Math 193 Statistics Examples

1. Centre and Spread of Data

1. Find the mean and median for the population 1, 6, 1, 8, 1, 1, 9.

2. Find the mean and median for the sample 92, 99, 96, 97.

3. A student has test marks 58, 63, 71. What mark on his 4th test gives him an average of 70?

4. If the measurements for the following two populations are combined into one population, find the mean.

	# of measurements	μ
Population 1	43	71
Population 2	26	68

5. Find the mean and median for the following sample:

Temperature (°C)	Frequency
22	11
23	6
25	3

6. Find the mean and median for the following sample:

mass (g)	relative frequency
84	0.1
85	0.85
86	0.05

7. Which sample is more spread out?

   (a) 1, 4, 10

   (b) 31, 36, 38

8. Two machines are filling 355 mL cans of pop. A sample of volumes has the following means and variances (in mL).

	Machine 1	Machine 2
$\overline{x}$	355.8	355.2
s	0.3	1.4

(a) Which machine is more accurate?

(b) Which machine is more precise?

9. Let a population consist of the salaries at a small engineering firm. What happens to the mean, median and SD in each situation:

(a) Each employee get a \$2,000 raise.

(b) Each employee's salary is doubled.

(c) The highest salary is decreased by \$10,000.

2. Probability

1. An experiment consists of flipping a fair coin 3 times and noting if it's heads (H) or tails (T). What is the probability of getting one or two heads?

2. An experiment consists of randomly selecting a number between 1 and 40 (inclusive). Find the probability of getting a multiple of 5 or 7.

3. An experiment consists of rolling a pair of fair die and noting the number. Find the probability of getting a sum of at most 5.

4. Four students own the following number of textbooks: 4, 6, 7, 9. Pick two of the students at random (order doesn't matter). Find the probability that they have at least 15 books in total.

5. Forty Math 193 students were asked 2 questions:
Do you like stats?
Do you understand stats?

The results are:

	like	don't like
understand	19	1
don't understand	3	17

(a) Represent this data in a Venn diagram.

(b) Find the probability that a student likes stats.

(c) Find the probability that a student likes stats but doesn't understand it.

(d) Find the probability that a student likes stats or doesn't understand it.

6. In a class of 45 students, 26 have jobs and 17 have cars. Of those who don't have a car, 10 have jobs. Find the probability that a student has:

   (a) a car or a job.

   (b) a car but not a job.

7. On an given day, the probability that Machine I breaks down is 4%, the probability that Machine II breaks down is 7%, and the probability that both machines break down is 2%. Find the probability that Machine II breaks down and Machine I doesn't.

8. A password consists of 7 digits, each chosen from 0,1,2, $\cdots$, 9. Find the

   (a) total number of passwords possible.

   (b) number of passwords that end with 3.

   (c) number of passwords that don't end with 3.

   (d) the probability that a password starts with 4.

   (e) the probability that a password doesn't start with 4.

   (f) the probability that a password doesn't start with 32.

   (g) the probability that a password contains a least one 4.

   (h) the probability that a password starts with 29 or ends with 1.

3. Discrete Random Variables

1. Given the following probability distribution

x	$P(x)$
-5	0.15
-2	0.2
1	0.4
6	0.25

find:

(a) $P(-2.5 \leq X \leq 2.5)$.

(b) the mean of X.

(c) the population variance of X.

(d) the standard deviation of X.

(e) the probability that an x-value lies within one standard deviation of the mean.

2. Project 1 has a 35% chance of earning \$0, a 50% chance of earning \$300,000, and a 15% chance of earning \$800,000.

   Project 2 has a 60% chance of earning \$0 and a 40% chance of earning \$1,000,000.

   (a) Find the probability distributions of the earnings for each project.

   (b) Find the expected earnings for each project.

   (c) Find the standard deviation of earnings for each project.

   (d) Which project has higher expected earnings?

   (e) In terms of earnings, which project is riskier?

3. Suppose you want to insure a \$2,000 tablet against theft for one year by paying a premium m, and that the probability of theft is 4.7%.

   (a) Find the probability distribution of the insurance company's gain.

   (b) Find the premium (i.e. the value of m) if the insurance company expects to gain \$40.

4. Binomial, Hypergeometric and Poisson Distributions

1. Given 4 objects A, B, C, D, how many ways are there to choose 2 of the objects?

2. How many ways are there to

   (a) choose 7 objects from a group of 12 different objects?

   (b) choose 4 students out of 45 to be on a committee?

3. A drilling company is successful on 82% of its drilling attempts. Find the probability of having at least 7 successes in the next 8 attempts.

4. Roll a die 13 times. Find the probability of rolling at most three 2s or 3s.

5. A dart-thrower hits the target 36% of the time. He does not improve with practice. Find the probability that he hits the target 2 or 3 times in 10 throws.

6. An engineering firm has 37 employees. 12 are part-time and the rest are full-time. 8 employees are randomly selected for a committee. Find the probability that exactly 3 part-time employees are selected.

7. A civil engineer collects 12 soil samples: 7 deep and 5 shallow. Four samples are randomly selected for further inspection. Find the probability that exactly 2 shallow samples are selected.

8. A shipment of 25 parts contains 4 defective parts. 3 parts are randomly selected from the shipment. Let X be the number of defective parts selected. Find the probability distribution of X.

9. At a college with 5,000 students, 28% are technology students. 80 students from this college are randomly selected. Find the probability that between 21 and 23 (inclusive) are technology students.

10. In a certain town of 2,000 residents, 37% work. 40 residents are randomly selected for a survey. Find the probability that between 12 and 14 (inclusive) of the selected residents work.

11. A certain website receives an average of 7 visits per hour. Find the probability that there are at most 3 visits in the next hour.

12. In a certain city, the average number of cracks per square meter of sidewalk is 1.9. Find the probability that a randomly chosen square meter of sidewalk has:

    (a) 2 or 3 cracks.

    (b) at least 3 cracks.

13. Suppose that the concentration of bacteria in the inner harbour is 3 per 100 mL of water. Find the probability that there are at most 2 bacteria in a 50 mL sample of water.

14. There are an average of 1.8 accidents per week on a certain highway. Find the probability that there will be at least 4 accidents in the next 2 weeks.

5. Continuous Random Variables

1. The p.d.f. for X is

$$f(x) = \begin{cases} \frac{1}{8}x & \text{if } 0 < x \leq 2 \\ \frac{1}{4} & \text{if } 2 < x \leq 5 \\ 0 & \text{otherwise} \end{cases}$$

Find:

(a) $P(X = 2.2)$

(b) $P(1 \leq X \leq 3)$

(c) $P(1 < X < 3)$

(d) $P(X > 1.2)$

(e) $P(X < 0.6)$

2. Find the value of k that makes $f(x)$ a valid p.d.f.:

$$f(x) = \begin{cases} kx^7 & 1 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

3. Let X be the number of hours of TV that a certain student watches per week, with

$$f(x) = \begin{cases} \frac{1}{(\ln 12)(x+1)} & 0 \leq x \leq 11 \\ 0 & \text{otherwise} \end{cases}$$

Find the probability that the student watches:

(a) exactly 3 hours of TV.

(b) between 2 and 4 (inclusive) hours of TV.

4. A student's study time for a test (in hours) is a uniform continuous random variable with non-zero values between 0 and 8 hours. Find:

(a) the p.d.f. for the study time.

(b) the probability that the student studies less than 3 hours.

5. The lifetime of a certain machine part (in years) has p.d.f.

$$f(x) = \begin{cases} 3e^{-3x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Find the probability that the part lasts more than 0.2 years.

6. Find the mean and standard deviation of X with p.d.f.

$$f(x) = \begin{cases} x & 0 < x \leq 1 \\ \frac{1}{2} & 1 < x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

6. The Normal Distribution

1. The volume in bottles of gingerale is normally distributed with a mean of 2.01 L and a SD of 0.13 L. Find the probability that a bottle has a volume

   (a) between 1.77 and 2.29 L.

   (b) between 1.59 and 1.73 L.

(c) less than 1.81 L.

(d) more than 1.91 L.

2. The mass of a certain brand of chocolate bar is normally distributed with a mean of 85 g and a SD of 1.5 g. Find the mass that is less than the top 32% of chocolate bar masses.

3. The length of a certain type of drill bit is normally distributed with a mean of 4.2 cm and a SD of 1.1 cm. Find the length that is longer than the shortest 15% of drill bit lengths.

4. The time it takes to inspect a ball bearing is normally distributed with a mean of 6.8 s. Find the SD of the inspection times if 26.62% of inspection times are between 6.2 and 7.4 s.

7. Sampling Plans and The Central Limit Theorem

1. A large class has a test average of 72 with a SD of 8. Take a random sample of n tests. Find the probability that the n tests average to more than 75 if:

   (a) $n = 30$.

   (b) $n = 80$.

2. Checked baggage has a mean mass of 21 kg with a SD of 4 kg. 40 bags are randomly selected. Find the probability that the average mass is:

   (a) between 20 and 23 kg.

   (b) less than 20 kg or more than 22 kg.

3. Checked baggage has a mean mass of 21 kg with a SD of 4 kg. Find the probability that the total mass of 50 randomly selected bags is greater than 1130 kg.

4. A certain type of ball bearings have an average radius of 9.9 mm with a SD of 1.4 mm. Take a random sample of 60 ball bearings. Find c such that $P(\overline{x} \geq c) = 0.97$.

8. Inferences about the Population Mean

1. Volumes in cans of Coke hava a SD of 2.5 mL. A random sample of 60 cans had an average volume of 355.3 mL. Find a 95% confidence interval for the average volume among all cans of Coke.

2. If we use the same σ, n, and $\overline{x}$ from the previous question, would a 99% confidence interval be wider or narrower?

3. 50 randomly selected Camosun students were polled on their number of Google searches per week. The average was 23.4 with a SD of 3.7. Find a 90% CI for the average number of searches among all Camosun students.

4. The thickness of sheet metal has a standard deviation of 2.7 mm. We want to estimate μ with a 95% margin of error of less than 0.02 mm. What is the minimum sample size n required?

5. A sample of lightbulb lifetimes has a SD of 8.9 months. We want to estimate μ with a 90% margin of error of less than 0.1 months. What is the minimum sample size n required?

6. 30 randomly selected water samples have a mean pollution concentration of 48.1 ppm with a standard deviation of 6.2 ppm. Find a 99% UCB for the mean pollution concentration in the body of water.

7. In a large class, test marks have a SD of 10.3. A random sample of 40 tests has an average mark of 69.1. Find a 98% LCB for the class average.

8. The radius of a certain type of ball bearing is normally distributed. A random sample of 10 ball bearings has a mean radius of 4.9 cm with a SD of 0.9 cm. Find a 95% confidence interval for the mean radius of all ball bearings of this type.

9. The mass of Sirene chocolate bar is normally distributed. A random sample of 18 bars had a mean mass of 84.7 g with a SD of 2.6 g. Find a 99% UCB for the mean mass among all Sirene bars.

10. The breaking strength of a brand of rope is normally distributed. A random sample of 8 ropes had a mean of 62.1 lbs with a SD of 2.5 lbs. Find a 90% LCB for the mean breaking strength of this brand of rope.

9. Linear Regression

1. The following bivariate data set has $\hat{y} = -0.13x + 5.61$ and a coefficient of determination of 0.9522:

x = age of Corolla (years)	y = resale value ($1,000)
2	5.4
3	5.1
5	4.9
7	4.8
10	4.2

   (a) Is the linear association positive or negative?

   (b) Find the correlation coefficient.

   (c) What % of the variation in y is accounted for by the best-fit line?

   (d) What resale value is predicted for a 4-year-old Corolla?

   (e) Why should we not predict the resale value for a 1-year-old Corolla?

   (f) What age corresponds to a resale value of $4,500?

2. Find $\hat{y}$ and r^2 for the following data set:

x	y
2	5
8	4
9	2