STA 032 Homework 1

Hardy Jones
999397426
Professor Melcon
Fall 2014

§ 1.1

3
- i. False
- ii. True

8
- i. This interview is an observational study.
- ii. No, the conclusion is not well-justified given the information presented. It does not state how many were in each category. If there was only one person in the low exercise category, then that category is underrepresented.

§ 1.2

4 No, the sample median can differ from one of the values in the sample. For instance, the sample median of the following data, $D = \{1, 9\}$ is $\frac{1+9}{2} = \frac{10}{2} = 5 \notin D$.

6 Yes, it is possible for the standard deviation to be more than the mean for positive data. For instance:

x	$\overline{x}$	σ
$\{1, 9\}$	$\frac{1+9}{2} = 5$	$\sqrt{\frac{(1-5)^2+(9-5)^2}{1}} = \sqrt{(-4)^2 + 4^2} = \sqrt{16+16} = \sqrt{32}$

And $\sqrt{32} > 5$.

10 1)

$$\begin{aligned}
\overline{x} &= \frac{0(27) + 1(22) + 2(30) + 3(12) + 4(7) + 5(2)}{27 + 22 + 30 + 12 + 7 + 2} \\
&= \frac{22 + 60 + 36 + 28 + 10}{100} \\
&= \frac{156}{100} \\
&= 1.56
\end{aligned}$$

So of the sample, the mean number of children had was 1.56.

2)

$$\begin{aligned}
\sigma &= \sqrt{\frac{27(-1.56)^2 + 22(-0.56)^2 + 30(0.44)^2 + 12(1.44)^2 + 7(2.44)^2 + 2(3.44)^2}{99}} \\
&= \sqrt{\frac{27(2.4336) + 22(0.3136) + 30(0.1936) + 12(2.0736) + 7(5.9536) + 2(11.8336)}{99}} \\
&= \sqrt{\frac{65.7072 + 6.8992 + 5.8080 + 24.8832 + 41.6752 + 23.6672}{99}} \\
&= \sqrt{\frac{168.64}{99}} \\
&= \sqrt{1.70\overline{34}} \\
&\approx 1.31
\end{aligned}$$

So of the sample, the standard deviation is ≈ 1.31

3) Since there are 100 samples, we want to find the average of the 50^{th} and 51^{st} samples in ascending order. There are 27 samples of value 0 and 22 samples of value 1. Since there are 30 samples of value 2, the 50^{th} and 51^{st} samples are both 2.
So the median value is $\frac{2+2}{2} = 2$

4) Since there are 100 samples, the first quartile is the 25^{th} sample in ascending order. As there are 27 samples of value 0, the first quartile is 0.

5) The mean in 1.56, so the number of women that had more children than the mean is the total of samples with values 2, 3, 4, and $5 = 30+12+7+2 = 51$.
So 51 of the 100 women had more children than the mean.

6) The standard deviation is ≈ 1.31, so we want all samples where the value is greater than the mean plus the standard deviation. In other words we want all samples with value $> 1.56 + 1.31 = 2.87$. We want to total the number of samples with values 3, 4, and $5 = 12 + 7 + 2 = 21$
So 21 of the 100 women had more than one standard deviation's worth more children.

7) WAT!!

14 1) We can calculate the total from the information given, and use that total to calculate the change.

$$
\begin{aligned}
\text{total} &= \$70,000 \cdot 10 = \$700,000 \\
\overline{x_0} &= \frac{\text{total} - \$100,000 + \$1,000,000}{10} \\
&= \frac{\text{total} + \$900,000}{10} \\
&= \frac{\$700,000 + \$900,000}{10} \\
&= \frac{\$1,600,000}{10} \\
&= \$160,000
\end{aligned}
$$

The new mean is \$160,000.

2) Since only the largest datum was changed, the value of the median does not change.
The median is still \$55,000.

3) We can calculate the variance from the given information, and use that variance to calculate the change.

$$\sigma^2 = \frac{1}{n-1}\left[\left(\sum_i x_i^2\right) - n\overline{x}^2\right]$$

We want to initially find this term: $\sum_i x_i^2$, as the rest we already know.

$$
\begin{aligned}
\sigma^2 &= \frac{1}{n-1}\left[\left(\sum_i x_i^2\right) - n\overline{x}^2\right] \\
(n-1)\sigma^2 &= \left(\sum_i x_i^2\right) - n\overline{x}^2 \\
(n-1)\sigma^2 + n\overline{x}^2 &= \sum_i x_i^2 \\
9(20,000)^2 + 10(70,000)^2 &= \sum_i x_i^2 \\
9(400,000,000) + 10(4,900,000,000) &= \sum_i x_i^2 \\
3,600,000,000 + 49,000,000,000 &= \sum_i x_i^2 \\
52,600,000,000 &= \sum_i x_i^2
\end{aligned}
$$

Now that we have this value, we can perform the adjustment.

$$\begin{aligned}
\text{new sum} &= \sum_i x_i^2 - 100,000^2 + 1,000,000^2 \\
&= 52,600,000,000 - 10,000,000,000 + 1,000,000,000,000 \\
&= 1,042,600,000,000
\end{aligned}$$

Finally, we calculate our new standard deviation σ_0

$$\begin{aligned}
\sigma_0 &= \sqrt{\frac{1}{9}\left(\text{new sum} - 10\overline{x_0}^2\right)} \\
&= \sqrt{\frac{1}{9}\left(1,042,600,000,000 - 10(160,000)^2\right)} \\
&= \sqrt{\frac{1}{9}\left(1,042,600,000,000 - 256,000,000,000\right)} \\
&= \sqrt{87,400,000,000} \\
&\approx 295,634.91
\end{aligned}$$

So, the new standard deviation $\sigma_0 \approx \$295,634.91$