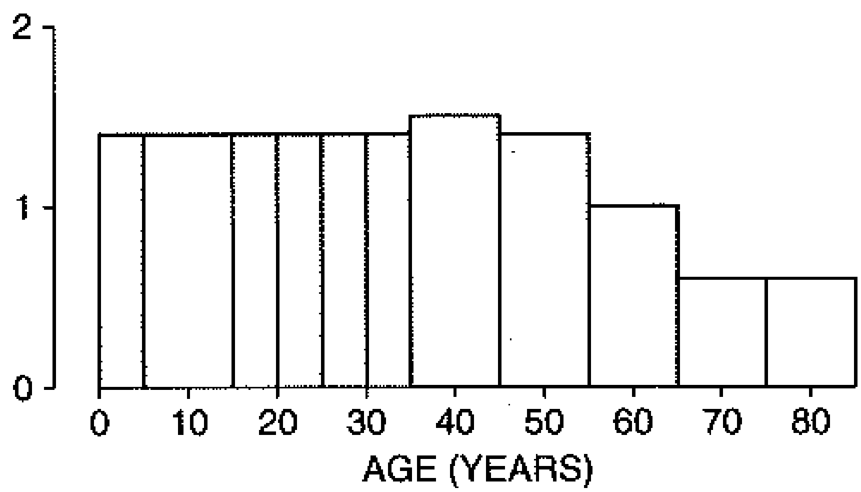
Homework 1 Solutions

by the esteemed Asa Levi

p50 # 2 (10 pts)

Histogram



Part c

There are more people age 0-4.

Part d

The percentage of people age 35 and over is around 50

p74 # 1 (10 pts)

Part a

The average of the list is:

$$\begin{aligned} \text{average} &= \tfrac{41+48+50+50+54+57}{6} \\ &= \tfrac{300}{6} \\ &= 50 \end{aligned}$$

The standard deviation is the root mean square of the list {-9,-2,0,0,4,7}, which is obtained by subtracting the average from each element in the previous list. So we get:

$$\begin{aligned}\sqrt{\frac{9^2+(-2)^2+0^2+0^2+4^2+7^2}{6}} &= \sqrt{\tfrac{81+4+16+49}{6}}\\ &= \sqrt{\tfrac{150}{6}}\\ &= \sqrt{25}\\ &= 5\end{aligned}$$

Part b

0.5 SDs is 2.5, so we just pick which elements are within 2.5 of 50. These are 48, 50 and 50.
1.5 SDs is 7.5, and 48, 50, 50, 54 and 57 are within 7.5 of 50.

p75 # 7 (10 pts)

Part a

This question has nothing to do with statistics, just arithmetic. The average wieght for men is 66kg with an SD of 9kg. 1kg = 2.2lbs, so the average weight for men is 66*2.2 = 145.2 lbs with an SD of 9*2.2 = 19.8 lbs. Similarly women have an average weight of 121 lbs with an SD of 19.8 lbs.

Part b

To estimate the percentage we will approximate the data by the normal distribution and use a z-table. First, we need to get our range in standard units. $57\text{kg} = 66\text{kg} - 9\text{kg} = -1\text{SU}$, Similarly 75 kg is 1 SU. So we need to look for the area in the line $z = 1$ in the z-table in the back of the book, which gives us 68.27%. Thus 68.27% of men weighed between 57 and 75 kg.

Part c

If you took the men and women together the SD of their weights would be bigger than 9kg. The SD would go up because the spread of the histogram would increase; individually the men's histogram had a peak at 66kg, while the women's had a peak at 55kg. But when you combine these it looks more like a flat plateau between 55kg and 66kg. This flatness means that the data has spread out, and is now less concentrated around the mean. Thus the SD will increase.

p94 # 3 (10 pts)

Part a

First we need to figure out what 700 is in standard units. To do this we compute $\frac{700-543}{110} \approx 1.42$. Now we look at a z-table and find that for $z = 1.40 \approx 1.42$ the area is about 83.85%. If we take half of this area plus half of the total area under the graph this will be the percentage of students who scored less than 700. This is 91.925%. Then the students who scored more than 700 are $100\% - 91.925\% = 8.075\%$.

Part b

Using the same approach as in (a) you get that 3.595% of students scored above 700.

p95 # 6 (8 pts)

No. 178 is only 1 SD above 169, and there are 0% of people above 1 SD away from the average. In the normal curve about 16% of people are more than 1 SD above the average, so this data doesn't follow the normal curve.

p95 # 8 (12 pts)

Part a

True. If your list is $\{x_1, x_2, x_3, \ldots, x_n\}$ then the average is $\frac{x_1+x_2+x_3+\ldots+x_n}{n}$. After adding 7 to each entry the list would be $\{x_1+7, x_2+7, x_3+7, \ldots, x_n+7\}$ and the average would be:

$$\frac{x_1+7+x_2+7+x_3+7+\ldots+x_n+7}{n} = \frac{x_1+x_2+x_3+\ldots+x_n+7n}{n} = \frac{x_1+x_2+x_3+\ldots+x_n}{n} + \frac{7n}{n} = \frac{x_1+x_2+x_3+\ldots+x_n}{n} + 7$$

Which is your previous average plus 7.

Part b

False. Consider the list {-1, 1}. The SD of this list is 1. Now, consider the list $\{-1+7, 1+7\} = \{6, 8\}$. The average of this new list is 7, so the SD is the RMS of the list $\{6-7, 8-7\} = \{-1, 1\}$, which is also 1.

Part c

True. If your list is $\{x_1, x_2, x_3, \ldots, x_n\}$ then the average is $\frac{x_1+x_2+x_3+\ldots+x_n}{n}$. After doubling each entry the list would be $\{2x_1, 2x_2, 2x_3, \ldots, 2x_n\}$ and the average would be:

$$\frac{2x_1+2x_2+2x_3+\ldots+2x_n}{n} = \frac{2(x_1+x_2+x_3+\ldots+x_n)}{n} = 2\frac{x_1+x_2+x_3+\ldots+x_n}{n}$$

Which is twice your previous average.

Part d

True. Let your list be $\{x_1, x_2, \ldots, x_n\}$ and the average be M. Then if you double each entry in the list you have $\{2x_1, 2x_2, \ldots, 2x_n\}$ with average $2M$. The SD of the first list is $\sqrt{\frac{(x_1-M)^2+(x_2-M)^2+\ldots+(x_n-M)^2}{n}}$. The SD of the second list is:

$$\sqrt{\frac{(2x_1-2M)^2+(2x_2-2M)^2+\ldots+(2x_n-2M)^2}{n}} = \sqrt{\frac{4((x_1-M)^2+(x_2-M)^2+\ldots+(x_n-M)^2)}{n}} = 2\sqrt{\frac{(x_1-M)^2+(x_2-M)^2+\ldots+(x_n-M)^2}{n}}$$

Which is twice the SD of the first list.

Part e

True. If your list is $\{x_1, x_2, x_3, \ldots, x_n\}$ then the average is $\frac{x_1+x_2+x_3+\ldots+x_n}{n}$. After negating each entry the list would be $\{-x_1, -x_2, -x_3, \ldots, -x_n\}$ and the average would be:

$$\begin{aligned} \frac{-x_1 - x_2 - x_3 + \ldots - x_n}{n} &= \frac{-(x_1+x_2+x_3+\ldots+x_n)}{n} \\ &= -\frac{x_1+x_2+x_3+\ldots+x_n}{n} \end{aligned}$$

Which is the opposite of your previous average.

Part f

False. The SD is always positive, but to see this easily you can consider the list {-1, 1}. It has SD 1, and when you change the sign of each entry you get the list {1, -1}, which also has SD 1.

p95 # 9 (10 pts)

Part a

False. For example take the list {1,1,100}. Its median is 1, while its average is 34. I'd say 1 and 34 aren't close together.

Part b

False. For example consider the list {2, 50, 50}. Its average is 34, and only a third of the list is less than 34.

Part c

False. The income histogram we discussed in class has a long right tail, and does not follow the normal distribution. There are many other examples of data sets that don't follow the normal distibution as well.

Part d

False. Consider the two lists $L_1 = \{40, 60\}$ and $L_2 = \{50 - 10\sqrt{2}, 50, 50, 50 + 10\sqrt{2}\}$. They both have average 50 and SD 10, but in L_1 all of the list is between 40 and 60, while in L_2 only half of the list is.

Worksheet Q1 (10 pts)

Part a

Average age of husband: 48.18

Part b

Average age of wife: 45.59

Part c

Median income of husband: 27333

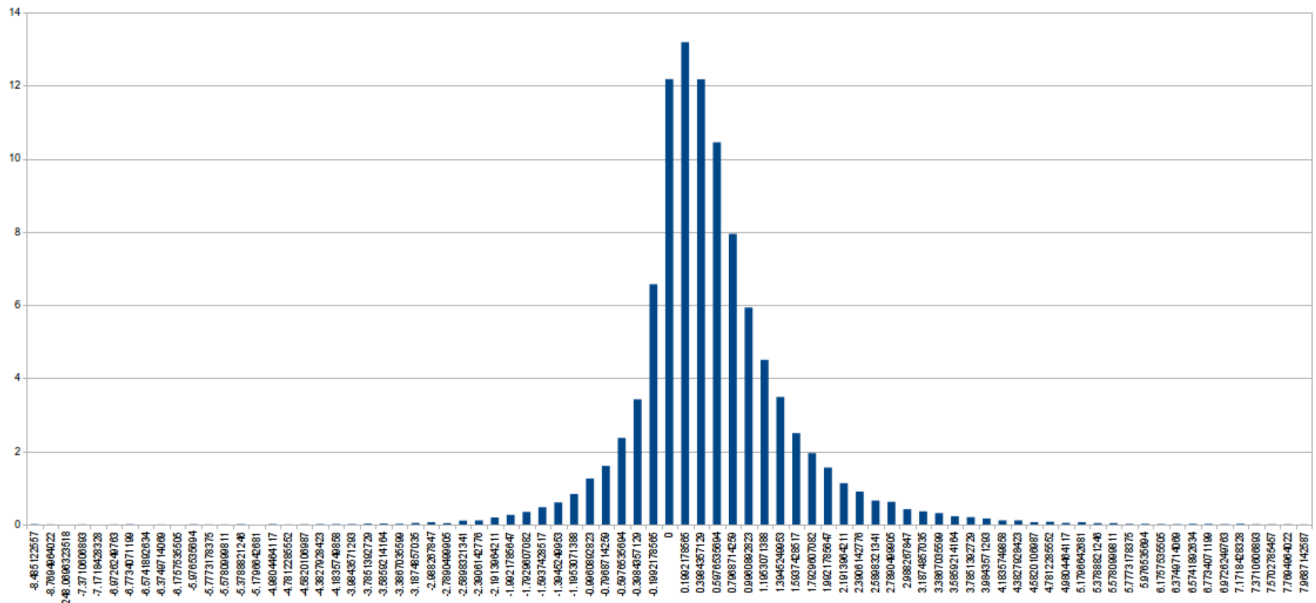
Part d

Median total household income: 41151

Part e

Maximum total household income: 449996

Worksheet Q2 (10 pts)



Worksheet Q3 (10 pts)

It sure looks like a normal curve, but by all metrics it fails to behave like one. Here are a few examples:

- Over 80% of the age differences are within 1 SD from the average
- The histogram is not centered about the line $x = 0$ in standard units.
- About 70% of the area in the histogram is to the right of 0.