SPRING 2017

HUNTER COLLEGE

STAT 213 Section 05

Introduction to Applied Statistics

# Midterm Exam

Last Name: ________________________________

First Name: ________________________________

Graduation Year: __________________________

April 19, 2017

1. Please do not leave blank for any question.
2. There are 8 questions, each question is 5 points. A perfect score is 40 points.
3. You have 70 minutes for this exam (4:15 pm - 5:25 pm).
4. Explain briefly = Explain in one sentence or several phrases.

## Formulas

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

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

$z = \frac{x-\mu}{\sigma}, \quad x = \mu + \sigma z$

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

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

$\Pr\left(|X-\mu| \le k\sigma\right) \ge 1 - \frac{1}{k^2}$, for any distribution

$\Pr\left(|X-\mu| \le \sigma\right) \approx 0.68$, $\Pr\left(|X-\mu| \le 2\sigma\right) \approx 0.95$, $\Pr\left(|X-\mu| \le 3\sigma\right) \approx 0.997$, for normal distribution

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

$\widehat{y} = a + bx$

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

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

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

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

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

$SSResid = \sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2$
$SSResid = \sum_{i=1}^{n} y_i^2 - a\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i y_i$

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

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

If events $E_1, \cdots, E_k$ are all mutually exclusive, then $\Pr(E1 \cup \cdots \cup E_k) = \Pr(E_1) + \cdots + \Pr(E_k)$

The event $E$ and $F$ are independent if and only if $\Pr(E \cap F) = \Pr(E)\Pr(F)$

$\Pr(E \cup F) = \Pr(E) + \Pr(F) - \Pr(E \cap F)$

$\Pr(E \cap F) = \Pr(E|F)\Pr(F) = \Pr(F|E)\Pr(E)$

If events $B_1, \cdots, B_k$ are all mutually exclusive with $\Pr(B_1) + \cdots + \Pr(B_k) = 1$, then for any event $E$, $\Pr(E) = \Pr(E \mid B_1)\Pr(B_1) + \cdots + \Pr(E \mid B_k)\Pr(B_k)$

If events $B_1, \cdots, B_k$ are all mutually exclusive with $\Pr(B_1) + \cdots + \Pr(B_k) = 1$, then for any event $E$, $\Pr(B_i \mid E) = \frac{\Pr(E|B_i)\Pr(B_i)}{\Pr(E|B_1)\Pr(B_1) + \cdots + \Pr(E|B_k)\Pr(B_k)}$

## 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 three students from a group of fifteen by using Simple Random Sampling without Replacement (SRS):

   1 Alan
   2 Lucy
   3 Tom
   4 Azar
   5 Jayne
   6 Nadima
   7 Matthew
   8 Sushi
   9 Mohammed
   10 Rachel
   11 Ben
   12 Emma
   13 Ada
   14 Alex
   15 Mary

   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 to high-dose group and 4 subjects to low-dose group at random. Briefly describe the allocation method you use.

3. Plots for Quantitative Variables

Zlatan Ibrahimovic is a Swedish professional soccer player.

The stem-and-leaf display shows the numbers of goals by season by Zlatan Ibrahimovic from season 1999 to season 2017.

Table 1: Number of Goals (5|1 means 51)

| stem | leaf |
|---|---|
| 0 | 6 |
| 1 | 2 |
| 2 | 9 |
| 3 | 13467 |
| 4 | 2224566779 |
| 5 | 1 |

Find out (a) the minimum, (b) the maximum, (c) the median, (d) the mean, and (e) the interquartile range. (f) Draw the boxplot, and (g) report whether outliers are observed. (h) Discuss the skewness of the distribution of the numbers of goals (symmetric, positively or negatively skewed).

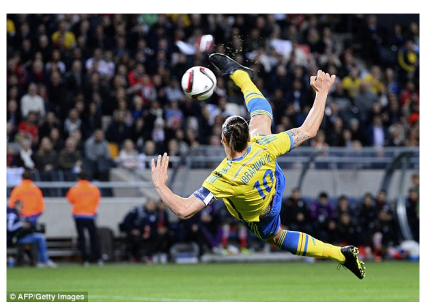



Figure 1: Zlatan Ibrahimovic

4. 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 6 All-Star Games (There is no game held in 2008, 2012 and 2016 due to the 2008, 2012 and 2016 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 6 games (Calculate the average of yearly total)

(b) Find the Standard Deviation (SD) of the total number of points scored by both teams during the recent 6 games.

(c) Find the actual percentages of the observations which are with in 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 (2007-2015)

| Year | West | East | Total | Deviation | Deviation-squared |
|---|---|---|---|---|---|
| 2007 | 99 | 103 | 202 | -24.5 | 600.25 |
| 2009 | 130 | 118 | 248 | 21.5 | 462.25 |
| 2011 | 113 | 118 | 231 | 4.5 | 20.25 |
| 2013 | 102 | 98 | 200 | -26.5 | 702.25 |
| 2014 | 124 | 125 | 249 | 22.5 | 506.25 |
| 2015 | 117 | 112 | 229 | 2.5 | 6.25 |
| Sum | 685 | 674 | 1359 | 0.0 | 2297.50 |

Figure 2: Logo for the inaugural WNBA All-Star Game, held in 1999

5. Correlation and Simple Linear Regression

   The data in Table 3 are the geographic latitude and the average low January temperatures (Fahrenheit) for 5 cities in the United States.

Table 3: Geographic Latitude and Average Low January Temperature

| | Latitude | Temperature |
|---|---|---|
| Miami FL | 26 | 59.9 |
| Philadelphia PA | 40 | 25.6 |
| New York NY | 41 | 26.9 |
| Chicago IL | 42 | 18.2 |
| Portland OR | 46 | 35.8 |

Table 4: Summary statistics for Geographic Latitude and Average Low January Temperature

| | Mean | SD |
|---|---|---|
| Latitude | 39.00 | 7.616 |
| Temperature | 33.28 | 16.141 |

   Summary statistics for the two variables are shown in Table 4.

   (1) Find the correlation between Geographic Latitude and average low January Temperature

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

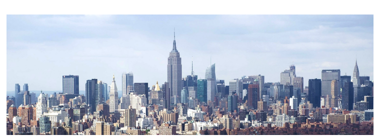


Figure 3: New York City

6. Simple Linear Regression

We made a survey in class, and collected shoe sizes (US system) and heights (feet-inches). Nineteen responses were randomly chosen.

In the United States, the system for shoe size is

$$\text{man shoe size} = 3 \times \text{foot length in inches} - 24$$
$$\text{woman shoe size} = 3 \times \text{foot length in inches} - 22.5$$

The least-square regression line based on the responses is

$$\text{height (inch)} = 24.67 + 3.903 \times \text{foot\_length (inch)}$$

(a) Student "H" in our class only reported his shoe size during this survey. His shoe size is 9.5 (male). Apply the linear model to predict Student "H"'s height from his shoe size (Calculate the predicted height). Is this prediction an interpolation or an extrapolation?

(b) The residual sum of squares (SSResid) is 64.545. Calculate the standard deviation about the least-square line $s_e$. Note that the sample size is 19.

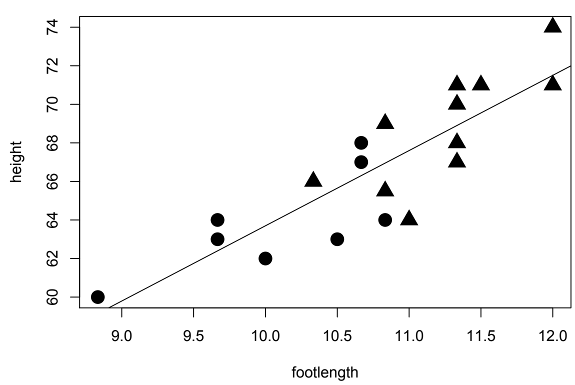
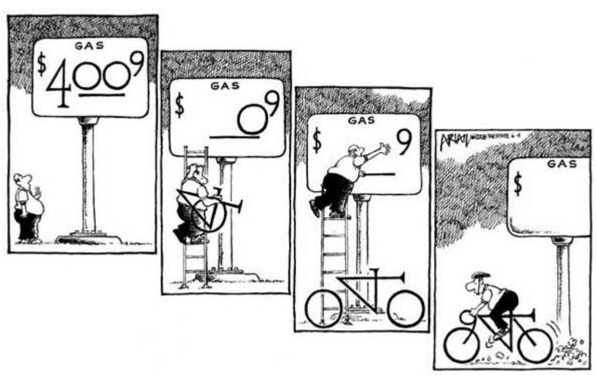



Figure 4: Height (inch) versus Foot length (inch)

7. Probability
At a certain gas station 60% of the customers request regular gas, 15% request plus gas, and 25% request premium gas. Of those customers requesting regular gas, 25% fill up their tanks. Of those customers requesting plus gas, 45% fill up their tanks, while of those requesting premium, 40% fill up their tanks. If the next customer fills up the tank, what is the probability that regular gas is requested? Compare this probability with 60%, and discuss the difference between these two probabilities.



8. Simulation
   One-Boy-or-Three-Child Family Planning:
   Suppose that couples who wanted children were to continue having children until either one boy was born, or three children in total were born. Assuming that each newborn child is equally likely to be a boy or a girl, would this behavior change the proportion of boys in the population? Design a simulation to answer this question. At least simulate 10 families.

More Space

End of the Midterm of Stat 213 Sec 05 (Instructor: Jiangtao Gou)