

STAT 101

Assignment 1

1. From the text: # 1.30 on page 29.

A: For the centre the median is 2, the mean is 2.62. I am happy with either for an answer and I am happy to have these read off roughly by eye. The range is 8 which is an easy measure of spread. In a later question you are supposed to get accurate numbers for quartiles. It is fine if you used those numbers here.

The shape of the histogram is skewed to the right. There were $15+11=26$ girls who ate fewer than 2 servings per day. This is $(26/74)*100\% = 35\%$. Note: I do not care how many digits you rounded off to.

2. From the text: # 1.33 on page 31.

A: Graph c is for gender, probably: more women than men in the course. Graph b is handedness – only two possibilities but left handed is much less common than right. Only judgement about whether left handed is less common than male in a typical university course tells you which is which. The heights must be d because that histogram will be more symmetric than the histogram for time spent studying.

3. From the text: # 2.25 on page 59.

A: Income distributions are skewed to the right so means are larger than medians. The mean is \$58,886 while the median is the lower figure of \$46,453.

4. From the text: # 2.30 on page 60.

A: See my answer to the first question. There are 15 zeroes, 11 ones, 15 twos, 11 threes and so on out to 3 eights in the list of 74 numbers. That is an even number so the median is halfway between the 37th and 38th number. Starting from the bottom the 37th number up is in the group of twos and so is the 38th. So the median is two. The first quartile is the median of the first 37 numbers which is the 19th number. This is a one so the first quartile is one. The third quartile is the 19th number in the list from the 38th to the 74th. Count down from the top: 3 eights, 6 are 7 or more (3 sevens and 3 eights) 9 are six or more, 14 are five or more and 22 are four or more. The third

quartile is four. The IQR is $4-1=3$. (The question does not ask for the IQR.)

5. From the text: # 3.41 on page 90.

A: In the class I used the notation $N(5,100)$ to indicate the mean is 5, the SD is 10 and the variance is 100. In this question the SDs are given not the means. Students who took square roots to convert variances to SDs should not be penalized. You need to find the area to the right of 69.3 under a $N(64,2.7)$ curve – an approximation to the proportion of women over 69.3 inches in height. Convert 69.3 to standard deviation units:

$$z = \frac{69.3 - 64}{2.7} = 1.96.$$

In the tables the area to the left of 1.96 is 0.9750 or 97.5%. Students who took the square root of 2.7 get $z = 3.29$ and an area of 99.95%. The needed area is the area to the right: 2.5% or 0.05% for students who took the square root.

6. From the text: # 3.36 on page 76.

A: You are told the area to the right of the mileage you are to find is 10%. So the area to the left of this mileage is 0.9. In the Table A this corresponds to $z = 1.28$. Now you know the value in standard units so you use $x = s * z + \bar{x}$ to get

$$4.3 * 1.28 + 18.7 = 24.2 \text{ mpg.}$$

7. From the text: # 3.37 on page 89.

A: The area to the left of any first quartile is $1/4$ or 0.25. In the tables this corresponds to z being -0.67 or -0.68. (Students may use either value.) The third quartile as in class is 0.67 or 0.68. The first quartile is

$$4.3 * (-0.68) + 18.7 = 15.78 \text{ mpg.}$$

The third quartile is

$$4.3 * (0.68) + 18.7 = 21.62 \text{ mpg.}$$

The estimated median is the same as the mean because an area of 0.5 corresponds to $z = 0$. That makes the median 18.7.

Notice that I gave you those numbers (0.67 and -0.67) in class working with Father's heights.

8. For adults aged 15 to 24 with income in BC in 2000 the census shows an average income of about \$10,200 with a standard deviation of approximately \$10,100. Make a normal approximation to estimate the proportion of such adults whose income is negative. Is a normal approximation wise here? Explain.

A: Convert 0 to standard deviation units:

$$z = \frac{0 - 10200}{10100} = -1.01$$

The area to the left of -1.01 is 15.62% which is a gross overestimate since the true proportion reported in the data set is 0%. The normal approximation is bad here because the distribution is badly skewed to the right.

9. Is the median income likely higher or lower than \$10,000 in the group in the previous question? Explain your answer.

A: For skewed to the right data the mean is typically larger than the median. In this case the skewing is substantial so the difference is probably a lot more than \$200 (the amount by which the mean exceeds the number — \$10,000 — I asked about).

10. The standard deviation for heights in the data set I collected for this class is 3.7 inches. Do you expect the standard deviation of height for women to be more or less than 3.7 inches? Explain.

A: When you put together two histograms of similar widths to make one histogram and when the two histograms have different centres the overall histogram is more spread out than either one of the components. Another way to say this is to say that two people of the same sex tend to be more similar than two people picked without regard to sex. In either case, the SD for the individual groups should be smaller than the overall SD because the two subgroups are less spread out than the overall group.

11. The mean height is 66.8 inches. Make normal approximations to estimate the 90th percentile of heights and the interquartile range.

A: The 90th percentile of the normal curve is the z for which the area in the table is 0.90. This is $z = 1.28$. Convert back to original units:

$$z * SD + \text{mean} = 1.28 * 3.7 + 66.8 = 71.5 \text{ (inches)}$$

The 75th percentile is

$$0.67 * 3.7 + 66.8 = 69.3$$

while the 25th percentile is

$$-0.67 * 3.7 + 66.8 = 64.3$$

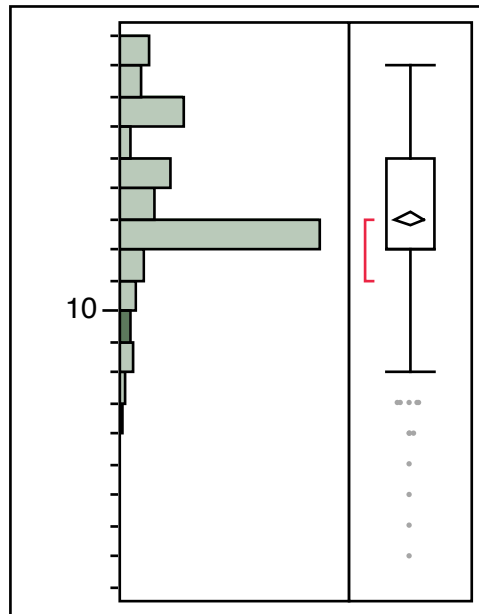
Subtract these to find the IQR is 5 inches.

12. JMP practice question. Graphs to follow.

CPS: Distribution of EDUCATION

Distributions

EDUCATION



Quantiles

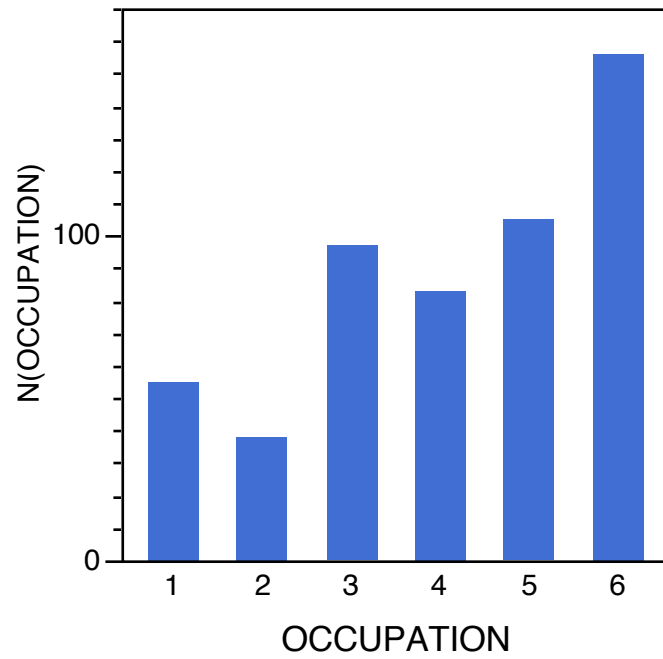
100.0%	maximum	18
99.5%		18
97.5%		18
90.0%		17
75.0%	quartile	15
50.0%	median	12
25.0%	quartile	12
10.0%		10
2.5%		8
0.5%		3.675
0.0%	minimum	2

Moments

Mean	13.018727
Std Dev	2.6153726
Std Err Mean	0.1131782
Upper 95% Mean	13.241057
Lower 95% Mean	12.796396
N	534

CPS: Chart of OCCUPATION

Chart



Scatterplot Matrix

