STAT 260 Summer 2024: Written Assignment 2

Due: Upload your solutions to Crowdmark BEFORE 6pm (PT) Friday May 24.

You may upload and change your files at any point up until the due date of Friday May 24 at 6pm (PT).

A 2% per hour late penalty will be automatically applied within Crowdmark. The penalty is applied in such a way so that assignments submitted 6pm to 6:59pm will have 2% deducted, assignments submitted 7pm-7:59pm will have 4% deducted, etc.

Note that if you submit any portion of your assignment before the deadline, Crowdmark will NOT permit you to edit your submission (including make additional uploads) after the 6pm deadline passes. This means that if, for example, you upload only Question 1 before the deadline, you will not be able to upload Question 2 after the deadline. If you intend to submit late (with penalty) you must submit the entire assignment late.

Submission: Solutions are to be uploaded to Crowdmark. Here you will be asked to upload your solutions to each question separately. Your solution to Question 1 must be uploaded in the location for Question 1, your solution to Question 2 must be uploaded in the location for Question 2, etc. If your work is uploaded to the wrong location, the marker will not be able to grade it.

You may hand-write your solution on a piece of paper or tablet. If you wish to use this question sheet and write your solutions on the page, space has been provided below. One of the quickest ways to upload work is by accessing Crowdmark from within a web browser on a smartphone. In the area where you upload work, press the "+" button. This will give you the option of using a file already on your phone, or you can use the phone camera to photograph your work. If you complete your work on a tablet, save the file as a PDF or each question as a jpeg and drag/drop the file into the Crowdmark box. ***Photographs of laptop/tablet screens will not be graded***; take a proper screenshot.

Instructions: For full marks, your work must be neatly written, and contain enough detail that it is clear how you arrived at your solutions. ***You will be graded on correct notation***. Messy, unclear, or poorly formatted work may receive deductions, or may not be graded at all. Only resources presented in lecture or linked to on the Stat 260 Brightspace page are permitted for use in solving these assignments; using outside editors/tutors, and/or software (include AIs) is strictly forbidden. Talking to your classmates about assigned work is a healthy practice that is encouraged. However, in the end, each person is expected to write their own solutions, in their own words, and in a way that reflects their own understanding.

1. **[6 marks]** Polymethylmethacrylate (PMMA) is a bone cement used in the surgical fixation of artificial joints. As its primary function is to transfer force from bone to prosthesis, it is important to consider changes to the strength of the material when contaminates (ex. blood) are introduced.

   Surgical standard 20g packets of PMMA with and without human blood were extruded into cylindrical molds and allowed to set. After curing, the samples were tested for their flexural strength (in MPa), and the results were recorded below. Determine the **correlation coefficient** for the data. Then, in a brief sentence or two, interpret your result.

   Be sure to explicitly state (with proper notation) the covariance of the bivariate data, as well as the two sample means and standard deviations. You do not need to show the full calculation for the means and standard deviations; it is enough to indicate that you used your calculator STAT functions.

Blood in sample (ml)	Flexural strength (MPa)
0	90.3
0.5	71.2
1	68.4
1.5	52.8
2	47.2
3	40.3

$\bar{x} = 1.3\dot{3}$

$S_x = 1.08$

$\bar{y} = 61.70$

$S_y = 18.44$

Covariance:

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$
0	90.3	−1.33	28.6	−38.038
0.5	71.2	−0.83	9.5	−7.885
1	68.4	−0.33	6.7	−2.211
1.5	52.8	0.17	−8.9	−1.513
2	47.2	0.67	−14.5	−9.715
3	40.3	1.67	−21.4	−35.738
				$= -95.1$

$$S_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

$$S_{xy} = \frac{-95.1}{6}$$

$$S_{xy} = -19.02$$

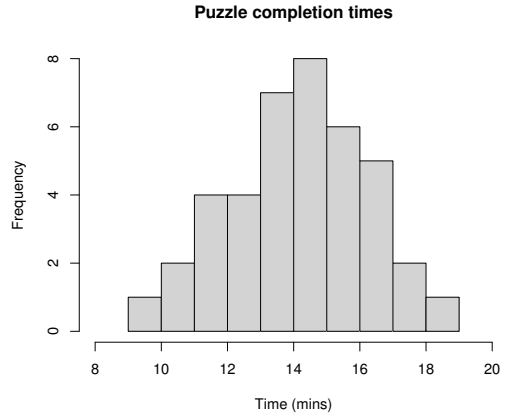
$$\text{Corelation coefficient } (r) = \frac{S_{xy}}{S_x \cdot S_y}$$

$$= \frac{-19.02}{(1.08)(18.44)}$$

$$r = -0.96$$

An r of −0.96 represents a strong negative linear relationship between amount of blood in the PMMA samples and their flexural strength. There is a substantial decrease in flexural strength as the amount of blood increases.

2. [**4 marks**] In a study of aging and learning, subjects were given a sudoku puzzle and their completion times for the puzzle were recorded in minutes. The results of the study were recorded in the histogram below.



Two statistics students later observe the histogram and disagree upon its interpretation:

Student A claims that the sample mean completion time is 14.5 minutes, because the tallest bar is at 14.5.

Student B claims the mean is 4.0 minutes because:

$$\frac{1+2+4+4+7+8+6+5+2+1}{10} = 4.0$$

Which student (if any) is correct? In a few brief sentences, explain why each student is correct or incorrect.

Marks will be deducted for unclear, imprecise, or poorly written explanations.

Both student A and student B have misunderstood how to calculate the mean from a histogram.

Student A is incorrect because the tallest bar in a histogram does not represent the mean. The tallest bar represents the most frequent data point (the mode), not the average. In this case, the mode is 14.5 minutes, but that does not mean the average completion time is also 14.5 minutes.

Student B is also incorrect because they have not considered the frequency of each time interval. The mean is not simply the sum of the time intervals divided by the number of intervals. Instead, each time interval should be multiplied by its frequency (the height of the bar in the histogram), and these products should be summed. The sum should then be divided by the total number of observations (the sum of the frequencies) to find the mean.

Student B's calculation seems to be adding the frequencies and dividing by the number of bars in the histogram. This does not represent the mean, and it doesn't provide any meaningful statistical information. The sum of the frequencies represents the total number of subjects that took part in the study but dividing this total by the number of bars (10 in this case) doesn't provide any useful information.