MATH 193
STATISTICS PROBLEMS

CONTENTS

Section 1. Centre and Spread of Data

1. Calculate the mean and median for the following population of test scores:
49, 61, 67, 68, 74, 77, 79, 82, 91

2. Calculate the mean and median for the following sample of masses (in grams): 22, 25, 28, 23, 22, 27, 27, 29

3. A math class has four tests. A student has earned the following marks on the first three tests: 52, 69, 73. What mark does the student need on the fourth test in order to have an average of 70 on the four tests?

4. Leah's class of 29 students has a mean test score of 78. Pat's class of 36 students has a mean test score of 72. Find the mean test score if the two sets of test scores are combined into one population.

5. Calculate the mean and median for the sample of temperature readings:

Temperature ($°C$)	Frequency
36.8	4
37.1	6
37.2	2
37.4	5
37.7	8

6. Find the mean and median for the following sample of house prices:

Price ($)	Relative Frequency
600,000	0.1
800,000	0.45
1,000,000	0.3
1,200,000	0.1
1,400,000	0.05

7. Calculate the variance and standard deviation for the following population of test scores: $71, 76, 76, 79, 83$

8. Calculate the variance and standard deviation for the following sample of masses (in grams): $22, 25, 27, 28$

9. Which data set is more spread out, or are they equally spread out?

a) Set 1: $5, 7, 11, 13, 19$ Set 2: $15, 17, 21, 23, 29$

b) Set 1: $5, 7, 11, 13, 19$ Set 2: $10, 14, 22, 26, 38$

10. A small engineering firm has three employees with the following salaries: $\$35,000$, $\$60,000$ and $\$100,000$. State what happens to the mean, median and standard deviation of the salaries in each situation below (i.e. do they increase, decrease or stay the same?)

a) Each employee gets a $\$5,000$ raise
b) Each employee gets a 10% raise
c) The lowest salary is bumped up to $\$50,000$

Section 2. Probability

1. You flip a fair coin three times. Find the probability that you get:
a) at most one head
b) exactly two tails

2. Five employees have the following years of experience: $1, 3, 7, 11$ and 13. If two of these employees are randomly selected for a project, what is the probability that they have at least 15 years of experience in total?

3. You roll a pair of fair 4-sided dice. Find the probability that the two rolls:
a) sum to 4
b) sum to 3 or 4
c) don't sum to 6

4. You will be assigned two of four different products to analyze. Call the products A,B,C and D. Find the probability that:
a) you are assigned products A and C
b) product B is assigned to you
c) product D is not assigned to you

5. We randomly select a number from 1 to 30 inclusive. What is the probability that the number is divisible by 3 or 5?

6. Below is the make-up of employees at an engineering firm.

	Male	Female
Contract	37	41
Permanent	98	55

Find the probability that an employee is:
a) female
b) male or on contract
c) female and permanent

7. The probability that Device A fails is 2.3%. The probability that Device B fails is 3.1%. The probability that both devices fail is 0.3%. Find the probability that neither device fails.

8. Out of 62 job applicants, 35 have their P.Eng. qualification and 23 are fluent in French. Of those who are fluent in French, 17 have their P.Eng. qualification. What is the probability that an applicant has their P.Eng. but does not speak French?

9. Canadian postal codes have the following format:
letter-number-letter number-letter-number, where numbers 0-9 are used. The letters D,F,I,O,Q,U are never used; the letters W and Z cannot be used in the first position.
a) How many postal codes are possible?
b) How many postal codes begin with A?
c) How many postal codes don't end with 0?
d) How many postal codes begin with B or end with 9?

10. Your employer's computer network requires a case-sensitive alphanumeric password that is four symbols long. Find the probability that a password:
a) contains at least one number
b) starts or ends with d

Section 3. Discrete Random Variables

1. Consider the random variable X below:

X	$P(X)$
9.12	0.41
10.89	0.28
12.31	0.13
14.22	0.11
15.06	0.07

Find:
a) $P(10 \leq X \leq 15)$
b) the mean (or expected value) of X
c) the variance of X
d) the standard deviation of X
e) the probability that a value of X lies within two standard deviations of the mean

2. Consider the random variable Y below:

Y	$P(Y)$
-4	0.4
-3	0.2
2	0.1
7	0.3

Find:
a) $P(Y < 2)$
b) the mean (or expected value) of Y
c) the variance of Y
d) the standard deviation of Y
e) the probability that a value of Y lies within one standard deviation of the mean

3. Project A has a 60% probability of earning \$10,000, a 30% probability of earning \$5,000 and a 10% probability of earning nothing. Project B has an 80% probability of earning \$25,000 and a 20% probability of earning nothing.
a) Find the expected earnings of Project A
b) Find the expected earnings of Project B
c) Find the variance of earnings for Project A
d) Find the variance of earnings for Project B
e) In terms of earnings, which project is more risky? (Which project's earnings has a larger variance?)

4. Your engineering firm is considering competing for Project Alpha. The cost of competing for Project Alpha is $\$10,000$. The firm has a 35% probability of success, which will mean revenue of $\$80,000$.
a) Find the probability distribution of the earnings from Project Alpha, where earnings=revenue−cost
b) Find the expected earnings
c) Find the standard deviation of the earnings
d) Project Beta's earnings has a standard deviation of $\$25,000$. In terms of earnings, which project is more risky: Project Alpha, or Project Beta?

5. You want to insure a used car worth $\$4,000$ against theft (not damage) for one year by paying a premium m. The probability of theft during the year is 1.3%.
a) Find the probability distribution of the insurance company's gain.
b) Find the premium if the insurance company expects to gain $60.

Section 4. Binomial, Hypergeometric and Poisson Distributions

1. How many ways are there to do the following:
a) Choose three objects from a group of eight different objects if order does not matter?
b) Choose five employees out of a group of 20 to form a team?

2. A basketball player makes 72% of his free throws. He does not improve with practice. Find the probability that in his next six free throw attempts:
a) he makes exactly five of them
b) he makes at least four of them
c) he makes at least two of them

3. A multiple choice test has 20 questions, each of which has 3 possible answers. A student guesses randomly on each question. What is the probability that the student gets:
a) exactly 6 questions right?
b) between 5 and 7 questions right (inclusive)?
c) at most 3 questions right?

4. In Lotto 6/49, a player chooses 6 different numbers from a set of 49 numbers. Six winning numbers are drawn at the end of the week. Find the probability of picking exactly three winning numbers.

5. A package of 30 drill bits contains five defective drill bits. Three drill bits are randomly selected from the package.
a) Find the probability distribution for $X =$ the number of defective drill bits selected.
b) The batch of drill bits is rejected by quality control if at least two of the three selected drill bits are defective. Find the probability that the batch is rejected.

6. In a town of $100,000$ residents, 65% of residents got a flu shot last year. Twenty residents are randomly selected. Find the probability that between 10 and 12 of the selected residents (inclusive) got a flu shot last year.

7. At a college with $10,000$ students, 12% are technology students. Fifty students are randomly selected. Find the probability that at most three of the selected students are technology students.

8. For a particular cement mix, the average number of cracks per cubic metre of concrete is 1.7. Find the probability that a randomly-chosen cubic metre of concrete has:
a) at least one crack
b) at most three cracks

9. Suppose 400 typos are distributed randomly throughout a textbook that is 1000 pages long. Find the probability that a given page contains:
a) exactly two typos
b) more than one typo

10. A web server receives an average of 3 requests per 15-minute interval. What is the probability that the server receives at most 4 requests in the next hour?

Section 5. Continuous Random Variables

1. The probability density function for X is $f(x) = \begin{cases} x, & 0 < x \leq 1 \\ 1/4, & 1 < x \leq 3 \\ 0, & \text{otherwise} \end{cases}$

Find:
a) $P(X = 1.5)$
b) $P(0.5 < X < 1.5)$
c) $P(0.5 \leq X \leq 1.5)$
d) $P(X > 1.5)$
e) $P(X < 0.5)$

2. Find the mean and standard deviation of X with probability density function:

$$f(x) = \begin{cases} x, & 0 < x \leq 1 \\ 1/4, & 1 < x \leq 3 \\ 0, & \text{otherwise} \end{cases}$$

3. a) Find the value of k that makes $f(x)$ a valid probability density function:
$f(x) = \begin{cases} kx^4, & 0 \leq x \leq 2 \\ 0, & \text{otherwise} \end{cases}$

b) Find the mean of X with probability density function $f(x)$
c) Find the standard deviation of X with probability density function $f(x)$

4. Let X represent how often a student studies alone, as a proportion of their total study time. (For example, $X = 0.35$ indicates that a student spends 35% of their total study time alone.) The probability density function of X is:
$f(x) = \begin{cases} \frac{1}{(\ln 2)(x+1)}, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$
Find the probability that a student studies alone:
a) exactly 20% of the time
b) less than 20% of the time
c) at least 30% of the time
d) between 50% and 75% of the time

5. The time it takes students to complete a certain project is a uniform continuous random variable with values between 2 and 11 hours. Find:
a) the probability density function for the completion time
b) the probability that a student takes between 3 and 8 hours to complete their project
c) the probability that a student takes more than 7 hours to complete their project
d) the probability that a student takes less than 4 hours to complete their project

6. The lifetime of a certain machine part (in years) has probability density function $f(x) = \begin{cases} 2e^{-2x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$
a) Find the probability that a part lasts less than 0.1 years.
b) Using part a), find the probability that a part lasts at least 0.1 years.

7. The shelf life of a brand of tomato soup (in months) has probability density function $f(x) = \begin{cases} 0.1e^{-0.1x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$
a) Find the probability that a can of soup has a shelf life between two and five months.
b) Find the average shelf life given that $\int_0^\infty xe^{-kx}\, dx = \frac{1}{k^2}$ for $k > 0$.

Section 6. The Normal Distribution

1. Let z be the standard normal random variable. Find:
a) $P(0 \leq z \leq 1.20)$
b) $P(-2.81 \leq z \leq 0)$
c) $P(-1.98 \leq z \leq 3.41)$
d) $P(1.35 \leq z \leq 1.85)$
e) $P(-2.93 \leq z \leq -1.90)$
f) $P(z \geq 2.46)$
g) $P(z \leq -1.34)$

2. In a certain town, the weights of adult males are normally distributed with a mean of 180 pounds and a standard deviation of 22 pounds. Find the probability that an adult male weighs:
a) less than 160 pounds
b) more than 170 pounds

3. The mass of a certain brand of chocolate bar is normally distributed with a mean of 85 grams and a standard deviation of 0.8 grams. Find the probability that a chocolate bar has a mass between:
a) 84 and 86.2 grams
b) 83.5 and 84.5 grams

4. The lengths of drill bits are normally distributed with a mean of 3.01 cm and a SD of 0.12 cm. Find the length that is shorter than the longest 23% of drill bit lengths.

5. The volume of juice in juice cartons is normally distributed with a mean of 250mL and a standard deviation of 12.0mL. Find the volume that is larger than the smallest 13% of juice carton volumes.

6. At a manufacturing plant the diameters of ball bearings are normally distributed with a mean of 10mm. Find the standard deviation of the diameters if 78.88% of ball bearings have diameters between 9mm and 11mm.

Section 7. Sampling Plans and The Central Limit Theorem

1. In the context of a random sample from a population, what does the word **random** mean?

2. Consider the following population: $1, 2, 4, 8, 9$.
a) How many possible samples of $n = 3$ measurements can be chosen from this population?
b) Write out all the possible samples of $n = 3$ measurements.

3. For quality control purposes, 40 machines need to be split randomly into two groups of size 20. Explain how you could use a simple random sample to accomplish this.

4. A beverage company produces one brand of pop, in regular and diet versions. Today the company produced 10,650 cans of regular pop and 4,350 cans of diet pop. A random sample of 600 cans is required from today's production. How many of each type of pop should be included in the sample?

5. Name each sampling method described below:
a) Starting with a random ball bearing, every 50th ball bearing coming off the manufacturing line is selected for further inspection.
b) Twenty soil samples are in test tubes labelled $1, 2, \ldots, 20$. A random number generator is used to select 5 soil samples for analysis.
c) A manufacturing company's drill bits are divided into two sizes: 40% are long and 60% are short. A random sample of 12 long drill bits and a random sample of 18 short drill bits are selected for further inspection.
d) A mining company's operations are divided over 23 sites, each containing several mines. A random sample of 3 sites is selected and every mine at the selected sites is investigated.

6. At a walk-in clinic the average amount of time patients spend with a doctor is 7.1 minutes. The standard deviation is 5.2 minutes. Sixty patient records are selected at random. Find the probability that the mean visit time for these 60 patients is between 6 and 8 minutes.

7. The test scores in a large Calculus class have a mean of 71 and a standard deviation of 9. Find the probability that a random sample of n tests have an average score of less than 70 if:
a) $n = 40$
b) $n = 250$

8. The mean radius of ball bearings at a manufacturing plant is 10mm, with a standard deviation of 1mm. Forty ball bearings are randomly selected. If their mean radius is less than 9.5mm or more than 10.5mm, the manufacturing process is stopped for adjustments. What is the probability that the manufacturing process is stopped?

9. At an engineering firm, employees worked a mean of 45.5 hours last week, with a standard deviation of 6 hours. One hundred and sixty employees are selected at random. What is the probability that their work hours totalled more than 7344 hours?

10. A machine is filling cans of pop. The volume per can has a standard deviation of 1.9 mL. What should the volume be set to on the machine (this is μ) in order to ensure that in a random sample of 30 cans, there is a 99% probability that the mean is at least 355.0 mL?

Section 8. Inferences about the Population Mean

1. A random sample of 40 test marks from a large Calculus class had an average of 78.0 and a standard deviation of 5.0. Find a 95% confidence interval for the average test mark.

2. A random sample of 100 ball bearings is pulled from a manufacturing line today. The sample has an average radius of 9.9mm and a standard deviation of 0.4mm. Find a 90% confidence interval for the average radius among all ball bearings manufactured today.

3. A random sample of 50 temperature readings are taken in a lab. The mean temperature is 28.00°C with a standard deviation of 3.32°C. Find an upper confidence bound for the mean temperature in the lab with:
a) 95% confidence
b) 98% confidence

4. At a paper factory, the paper length is known to have a standard deviation of 0.08 inches. In a random sample of 100 sheets, the mean length is found to be 11.00 inches. Find a lower confidence bound for the mean length among all sheets of paper produced at the factory with:
a) 90% confidence
b) 99% confidence

5. Based on a company's historical data, the thickness of their sheet metal is known to have a standard deviation of 0.02 inches. We want a 98% confidence interval for the mean thickness among all their sheet metal with a margin of error less than 0.003 inches. Find the minimum sample size.

6. At ABC Cereal Company, the mass of cereal boxes has a standard deviation of 13 grams. We want a 95% confidence interval for the mean mass among all their cereal boxes with a margin of error less than 2 grams. Find the minimum sample size.

7. Consider a large-sample confidence interval for the population mean. Describe the effect of the following on the margin of error, assuming that the other quantities remain unchanged:
a) the sample size increases
b) the standard deviation increases
c) the confidence level increases
d) the sample mean increases

8. Ten random water samples taken from the inner harbour yield a mean nitrate ion concentration of 25.0ppm with a standard deviation of 5.1ppm. The ion concentrations are normally distributed throughout the inner harbour. Find a 95% confidence interval for the mean ion concentration in the inner harbour.

9. Fifteen randomly selected ropes had a mean breaking strength of 69.1 pounds, with a standard deviation of 3.5 pounds. The breaking strengths of this brand of rope are known to be normally distributed. Find a 99% confidence interval for the mean breaking strength of this brand of rope.

10. Fuel efficiencies for city driving are measured for a random sample of twelve 2012 Prius cars. The mean fuel efficiency was 51 miles per gallon, with a standard deviation of 2 miles per gallon. The fuel efficiencies of 2012 Prius cars are normally distributed. Find an upper confidence bound for the mean fuel efficiency among all 2012 Prius cars with:
a) 90% confidence
b) 97.5% confidence

11. At a bottling plant, volumes are measured for a random sample of 20 cans of pop. The mean volume was 356.1mL, with a standard deviation of 1.9mL. The volumes among all cans at the bottling plant are normally distributed. Find a lower confidence bound for the mean volume among all cans at the bottling plant with:
a) 95% confidence
b) 99% confidence

Section 9. Linear Regression

1. For each data set below, use a scatterplot to decide whether the association is best described as positive, negative, zero, or nonlinear.

a)

x	10	15	25	30	40	50
y	78	71	45	41	20	3

b)

x	1	5	7	11	13	14
y	33	18	2	5	19	24

c)

x	1.1	1.4	1.6	1.6	1.8	1.9
y	1.3	1.7	2.1	2.3	2.6	2.9

d)

x	5	10	15	20	25	30	35
y	7	9	6	8	7	9	6

2. Consider a bivariate data set. What percentage of the variation in y is accounted for by the best-fit line if the correlation coefficient is:
a) 0.8?
b) -0.9?

3. Bivariate Data Set A has a correlation coefficient of 0.84. Bivariate Data Set B has a correlation coefficient of -0.96. Which data set has a stronger linear association? Explain.

4. International Travel to the USA
(Data from the 2011 World Almanac)

x = Year	y = Number of Visitors (millions)
1991	42.7
1994	44.8
1997	47.8
2000	50.9

The equation of the best-fit line is $y = 0.92x - 1789.31$ and the coefficient of determination is 0.9927.

a) Is the linear association positive or negative?
b) Find the correlation coefficient.
c) What percentage of the variation in y is accounted for by the best-fit line?
d) How many visitors does the best-fit line predict for 1992?
e) Why should we not use the data to predict the number of visitors for 2001?
f) According to the best-fit line, which year corresponds to 48.85 million visitors?

5. Consider the bivariate data set below:

x = Hours of TV Watched Last Week	y = Test Mark
1	91
2	93
2	85
3	72
5	71
7	62
12	53

The equation of the best-fit line is $y = -3.58x + 91.65$ and the coefficient of determination is 0.8469.

a) Is the linear association positive or negative?
b) Find the correlation coefficient.
c) What percentage of the variation in y is accounted for by the best-fit line?
d) What mark does the best-fit line predict for a student who watched 4 hours of TV?
e) Why should we not predict the mark for a student who watched no TV last week?
f) According to the best-fit line, what amount of TV viewing corresponds to a test mark of 75?

6. Consider the following bivariate data set:

x	y
1	1
2	2
3	5

a) Fill in the table below:

*	x	y	x^2	xy	y^2
*					
*					
*					
SUM					
MEAN			*	*	*

b) Compute $S_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$

c) Compute $S_{xy} = \Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n}$

d) Compute $S_{yy} = \Sigma y^2 - \frac{(\Sigma y)^2}{n}$

e) Compute $b_1 = \frac{S_{xy}}{S_{xx}}$

f) Compute $b_0 = \overline{y} - b_1\overline{x}$

g) Find the best-fit line $\hat{y} = b_0 + b_1 x$

h) Find the coefficient of determination $r^2 = \frac{(S_{xy})^2}{S_{xx}S_{yy}}$

ANSWERS

1. Centre and Spread of Data

1. $\mu = 72$; median $= 74$
2. $\overline{x} \approx 25$ grams ; median $= 26$ grams
3. 86
4. $\mu \approx 74.7$
5. $\overline{x} \approx 37.3°C$; median $= 37.4°C$
6. $\overline{x} = \$910,000$; median $= \$800,000$
7. $\sigma^2 = 15.6$; $\sigma \approx 3.9$
8. $s^2 = 7$ g^2; $s \approx 2.6$ g
9. a) Data sets are equally spread out.
b) Set 2 is more spread out.
10. a) Mean and median both increase (by $\$5,000$); SD stays the same.
b) Mean, median and SD all increase (by 10%)
c) Mean increases. Median stays the same. SD decreases.

2. Probability

1. a) $\frac{4}{8}$ b) $\frac{3}{8}$
2. $\frac{4}{10}$
3. a) $\frac{3}{16}$ b) $\frac{5}{16}$ c) $\frac{13}{16}$
4. a) $\frac{1}{6}$ b) $\frac{3}{6}$ c) $\frac{3}{6}$
5. $\frac{14}{30}$
6. a) $\frac{96}{231}$ b) $\frac{176}{231}$ c) $\frac{55}{231}$
7. 0.949 or 94.9%
8. $\frac{18}{62}$
9. a) 7.2 million b) $400,000$ c) 6.48 million d) 1.08 million
10. a) 0.51 or 51% b) 0.03 or 3%

ANSWERS

3. Discrete Random Variables

1. a) 0.52 b) μ or $E(X) = 11.0071$ c) $\sigma^2 \approx 3.9699$ d) $\sigma \approx 1.99$ e) 0.93

2. a) 0.6 b) μ or $E(Y) = 0.1$ c) $\sigma^2 = 23.29$ d) $\sigma \approx 4.8$ e) 0.7

3. a) \$7,500 b) \$20,000 c) 11,250,000 $\$^2$ d) 100,000,000 $\$^2$
e) In terms of earnings, Project B is riskier.

4. a)

X	$P(X)$
70,000	0.35
-10,000	0.65

b) \$18,000 c) $\sigma \approx \$\ 38,000$
d) In terms of earnings, Project Alpha is riskier.

5. a)

X	$P(X)$
m	0.987
m-4,000	0.013

b) \$112

4. Binomial, Hypergeometric and Poisson Distributions

1. a) $8C3 = 56$ b) $20C5 = 15,504$

2. a) 0.33 b) 0.78 c) 0.99

3. a) 0.18 b) 0.51 c) 0.06

4. 0.02

5. a)

X	$P(X)$
0	0.567
1	0.369
2	0.062
3	0.002

b) 0.064

6. 0.35

7. 0.13

8. a) 0.82 b) 0.91

9. a) 0.05 b) 0.06

10. 0.008

ANSWERS

5. Continuous Random Variables

1. a) 0 b) 0.5 c) 0.5 d) 0.375 e) 0.125
2. μ or $E(X) = \frac{4}{3}$ and $\sigma \approx 0.80$
3. a) $k = \frac{5}{32}$ b) $\frac{320}{192} = \frac{5}{3}$ c) $\sigma \approx 0.28$
4. a) 0 b) 0.26 c) 0.62 d) 0.22
5. a) $f(x) = \begin{cases} \frac{1}{9}, & 2 \le x \le 11 \\ 0, & \text{otherwise} \end{cases}$ b) $\frac{5}{9}$ c) $\frac{4}{9}$ d) $\frac{2}{9}$
6. a) 0.18 b) 0.82
7. a) 0.21 b) 10 months

6. The Normal Distribution

1. a) 0.3849 b) 0.4975 c) 0.9758 d) 0.0563 e) 0.0270 f) 0.0069 g) 0.0901
2. a) 0.1814 b) 0.6736
3. a) 0.8276 b) 0.2342
4. 3.10 cm
5. 236 mL
6. 0.8 mm

ANSWERS

7. **Sampling Plans and The Central Limit Theorem**

1. Every measurement in the population has an equal probability of being selected in the sample.

2. a) $5C3 = 10$
b) $\{1, 2, 4\}, \{1, 2, 8\}, \{1, 2, 9\}, \{1, 4, 8\}, \{1, 4, 9\},$
$\{1, 8, 9\}, \{2, 4, 8\}, \{2, 4, 9\}, \{2, 8, 9\}, \{4, 8, 9\}$

3. Select a simple random sample of 20 machines. The selected machines will be Group 1; the unselected machines will be Group 2.

4. 426 cans of regular pop and 174 cans of diet pop

5. a) 1-in-50 systematic sample b) simple random sample
c) stratified random sample d) cluster sample

6. 0.8594

7. a) 0.2420 b) 0.0392

8. 0.0016

9. 0.2005

10. 355.8 mL

ANSWERS

8. **Inferences about the Population Mean**

1. $76.5 \leq \mu \leq 79.5$

2. $9.8 \leq \mu \leq 10.0$

3. a) $\mu \leq 28.77$ b) $\mu \leq 28.96$

4. a) $10.99 \leq \mu$ b) $10.98 \leq \mu$

5. $n = 241$ is the minimum sample size

6. $n = 163$ is the minimum sample size

7. a) margin of error decreases b) margin of error increases
c) margin of error increases d) margin of error doesn't change

8. $21.4 \leq \mu \leq 28.6$

9. $66.4 \leq \mu \leq 71.8$

10. a) $\mu \leq 52$ b) $\mu \leq 52$

11. a) $355.4 \leq \mu$ b) $355.0 \leq \mu$

ANSWERS

9. Linear Regression

1. a) negative b) nonlinear c) positive d) zero

2. a) 64% b) 81%

3. Data Set B because $|r_B| > |r_A|$
Alternatively: Data Set B because $(r_B)^2 > (r_A)^2$

4. a) positive b) $r = 0.996$ c) 99.27% d) 43.3 million e) We should not predict outside the interval $1991 \leq x \leq 2000$ f) 1998

5. a) negative b) $r = -0.92$ c) 84.69% d) 77 e) We should not predict outside the interval $1 \leq x \leq 12$ f) 5 hours

6. a)

*	x	y	x^2	xy	y^2
*	1	1	1	1	1
*	2	2	4	4	4
*	3	5	9	15	25
SUM	6	8	14	20	30
MEAN	2	$\frac{8}{3}$	*	*	*

b) $S_{xx} = 2$ c) $S_{xy} = 4$ d) $S_{yy} = \frac{26}{3}$e) $b_1 = 2$

f) $b_0 = -\frac{4}{3}$ g) $\hat{y} = -\frac{4}{3} + 2x$ h) $r^2 = \frac{48}{52} \approx 0.92$