MONT 106N – Identifying Patterns
Solutions for Midterm Exam
October 19, 2009

I. The following quotation is taken from an article that appeared in the *New York Times* on December 3, 1991.

> "... new research also demonstrates that spanking, which remains parents' main disciplinary weapon, is not only useless in the long run but also potentially harmful to the child's emerging personality. In experimental studies with toddlers placed in a room with their mothers and breakable objects they were told not to touch, the children who were frequently spanked were less obedient than the children who were disciplined in nonphysical ways."

A) (5) True or False: The author is claiming that the studies show that children who are spanked frequently tend to become more disobedient than children who are disciplined in nonphysical ways.

Solution: This is True – note the first sentence says that the research to which the author is referring "demonstrates that spanking [is] ... potentially harmful to the child's emerging personality" because it causes them to be less obedient.

B) (10) What questions should one ask about the way these studies were performed before deciding whether they provide good evidence for a relation between spanking and obedience? Explain.

Solution: This kind of study was almost inevitably observational in nature. It would be very hard, if not impossible, to design a randomized double-blind trial on the effects of spanking. So while a randomized double-blind trial would give the strongest evidence, it is not really practical in this situation. Hence the main issue would be to determine how well the designers of the study tried to *control for* underlying confounders like conditions that might predispose children to be less obedient in response to instructions like "don't touch" *and* that might lead to spanking if the parents were inclined to spank. Some possible conditions are hyperactivity, attention deficit, and so on.

The scoring of this problem was done as follows:

- full 10 points only for a complete discussion of the possibility of doing randomized double-blind studies, and for raising issues of controlling for at least one specific and reasonably probable confounder,
- 7 points for raising questions about randomization or double-blinding and possible confounders, without a specific confounder (this is a good answer but not the best one!)
- 3 points for extraneous points not addressing the question of the experimental design.

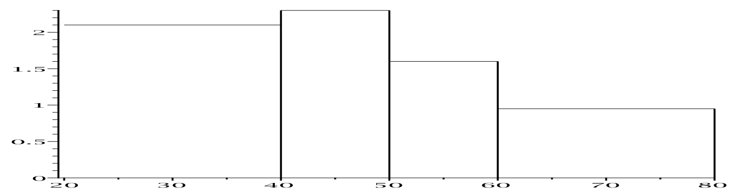

Figure 1: Histogram for II A – Countywide population

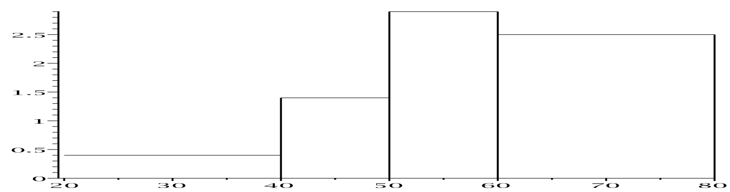

Figure 2: Histogram for II A – Grand Jurors

II. A study of grand juries in Alameda County, California, compared the demographic characteristics of grand jurors in one year with the general population. Here are the results for age. Only persons 21 years and older are included.

Age	Countywide Percentages	Number of Grand Jurors
$20 < \text{age} \leq 40$	42	5
$40 < \text{age} \leq 50$	23	9
$50 < \text{age} \leq 60$	16	19
$60 < \text{age} \leq 80$	19	33
Totals	100%	66 jurors

A) (15) Construct two age histograms, one for the countywide population and one for the actual grand jurors. (Use the graph paper provided.)

Solution: Note that the first column is given in units of percents, while the second is given in raw numbers of jurors. To construct the second histogram, we should start by computing what percentage of the total number of jurors falls into each age class interval:

Age	Percent of Grand Jurors
$20 < \text{age} \leq 40$	$5/66 \doteq 8\%$
$40 < \text{age} \leq 50$	$9/66 \doteq 14\%$
$50 < \text{age} \leq 60$	$19/66 \doteq 29\%$
$60 < \text{age} \leq 80$	$33/66 = 50\%$

Then the histograms are shown in Figure 1 and 2. The units on the horizontal axis are years and the units on the vertical axis are % per year. For example, the histogram for the actual grand jurors has a box of height 2.5 on the interval 60 to 80 since $50/20 = 2.5$.

B) (5) By comparing your histograms, what appears to be true about the age distributions of the two groups?

1. *Solution:* It seems that the grand jurors were older on average than the general population.

III. The grades in a large university calculus class were normally distributed with an average of 68 (out of a possible 100 points) and an SD of 10.

A) (15) If 400 students took the test, estimate how many scored between 58 and 83 inclusive. Show all work and draw normal curves as needed.

Solution: The interval from 58 to 83 is $ave - 1SD$ to $ave + 1.5SD$, so it corresponds to the standard unit interval $z = -1$ to $z = 1.5$. The area of this region can be computed from the standard normal table:

$$\frac{1}{2}(Area(1) + Area(1.5)) = \frac{1}{2}(68.27 + 86.64) \doteq 74.46\%$$

So about $.7446 \times 400 \doteq 310$ of the students scored in this range.

B) (15) The professor decided to set the cutoff numerical grade for an A letter grade at the 90th percentile of the grade distribution. What grade did a student need in order to earn an A in calculus?

Solution: For the 90th percentile, we want the z so that the the area between $-z$ and z is 80% (so 10% in both upper and lower tails). From the normal table this happens between $z = 1.25$ and $z = 1.3$. Using $z = 1.3$ to estimate, we get $ave + 1.3SD = 68 + 1.3(10) = 81$.

IV. The US Department of Justice made a study of civil cases during the year 1992. Juries awarded money damages to the plaintiffs in 5949 of the cases. The median damage award was $50,000 and the average damage award was $450,000.

A) (10) Sketch a possible (smoothed) histogram of the damage awards. Does your histogram have a long left tail, a long right tail, or is it symmetric?

Solution: Because the average is so much larger than the median, the histogram should show a long right tail (and no area for awards less than zero). That is, there had to be a few very large awards to make the average come out so large, while the median was relatively small. Any such histogram is OK.

B) (5) Which is larger, the difference between the award at the 10th percentile and the median award, or the difference between the median award and the award at the 90th percentile? Or are those differences about the same? Explain briefly.

For any histogram with a long right tail as in part A, the difference between the awards at the 90th percentile and the median would tend to be larger than the difference between the awards at the 10th percentile and the median. The distribution of awards is skewed to the larger values.

V.

A) (10) Find the equation of the line passing through the points $P = (1, -2)$ and $Q = (4, 3)$ in the plane.

Solution: The slope is $\dfrac{3-(-2)}{4-1} = \dfrac{5}{3}$. So the equation of the line is $y + 2 = \frac{5}{3}(x-1)$ or $y = \frac{5}{3}x - \frac{11}{3}$.

B) (10) Find the equation of the line parallel to the line in part A and passing through the point $R = (1, 2)$.

Solution: Parallel lines have the same slope, so the line is $y - 2 = \frac{5}{3}(x-1)$, or $y = \frac{5}{3}x + \frac{1}{3}$.