

1. US researchers report that walking two miles or more per day can cut the overall risk of dying in half, and also reduce the risk of dying from cancer. Researchers in the Honolulu Heart Program studied 707 nonsmoking, retired men, aged 61 to 81 years, and collected mortality data on these men for 12 years. During the study, 208 of the men died. The study results show that while 43.1% of men who walked less than one mile per day died, only 21.5% of the men who walked more than two miles per day died.

a) Is the research described above a controlled experiment or an observational study?

(4 points)

observational study since the men decided whether or not to walk & hence determined the treatment & control groups.

b) List two different possible confounding factors that are likely to have an effect on the outcome of this study. (6 points)

1. Healthier men are more likely to walk 2 miles or more & also not die.
2. The older men at the beginning of the study are more likely to not continue to walk & more likely to die.

2. Suppose you are given a list of the heights for all of the second-grade students in an elementary school.

a) How would you find the average of this list? (5 points)

Add them up & divide by the total.

b) How would you find the median of this list? (5 points)

If the number of students is odd, choose the middle score. If not, average the two middle scores.

c) How would you find the SD of this list? What does the SD of this list measure?

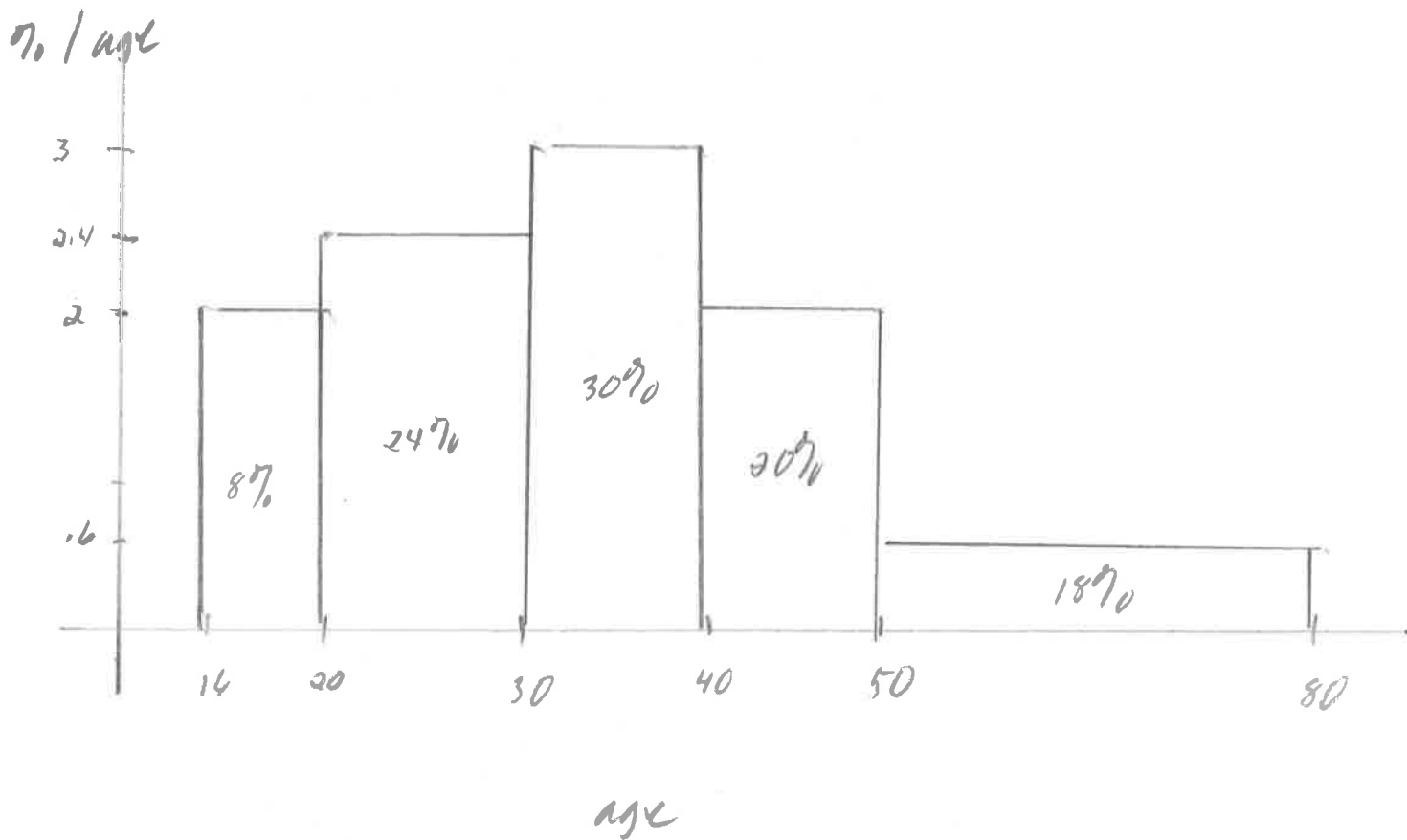
(5 points)

② I would use a calculator. The SD is a measure of the spread in the data.

③ (The SD is the r.m.s. of the deviations from \bar{AV})

3. A study in a neighboring state showed that for a variety of reasons police stopped 4,750 drivers during a particular time period. The following table gives the ages of these drivers. Age intervals include the left endpoint but not the right. Draw the histogram. (15 points)

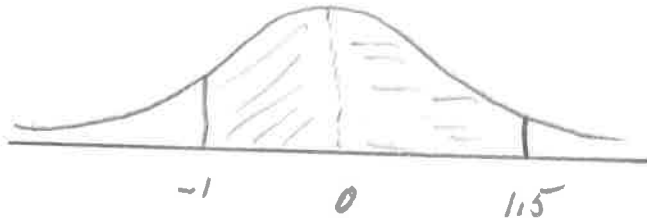
| Age | Number | % | width | height |
|---------|--------|-----|-------|--------|
| 16 - 20 | 380 | 8% | 4 | 2 |
| 20 - 30 | 1140 | 24% | 10 | 2.4 |
| 30 - 40 | 1425 | 30% | 10 | 3 |
| 40 - 50 | 950 | 20% | 10 | 2 |
| 50 - 80 | 855 | 18% | 30 | .6 |



4. Verbal SAT scores for a large high-school senior class follow the normal curve with an average of 500 and a standard deviation of 100.

a) Use the normal approximation to estimate the percentage of the students who scored between 400 and 650? (15 points)

$$\frac{400-500}{100} = -1, \quad \frac{650-500}{100} = 1.5$$

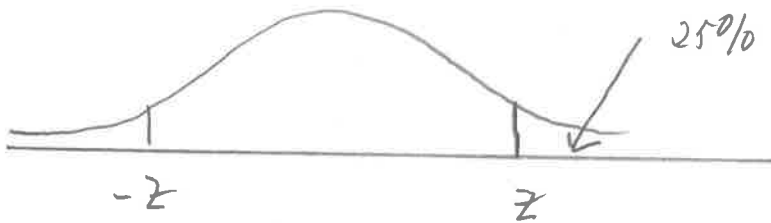


$$\frac{A(-1)}{2} + \frac{A(1.5)}{2}$$

$$\approx 34\% + 43\%$$

$$\approx 77\%$$

b) If Alex is at the 75th percentile, what is his score? (5 points)



$$A(z) = 50\%$$

$$z \approx .68$$

$$\frac{\text{score} - 500}{100} = .68$$

$$\text{score} = 100(.68) + 500 \approx 568$$

500 5. For the graduating class in question 4, their math SAT scores also had an average of ~~600~~ 500 and an SD of 100. The correlation coefficient was .8 and the scatter diagram was football-shaped.

a) What does the correlation coefficient measure? (5 points)

The degree of clustering about the SD line.

The linear association between x & y .

b) Find the equation of the regression line for predicting the math SAT from the verbal SAT. (10 points)

$$(500, 500), \quad m = .8 \cdot \frac{100}{100} = .8$$

$$y - 500 = .8(x - 500)$$

$$y - 500 = .8x - 400, \quad y = .8x + 100$$

c) If Leslie scored 750 on the verbal SAT, use the regression line or 6-step method to estimate her score on the math SAT? (15 points)

$$\text{when } x = 750, \quad y = (.8)750 + 100 = \boxed{700}$$

1) verbal SAT

2) 750

$$3) \frac{750 - 500}{100} = 2.5$$

$$4) (2.5)(.8) = 2$$

$$5) 100 \times 2 = 200$$

$$6) 500 + 200 = \boxed{700}$$

d) If Matthew scored 600 on the verbal SAT, why would you expect him to score less than 600 on the math SAT? (5 points)

Regression effect

e) The rms error for estimating the math SAT from the verbal SAT is equal to

$SD_y \times \sqrt{1 - r^2} = 100 \times \sqrt{1 - (.8)^2} = 60$. What does the rms error measure? Consider all those students who scored 750 on the verbal SAT. Estimate the SD for their math SAT scores. (5 points)

a) the overall size of the error in regression

b) 60