Death) and non-death-penalty sentences (No death) in murder convictions in the state of Florida. Data are stratified according to the race of the victim of the murder.

Table 9: Death-penalty sentences in Florida murder cases

Victim's race	Defendant's race	Death	No Death
Caucasian	Caucasian	53	414
Caucasian	African-American	?	37
African-American	Caucasian	?	16
African-American	African-American	4	139

(a) What is the percent of Caucasian defendants who received the death penalty when considering the cases only involving Caucasian victims?

(b) Two numbers are missing in Table 9. Try to figure out two suitable non-negative integers (the solutions may not be unique, and you only need to find one set of two suitable non-negative integers).

(Hint: First, summarize the data in Table 9 by only considering the defendant's race, as shown in Table 10. Second, try to write three inequalities, and figure out possible solutions.)

Table 10: Effect of race on death-penalty sentences

Defendant's race	Death	No Death
Caucasian	53+?	430
African-American	4+?	176

More Space

More Space

End of the Midterm of Stat 202-0-23