MATH 321: Mathematical Statistics
Assessments

Colton Gearhart

April 23, 2024

Contents

1 In-Class Assignments

Name:

MATH 321: In-Class 0

1. Let X have the following pmf:

x	0	1	2	3
$f(x)$	0.3	0.15	0.24	0.31

   (a) Find $P(X < 2)$.

   (b) Find $E(X)$.

   (c) Find $V(X)$.

   (d) Find $E[(X + 3)(X - 1)]$

2. Let X have the following pdf: $f(x) = cx^2$, for $0 < x < 1$.

   (a) Find c so that $f(x)$ is a valid pdf.

   (b) Find $P(0.25 < X < 0.55)$.

   (c) Find $E(X^2 + \sqrt{X})$.

Name:

MATH 321: In-Class 14

1. Let $f(x,y) = \dfrac{xy+y}{c}$ for $x = 1,2,3$ and $y = 1,2$ be the joint pmf for the random vector (X,Y).

   (a) Find the value c that makes this a valid pmf.

   (b) Construct the joint pmf table for (X,Y) using your answer from part (a); add the marginal pmfs to the table as well.

   (c) Find the following probabilities: $P(X=Y)$, $P(X-Y=1)$ and $P(X^2 \leq 4)$.

2. Let (X,Y) be a bivariate continuous random vector with joint pdf
   $f(x,y) = \dfrac{1}{4} + \dfrac{x}{2} + \dfrac{y}{2} + xy$ for $0 \leq x \leq 1$ and $0 \leq y \leq 1$.

   (a) Find $P(0 \leq X \leq 0.5, 0.5 \leq Y \leq 1)$.

(b) Find the marginal pdf of X, $f_X(x)$ (Reminder: $f(x,y) = \frac{1}{4} + \frac{x}{2} + \frac{y}{2} + xy; \quad 0 \le x \le 1, 0 \le y \le 1$).

3. Let (X, Y) be a bivariate continuous random vector with joint pdf $f(x,y) = 2x^2 + 3y \quad$ for $0 \le y \le x \le 1$.

   Find $P(X + Y > 1)$.

Name:

MATH 321: In-Class 15

1. Let $f(x, y) = \dfrac{xy + y}{27}$ for $x = 1, 2, 3$ and $y = 1, 2$ be the joint pmf for the random vector (X, Y).

   (a) Find the marginal pmfs $f_X(x)$ and $f_Y(y)$ (keep in functional form).

   (b) Find the conditional pmfs $f(x \mid Y = 1)$ and $f(y \mid X = 2)$.

   (c) Find $E(X \mid Y = 1)$ and $E(Y^2 \mid X = 2)$.

2. Let $f(x, y) = x + (3/2)y^2$ for $0 \leq x \leq 1$ and $0 \leq y \leq 1$.

   (a) Find the conditional pdf $f(y \mid x)$.

(b) Find $P(Y < 0.5 \mid X = 0.5)$.

(c) Find $E(Y \mid X = 0.5)$.

(d) Find $V(Y \mid X = 0.5)$.

Name:

MATH 321: In-Class 16

1. Let $f(x,y) = \dfrac{xy^2}{12}, \quad 0 \le x \le 3, 0 \le y \le 2.$

   (a) Show if X and Y are independent or dependent using the definition.

   (b) Using your results from part (a), write the integrals to find $P(X > 2, Y < 1.5)$ and $E(Y^3\sqrt{X})$ (don't actually solve, just set up).

2. Let $f(x,y) = \dfrac{x+2y}{18}$ for $x = 1,2$ and $y = 1,2$ be the joint pmf for the random vector (X,Y).

   (a) Show if X and Y are independent or dependent by inspection.

   (b) Using your calculator, find all of the following items:

   $E(X)$, $E(Y)$, $SD(X)$, $SD(Y)$ and $E(XY)$.

   (c) Using your results from part (c), calculate $\text{Cov}(X,Y)$ using the alternate formula and $\text{Corr}(X,Y)$.

Name:

MATH 321: In-Class 17

1. Find $V(W - 3X - 0.5Y + 4Z)$ in terms of the variances and covariances of W, X, Y and Z.

2. Let X_1, X_2, X_3 be mutually independent random variables where $X_1 \sim \text{Bin}\,(n = 3, p = 0.2)$, $X_2 \sim \text{Bin}\,(n = 4, p = 0.2)$, and $X_3 \sim \text{Bin}\,(n = 5, p = 0.2)$.

   (a) Find $P(X_1 = 2, X_2 = 1, X_3 = 3)$.

   (b) Find the distribution of $S = X_1 + X_2 + X_3$ using the mgf technique.

   (c) Find $P(S < 4)$.

3. Suppose $f(x,y,z) = \frac{1}{4}x, \quad 0 \leq x \leq 2, 0 \leq y \leq 1, 0 \leq z \leq 2$.

   (a) Find the marginal distribution $f(y)$.

   (b) Find the conditional distribution $f(x,z \mid y)$.

4. Let X_1, X_2, X_3 be mutually independent random variables where $X_1 \sim \text{Normal}\,(\mu = 150, \sigma^2 = 225)$, $X_2 \sim \text{Normal}\,(\mu = 100, \sigma^2 = 64)$, and $X_3 \sim \text{Normal}\,(\mu = 200, \sigma^2 = 81)$.

   Find $P(X_1 + 3X_2 < 2X_3)$.

   HINT: Rearrange the probability statement to see the distribution we need to find first.

Name:

MATH 321: In-Class 1

1. Suppose $X_i \overset{iid}{\sim} \text{Poisson}\,(\lambda = 5)$, for $i = 1, \ldots, 30$.

   (a) Find the joint pmf $f(x_1, \ldots, x_{30})$.

   (b) Let $\bar{X} = \dfrac{1}{30}\displaystyle\sum_{i=1}^{30} X_i$. Find $E(\bar{X})$ and $V(\bar{X})$.

   (c) Find the distribution of $\bar{X}$.

2. Let $Z_1, \ldots, Z_9$ be independent. Also assume $Z_i \sim \text{Normal}\,(0, 1)$ for $i = 1, \ldots, 9$.

   (a) If $Y_1 = Z_1^2 + \ldots + Z_9^2$ find $P(Y_1 < 15)$.

   (b) If $Y_2 = Z_3^2 + \ldots + Z_9^2$ find $P(2 < Y_2 < 13)$.

   (c) If $Y_3 = Z_3^2 + \ldots + Z_5^2$ find $P(Y_3 > 12)$.

3. (a) Suppose $X \sim \text{Normal}\,(\mu = 100, \sigma = 25)$. Find the 90^{th} percentile of X (i.e. find $x_{0.9}$ such that $P(X < x_{0.9}) = 0.90$).

   HINT: Use Z-table "backwards" or invNorm() on graphing calculator.

(b) Now use qnorm() in R to find the 90^{th} percentile of X. Answer should match part (a).

(c) Suppose $Y \sim \chi^2(15)$. Find the 30^{th} percentile of Y.

(d) Suppose $T \sim \mathrm{t}(10)$. Find the *upper* 25^{th} percentile of T.

4. Let $X_1, \ldots, X_{20}$ be a random sample from Normal($\mu = 30, \sigma^2 = 100$) and let $\bar{X}$ be the sample mean and S be the sample standard deviation.

   (a) Find $P(-2 < \dfrac{\bar{X} - 30}{S/\sqrt{20}} < 2)$.

   (b) Find $P(-2 < \dfrac{\bar{X} - 30}{10/\sqrt{20}} < 2)$.

Name:

MATH 321: In-Class 2

1. Suppose $X_1, X_2 \overset{iid}{\sim} f(x) = 3x^2 \quad 0 < x < 1$.
   (a) Find the survival function of $X_{(1)} = min(X_1, X_2)$.

   (b) Find the cdf of $X_{(1)} = min(X_1, X_2)$.

   (c) Find the cdf of $X_{(2)} = max(X_1, X_2)$.

2. Let X_1, X_2, X_3 be a random sample from Exponential ($\lambda = 2$).
   (a) Find the cdf of $X_{(3)} = max(X_1, X_2, X_3)$.

   (b) Find the pdf of $X_{(3)} = max(X_1, X_2, X_3)$ by taking the derivative of part (a).

   (c) Find the pdf of $X_{(3)} = max(X_1, X_2, X_3)$ using the pdf theorem (answer should match (b)).

(d) Find the cdf of $X_{(1)} = min(X_1, X_2, X_3)$. *HINT: Can start with the cdf written as a probability statement and then think about it with logic to continue (and use a complement).*

(e) Find $P(X_{(1)} < 1.5)$ and $P(X_{(3)} < 1.5)$.

(f) Find the cdf and the pdf of the sample median $X_{(2)}$ using the theorems.

3. Use R to create a qqplot using the following steps:

   (a) Generate (and save) a random sample of size $n = 100$ from $X \sim t\,(2)$.

   (b) Run the following two lines:

   qqnorm($<$ your sample $>$)

   qqline($<$ your sample $>$)

   (c) Roughly sketch the result, which is visually testing whether a t distribution matches the characteristics of a normal distribution.

   Is there a pattern? Draw / trace it. What does this pattern tell you about the t-distribution?

Name:

MATH 321: In-Class 4

1. Let $X_1, \ldots, X_m \overset{iid}{\sim} \text{Binomial}\,(n = 5, p)$.
   (a) Find the method of moments estimator for p.

   (b) Show that $\hat{p}_{MME}$ is an unbiased estimator.

2. Let $X_1, \ldots, X_n$ be a random sample from $f(x \mid \theta) = (\theta + 1)x^{\theta}, \quad 0 < x < 1,\, \theta > -1$.
   Find the MME of θ.

3. Let $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}\,(p)$. We are going to find the maximum likelihood estimator for p.
   (a) Find the likelihood function and log-likelihood function for p.

   (b) Optimize the log-likelihood function and solve for $\hat{p}$.

(c) Perform second derivative test to confirm if $\hat{p}$ is the MLE for p.

(d) Suppose we collected a random sample of size $n = 8$ and $\mathbf{x} = \{0, 1, 1, 1, 0, 1, 0, 0\}$. Compute $\hat{p}_{MLE}$.

(e) Now find the MLE for $V(X) = p(1 - p)$.

Name:

MATH 321: In-Class 5

1. Let $X_1, \dots, X_{75}$ be a random sample of size 75 from a random variable X with pdf

$$f(x) = \frac{3}{x^4}, \quad 1 < x < \infty$$

   (a) Given $\mu = E(X) = 3/2$, compute $\sigma^2 = V(X)$.

   (b) Using answer from (a), approximate $P(\bar{X} > 13/8)$.

2. A soft-drink vending machine is set so that the amount of drink dispensed is a random variable with a mean of 200 millimeters and a standard deviation of 15 millimeters.

   Find the approximate probability that the average (mean) amount dispensed in a random sample of size 36 is at least 204 millimeters.

3. In an analysis of healthcare data, ages have been rounded to the nearest multiple of five years. The difference between the true age and the rounded age is assumed to be uniformly distributed on the interval from -2.5 years to 2.5 years. The healthcare data are based on a random sample of 48 people.

   Find the approximate probability that the mean of the rounded ages is within 0.25 years of the mean of the true ages.

4. A company has a pet policy that divides its policyholders into two classes as follows (assume all policies are independent):

Class	Prob of Claim	Benefit	Number in class
A	0.01	500	1000
B	0.03	200	500

   (a) Find the mean and variance for the random variables of the amount paid for a claim *on a single policy* by the insurance company for each class.

   (b) Find the distributions for the total amount paid in claims for *all of the policies* for each class.

   (c) The insurance company wants to collect a premium that equals the 90^{th} percentile of the distribution of the total claims (for both classes combined). Find what this premium be.

5. Suppose you are in a competition for shooting 50 free throws, and you have an 80% chance of making each free throw (assuming independence of shots). Approximate the probability of making less than 40 shots using the continuity correction.

Name:

MATH 321: In-Class 6

1. A NatGeo Poll interviewed 1200 hiking enthusiasts and asked "Are you more afraid of spiders or snakes??" Out of the 1200 people, 768 responded "Ewww, snakes..."

   (a) Check the conditions for a confidence interval for the true proportion p of hikers who are more afraid of snakes.

   (b) Calculate a 95% confidence interval for p.

   (c) Calculate a 90% lower-bound confidence interval for p and a 90% upper-bound confidence interval for p.

2. The poll from (1) also asked 1100 climbers the same question. 662 of the 1100 climbers, responded "Ewww, snakes...."

   Calculate a 92% confidence interval for the difference in proportion of climbers vs hikers who are more afraid of snakes than spiders. State the conclusion as well.

3. 15 out of 23 people from a random sample said their National Championship team is still remaining in their NCAA March Madness Bracket.

   Calculate a 85% confidence interval for the true proportion p of brackets that still have their National Championship team remaining.

4. For a comparison of the rates of defectives produced by two assembly lines, independent random samples of 100 items were selected from each line. Line A yielded 18 defectives in the sample, and line B yielded 12 defectives.

   Find a 98% confidence interval for the true difference in proportions of defectives for the two lines AND state a conclusion if one line produces a higher proportion of defectives than the other.

5. From a random sample 500 people, 64% said they prefer to vacation at the beach compared to the mountains.

   (a) Calculate a 93% confidence interval for the true proportion p of people who prefer beaches over mountains for vacation.

   (b) If the sample size was increased to 600 people and all else remains constant, what will happen to the new confidence interval? Calculate this new interval.

   (c) If the confidence level for the interval from (a) changed to 90% confident, what will happen to the new confidence interval? Calculate this new interval.

Name:

MATH 321: In-Class 7-6

1. Two methods for teaching reading were applied to two randomly selected groups of elementary schoolchildren and then compared on the basis of a reading comprehension test given at the end of the learning period. The sample means and variances computed from the test scores are shown below. Assume scores for both methods are normally distributed with unknown, common variance σ^2.

   Method I : $n_1 = 11, \bar{x}_1 = 64, s_1^2 = 52$ Method II : $n_2 = 14, \bar{x}_2 = 69, s_2^2 = 71$

   (a) Do the data present sufficient evidence to indicate a difference in the mean scores for the populations associated with the two teaching methods? Use $\alpha = 0.10$ and make the conclusion using the "traditional" method (RR).

   (b) Confirm the result in (a) by constructing a two-sided 90% confidence interval for $\mu_1 - \mu_2$ and seeing if $D_0 = 0$ is in the interval.

2. According to the Washington Post, nearly 45% of all Americans are born with brown eyes, although their eyes don't necessarily stay brown. A random sample of 80 adults found 30 with brown eyes.

   (a) Is there sufficient evidence at the $\alpha = 0.05$ level to indicate that the proportion of brown-eyed adults is less from the proportion of Americans who are born with brown eyes? Make the conclusion using the "p-value" method.

   (b) If $\alpha = 0.01$ instead, would the same conclusion be made? How about $\alpha = 0.10$? Explain.

(c) Confirm your conclusion from (b) when $\alpha = 0.10$ by constructing the appropriate one-sided 90% confidence interval for p and seeing if $p_0 = 0.45$ is in the interval.

3. To test whether a golf ball of brand A is better or worse than Brand B, 9 golfers hit a ball of each brand off the tee and measured the distance. The results for the difference in distances $(A - B)$ in yards are shown below. Assume that the paired differences in distance are approximately normally distributed.

   With $\alpha = 0.05$, test if there is a difference in average distance between brand A and brand B. Make sure to find the RR and the p-value and state any additional insights in the conclusion.

Golfer	Difference Brand A - Brand B
1	13
2	-4
3	3
4	14
5	-1
6	17
7	11
8	13
9	17

4. Some college professors and students were interested in studying a certain characteristic in Canadian geese and would like to estimate p, the proportion of birds with this characteristic.

   (a) Determine the minimum sample size needed to estimate p within $\epsilon = 0.04$ with 90% confidence.

   (b) A previous study took a random sample of 137 Canadian geese and found that 54 had this characteristic. Taking this into account, determine the minimum sample size needed to estimate the unknown p within $\epsilon = 0.04$ with 90% confidence.

5. Let X equal the tarsus length for a male grackle. Assume that the distribution of $X \sim \mathrm{N}(\mu, \sigma^2 = 4.84)$. Find the sample size n that is needed to achieve a maximum error of the estimate of

   (a) $\epsilon = 0.04$ for a 95% CI for μ

   (b) $\epsilon = 0.08$ for a 85% CI for μ

2 Homework

Name:

MATH 321: Homework 14

Due ____________________: Turn in a hard copy, neat and stapled.

1. A fair coin is tossed. If heads is tossed then one fair 4-sided die is thrown and if tails is tossed two fair 4-sided dice are thrown. Let $X = 1$ for heads and $X = 2$ for tails and let Y be the total number of dots on the dice.

   (a) Plot the range of the joint pmf of (X, Y), then find the corresponding joint probabilities.

   (b) Find the following probabilities: $P(X = Y)$, $P(2X < Y)$, and $P(X + Y \leq 7)$.

   (c) Find the marginal pmfs of X and Y, $f_X(x)$ and $f_Y(y)$, respectively.

   (d) Find the following probabilities: $P(X = 1)$ and $P(3 \leq Y \leq 5)$.

2. A basketball team has 3 players from Ohio, 5 from Indiana and 2 from Kentucky. Two of these players are selected at random for an interview. Let X be the random variable for the number of players from Ohio chosen and let Y be the random variable for the number of players from Indiana chosen.

   (a) Construct the joint pmf table for (X, Y).

   (b) Let $g_1(X, Y) = 2X$, $g_2(X, Y) = Y^2$ and $g_3(X, Y) = XY$.

   Find the expected values of each $g_i(X, Y)$, $i = 1, 2, 3$.

3. A home insurance company separates its claims into two parts: losses due to wind damage and losses due to water damage. If X is the random variable for losses due to wind damage and Y is the random variable for losses due to water damage,

$$f(x, y) = \frac{30 - x - y}{1875} \quad \text{for } 0 \leq x \leq 5, 0 \leq y \leq 25$$

   (a) If a claim is filed after a storm, find the probability that there is more loss due to water damage than wind damage.

   (b) Find the expected value of the total loss for a claim, i.e. wind damage plus water damage.

4. Let (X, Y) be a bivariate continuous random vector with joint pdf

$$f(x, y) = 2x \quad \text{for } 0 \leq x \leq 1, 0 \leq y \leq 1$$

   Find $P(X^2 < Y < X)$.

Select answers

1. (a)
   (b) $P(X + Y \leq 7) = 0.8125$
   (c) $P(3 \leq Y \leq 5) = 0.53125$
2. (a)
   (b) $E[g_3(X, Y)] = 1/3$
3. (a) Prob ≈ 0.8333
   (b) $E(X + Y) \approx 11.3889$
4. $P(X^2 < Y < X) = 1/6$

Name:

MATH 321: Homework 15

Due ____________________: Turn in a hard copy, neat and stapled.

1. An actuary determines that the annual number of tornadoes in counties X and Y are jointly distributed as follows:

y \ x	0	1	2	3
0	.12	.06	.05	.02
1	.13	.15	.12	.03
2	.05	.15	.10	.02

   (a) Construct pmf tables for the following conditional random variables: $X \mid Y = 1$ and $Y \mid X = 2$.

   ** Round to 2 decimals.

   (b) Find the following conditional probabilities:

   $P(X \geq 1 \mid Y = 1)$, $P(X + Y \leq 2 \mid Y = 1)$ and $P(Y^2 \leq 2 \mid X = 2)$.

2. Let (X, Y) be a bivariate continuous random vector with joint pdf

$$f(x,y) = \frac{1}{4}(3x^2 + 2 - y) \quad \text{for } 0 \leq x \leq 1, 0 \leq y \leq 2$$

   Find the conditional expected value $E(X^3 \mid Y = 0.6)$.

3. The stock prices of two companies at the end of any given year are modeled with random variables X and Y that follow a distribution with joint density function

$$f(x,y) = 2x \quad \text{for } 0 < x < 1, x < y < x + 1$$

   Find the conditional variance $V(Y \mid X = x)$.

Select answers

1. (a)

   (b) $P(X \geq 1 \mid Y = 1) = 0.70$, $P(X + Y \leq 2 \mid Y = 1) = 0.65$ and $P(Y^2 \leq 2 \mid X = 2) = 0.63$

2. $E(X^3 \mid Y = 0.6) = 17/48$

3. $V(Y \mid X = x) = 1/12$

Name:

MATH 321: Homework 16

Due ____________________: Turn in a hard copy, neat and stapled.

1. An insurance company sells two types of auto insurance policies: Basic and Deluxe.

   The time until the next Basic Policy (X) claim is an exponential random variable with mean two days. The time until the next Deluxe Policy claim (Y) is an independent exponential random variable with mean three days.

   (a) Find the density function needed to solve $P(Y < X)$.

   (b) Calculate the probability from part (a).

   (c) Find the variance of the combined waiting time for the two different policy types $X + Y$.

2. Let $f(x, y) = \dfrac{4(1 - xy)}{3}$ for $0 \le x \le 1$ and $0 \le y \le 1$.

   (a) Show if X and Y are independent or dependent.

   (b) Find $\text{Cov}(X, Y)$.

3. A joint density function is given by

$$f(x, y) = kx \quad 0 < x < 1, 0 < y < 1$$

   where k is a constant.

   (a) Find $\text{Corr}(X, Y)$.

   (b) Find $f(x \mid y)$.

4. Let X and Y be the number of hours that a randomly selected person watches movies and sporting events, respectively, during a month. The following information is known about X and Y.

$$E(X) = 50 \quad V(X) = 50 \quad E(Y) = 20 \quad V(Y) = 30 \quad \text{Cov}(X, Y) = 10$$

   (a) Find $\text{Corr}(X, Y)$.

   (b) Let X' and Y' now represent the number of minutes spent watching movies and sporting events respectively. Find $\text{Cov}(X', Y')$.

   (c) Find $\text{Corr}(X', Y')$.

   (d) Find $V(X' + Y')$.

Select answers

1. (a)
   (b) $P(Y < X) = 0.4$
   (c) $V(X + Y) = 13$
2. (a)
   (b) $\text{Cov}(X, Y) = -1/81$
3. (a)
   (b)
4. (a) $\text{Corr}(X, Y) \approx 0.2582$
   (b) $\text{Cov}(X', Y') = 36,000$
   (c)
   (d) $V(X' + Y') = 360,000$

Name:

MATH 321: Homework 1

Due __________________: Turn in a hard copy, neat and stapled.

1. Entrance exam scores of students at a certain college, X, are normally distributed with $X \sim \mathrm{N}(\mu = 1300, \sigma^2 = 2500)$. Suppose an admissions officer selects a random sample of 30 students at this college and checks their entrance exam scores.

   (a) If $\bar{X}$ is the sample mean of the 30 exam scores, find the distribution of $\bar{X}$.

   (b) Find $P(1290 \leq \bar{X} \leq 1310)$.

   (c) Find $P(X \leq 1230)$.

   (d) Let Y be the number of random variables (exam scores) in the sample that have values of at most 1230. Find the probability that less than 5 of the random variables in the sample have scores of at most 1230, that is, $P(Y < 5)$.

   HINT: Think of $P(X \leq 1230)$ as a success probability, and the result of checking this is for each random variable is either a success or failure.

   (e) Let S^2 be the sample variance of the 30 exam scores. Find $P(S^2 > 2000)$.

   HINT: Think about how to get the random variable of interest to follow a distribution we know.

2. Let $X_1 \sim \chi^2(10)$ and $X_2 \sim \chi^2(13)$ and $X_1 \perp\!\!\!\perp X_2$.

   (a) Let $Y_1 = \dfrac{X_1/10}{X_2/13}$. Find the distribution of Y_1 and $P(Y_1 > 5)$.

   (b) Let $Y_2 = 1/Y_1$. Find the distribution of Y_2 and the IQR of Y_2.

3. Suppose we take independent random samples of sizes $n_1 = 6$ and $n_2 = 10$ from two normal populations with equal population variances ($\sigma_1^2 = \sigma_2^2$). Let S_1^2 and S_2^2 be the sample variances from populations 1 and 2, respectively.

   Find $P(S_1^2/S_2^2 > 2)$.

4. **R simulation**: We are going to simulate a sampling distribution for a statistic and estimate some probabilities using empirical methods.

   Guidelines: Complete each of the following steps for the following pairs of distribution and statistic: 1) Gamma distribution and median and 2) Your choice (be creative!). For example, Binomial / sample variance or χ^2 / minimum.

   (a) Plot the population distribution of interest that you will be sampling from.

   (b) Generate $i = 10,000$ random samples of size $n = 30$ from the population distribution.

   (c) Calculate the sample statistic $Y = T(X_1, \ldots, X_{30})$ for each of the random samples.

   (d) Plot a histogram of the simulated sampling distribution of Y.

   (e) Calculate $\hat{\mu}_Y$ and $\hat{\sigma}_Y$, the estimated mean and standard deviation of the sampling distribution of the sample statistic based on the simulated results, respectively.

   (f) Calculate the estimated probability the sample statistic is within two standard deviations of its mean:
   $P(\hat{\mu}_Y - 2\hat{\sigma}_Y < Y < \hat{\mu}_Y + 2\hat{\sigma}_Y)$.

   Restrictions: Do not use the normal distribution or the sample mean $\bar{X}$.

   Submission: This problem will be worth 15 of the 30 points. Please submit your completed version of the starter .qmd file on canvas and rendered .html file.

Select answers

1. (a)

   (b) $P(1290 \leq \bar{X} \leq 1310) \approx 0.7267$

   (c) $P(X \leq 1230) \approx 0.0808$

   (d) $P(Y < 5) \approx 0.9097$

   (e) $P(S^2 > 2000) \approx 0.7673$

2. (a) $P(Y_1 > 5) \approx 0.0042$

   (b) IQR of $Y_2 \approx 0.8639$

3. $P(S_1^2/S_2^2 > 2) \approx 0.1727$

4.

Name:

MATH 321: Homework 2

Due ________________: Turn in a hard copy, neat and stapled.

1. Let $X_1, \ldots, X_9$ be a random sample from Exponential ($\lambda = 5$).
   (a) Find the cdf of $X_{(7)}$.
   (b) Find the pdf of $X_{(3)}$.

2. Let $X_{(1)}, \ldots, X_{(10)}$ be the order statistics from a continuous distribution with 70th percentile $x_{0.7} = 24.3$.

   Determine $P(X_{(8)} \leq 24.3)$ and $P(X_{(3)} \leq 24.3)$.

   HINT: Think about the information $x_{0.7} = 24.3$ *tells us.*

3. A pharmaceutical researcher is testing the effect of a medication measured on a continuous scale. Effects from the medication are independent and have continuous uniform distributions on (3,10). The researcher randomly selects five patients to examine.
   (a) Find the probability that the smallest effect is between 3 and 5.
   (b) Find is the expected value of the smallest effect and of the second largest effect.

      CAN USE TECHNOLOGY to solve the expected values.

4. **Excel q–q plots**: We are going to investigate some real data and try to figure out what potential distribution a dataset came from. Data can be found in the the starter file, which is where your work should be done as well.

   Guidelines:
   (a) Create a histogram of the data. Be sure to add axis labels and a title.
   (b) Complete discussions for (1) and (2) located in the template based on the histogram.
   (c) Create two q–q plots, one for testing for the standard normal distribution and one testing another distribution we have learned. The goal for the latter is to find the best model (or at least a better model) for the data.

      Feel free to try several distributions (NOTE: will need to lookup the appropriate <dist>.inv() functions in Excel).
   (d) Once you are satisfied with a model, complete discussion (3) located in the starter file based on the q–q plots.

   Submission: This problem will be worth 15 of the 30 points. Please submit a completed version of the starter excel file.

Select answers

1. (a)

   (b)

2. $P(X_{(8)} \leq 24.3) \approx 0.3828$ and $P(X_{(3)} \leq 24.3) \approx 0.9984$

3. (a) $P(3 \leq X_{(1)} \leq 5) \approx 0.81406$

   (b) $E(X_{(1)}) = 25/6$ and $E(X_{(4)}) = 23/3$

4.

Name:

MATH 321: Homework 3

Due __________________: Turn in a hard copy, neat and stapled.

1. **EDA**: We are going to explore the 'Factors that Affect High School Completion Rates for 2013-2014' dataset. The goal is to summarize, describe and display information related to the included variables. In doing so, see if you can find any trends, patterns, interesting observations and try to relate it to the context. Try to find a narrative or a question to answer and use your EDA to investigate this.

   This is intended to be a very open-ended assignment. You may explore as much or as little (as long as you meet the minimum requirements below) as you want. In addition you may do the work in Excel, R or both.

   Files:

   - If working in Excel, start with the 'high-school-data.xlsx' file on Canvas.
   - If working in R, read the 'high-school-data.csv' file on Canvas into a .qmd file.

   Requirements:

   - Must create a minimum of two:
     - Set of descriptive statistics → mean, st dev (sample), five-number summary, range and IQR.
     - Frequency and relative frequency table.
     - Histograms (of some kind: frequency, relative frequency, density).
     - Boxplots, if outliers are present extract / make note of them.
     - Scatterplots and correlation calculation (between a pair of variables).
   - In addition:
     - Your exploration must include an analysis by region in some capacity (e.g. comparative boxplots, summary statistics by region, or anything else you can come up with).
     - These count towards the above requirements.

   Submission: There will be two parts to the submission.

   - R / Excel work (20 points): Submit the ORGANIZED Excel / .qmd file with all of the summaries and displays you created.
   - (Informal) Write-Up (10 points): Based on the narrative you explored, copy the relevant summaries and displays from your Excel / R work into a word document.

     Then write a summary of your findings, relating it back to the context of the data (this may include unanswered questions that perhaps more / different data would be needed to explore, etc.)

     (A bulleted-list is perfectly fine, we will say 4 bullets minimum :)

Select answers

1. Answers may vary

Name:

MATH 321: Homework 4

Due ____________________: Turn in a hard copy, neat and stapled.

1. Let $X \sim \text{Gamma}\,(\alpha, \beta)$. Find the MMEs for α and β.

2. Let $X_1, \ldots, X_n \overset{iid}{\sim} \text{Poisson}\,(\lambda)$. Find the MLE for $P(X = 0)$.

3. Let $X_1, \ldots, X_n$ be a random sample from $f(x \mid \theta) = \theta x^{\theta - 1} \quad 0 < x < 1,\, 0 < \theta < \infty$.
   (a) Find the MLE for θ.
   (b) Find the MME for θ.

4. Let $X_1, \ldots, X_4 \overset{iid}{\sim}$ (continuous) $\text{Uniform}\,(0, \theta), \quad 0 < \theta < \infty$.
   (a) Find the MLE for θ. We are going to "logic" our way to this MLE (without taking any derivatives).

   Find the likelihood function like usual. Then think about think about what the range of θ must be if we actually have collected data (say $\mathbf{x} = \{0.5, 0.1, 2, 3\}$). Then sketch a plot of the likelihood, and it should be clear what the MLE is :)

   (b) Show that $\hat{\theta}_{MLE}$ is a biased estimator of θ.
   (c) Find an unbiased estimator of θ as a function of $\hat{\theta}_{MLE}$.

Select answers

1. $\hat{\alpha} = \dfrac{\bar{X}^2}{v}$ and $\hat{\beta} = \dfrac{\bar{X}}{v}$
2. $MLE = \mathrm{e}^{-\bar{X}}$
3. (a) $MLE = \dfrac{-n}{\sum \ln(X_i)}$
   (b) $MME = \dfrac{\bar{X}}{1 - \bar{X}}$
4. (a)
   (b)
   (c)

Name:

MATH 321: Homework 6

Due ___________________: Turn in a hard copy, neat and stapled.

Find the .xlsx file to work in and/or the .csv file to load into R on Canvas.

For all the questions below, do the following:

- Choose the correct form of the interval, must provide justification for this choice (i.e. reference the variables that affect the scenarios we have discussed).
- Calculate the point estimate, critical value and standard error using technology (Excel / R).
- Give the final lower and upper bounds.

1. Assume that the yield per acre for a particular variety of soybeans is $N(\mu, \sigma^2)$ with unknown variance.

   Using the 'Soybean' data, find an 80% confidence interval for μ.

2. Let X_1 and X_2 equal be the earnings of two different stores owned by a toy company during the Christmas season. Assume that $X_1 \sim \mathrm{N}(\mu_1, \sigma^2)$ and $X_2 \sim \mathrm{N}(\mu_2, \sigma^2)$. Data is given in the excel file.

   Using the 'Earnings' data, find a 95% confidence interval for $\mu_1 - \mu_2$.

3. As a clue to the amount of organic waste in Lake Macatawa, several samples of water were taken from the east basin and the count of the number of bacteria colonies in 100 milliliters of water was recorded.

   Using the 'Bacteria' data, find a 90% confidence interval for the mean number μ of colonies in 100 milliliters of water in the east basin.

4. SAT scores for 2010 are shown below. Separate random samples size 45 (so not the same test-taker's Verbal and Math) and produced the means and standard deviations listed in the accompanying table:

	Verbal	Mathematics
Sample mean	505	495
Sample standard deviation	57	69

   (a) Construct a 92% confidence interval for the mean Verbal score.

   (b) Construct a 92% confidence interval for the mean Math score.

   (c) Construct a 92% confidence interval for the mean difference in Verbal and Math scores.

   State a conclusion about how the mean Verbal and Math scores compare.

   (d) Suppose the average scores in 2005 for Verbal and Math were 508 and 520, respectively. State a conclusion about how each score compares to the respective 2005 average.

5. Researchers investigated a muscle condition between two types of exercise enthusiasts, runners and cyclists. Compartment pressure measurements were taken at for both groups at rest and during exercise (80% max O_2 consumption). The data from random samples of 10 runners and 10 cyclists for compartment pressure (in millimeters of mercury) are summarized in the following table:

	Runners		Cyclists	
Condition	Mean	s	Mean	s
Rest	14.5	3.92	11.1	3.98
80% maximal O_2 consumption	12.2	3.49	11.5	4.95

   (a) Construct a 95% confidence interval for the difference in mean compartment pressures between runners and cyclists under the resting condition.

   (b) Construct a 90% confidence interval for the difference in mean compartment pressures between runners and cyclists who exercise at 80% of maximal oxygen (O2) consumption.

6. Twenty-four 9^{th} and 10^{th} grade high school girls were put on an ultraheavy rope-jumping program where their 40-yard dash times were measured before and after the program. Use the '40 yard dash time' data to:

   (a) Construct a 85% confidence interval for the mean difference in before and after times for the 40-yard dash. State a conclusion about whether or not the jump rope program was effective.

   (b) Construct a 98% lower-bound confidence interval AND a 98% upper-bound confidence interval for mean difference in before and after times for the 40-yard dash.

   (c) Combine the intervals from (b) to form a two-sided confidence interval. State the new confidence and the conclusion about whether or not the jump rope program was effective. Is this a different conclusion than in (a)?

Select answers

1. $\approx [43.24, 53.16]$
2. $\approx [218.76, 546.42]$
3. $\approx [50.00, 75.47]$
4. (a) $\approx [490.12, 519.88]$

   (b) $\approx [476.99, 513.01]$

   (c) $\approx [-13.36, 33.36]$

   (d)
5. (a) $\approx [-0.31, 7.11]$

   (b) $\approx [-2.62, 4.02]$
6. (a) $\approx [0.00, 0.16]$

   (b)

   (c) $\approx [-0.03, 0.19]$

Name:

MATH 321: Homework 7

Due ___________________: Turn in a hard copy, neat and stapled.

For all the questions below, do the following:

- Define the parameter(s) of interest AND state the null and alternative hypotheses.
- Define the test statistic and calculate it.
- Determine the rejection region and/or calculate the p-value as necessary.
- State the conclusion and any additional insights that can be drawn.

1. A survey published in the American Journal of Sports Medicine reported the number of meters (m) per week swum by two groups of swimmers – those who competed exclusively in breaststroke and those who competed in the individual medley (which includes breaststroke). The number of meters per week practicing the breaststroke was recorded for each swimmer, and the summary statistics are given below.

   Is there sufficient evidence to indicate that the average number of meters per week spent practicing breaststroke is greater for exclusive breaststrokers than it is for those swimming individual medley? Use $\alpha = 0.05$ and the "traditional method" to make your conclusion.

	Specialty	
	Exclusively Breaststroke	Individual Medley
Sample size	130	80
Sample mean (m)	9017	5853
Sample standard deviation (m)	7162	1961
Population mean	μ_1	μ_2

2. The hourly wages in a particular industry are normally distributed with mean \$13.20 and standard deviation \$2.50. A company in this industry employs 40 workers, paying them an average of \$12.20 per hour.

   Can this company be accused of paying substandard wages? Use an $\alpha = 0.01$ level test and the "p-value" method to make your conclusion.

3. The commercialism of the U.S. space program has been a topic of great interest since Dennis Tito paid \$20 million to ride along with the Russian cosmonauts on the space shuttle. In a survey of 500 men and 500 women, 20% of the men and 26% of the women responded that space should remain commercial free.

   Does statistically significant evidence exist to suggest that there is a difference in the population proportions of men and women who think that space should remain commercial free? Use $\alpha = 0.05$.

4. Operators of gasoline-fueled vehicles complain about the price of gasoline in gas stations. The total tax per gallon for gasoline at each of these 18 locations is given below (in cents). Suppose that these measurements constitute a random sample of size 18:

42.89	53.91	48.55	47.90	47.73	46.61
40.45	39.65	38.65	37.95	36.80	35.95
35.09	35.04	34.95	33.45	28.99	27.45

Is there sufficient evidence to claim that the average per gallon gas tax is less than 41 cents? Use $\alpha = 0.05$.

Select answers

1. $TS \approx 4.756 \implies$ Reject H_0
2. p-value $\approx 0.0057 \implies$ Reject H_0
3. $TS \approx -2.254$, p-value $\approx 0.024 \implies$ Reject H_0
4. $TS \approx -0.858$, p-value $\approx 0.201 \implies$ Fail to reject H_0

3 Finals Reviews

Name:

MATH 321: Review Part 1

1. Let (X, Y) be a bivariate continuous random vector with joint pdf
   $f(x, y) = 3x \quad$ for $0 \leq y \leq x \leq 1$.

   (a) Find $P(X \geq 0.5, Y \geq 0.5)$.

   (b) Show if $X \perp\!\!\!\perp Y$ using the definition.

   (c) Find the conditional pdfs $f(x \mid y)$ and $f(y \mid x)$.

   (d) Using result from (c), find $E(X \mid Y = 0.5)$.

2. Let (X, Y) be a bivariate continuous random vector with joint pdf
$f(x, y) = 4xy \quad$ for $0 \leq x \leq 1, 0 \leq y \leq 1$.

(a) Find $P(Y \geq X)$.

(b) Show if $X \perp\!\!\!\perp Y$ by inspection.

(c) Using result from (b), find $E(X^4 Y)$

3. Let (X, Y) be a bivariate discrete random vector with joint pmf table:

$y \backslash x$	0	1
0	2/9	3/9
1	2/9	1/9
2	1/9	0

(a) Find the following probabilities: $P(X \leq 1, Y = 0)$, $P(X + 1 \leq Y)$, and $P(Y^2 = X)$.

(b) Find the marginal pmfs of X and Y. Also find the conditional pmfs of $f(x \mid Y = 0)$ and $f(x \mid Y = 1)$.

(c) Find the following: $E(X \mid Y = 1)$, and $E(X + 1 \mid Y = 0)$.

(d) Find $\text{Cov}(X, Y)$.

(e) Find $\text{Corr}(X, Y)$.

(f) Find $V(X + Y)$.

Name:

MATH 321: Review Part 2

1. Let X_1 and X_2 be independent random variables where $X_1 \sim \text{Normal}\,(\mu = 1, \sigma = 2)$, and $X_2 \sim \text{Normal}\,(\mu = 5, \sigma = 3)$.

   Find $P(3X_1 - X_2 < 2)$.

2. Let $X_{(1)}, \dots, X_{(8)}$ be the order statistics of a random sample of size $n = 8$ from a random variable X with pdf
$$f(x) = \frac{3}{4}(x^2 + 1), \quad 0 \le x \le 1$$

   (a) Find the cdf of X, $F_X(x)$, and $P(X \le 0.25)$.

   (b) Find the cdfs $F_{X_{(1)}}(x)$ and $F_{X_{(8)}}(x)$.

   (c) Find the pdfs $f_{X_{(1)}}(x)$ and $f_{X_{(8)}}(x)$.

(d) Find the cdf $F_{X_{(6)}}(x)$

(e) Find $P(X_{(6)} \leq 0.25)$.

(f) Find the pdf $f_{X_{(5)}}(x)$.

(g) Write the integrals to find $E(X_{(5)})$ and $P(X_{(5)} \geq 0.75)$

3. Let $X_1, \ldots, X_n$ be a random sample from $f(x \mid \theta) = \theta x, \quad 0 < x < 1, \theta > 0$.

   (a) Find the MME of θ.

   (b) Show if $\hat{\theta}_{MME}$ is unbiased.

4. Let $X_1, \ldots, X_n$ be a random sample from $f(x \mid \theta) = (\theta + 1)x^{\theta}, \quad 0 < x < 1, \theta > -1$.

   (a) Find the MLE of θ.

   (b) Find the MLE of $E(X) = \dfrac{\theta + 1}{\theta + 2}$.

5. Let $X_1, \ldots, X_{35}$ be a random sample from a Geometric ($p = 0.45$) experiment.

   (a) Let Y_1 be the average number of trials to get the first success in the $n = 35$ runs of this experiment, i.e. $Y_1 = \bar{X}$. Find the distribution of Y_1.

   (b) Let Y_2 be the total number of runs (across all trials) to get the first success in $n = 35$ trials of this experiment, i.e. $Y_2 = \sum_{i=1}^{35} X_i$. Find the distribution of Y_2.

6. Let $X_1, \ldots, X_{40}$ be a random sample from a $f(x)$, where $\mu = 15$ and σ^2 is unknown. Let $\bar{X}$ be the sample mean and S be the sample standard deviation.

   (a) Find the distribution of $\bar{X}$.

   (b) Find the distribution of $\dfrac{\bar{X} - 15}{S/\sqrt{40}}$.

   (c) Suppose we know the population variance $\sigma^2 = 4$. Find the distribution of $\dfrac{\bar{X} - 15}{2/\sqrt{40}}$.

Name:

MATH 321: Review Part 3

1. The University is investigating the safety of MATH courses for their students. To do this, they plan a study to compare the blood pressure of upper level MATH courses students before the final exam and after completing the final exam. Data are shown below (in mm Hg).

   Construct and interpret a 95% confidence interval for the difference in blood pressure before and after the final exam for math students. State a conclusion if final exams increased blood pressure on average.

BP Before	BP After
120	125
110	112
208	207
121	128
115	119
119	115
123	128
117	125
124	124

2. The campus bookstore is determining if they need to increase their marketing budget. They would like at least 65% of students to buy their textbooks directly from them rather than off-campus stores. In order to check this, they took a random sample of 137 students in which 81 students said they buy their books at the campus bookstore.

   (a) Is there enough evidence to conclude the true proportion of students who buy their books at the campus bookstore is greater than 65%? Use $\alpha = 0.05$.

(b) Construct the corresponding 90% (two-sided) confidence interval. State a conclusion if this proportion is greater at least the desired 65%.

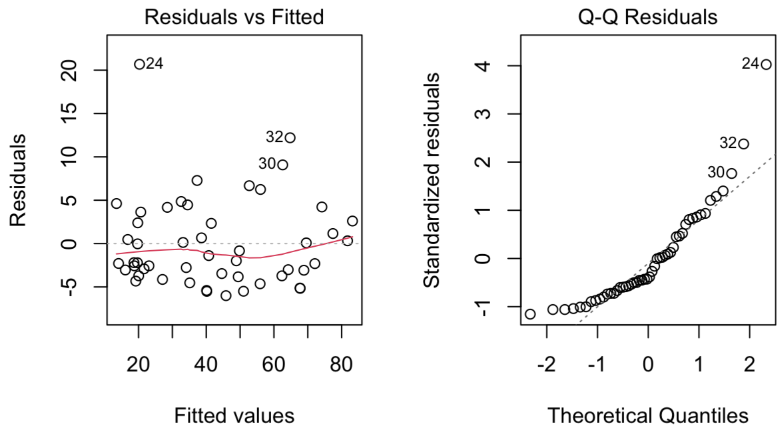
3. Scientists discovered a new mountain range under the sea. Lets assume the sea mountain heights are normally distributed with unknown standard deviation.

   From a random sample of 13 peaks, there was an average height of 11,308 ft and standard deviation of 5,287 ft.

   (a) Is there enough evidence to conclude the average heights of these new sea mountains is different than the Rocky Mountains, which average 14,400 ft? Use $\alpha = 0.12$.

   (b) Construct the corresponding (two-sided) confidence interval for this test and confirm your conclusion from part (a).

4. Given the following residual plots for the regression of $Y \sim X$, assess the assumptions of the normal error regression model.



5. (Extra R practice, NOT ON FINAL)

   (a) Using the data from question 1, perform EDA for the regression on After as the response variable and Before as the explanatory variable. Is there anything we should be concerned about beforehand?

   (b) Fit the regression from part (a) and get the estimated coefficients.

(c) Perform a t-test on the slope coefficient. Use $\alpha = 0.01$

(d) Predict the after blood pressure for a student with a before blood pressure of 123.