

Quiz 1
January 25

Your section: _____ Print your name: _____

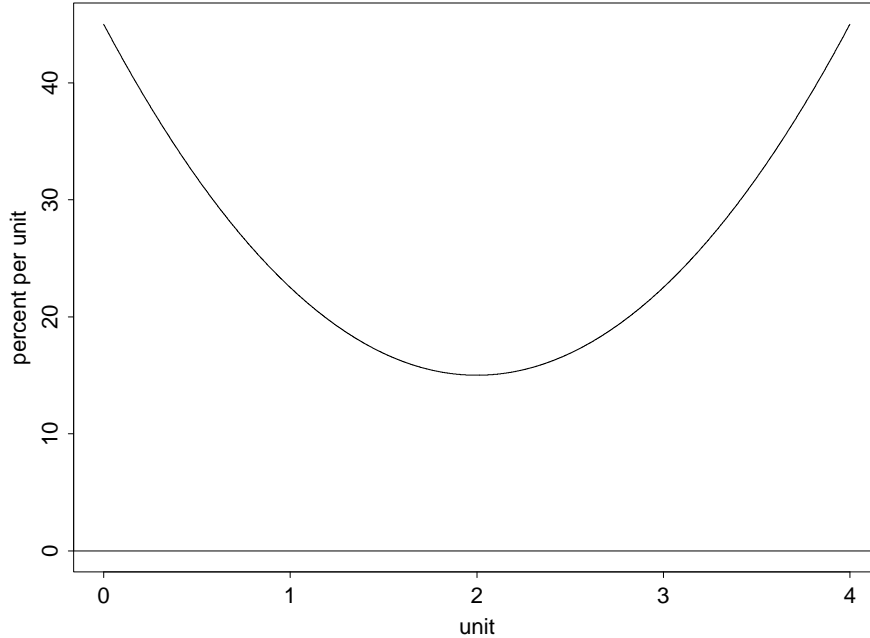
Sign your name: _____

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. Four divisions within a company are separately hiring employees who are all college graduates. It is found that the overall percentage of Ivy League graduates hired is much larger than the overall percentage of non Ivy League graduates hired. I conclude that the company definitely prefers Ivy League graduates a lot more than non Ivy League graduates. Is my inference valid? Explain briefly. (6 points)

Solution: False, because of Simpson's paradox. Relationships between percentages in subgroups can be reversed if the subgroups are combined. So, there could be a higher percentage of non Ivy League graduates hired in each division of the company but when you combine the four divisions the percentage of Ivy League graduates hired could be much higher than the percentage of non Ivy League graduates. For a concrete example of this kind of situation, 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) Is the 25'th percentile greater than or less than 1 ?
- (c) Is more than 75% of the area to the left of 3 or less ?
- (d) Is the interquartile range greater than or less than 2 ?

Solution: (a) The histogram is symmetric about the vertical line drawn through the point 2. Hence 2 is the median as well as the mean of the above histogram.

(b) If you draw the vertical line through 1, the area under the histogram between 0 and 1 is clearly seen to be less than the area under the histogram between 1 and 2. Since 50 % of the area of the histogram is to the left of 2, it follows that more than 25 % of the area is between 0 and 1. Consequently the 25'th percentile is less than 1.

(c) Reasoning as in (b), it is seen that less than 75% of the area is to the left of 3. Thus, the 75'th percentile must be larger than 3.

(d) From the definition of the interquartile range it follows immediately that it is larger than 2.

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

- (a) Construct a list of numbers with mean 5 and standard deviation 3. (Hint: Think of a list of two numbers.)
- (b) The average height of 20 students in a class is 65 inches. One student drops out and the average height of the remaining students drops to 64.9 inches. Was the drop-out student taller than 65 inches and if so, by how much ?
- (c) The wages of employees in a firm go up by 5 %. How does this change (i) the average wage (ii) the SD of the wages ?

Solution: (a) Take the average and add and subtract off the standard deviation; that gives you the desired list. In this case the list $2(5 - 3), 8(5 + 3)$. Notice how this relates to problem 3 (i) of the sample quiz.

(b) Since the average height of the class dropped when the student left, he/she must have been taller than the average height with him/her included, that is 65 inches. The sum of the heights of all 20 students combined is $65 \times 20 = 1300$ inches. The sum of the heights of the remaining 19 students is $64.9 \times 19 = 1233.1$ inches. So the height of the drop-out student is $1300 - 1233.1 = 66.9$ inches. Hence the student was $66.9 - 65 = 1.9$ inches taller than the average.

(c) Here

$$\text{New wage} = \frac{105}{100} \times \text{old wage} .$$

So

$$\text{New average wage} = \frac{105}{100} \times \text{old average wage} ,$$

and

$$\text{New } SD = \frac{105}{100} \times \text{old } SD .$$

In other words, both average and standard deviation go up by 5 %.

Problem 4 The heights of women aged 18-74 in the HANES sample averaged 63.5 inches; the SD was 2.5 inches. Assuming that the histogram followed the normal curve, estimate the percentage of women with heights between 61 and 68.5 inches. Also estimate the height below which 75 % of the women 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 61 and 68.5 inches: 61 in standard units is

$$\frac{61 - 63.5}{2.5} = -1.0,$$

and 72 in standard units is

$$\frac{68.5 - 63.5}{2.5} = 2.0.$$

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

Estimating the 75'th percentile of the height distribution: The 75'th percentile of the normal curve is approximately 0.7 (using normal tables at the back of the book). If the 75'th percentile of the height distribution is P then, P converted to standard units, must be close to 0.7. So

$$\frac{P - 63.5}{3} = 0.7$$

showing that

$$P = 63.5 + 0.7 \times 3 = 63.5 + 2.1 = 65.6.$$