Math 10

Midterm Exam ANSWER KEY

Instructions

1. Do not start reading the exam before your instructor says you can start.

2. Write your name, and the name of your instructor below.

3. Read academic honor principle carefully, make sure you understood it, and then sign it.

4. If a question asks you to give an answer to X significant figures, only do so for the final answer. Please do the calculations without rounding, **unless** the question says you may do your calculations with X significant figures.

5. This exam consists of 5 questions, which adds up to a total of 40 point.

Name :

Instructor :

Academic Honor Principle

On exams, you may not give or receive help from anyone. Exams in this course are closed book, and no notes, calculators or other electronic devices are permitted. Any student giving or receiving assistance during an examination violates the Academic Honor Principle.

Signature: ______________________

Question 1 (6 points)

Suppose that **X**, **Y**, **W** are normal random variables, with (population) means and variances as shown below.

Variable	Mean	Variance
X	2	4
Y	0	9
W	15	16

Rank the probabilities from smallest to largest. Show your work/explanation. Guesses without any work/explanation will not be given credit. *(6 points)*

$$P(X \leq 1),\ \ P(Y \leq -2),\ \ P(W \leq 12)$$

2 points for each probability. Must provide some kind of explanation. E.g. you could have ranked these using the z-score alone, without the finding the actual probabilities, and that will be given full credit.

$P(X \leq 1) = P(z\text{-score} \leq -0.5) = 0.3085.$

$P(Y \leq -2) = P(z\text{-score} \leq -0.67) = 0.2514$

$P(W \leq 12) = P(z\text{-score} \leq -0.75) = 0.2266.$

So, $P(W \leq 12) \leq P(Y \leq -2) \leq P(X \leq 1)$.

Question 2 (4 points)

Please show calculations or briefly explain your answer. Answers without any work will not be given credit. Rigorous proofs are not required.

a) A couple has two children. You are given that the **older** child is a boy. If the independent probabilities of having a boy or a girl are both 50 %, what is the probability that the couple has two boys? *(2 points)*

Several ways to do this. Here are a few.

First, enumerate all ordered pairs (younger,older) of children: GG, GB, BG, BB. Then, realize that the probability is $\frac{1}{2}$, since the condition that the older child is a boy restricts us to either BG or BB, both of which have the same probability.

Alternatively, using Bayes' Theorem: $P(BB|\text{ older } B) = \frac{P(\text{ older } B|BB)P(BB)}{P(\text{ older } B)} = \frac{1\cdot\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$.

Another alternative: full credit for saying the probabilities are independent, so 0.50 regardless of whether the older is boy or girl.

b) A couple has two children. You are given that **at least one** is a boy. If the independent probabilities of having a boy or a girl are both 50 %, what is the probability that the couple has two boys? *(2 points)*

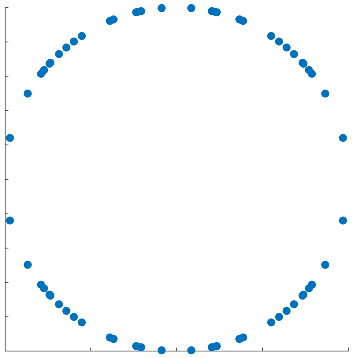
Same as before, enumerate all ordered pairs (younger,older) of children: GG, GB, BG, BB. Then, realize that the probability is $\frac{1}{3}$, since the condition that at least one child is a boy restricts us to either GB, BG or BB, which has the same probabilities.

Alternatively, using Bayes' Theorem: $P(BB|\text{ at least one } B) = \frac{P(\text{ at least one } B|BB)P(BB)}{P(\text{ at least one } B)} = \frac{1\cdot\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$.

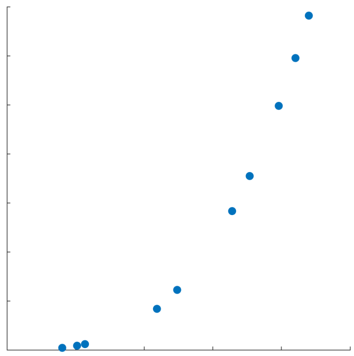
This is not required for the course. So please skip this answer if it confuses you. The interesting boy-girl paradox arises if we say we pick one of their children at random and it turns out to be a boy, then the BB case has a higher chance of being picked, resulting in the answer of $\frac{1}{2}$ instead. This gets full credit it is explained correctly.

Question 3 (10 points)

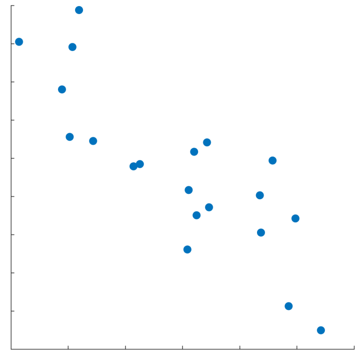
a) Match these Pearson correlation coefficients to their respective scatter plots: $r = -0.83$, $r = 0$, $r = 0.95$. *(3 points)*



r = 0

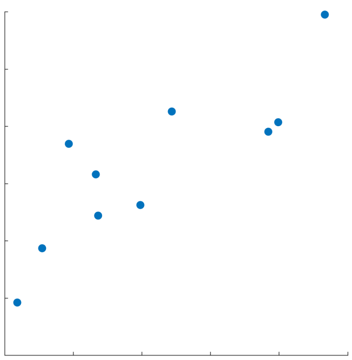


r = 0.95



r = -0.83

For part b,c and d, we will be working with a set of bivariate data as shown by the scatter plot below.



Variable x is plotted on the horizontal axis, and variable y on the vertical axis. The Pearson correlation coefficient is $r(x, y) = r = 0.8461$.

b) A researcher claims that if she switches the axes so that x is on the vertical axis, and y is on the horizontal axis (i.e. switched the variables x and y), the Pearson correlation coefficient is now $r(y, x) = -0.8461$ since the relationship is "reversed". True or false? Explain. *(2 points)*

False, 1 pt. Pearson's r is symmetric and will be unchanged when the variables are switched like this, 1 pt.

c) Suppose you change the unit of measurement of variable x (the horizontal axis) by replacing every x value with $z = 5x - 5$ instead. You then plot z on the horizontal axis, and the same corresponding y on the veritcal axis. Would the correlation coefficient of z, y be $r = 0.8461$, or will it be different? Explain. Do not calculate the new correlation cofficient. *(2 points)*

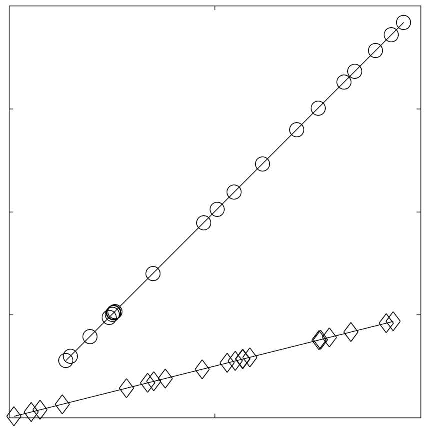
Remains the same, 1 pt. Pearson's r will be unchanged when these kind of linear transformations are applied to the variables, 1 pt.

d) Based only on the correlation coefficient of $r = 0.8461$ on this set of data, and no other information regarding what x and y actually are, a researcher concludes that there must be a strong relationship between x and y. She also concludes that either an increase in x will cause an increase in y, or that an increase in y will cause an increase in x. Do you agree with her conclusion? Explain. *(2 points)*

Disagree with "an increase in one **WILL** cause an increase in the other". 1 pt. Correlation does not imply causation. We could have gotten the r value by chance. We need to know more about what exactly x and y are before we can conclude that an increase in one will lead to an increase in the other. 1 pt

Note: it is ok to say such a high r *MIGHT* show that there is a strong linear correlation. Here, we want to know if that relationship is real or did it occur by chance. Without knowing what x and y really are, we cannot tell.

e) Consider the scatter plot of two sets of bivariate data below. Dataset 1 is drawn with **circles**, while dataset 2 is drawn with **diamonds**.



Which of these two sets of data, if any, would have a higher Pearson's correlation coefficient r? Explain. *(1 point)*

Both will be equal (1 pt). Full credit also given for saying they're both very close to 1, as they are close to being on a line. Or saying that they are likely to be the same, and close to 1 or have very high r values etc.

General Grading Stuff for Qns 4 and 5

1) I take away 1 point the first time a particular calculation error or "wrong formula used" error.

2) I try not to double penalize errors that I have already taken points from due to part 1).

3) Very minor errors are ignored. E.g. writing "Central Tendency Theorem" instead of "Central Limit Theorem".

4) I don't penalize if the final answer is simplified enough, but then a mistake is made to simplify more than required.

5) I don't actually penalize if the answers were not simplified answers enough UNLESS no effort were made to simplify.

Question 4 (10 points)

A population has mean age 30, with variance 64. The distribution of ages is highly negatively skewed, with a long tail to the left.

a) Are the mean age and the standard deviation good summary statistics for the distribution of ages in this population? Explain. *(2 points)*

No, 1 pt (reasons must be right). The distribution is highly negatively skewed 1 pt. Alternative: This is not symmetric. Mean and standard deviation would be good if it was symmetric. Will take away 1 pt for saying anything that is false but still got the general idea.

b) Which distribution would be a good approximation for the sampling distribution of the means in samples of size 16? Also state the mean and standard error of this approximation. Simplify your answer as much as possible. *(3 points)*

Normal distribution (1 pt) with mean 30 (1 pt) and standard error 2 (1 pt).

c) What theorem allowed us to make the approximation in the previous part b)? *(1 point)*

Central Limit Theorem. Get the 1 pt for stating this, even if did not apply it to part b), or applied it incorrectly.

d) If someone took a sample of size 16 from this population, and calculated the mean age in the sample, what is the approximate probability that the mean age would be in the interval $[28.6, 34.4]$? Show your work. You may do your calculations with 3 significant figures. *(4 points)*

Hint: illustrations are allowed. It might be helpful to draw the curve and shade the relevant area underneath.

Using the z-table: $P(z\text{-score} \leq 2.2) = 0.986$ 1 pt, $P(z\text{-score} \leq -0.7) = 0.242$, 1 pt. Awarded for each correct z-score calculation.

Then, $0.986 - 0.242 = 0.744$, 2 pts. If you managed to explain or show that you understand the logic of this last step, you get 2 pts even if the z-score calculations above are not correct, due to calculation errors or careless mistakes etc.

Question 5 (10 points)

a) A truck is full of watermelons that has population mean weight of 9 kilograms (kg). You do not know the variance. However, you know that the weights of the watermelons are normally distributed.

Suppose you take a simple random sample of $n = 4$ watermelons, and calculated a sample mean of $M = 8$ kg from this sample. Using this sample, you got an estimate of the standard deviation $s = 2$ kg.

Construct a 90 % confidence interval for the mean weight of watermelons.

Show your work. Simplify your answer as much as possible so that the interval is of the form $[a, b]$, where a and b are numbers to 2 signficant figures *(6 points)*.

Degrees of freedom $= 4 - 1 = 3$, 1 pt.

t-value from the table for 90 % confidence interval and degrees of freedom 3 is $t = 2.35$, 2 pts.

Standard error, $\frac{s}{\sqrt{n}} = \frac{2}{\sqrt{4}} = \frac{2}{2} = 1$, 1 pts.

90 % confidence interval, $[8 - 2.35 \cdot \frac{2}{\sqrt{4}}, 8 + 2.35 \cdot \frac{2}{\sqrt{4}}] = [5.65, 10.35]$, 2 pts.

You get points for calculating the correct standard error, even if the entire set up is wrong. E.g. using the normal distribution here. BUT only if the standard error is correct.

However, you need to be using the correct formula and set up to get the 2 pts for the confidence interval. You will get these 2 pts even if you made calculation errors previously.

b) Consider a population of males and females. You do not know the proportion of females in the population.

Suppose you take a simple random sample of $n = 4$ people from this population, and found that the proportion of females in this sample is $p = \frac{1}{5}$.

Construct the 95 % confidence interval for the population proportion of females.

Show your work. Simplify your answer as much as possible, so that the interval is of the form $[a, b]$, where a and b are might contain a mix of numbers, and fractions that are reduced to lowest terms. You do not have to add/subtract $\frac{0.5}{n}$ from the upper and lower limit of the interval. *(4 points)*

Normal approximation to the binomial gives the z-value as $z = 1.96$. You can ue the 68-95 heuristic to get $z = 2$, which is fine. 1 pt.

Standard error, $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{\frac{1}{5}\frac{4}{5}}{4}} = \sqrt{\frac{1}{25}} = \frac{1}{5}$, 2 pts.

$[\frac{1}{5} - 1.96 \cdot \frac{1}{5}, \frac{1}{5} + 1.96 \cdot \frac{1}{5}]$, 1 pt

Same as before, you get points for correct standard error, even if the entire set up is wrong. E.g. using the normal distribution here. BUT only if the standard error is correct.

Likewise, you need to be using the correct formula and set up to get the 1 pt for the confidence interval. You will get these 1 pt even if you made calculation errors previously.