

Midterm

May 2

Your section: _____ Print your name: _____

Sign your name: _____

This is a closed book exam. However, you are allowed to bring two sheets (double-sided) of 8.5" × 11" paper with notes. The midterm exam consists of three problems, and there is a normal table provided at the end. The exam carries 54 points but the maximum you can score is 50. Good luck!

Problem:	.. 1 2 3	Sum
Points:	20	16	18	56

Problem 1. The cuckoo is a common European bird noted for its characteristic call and its habit of laying its eggs in the nests of other birds, which hatch and rear the young cuckoos. The following table shows the distribution of the length of 243 cuckoos' eggs measured in millimeters.

<u>Length (in millimeters)</u>	<u>Percent</u>
19-21	8.0
21-22	22.0
22-23	40.0
23-24	22.0
24-26	8.0

- Plot the histogram for this data set. Show all intermediate work. Mark the horizontal and vertical scales carefully. Label the axes.
- You find a cuckoo's egg and its length is 23.2 millimeters. Is that above or below the median? Explain your answer.
- Estimate the mean length of a cuckoo's egg from the histogram.

(10 + 5 + 5 = 20 points)

Solution: (a) Millimeters ought to be plotted on the horizontal axis and percentage per millimeter on the vertical axis. The heights of the blocks over the class intervals are computed as follows:

<u>Length (in millimeters)</u>	<u>Percent/millimeter</u>
19-21	$8.0/2 = 4.0$
21-22	$22.0/1 = 22.0$
22-23	$40.0/1 = 40.0$
23-24	$22.0/1 = 22.0$
24-26	$8.0/2 = 4.0$

The actual histogram is not drawn here.

(b) The histogram is seen to be symmetric around the value 22.5 millimeters; thus 50 % of the area of the histogram lies to the left of 22.5 millimeters and 50 % to the right. So the median is 22.5 millimeters, which is less than 23.2 millimeters.

(c) For a symmetric histogram the mean and the median coincide; so the mean length of a cuckoo's egg is estimated as 22.5 millimeters as well.

Problem 2. For each of the situations described below, fill in the blank with one of the following five options:

exactly -1 somewhat negative exactly 0 somewhat positive exactly 1

Then explain briefly.

- (a) For the students taking Statistics 220 in Winter quarter 2001, the correlation coefficient between the score on the midterm and the score on the final will be _____.

Brief explanation: somewhat positive. Students who score higher on the midterm will also in general score higher on the final.

- (b) Suppose that a class meets 30 days in a quarter. For each student in the class we record the number of days they were present and the number of days they were absent. The correlation co-efficient between the number of days present and the number of days absent is _____.

Brief explanation: exactly -1 . Since no. of days present = $30 -$ no. of days absent.

- (c) For the data set shown below, the correlation coefficient is _____.

x	y
1	10
1	-10
5	8
5	-8

Brief explanation: exactly 0 . Because for each x value, you get two y values which are equal in magnitude and of opposite signs, the sum of the products when the variables are expressed in standard units is exactly 0 (the average of the y values is 0).

- (d) The correlation between temperature readings in Fahrenheit and temperature in Centigrade on a particular day at state capitals throughout the U.S will be _____ .

Brief explanation: exactly 1. The Celsius reading C is related to the Fahrenheit reading F by the relation

$$\frac{C}{5} = \frac{F - 32}{9} .$$

Problem 3. In a very large class, the midterm had an average of 50 points with an SD of 20. The final scores averaged out to 60 with an SD of 15. The correlation between midterm and final scores was 0.5, and the scatter diagram was football-shaped.

- (a) Solly scores 60 on the midterm and 90 on the final. However, the instructor loses her midterm score and asks her to provide the score. Solly knows that the instructor will estimate her midterm score by the regression method if she refuses to cooperate. In order to maximize her score, should she cooperate or not? Answer with “yes” or “no”, and explain carefully. Show all your work.

Solution: The regression method gives Solly’s predicted score for the midterm as

$$50 + 20 \times 0.5 \times \frac{90 - 60}{15} = 50 + 20 = 70.$$

This is more than what she actually got on the midterm. So, it does not make sense for Solly to cooperate.

- (b) If you considered the subpopulation of students that scored 50 on the midterm to the subpopulation of students that scored 52 on the midterm, how would the average score on the final for the first subpopulation be related to the average score on the final for the second subpopulation, in terms of the slope of the regression line of the final score on the midterm score ?

Solution: The slope of the regression line of the final score on the midterm score is positive (since the correlation is positive) and gives you the change (increase) in the average score on the final when you compare two subpopulations of students, the first who scored some fixed score on the midterm and the second who scored 1 point more on the midterm. Here, the average score on the final for the subpopulation that scored 52 is going to be higher than the average score for the subpopulation that scored 50 by twice the regression coefficient.

- (c) Estimate the percentage of students that scored between 30 and 70 on the midterm, assuming that the histogram for the midterm scores approximately follows the normal curve. **Solution:** Converting 30 to standard units we get $(30 - 50)/20 = -1$. Converting 70 to standard units we get $(70 - 50)/20 = 1$. So we need to find the area under the normal curve between -1 and 1 which is 68%. This gives the percentage of students who scored between 30 and 70 on the midterm.

(7 + 5 + 6 = 18 points)