Statistics 220/301 **Banerjee**
Winter 2001

Sample Quiz 1

January 25

Your section: _____ Print your name: _____

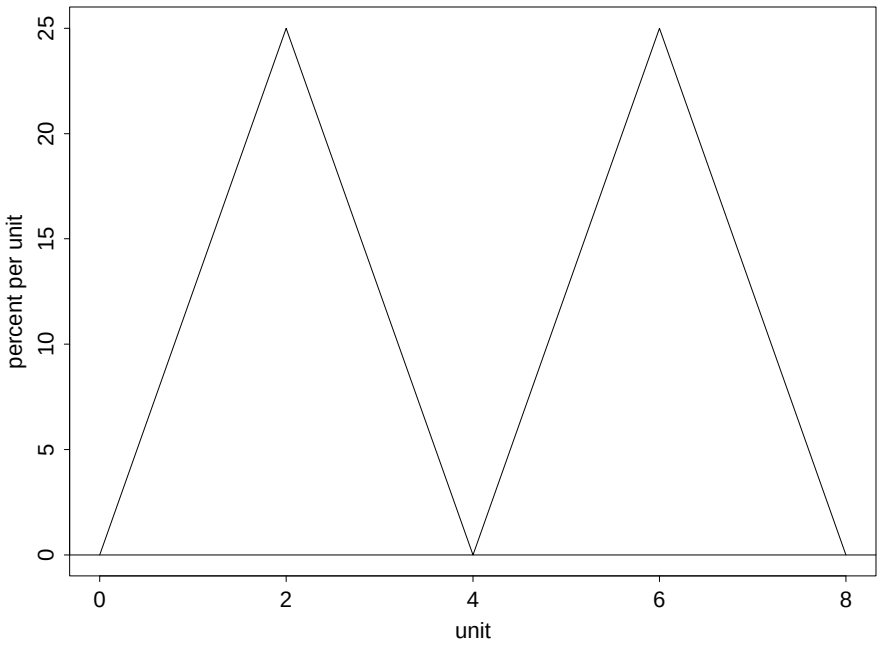
Sign your name: _____

This is an open book exam. However, you are not allowed to pass any material (such as books, notes, or calculators) to each other. This quiz consists of *four problems* carrying 31 points. The maximum you can score is 25. You have 35 minutes. There is a lot of space provided, so try to show as much work as possible. All the best!

Problem 1. There are two kinds of Schools in Northern Ireland; Grammar Schools and Secondary Intermediate Schools. Before graduation, students in both types of schools take a Standard Proficiency Exam. At Grammar Schools, Catholic students do somewhat better in this exam than the Protestant students. At the Secondary Intermediate Schools too Catholics do somewhat better. True or False and explain: If you combine the results from both kinds of schools, the Catholic students must do somewhat better in the Proficiency Exam than the Protestant students.

Solution: False, because of Simpson's paradox. Relationships between percentages in subgroups can be reversed if the subgroups are combined. So, there might be a higher percentage of Catholic students scoring well in each type of school, but when you combine the types of schools the percentage of Protestant students doing well might be higher than the percentage of Catholic students. For a concrete example, look at the solution to the last problem on Homework 1 (the Democrat/Republican example).

Problem 2. (2+2+2+2 = 8 points)



(a) What is the mean of the above histogram? What is the median ?

(b) Find the 25'th percentile of the above histogram.

(c) What is the interquartile range ?

(d) Can you use the normal approximation for this histogram ?

Solution: (a) The histogram consists of two identical triangles located symmetrically on either side of the point 4 on the horizontal axis. Hence 4 is the median as well as the mean of the above histogram.

(b) 50 % of the area of the histogram is under the triangle to the left of the point 4. This triangle is symmetric about the vertical line drawn through the point 2. Hence 50 % of the area of this triangle, which is 25 % of the area of the histogram is to the left of the point 2. It follows that the 25'th percentile is the point 2.

(c) The 75'th percentile of the histogram is the point 6 (arguing as in (b)). So the interquartile range = 75'th percentile - 25'th percentile = 6 - 2 = 4.

(d)No, you cannot. A histogram for which you can use the normal curve is peaked at its mean and then descends symmetrically on either side of the mean and for this histogram this is FAR from being the case (even though the histogram is symmetric about the mean).

Problem 3. (4+4+3=11 points)

(a) If I have a list of two different numbers, how is the deviation of the first number from the average related to the deviation of the second number from the average ? How is the standard deviation related to the deviation of the larger of the two numbers from the average ? (Recall that the deviation of a number from the average is given by: number - average)

(b) Twenty one people in a room have an average height of 5 feet 6 inches. A 22nd person enters the room. How tall would he have to be to raise the average height by 1 inch ?

(c) The wages of employees in a firm go up by $ 100. How does this change (i) the average wage (ii) the SD of the wages ?

Solution: (a) Here

$$Average = \frac{Smaller\ number + Larger\ number}{2}.$$

This sits at the middle between the smaller number and the larger number. So the deviation of the larger number from the average is

$$\frac{Larger\ number - Smaller\ number}{2}$$

and this is the same in absolute value but opposite in sign to the deviation of the smaller number from the average, this being

$$-\frac{Larger\ number - Smaller\ number}{2}.$$

The standard deviation equals the deviation of the larger number from the average, which is positive.

(b) The sum of the heights of the 21 persons is $66 \times 21 = 1386$ inches. If the average height increases by 1 inch after the 22'nd person enters the room, then the sum of the heights of the 22 persons is $67 \times 22 = 1474$ inches. So the height of the 22'nd person is $1474 - 1386 = 88$ inches which is 7 feet 4 inches..a basketball player ?? !!

(c) (i)

$$New\ wage = Old\ wage + 100\,,$$

so

$$New\ average\ wage = Old\ average\ wage + 100\,.$$

So the average wage goes up by 100 dollars.

(ii) The standard deviation remains unafffected, since

$$New\ wage - new\ average = Old\ wage + 100 - (Old\ average + 100) = Old\ wage - old\ average\,.$$

Problem 4 The heights of men aged 18-74 in the HANES sample averaged 69 inches; the SD was 3 inches. Assuming that the histogram followed the normal curve, estimate the percentage of men with heights between 63 and 72 inches. Also estimate the height below which 25 % of the men in the HANES sample fall (use the normal table at the back of the book). (3+3=6 points)

Solution: Estimating the percentage with heights between 63 and 72 inches: 63 in standard units is

$$\frac{63-69}{3} = -2.0\,,$$

and 72 in standard units is

$$\frac{72-69}{3} = 1.0\,.$$

So we need to find the area between -2 and 1 under the normal curve. This is approximately 47.5+34 = 81.5%. (using normal table on Page A - 105 of the text book).

Estimating the 25'th percentile of the height distribution: The 25'th percentile of the normal curve is approximately -0.7 (using normal tables at the back of the book). If the 25'th percentile of the height distribution is P then, P converted to standard units must be close to -0.7. So

$$\frac{P-69}{3} = -0.7$$

showing that

$$P = 69 - 0.7 \times 3 = 69 - 2.1 = 66.9\,.$$