Introduction to Statistics
Math 1040
Sample Exam I - Chapters 1-4
5 Problem Pages – 1 Formula/Table Page
Time Limit: 90 Minutes[1]   No Scratch Paper   Calculator Allowed: Scientific

Name:_____

The point value of each problem is in the left-hand margin. You must show your work to receive any credit, except on problems 1-3. Work neatly.

(8) 1. Fill in the blanks.

    (a) A matched-pairs design is an experimental design in which the experimental units are   .

    (b) The first quartile of the data 2, 4, 6, 6, 8, 9, 9, 10 is   .

    (c) For a bell-shaped data distribution,   of data are within one standard deviation of the mean.

    (d) The mean and standard deviation of a sample are 12 and 3, respectively. The z-score of the data 18 is   .

(8) 2. True or False.

(   ) (a) The 75th percentile is the value that separates the lower 75% of data from the upper 25%.

(   ) (b) The correlation coefficient r between two variables x and y can be $+2$.

(   ) (c) A designed experiment is a controlled study conducted to determine the effect that varying one or more explanatory variable has on a response variable.

(   ) (d) If the scatter diagram of the residues has a pattern, then the least-squares regression line is not the appropriate model.

(8) 3. A medical researcher wants to estimate the survival time of a patient after the onset of a particular type of cancer and after a particular regimen of radiotherapy.

    (a) What is the variable of interest to the medical researcher?

    (a) Is the variable in part a qualitative, quantitative discrete, or quantitative continuous?

---

[1]**If you exceed the time limit, you will receive a score of zero.**

(10) 4. The table below represents the educational attainment in 1990 and 2000 of adults 25 years and older who are residents of United States. Construct a side-by-side (relative frequency) bar graph to compare the educational attainment in 1990 and 2000.

Educational attainment	1990	2000
Less than 9th grade	16,502,211	12,327,601
9th-12th grade, no diploma	22,841,507	20,343,848
High school diploma	47,642,763	52,395,507
Some college, no degree	29,779,777	36,453,108
Associate's degree	9,791,925	11,487,194
Bachelor's degree	20,832,567	28,603,014
Graduate/professional degree	11,477,868	15,930,061
Totals	**158,868,436**	**177,540,333**

(8) 5. Draw the dot plot of scores in a quiz shown below.

12   18   15   13   15   20   12   15   11   19   18   19   14   12
10   15   18   20   13   14   16   20   16   19   12   11   20   17

(12) 6. Find and state the five-number summary of the following data, construct its box plot and identify its outliers. Show your work.

$$3\,,\ 9\,,\ 10\,,\ 2\,,\ 6\,,\ 7\,,\ 5\,,\ 8\,,\ 6\,,\ 6\,,\ 4\,,\ 4\,,\ 9\,,\ 22$$

(10) 7. The amount of different snack foods eaten during a recent Super Bowl is shown below. Construct a pie chart to describe the data. **Show your work, including angles.**

Snack	Millions of Pounds
Potato chips	11.2
Tortilla chips	8.2
Pretzels	4.3
Popcorn	3.8
Snack nuts	2.5

(8) 8. Construct a stem and leaf plot for the following 50 measurements. Show all your work.

$$
\begin{array}{cccccccccc}
3.1 & 4.9 & 2.8 & 3.6 & 2.5 & 4.5 & 3.5 & 3.7 & 4.1 & 4.9 \\
2.9 & 2.1 & 3.5 & 4.0 & 3.7 & 2.7 & 4.0 & 4.4 & 3.7 & 4.2 \\
3.8 & 6.2 & 2.5 & 2.9 & 2.8 & 5.1 & 1.8 & 5.6 & 2.2 & 3.4 \\
2.5 & 3.6 & 5.1 & 4.8 & 1.6 & 3.6 & 6.1 & 4.7 & 3.9 & 3.9 \\
4.3 & 5.7 & 3.7 & 4.6 & 4.0 & 5.6 & 4.9 & 4.2 & 3.1 & 3.9 \\
\end{array}
$$

(10) 9. The following is a sample of salaries for nine players in NBA for the 2000-2001 season, in thousands of dollars.

$19,290$, $10,130$, $6,500$, $4,800$, $3,400$, $3,380$, $2,250$, $1,760$, $1,200$

Find the sample mean and standard deviation of this data.

(8) 10. Use the random number table to choose five students in random from a class of 123 students. Clearly, describe the procedure you are using including how to get started, how to pick numbers and when a number is skipped, if any, or what to do when the chosen number is too large or is a repeat.

(10) 11. The following table represents the per capita gross domestic product x, in thousands of US dollars, and the average life expectancy y of the population for several countries.

	Austria	Belgium	Finland	France	Germany	Ireland	Italy	Netherlands	Switzerland	United Kingdom
x	21.4	23.2	20.0	22.7	20.8	18.6	21.5	22.0	23.8	21.2
y	77.48	77.53	77.32	78.63	77.17	76.39	78.51	78.15	78.99	77.37

Given that $\bar{x} = 21.52$, $\bar{y} = 77.754$, $s_x = 1.5317$ and $s_y = 0.7948$. Find the correlation coefficient r between x and y. Is there a linear relationship between x and y? If yes, find the least-squares regression line to describe y as a function of x.

$$\mu = \frac{\sum x_i}{N}, \quad \sigma^2 = \frac{\sum (x_i - \mu)^2}{N} = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{N}}{N}, \quad \overline{x} = \frac{\sum x_i}{n}, \quad s^2 = \frac{\sum (x_i - \overline{x})^2}{n-1} = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}$$

$$z\text{-score} = \frac{x - \overline{x}}{s}, \quad s_{xy} = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{n-1} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1}$$

$$r = \frac{\sum \left(\frac{x_i - \overline{x}}{s_x} \right) \left(\frac{y_i - \overline{y}}{s_y} \right)}{n-1} = \frac{s_{xy}}{s_x \, s_y} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sqrt{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} \sqrt{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}}$$

$$\hat{y} = b_1 x + b_0, \quad b_1 = r \left(\frac{s_y}{s_x} \right) = \frac{s_{xy}}{s_x^2} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}, \quad b_0 = \overline{y} - b_1 \overline{x}$$

Table I

Random Numbers
Column Number

| Row Number	01–05	06–10	11–15	16–20	21–25	26–30	31–35	36–40	41–45	46–50
01	89392	23212	74483	36590	25956	36544	68518	40805	09980	00467
02	61458	17639	96252	95649	73727	33912	72896	66218	52341	97141
03	11452	74197	81962	48443	90360	26480	73231	37740	26628	44690
04	27575	04429	31308	02241	01698	19191	18948	78871	36030	23980
05	36829	59109	88976	46845	28329	47460	88944	08264	00843	84592
06	81902	93458	42161	26099	09419	89073	82849	09160	61845	40906
07	59761	55212	33360	68751	86737	79743	85262	31887	37879	17525
08	46827	25906	64708	20307	78423	15910	86548	08763	47050	18513
09	24040	66449	32353	83668	13874	86741	81312	54185	78824	00718
10	98144	96372	50277	15571	82261	66628	31457	00377	63423	55141
11	14228	17930	30118	00438	49666	65189	62869	31304	17117	71489
12	55366	51057	90065	14791	62426	02957	85518	28822	30588	32798
13	96101	30646	35526	90389	73634	79304	96635	06626	94683	16696
14	38152	55474	30153	26525	83647	31988	82182	98377	33802	80471
15	85007	18416	24661	95581	45868	15662	28906	36392	07617	50248
16	85544	15890	80011	18160	33468	84106	40603	01315	74664	20553
17	10446	20699	98370	17684	16932	80449	92654	02084	19985	59321
18	67237	45509	17638	65115	29757	80705	82686	48565	72612	61760
19	23026	89817	05403	82209	30573	47501	00135	33955	50250	72592
20	67411	58542	18678	46491	13219	84084	27783	34508	55158	78742

Introduction to Statistics
Math 1040
Additional (not on the Sample Exam) Review Problems for Chapters 1-4
This is not an exhaustive list of all possible types of problems.

1. Fill in the blanks.

   (a) A _____ is a numerical summary of a sample.

   (b) A _____ is a numerical summary of a population.

   (c) A _____ is an explanatory variable that was not considered in a study, but that affects the value of the response variable in the study.

   (d) _____ is the process of using chance to select individuals from a population to be included in the sample.

   (e) In a relative frequency distribution, the sum of relative frequencies should be _____ .

   (f) The sum of the deviations about the mean always equals _____ .

   (g) A _____ is the difference between the corresponding observed and predicted values.

2. True or False.

   ( ) (a) Inferences based on voluntary response samples are generally not reliable.

   ( ) (b) If the results of the sample are not representative of the population, then the sample has bias.

   ( ) (c) Generally, the goal of an experiment is to determine the effect that treatments will have on the response variable.

   ( ) (d) A data set will always have exactly one mode.

3. Gallup News Service conducted a survey of 1,006 American adults aged 18 years or older, September 24-27, 2007. The respondents were asked, "What, if anything, worries you most about your personal financial situation in the long term?" Of the 1,006 adults surveyed, 18% said they were most worried about having enough money for retirement. Gallup reported that 18% of all adult Americans were most worried about not having enough money for retirement, with a 4% margin of error with 95% confidence.

   (a) What is the research objective?

   (b) What is the population?

   (c) What is the sample?

   (d) List the descriptive Statistics.

   (e) What can be inferred from this survey?

4. The following frequency histogram represents student scores on an exam in a college course. Each score class or subinterval excludes its left-endpoint and includes its right-endpoint, right-inclusion.
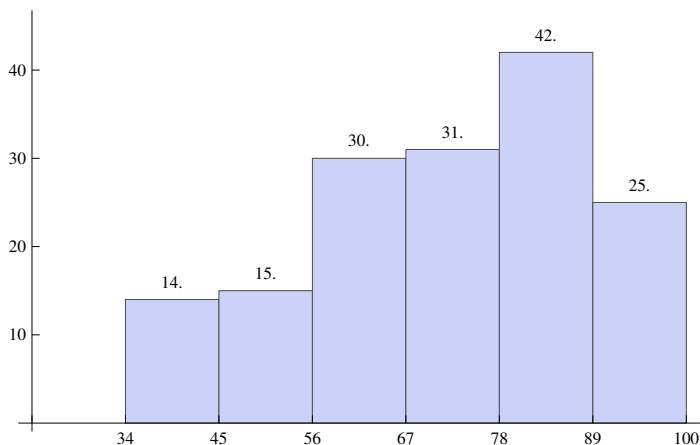
   (a) How many students took the exam?

   (b) What percent of student got a score of higher than 90?

   (c) If scores of 60% or less is considered failing, what percentage of students failed the exam?

5. The Gallup organization conducted a survey in December 2005 in which 1257 randomly selected adult Americans who use Internet were asked, "How often do you, yourself, use the Internet to buy or sell products in online auctions, such as e-Bay?" Construct a pie chart of the result of the survey shown below.

Response	Frequency
Frequently	541
Occasionally	123
Rarely	131
Never	462

6. Dr. Paul Oswiecmiski randomly selectes 40 of his 20- to 29-year-old patients and obtains data regarding their serum HDL cholestrol. Construct a frequency histogram of the data listed below.

73	70	70	69	66	63	62	60
60	58	56	56	56	55	54	53
53	52	52	51	51	51	50	49
48	48	48	46	46	45	45	44
44	39	38	36	35	33	32	28

7. Draw the dot plot of waiting times, in minutes, for getting a table in a popular restaurant shown below.

2	8	5	3	10	15	9	5	12	10	3	8	4	6
10	5	8	12	3	4	8	11	16	9	2	8	15	7

8.   Benjamin owns a small Internet business. Besides himself, he employs nine other people. The salaried earned by the employees, in thousands of dollars, are 30, 30, 45, 50, 50, 50, 55, 55, 60, and 75. Find and state the range, mean, median and mode of this data.

9.   SAT math scores have a bell-shaped distribution with a mean of 515 and a standard deviation of 114.

   (a) What percentage of SAT scores is between 401 and 629?

   (b) What percentage of SAT scores is less than 401 or greater than 629?

   (c) What percentage of SAT scores is greater than 743?

10.   According to the U.S. Census Bureau, the mean of the commute time to work for a resident of Boston, Massachusetts, is 27.3 minutes. Assume that the standard deviation of the commute time is 8.1 minutes to answer the following.

   (a) What minimum percentage of commuters in Boston has a commute time within 2 standard deviations of the mean?

   (b) What minimum percentage of commuters in Boston has a commute time within 1.5 standard deviations of the mean? What are the commute times within 1.5 standard deviations of the mean?

   (c) What is the minimum percentage of commuters who have commute time between 3 minutes and 51.6 minutes?

11.   The average 20- to 29-year old man is 69.6 inches tall, with a standard deviation of 3.0 inches, while the average 20- to 29-year old woman is 64.1 inches tall, with a standard deviation of 3.8 inches. Who is relatively taller, a 75-inch man or a 70-inch woman?

12.   Draw a scatter diagram of the following data.

x	-2	-1	2	4	6	6	7
y	-3	1	0	4	6	9	8

13.   The following data represent the number of days absent, x, and the final grade, y, for a sample of college students in a general education course at a large state university.

No. of absences, x	0	1	2	3	4	5	6	7	8	9
Final grade, y	89.2	86.4	83.5	81.1	78.2	73.9	64.3	71.8	65.5	66.2

Find the least-squares regression line treating number of absences as the explanatory variable and the final grade as the response variable.

14. The following gives amount spent on groceries per week y and the number of household members x.

x	2	2	3	4	1	5
y	$45.75	$60.19	$68.33	$ 100.92	$35.86	$130.62

(a) Given that $\bar{x} = 2.833$, $\bar{y} = 73.612$, $s_x = 1.472$, $s_y = 35.792$ and $r = 0.98$, find the least-squares regression line to describe y as a function of x.

(b) What is the predicted weekly grocery bill of a household of seven people?

(c) Should we use the value in part b to estimate the weekly grocery bill of a household of seven people? why or why not?

Just Answers To Additional (not on the Sample Exam) Review Problems for Chapters 1-4

The following are just answers, not complete solutions!

1. (a) statistic (b) parameter (c) lurking variable (d) Simple random sampling (e) one (f) zero (g) residual

2. (a) True (b) True (c) True (d) False

3. (a) Determine what worries American adults, in the long term, most about their personal finances. (b) American adults aged 18 years old or older (c) 1006 American adults aged 18 years old or older (d) 18% were most worried about having enough money for retirement (e) Answer varies

4. (a) 160 (b) 12.5% (c) 15%

5. Note: Be sure to show calculations of angles!   6. Left-inclusion is used in the histogram.
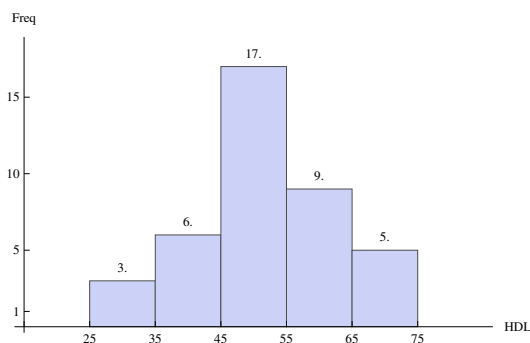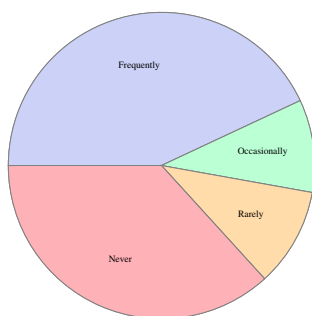
7.
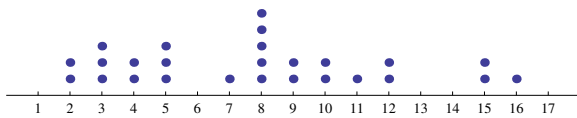
12.
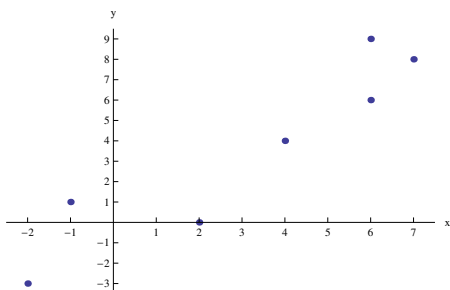
8. Range $= 45$, Mean $= 50$, Median $= 50$, Mode $= 50$

9. (a) 68% (b) 32% (c) 2.5%

10. (a) 75% (b) 55.6%; 15.15 minutes to 39.45 minutes (c) 88.9%

11. The 75-inch tall man is relatively taller since $Z_{\text{man}} = 1.8 > Z_{\text{woman}} = 1.55$.

13. $\hat{y} = -2.83x + 88.73$

14. (a) $\hat{y} = 23.826x + 6.106$ (b) 172.89 (c) Household of seven people is outside the x values of data. So, the estimate can only be used with caution!