STAT 254

Lecture Questions

Patricia R. Wrean

Department of Mathematics & Statistics
Camosun College
wrean@camosun.bc.ca

Acknowledgements

My thanks to Matt Heywood for typesetting these questions in LaTeX.

Contents

Chapter 1

Variables and Data

1.1 Intro to Statistics

There are no examples for this section.

1.2 Variables and Data

There are no examples for this section.

1.3 Graphs

There are no examples for this section.

1.4 Numerical Measures

1. Here are some inaccurate starting salaries for bridge students:

$$\begin{aligned} \$70,000 \\ \$66,000 \\ \$100,000 \\ \$55,000 \\ \$2,000,000 \end{aligned}$$

   (a) Find the mean.

   (b) Find the median.

   (c) Find the range.

1.5 Tchebysheff and the Empirical Rule

Refer to the handout for this section.

Chapter 2

Probability

2.1 Intro to Probability

1. What is the sample space for flipping a coin three times?

2. Two fair 4-sided die are rolled.
   Find the probability of rolling the following sums:

Sum	P(Sum)
2	
3	
4	
5	
6	
7	
8	

2.2 Calculating Probabilities

1. Consider 5-digit, case sensitive, alphanumeric passwords.

   (a) How many are there in total?

   (b) How many are there that contain at least one number and one letter?

2. Consider 4-digit PINs.

(a) How many are there that start with a 9?

(b) How many are there that end in a 4?

(c) How many are there that start with a 9 or end in a 4?

(d) How many are there that start with a 9 or a 4?

2.3 Combinations and Permutations

1. How many 4 digits PINs are there if repetition of digits is not allowed?

2. Pat assigns 10 sample problems to do for practice. Some overworked bridge students decide they only have time to do 3 of them. How many different groups of problems could they potentially complete?

3. The BC Lotto 6/49 draws at random 6 numbers from 49 choices and order of selection does not matter. (we are ignoring bonus/extra)

   (a) When you buy your ticket how many ways can you choose all 6 winning numbers?

   (b) When you buy your ticket how many ways can you choose 5 of the 6 winning numbers?

   (c) When you buy your ticket how many ways can you choose 4 of the 6 winning numbers?

   (d) When you buy your ticket what are the odds of choosing a winning ticket? (choosing all 6 winning numbers)

2.4 Conditional Probability and Independence

1. Suppose that there are only two coffee houses in Cook street Village, Starbucks and Moka House. Assume that Gilles goes to Starbucks 60% of the time. If Gilles goes to Starbucks, he will order a pumpkin spice latte 20% of the time. He thinks the pumpkin spice lattes are better at Moka House, so he will order them 25% of the time there.

   (a) Draw a tree diagram to represent all outcomes, including probabilities.

(b) What's the probability that Gilles orders a pumpkin spice latte?

(c) Are "ordering a pumpkin spice latte" and "going to Starbucks" independent?

2. Consider the following data on whether students at Camosun Interurban are taking a math class.

	Taking Math	Not Taking Math
Technology	50	50
Business	80	20

(a) Calculate the probability that a randomly selected student is a Business student taking a math course.

(b) Calculate the probability that a randomly selected Business student is taking a math course.

(c) Calculate the probability that if a randomly selected student is taking math, that they are in Business.

(d) Calculate the probability that a randomly selected student is not taking math.

(e) Calculate the probability that a randomly selected student is taking math or in Technology or both.

(f) Calculate (e) again using another method.

(g) Calculate (e) again using yet another method!

(h) Are the events "taking math" and "taking Business" independent?

3. consider the following with events A & B.

(a) If $P(A) = 0.3, P(B) = 0.5$, and $P(AB) = 0.12$, are A and B independent?

(b) If $P(A) = 0.1, P(B) = 0.5$, and $P(AorB) = 0.6$, are A and B mutually exclusive?

4. Are mutually exclusive exclusive events A and B independent? (Assume that P(A) and P(B) are both non-zero.)

2.5 Bayes' Theorem

1. The test for a rare disease has 99% reliability. Only one percent of the population has this rare disease.

   If the entire population is tested, then some who are healthy will have the test be positive (false positive) and some who have the disease will test negative (false negative).

   If an individual tests positive, what is the probability that they actually have the disease?

   Hint: If you prefer, consider the population to be 10,000 individuals and determine the number of individuals in each group.

Chapter 3

Discrete Random Variables

3.1 Discrete Random Variables

1. What is the expectation value for winning the jackpot in Lotto 6/49 when the pot is $2.2 million? Assume only one winning ticket and no other prize.

2. Five Civil Bridge students apply for two summer jobs at a local engineering firm. Suppose all five students are equally qualified and therefore equally likely to be chosen. If three of these students are from Alberta and two are from BC, let x be the number of Albertan Students hired for the job.

   (a) Calculate $p(x)$.

   (b) Calculate $E(x)$ and explain it's meaning.

   (c) Calculate σ^2 and σ.

3.2 Binomial Distribution

1. On Star Trek Voyager, the odds of crashing the shuttle on any away mission appear to be 75%.

   (a) If these crashes are independent, what are the odds of having exactly four crashes on five shuttle missions?

(b) What are the odds of having at least four crashes?

3.3 Poisson Distribution

1. For a particular cement mix, the average number of cracks per concrete specimen is 2.5. Assume that this number of cracks obeys a Poisson distribution.
   (a) Find the mean and standard deviation.
   (b) What's the probability of having at least one crack in a randomly chosen specimen?
2. In nuclear physics, the number of neutrons detected in a particular detector over a certain time period is a Poisson process. What average number of events should you measure so that your uncertainty (standard deviation) is 1% of the mean?

3.4 Hypergeometric Distribution

1. A case of wine (12 bottles in total) has 5 bottles that contain spoiled wine. Three bottles are randomly sampled.
   (a) What is the probability distribution for x, the number of spoiled bottles sampled?
   (b) What is the mean value of x?
   (c) What is the standard deviation of x?
   (d) A case of wine will be rejected if, when 3 bottles are randomly sampled, one or more bottles is found to be spoiled. What is the probability that a case with 5 spoiled bottles will be accepted.

Chapter 4

Continuous Random Variables

4.1 Continuous Probability Distributions

1. Let x denote the amount of time for a book which is checked out at the library if it is on two hour reserve.
   Suppose the density function is:

$$f(x) = \begin{cases} 0.5x & \text{For } 0 \leq x \leq 2 \\ 0 & \text{Otherwise} \end{cases}$$

   (a) Calculate $P(x \leq 1)$.

   (b) Calculate the mean value of x.

   (c) Verify that the area under the curve equals one.

4.2 Continuous Uniform Distribution

1. Suppose the research department of a steel manufacturer believes that one of the company's rolling machines is producing sheets of steel of varying thickness. The thickness is a uniform random variable with values between 150 and 200 mm. Any sheets less than 160 mm thick must be scrapped because they are unacceptable to buyers.

(a) Calculate the fraction of steel sheets produced by this machine that have to be scrapped.

(b) Calculate the mean and standard deviation of the thickness of sheets produced.

(c) What is the probability that a randomly selected sheet will lie within 1 standard deviation of the mean?

(d) What is the probability that a randomly selected sheet will lie within 2 standard deviations of the mean?

4.3 Exponential Distribution

1. The time between successive arrivals of trucks at a warehouse has an exponential distribution. On average, three trucks arrive per hour at the warehouse.

   (a) What is the probability that the time between arrivals of successive trucks will be less than 5 minutes?

   (b) What is the probability that the time between arrivals of successive trucks will be at least 45 minutes?

2. Suppose that a system containing a certain type of component has a "time to failure" of the component given by T. The random variable T is modeled nicely by the exponential distribution with a mean time to failure of 5 years.

   (a) What is the probability that a given component is still functioning after 8 years?

   (b) If five of these components are installed in different systems, what is the probability that at least 2 are functioning at the end of 8 years?

4.4 Normal Distribution

1. For a car travelling at 35 mph, the distance required to a stop is normally distributed with mean of 8 ft.

(a) Suppose you are travelling at that speed and a really big truck suddenly moves into your path 65 ft ahead. What's the probability that you'll smack into said truck?

(b) Suppose instead that the truck was a distance x away. Calculate x so that only 35% of the time you'll be able to brake and avoid the collision.

2. What is the probability that a normally distributed variable will have a value within

   (a) 1 standard deviation of the mean.

   (b) 2 standard deviations of the mean.

3. A grain loader can be set to discharge grain in amounts that are normally distributed with a standard deviation of 25.7 bushels. If a company wishes to use the loader to fill the containers that hold 2000 bushels of grain and wants to overfill only one container in 100, at what mean value should the company set the loader?

Chapter 5

The Central Limit Theorem

5.1 Sampling Plans and Experimental Design

There are no examples for this section.

5.2 Central Limit Theorem

1. Suppose you roll a fair 4-sided die. What's the probability distribution?

2. Suppose that at all campuses of a large university, there are 12 sections of intro calculus running concurrently. Consider the probability of an individual student getting greater than 90% on a test. Is this probability for an individual the same or different than the probability that one of the sections will have an average greater than 90%?

3. A handful of six-sided dice are loaded - the probability of rolling a six is significantly higher (and therefore the probability of rolling a one is significantly lower) than the other four numbers.

   (a) Describe the probability distribution of the number rolled on a single die (you can sketch it if you prefer).

   (b) Describe the probability distribution for the sum of fifty such dice.

4. The weight of luggage checked by airline passengers is a random variable with a mean of 50 lbs and a standard deviation of 30 lbs. The total baggage limit for 100 randomly selected passengers is 5750 lbs. What is the probability that the baggage limit will be exceeded?

5. The dean of admissions in a large university has determined that the scores of the first-year class on a math test are normally distributed with a mean of 82 and a standard deviation of 8.

   (a) Whats is the probability that any one student drawn at random from the class has a score of at least 80?

   (b) What is the probability that the mean score of a random sample of 64 students is at least 80?

5.3 The Sample Proportion

There are no examples for this section.

5.4 Statistical Process Control

1. A producer of brass rivets randomly samples 500 rivets every hour and calculates the proportion of defectives. The mean sample proportion calculated from 10,000 samples was 0.025.

   (a) Calculate the UCL and LCL for a control chart.

   (b) What does it mean if the next ten samples are above the UCL line?

Chapter 6

Confidence Intervals

6.1 Point Estimates and Confidence Intervals

1. A sample of 75 plots randomly chosen in BC's forests produced mean diameters for a certain species of tree to be 85 cm with a standard deviation of 8 cm.

   (a) Estimate the average diameter for that species of tree for forests in BC. Include a 95% margin or error.

   (b) Write your answer for part (a) in terms of a confidence interval.

2. What is $Z_{\alpha/2}$ for a confidence level of 92%?

3. A random sample of 500 car registrations is selected from a PEI database and 68 of these are classified as SUVs.

   (a) Use a 95% confidence interval to estimate the proportions of SUVs in PEI.

   (b) How can you estimate the proportion with greater precision?

6.2 One-sided Confidence Bounds

1. Forty samples of pollutants in Victoria's Inner Harbour yielded a mean concentration of the nitrate ion to be 25 ppm, with a standard devia-

tion of 5 ppm. Calculate a 98% upper confidence bound for the mean concentration of this pollutant in the Inner Harbour.

6.3 Estimating the Difference Between Two Populations

1. Police records for two different districts record the mean number of emergency calls per shift. Find a 90% confidence interval for the difference in mean numbers of emergency calls per shift and interpret your result.

	Region 1	Region 2
Sample Size	100	100
Mean	2.4	3.1
Variance	1.44	2.64

2. Five hundred students were randomly chosen and split into two equal groups, with laptops provided to the first group. 72% of the first group got a B or higher, while 78% of the non-laptop group got a B or higher.
   (a) Construct a point estimate for the difference in success rates between the two groups.
   (b) Construct a 95% confidence interval for the difference in success rates.
   (c) Interpret your result.

6.4 Choosing the Sample Size

1. Suppose you wish to estimate the mean time between failures for a certain brand of disk drive. From previous experience, you know that σ is in the neighbourhood of 200 hours. If you want your estimate of the mean to be precise (with 99% confidence) to within ± 50 hours of the true value, how many disk drives will you have to test?

2. Camosun wants to know the difference (if any) between rates of smoking for students on the two different campuses. Assume that equal numbers of students will be surveyed on each campus. The sampling

error in the difference between the two proportions is required to be no longer than ±3%.

(a) How large a sample size is required if the rate of smoking for young adults in Canada is known to be 28%?

(b) How large a sample size is required if the rate of smoking is initially unknown?

Chapter 7

Large-Sample Hypothesis Tests

7.1 Tests of Hypothesis for Large Samples

There are no examples for this section.

7.2 Large Sample Hypothesis Tests - Population Means

Refer to the handout for this section.

7.3 Large Sample Hypothesis Tests - Binomial Proportions

Refer to the handout for this section.

7.4 Type I and Type II Errors

1. Situation: Movement in grass.
   Null hypothesis: It's just the wind.

Alternate hypothesis: It's a jaguar!

(a) State in words what a Type I error would be.

(b) State in words what a Type II error would be.

Chapter 8

Small-Sample Hypothesis Tests

8.1 Student's t-distribution and the Population Mean

1. The duration for a random sample of six earthquakes in Southern California has been measured to be:

   1.1, 0.9, 1.5, 0.7, 1.4, and 1.3 minutes.

   (a) Given that the mean is 1.15 minutes with standard deviation 0.308 minutes, and assuming that the duration for these type of earthquakes is normally distributed, give a 98% confidence interval for the duration of Southern California earthquakes.

   (b) An expert claims that the duration of Southern California earthquakes is 0.8 minutes. Your sample mean is higher, but is it significantly higher? Use a test of hypothesis.

8.2 Small-Sample Differences in the Population Mean

1. On the most recent STAT 254 test, three HP users finished the quiz of average in 39 minutes with a standard deviation of 5 minutes. The twenty-one Ti-89 users finished on average in 43 minutes with a standard deviation of 4 minutes. Assume that the time taken to finish

the quiz for each group is normally distributed. At the 95% confidence level, perform a test of hypothesis to determine whether the two groups finished in the same amount of time.

8.3 Paired Difference

1. Pat's MATH 173 sections had the following results:

	Mech	Civil
Quiz 1	79%	84%
Quiz 2	67%	75%
Quiz 3	70%	73%
Quiz 4	74%	78%
Average	72.5%	77.5%
Std Dev	5.2%	4.8%
n	4	4

   Is this difference consistent with zero or are the Civil students doing consistently better? Use a p-test.

2. For an introductory physics lab, six students measured their height while standing up and their length while lying down, with the following results.

Lying down	Standing up
155 cm	153 cm
168 cm	165 cm
141 cm	141 cm
148 cm	147 cm
162 cm	161 cm
175 cm	173 cm

   Construct a 95% confidence interval for the difference in heights between lying down and standing up.

8.4 Inferences Concerning a Population Variance

1. How precise is your measuring instrument? Manufacturer claims that the digital scale reads to within ± 0.1 g. How can you verify this?

2. A research study involving 9 male triathletes reports that their maximum heart rate while swimming had a sample mean and standard deviation of 188.0 and 7.2 beats/minute, respectively. Calculate a 90% confidence interval for the standard deviation of the true mean heart rate while swimming.

3. In any canning factory, the manufacturer loses money if the cans are overfilled and risks fines if they are under filled. Therefore, a quality control inspector at the Gunk factory is interested in testing whether the amount of gunk dispensed into their 32-ounce cans has a variance of greater than 0.2 oz? Due to practical constraints, a random sample of 10 cans will have to do. Their weights have:

$$\bar{x} = 31.55 \text{ ounces}$$
$$s = 0.48 \text{ ounces}$$

Do the machines need adjusting? Perform a hypothesis test with 95% confidence.

8.5 Goodness of Fit

1. An urban economist wishes to determine whether the distribution of residents in the USA is the same today as it was in the year 2000. That year, 19.0% of the US population lived in the Northeast, 22.9% lived in the Midwest, 35.6% lived in the South, and 22.5% lived in the South, and 22.5% lived in the West (based on data from the Census Bureau). The economist randomly selects 1500 households in the US and obtains the frequancy data shown below. Does the evidence suggest that the distribution of US residents is different today from the 2000 distribution. Use $\alpha = 0.05$.

Region	Frequency
Northeast	274
Midwest	303
South	564
West	359

2. A researcher designs an experiment in which a rat is attracted to the end of a ramp that then divides leading to doors of three different colours. The researcher sends the rat down the ramp ninety times and

observes the rat's choice as tabulated below. Does the rat have (or acquire) a preference for one of the three doors? Use a p-test to draw your conclusion.

Colour	Frequency
Red	24
Blue	39
Green	27

Chapter 9

Linear Regression

9.1 Coefficients of Correlation and Determination

There are no examples for this section.

9.2 The Residuals Plot

There are no examples for this section.

9.3 Calculating the Best-Fit Line

There are no examples for this section.

Chapter 10

Review

1. A random sample of 400 Tech students were recently surveyed to find out how many of them had skipped at least one class. The results of the survey were that 326 of them has skipped at least one class. Construct a 90% confidence interval for the percentage of Tech students who have skipped at least one class.

2. In BC, it usually costs \$25,400 to repair a bridge for storm damage. This year, a random sample of 55 bridges showed an average cost per repaired bridge was \$25,950 with standard deviation of \$2,750. Using the p-value approach, determine the significance of this increase.

3. Suppose that the time it takes to fill a gas tank is an exponential distribution with a mean time of 42 seconds.

   (a) What's the probability that it will take you longer than a minute to fill your gas tank?

   (b) What's the probability that it will take you less than a minute to fill your gas tank?

   (c) What's the probability that it will take you exactly one minute to fill your gas tank?

4. The probability that the starting goalie for the Canucks is injured is 9%. The probability that the backup goalie is injured is 12%. The probability that both are injured is 2%.

   (a) What is the probability that neither of them is injured?

(b) What is the probability that at least one of them is injured?

5. The probability of a randomly selected driver wearing a seat belt is 77%. If the driver is then in an accident, the probability of being uninjured is 58% for an unseat belted one.

   (a) If the driver is injured in an accident, what is the probability that they weren't wearing a seat belt?

   (b) Are "not wearing a seat belt" and "being injured" independent?

6. At the end of the school year, the Bridge students held a raffle. Each of the fifty students bought a ticket and then three winning tickets were chosen at random. First prize was a 12 pack of beer, second prize was a new car, and third prize was a high-five from the dean.

   (a) What's the probability that Isaac wins first prize, Joel wins seconds prize, and Adam winds third prize?

   (b) What's the probability that Isaac, Joel, and Adam between them win all three prizes?

7. Suppose that some phenomenon has the following probability distribution:

$$f(x) = \begin{cases} \frac{k}{x} & \text{for } 1 \le x \le 10 \\ 0 & \text{otherwise} \end{cases}$$

   (a) Calculate k such that f(x) is indeed a probability distribution function. Give both an exact and an approximate answer.

   (b) Calculate the mean value of x. Give both an exact and an approximate answer.

8. Standards set by the government agencies indicate that Canadians should not exceed an average daily sodium intake of 3300 mg. To find out whether Canadians are exceeding this limit, a sample of 100 Canadians is selected and the mean and standard deviation of daily sodium intake are found to be 3400 mg and 1100 mg, respectively. Perform a p-test to determine your conclusion.

9. Previous enrollment records at a large university indicate that of the total number of persons who apply for admission, 60% are admitted unconditionally, 5% are admitted on a trial basis, and the remainder are refused admission. Of 500 applicants to apply for the coming year, 329 have been admitted unconditionally, 43 have been admitted on a

trial basis, and the remainder have been refused admission. Do these data indicate a departure from previous admission rates? Test using $\alpha = 0.05$.