

Fall 2015
HUNTER COLLEGE
STAT 213 Section 3
Introduction to Applied Statistics
Midterm Exam One

Name: _____

Graduation Year: _____

October 14, 2015

1. Please do not leave blank for any question.
2. There are 7 questions, each question is 10 points.
3. You have 70 minutes for this exam (5:40 pm - 6:50 pm).
4. Explain briefly = Explain in one sentence or several phrases.

Formulas

$$\sum_{i=1}^n x_i = x_1 + x_2 + \cdots + x_n$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$z = \frac{x - \mu}{\sigma}$$

$$x = \mu + \sigma z$$

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

$$s_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

$$\Pr(|X - \mu| \leq k\sigma) \geq 1 - \frac{1}{k^2}, \text{ for any distribution}$$

$$\Pr(|X - \mu| \leq \sigma) \approx 0.68, \Pr(|X - \mu| \leq 2\sigma) \approx 0.95, \Pr(|X - \mu| \leq 3\sigma) \approx 0.997, \text{ for normal distribution}$$

$$r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$$

$$\hat{y} = a + bx$$

$$b = r \frac{s_y}{s_x}$$

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$a = \bar{y} - b\bar{x}$$

$$SSTo = \sum_{i=1}^n (y_i - \bar{y})^2$$

$$SSTo = \sum_{i=1}^n y_i^2 - n\bar{y}^2$$

$$SSResid = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$SSResid = \sum_{i=1}^n y_i^2 - a \sum_{i=1}^n y_i - b \sum_{i=1}^n x_i y_i$$

$$r^2 = 1 - \frac{SSResid}{SSTo}$$

$$s_e = \sqrt{\frac{SSResid}{n-2}}$$

Random Number Table

12872 18212 02190 50256 79500 65210 34639 99795 04327 43848 98727 07531
34234 27566 94454 20349 69224 69483 21821 38248 62410 16481 54270 14344
60679 95118 44916 95522 17144 05395 40643 08340 52134 20753 41452 52797
45320 67751 00459 28894 43588 46388 64547 10072 00054 56665 60274 22889
35043 72024 87641 67346 28230 19021 20090 16885 26498 97659 10735 24621
56406 07936 06463 37439 17953 23294 07272 55338 11140 70292 66278 31434
09408 48929 30366 12613 39316 59206 26094 25430 00863 01122 53461 69887
94050 48120 85909 45984 92318 26757 49997 27162 22226 10476 45725 39980
83773 52393 73092 84437 71657 66721 54971 90220 84475 28268 70330 17587
07148 56945 07552 29174 17424 52673 46928 90721 32783 80040 64827 57350
79781 12488 40923 82176 58418 76576 22101 12084 68695 72304 34919 73631
84053 99671 79376 40260 57609 58677 55473 65086 09688 22765 36651 94994
19965 18493 49468 56541 61881 45860 93925 23170 08879 78308 43464 47996
87517 42396 51200 77903 71236 38123 64018 12893 13152 65490 81917 06079
60150 97939 58013 04348 38787 88585 39192 60813 49064 84312 52009 95803
64422 85121 96466 88989 11420 44128 72563 87258 90057 08216 53741 43723
00334 03943 66559 78713 15693 31310 11016 71899 62691 63759 60554 70167

1. Sampling

Please randomly draw two students from a group of twelve:

- 1 Alan
- 2 Lucy
- 3 Tom
- 4 Azar
- 5 Jayne
- 6 Nadima
- 7 Matthew
- 8 Sushi
- 9 Mohammed
- 10 Rachel
- 11 Ben
- 12 Emma

Briefly describe the sampling procedure you use.

2. Design of Experiments

A clinical trial compares two doses (high and low) of a new medicine. Here are the names of 8 subjects.

- (a) Grace
- (b) Anna
- (c) Sophie
- (d) Karen
- (e) Joshua
- (f) James
- (g) Helen
- (h) Joseph

Please assign 4 subjects at random to each of the two groups. Briefly describe the allocation method you use.

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 1.

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

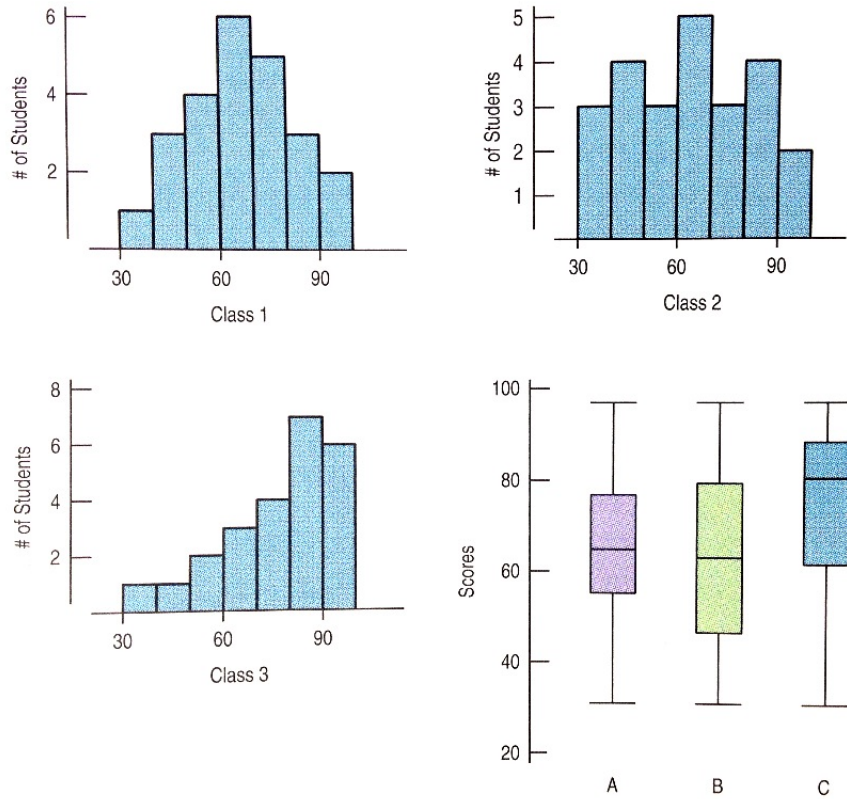


Figure 1: Histograms and boxplots of the scores for each Physics class

4. Plots for Quantitative Variables

The stem-and-leaf display shows the numbers of steals by Shaquille O'Neal during 19 regular seasons of National Basketball Association (NBA) games from season 1992-93 to season 2010-11.

Table 1: Number of Steals (6|0 means 60)

stem	leaf
0	8
1	56
2	3
3	24466689
4	1679
5	
6	0
7	36

Find out (a) the minimum, (b) the maximum, and (c) the median, and (d) the interquartile range, and (e) draw the boxplot.

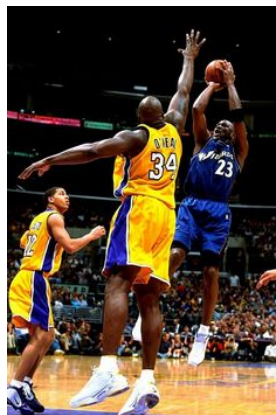


Figure 2: Shaq blocks Jordan.

5. Describing Distributions with Numbers

The Women's National Basketball Association All-Star Game is an annual exhibition basketball game played in the United States between the best players of the Eastern and Western Conference of the Women's National Basketball Association (WNBA).

Table 2 shows the numbers of points scored by both teams in the recent 5 All-Star Games (There is no game held in 2012 due to the 2012 Summer Olympics, and in 2010, the game is between USA women's national team and a team of WNBA All-Stars).

(a) Find the mean of the total number of points scored by both teams during the recent 5 games.

(b) Find the Standard Deviation (SD) of the total number of points scored by both teams during the recent 5 games.

(c) Find the actual percentages of the observations which are within 2 standard deviations of the mean, and find those obtained from Chebyshev's Rule and the Empirical Rule.

Table 2: WNBA All-Star Game results (2009-2015)

Year	West	East	Total	Deviation	Deviation-squared
2009	130	118	248	16.6	275.56
2011	113	118	231	-0.4	0.16
2013	102	98	200	-31.4	985.96
2014	124	125	249	17.6	309.76
2015	117	112	229	-2.4	5.76
Sum	586	571	1157	0.0	1577.20



Figure 3: Logo for the inaugural WNBA All-Star Game, held in 1999

6. Correlations

Total Fat versus Calories for 5 items on the Subway menu are shown in Table 3. (data source: <https://www.subway.com/nutrition/nutritionlist.aspx>)

Table 3: Nutrition Facts

	Fat (g)	Calories
6" Black Forest Ham	4.5	290
6" Roast Beef	5.0	320
6" Turkey Breast	3.5	280
6" Veggie Delite	2.5	230
6" Chicken Teriyaki	4.5	370

Table 4: Product of the deviations

	X_i	Y_i	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})(Y_i - \bar{Y})$
Black Forest Ham	4.5	290	0.5	-8	-4.0
Roast Beef	5.0	320	1.0	22	22.0
Turkey Breast	3.5	280	-0.5	-18	9.0
Veggie Delite	2.5	230	-1.5	-68	102.0
Chicken Teriyaki	4.5	370	0.5	72	36.0
Sum	20	1490	0.0	0	165.0

In Table 4, 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 1.000 ($s_X = 1.000$), and the standard deviation of Calories is 51.672 ($s_Y = 51.672$).

- Find the correlation between Total Fat content and Calories.
- Are total fat content and calories related to each other? Explain briefly.
- Find the coefficient of determination r^2 , the total sum of squares $SSTo$, and the residual sum of squares $SSResid$.



Figure 4: Subway Sandwich

7. Simple Linear Regression

The data in Table 5 are the geographic latitude and the average low January temperatures (Fahrenheit) for 4 cities in the United States.

Table 5: Geographic Latitude and Average Low January Temperature

	Latitude	Temperature
Miami FL	26	59.9
New York NY	41	26.9
Chicago IL	42	18.2
Portland OR	46	35.8

Table 6: Summary statistics for Geographic Latitude and Average Low January Temperature

	Mean	SD	Correlation
Latitude	38.75	8.770	−0.8134
Temperature	35.20	17.966	

The correlation between Geographic Latitude and average low January Temperature is -0.8134 . Summary statistics for the two variables are shown in Table 6.

(a) Find the linear regression equation for predicting average low January Temperature from Geographic Latitude.

(b) List predicted temperatures (\hat{y}_i) based on latitudes by using the least-squares line.



Figure 5: Chicago

More Space

More Space

End of the Midterm 1 of Stat 213 Sec 3 (Instructor: Jiangtao Gou)