Statistics test

Fri. 17th Nov. 23

exercises taken from https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Exercises_(Introductory_Statistics)/Exercises%3A_Shafer_and_Zhang/02.E%3A_Descriptive_Statistics_(Exercises)

Exercise 1. No unit is given for the cholesterol levels, which a pity, but we will do without[1]. Let's order the ten given levels[2]:

$$132 \quad 133 \quad 139 \quad 145 \quad \boxed{147 \quad 148} \quad 150 \quad 153 \quad 160 \quad 162.$$

Since the total frequency 10 is even and is the double of $\boxed{5}$, the median level is the average of the $\boxed{5}$th and 6th levels , *i. e.* $\frac{147+148}{2} = 147.5$.

The mean level is the sum of all levels divided by the total frequency, *i. e.* one tenth of the sum[3]

$$\begin{aligned} & \underline{132} + \widehat{133} + 139 + \underbrace{145} + \widehat{147} + \underline{148} + 150 + \underbrace{153} + 160 + \underbrace{162} \\ = \; & \underline{280} + \widehat{280} + 140 - 1 + \underbrace{460} + 310 = 700 - 1 + 770 = 1470 - 1 = 1469. \end{aligned}$$

The sought-after tenth therefore equals 146.9.

An easier way would be to notice all scores are gathered around 150: so first shift the whole list by -150, compute the (easier!) mean, then shift back by $+150$. The shifted (ordered) list is

$$\widehat{-18} \quad \widehat{-17} \quad -11 \quad \underline{\underline{-5}} \quad \underline{\underline{-3}} \quad \underline{\underline{-2}} \quad 0 \quad \widetilde{3} \quad \underline{\underline{10}} \quad \widetilde{12}.$$

When adding up, the 4 underlined terms cancel, which leaves only $\widehat{-35} - 11 + \widetilde{15} = -20 - 11 = -31$. Then dividing by 10 and adding 150 yields $150 - 3.1 = 146.9$.

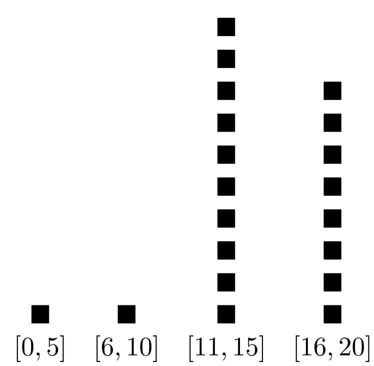
Last, the range is the difference between the maximal (rightmost) and minimal (leftmost) values, *i. e.* $162 - 132 = 30$.

Exercise 2. Again, no unit... :/ :/ :/ Since there are 20 numbers, the unit is probably *the number of* rice cookers sold each week of the 20 last (with such small numbers, amounts of money would most probably have digits after the dot, whatever the currency).

Let's re-order the sales list, with one line per given class[4]:

0									
9									
12	13	14	14	15	15	15	$\boxed{15}$	$\boxed{15}$	15
16	16	16	17	18	18	19	19		

(we have drawn a box around $\boxed{\text{the two middle values}}$ for the median sales – see below). Each class contains as many elements as the corresponding line, *i. e.* resp. 1, 1, 10 and 8, hence the frequency histogram:



[1] *challenge*: inquire by yourself which unit is most probably used

[2] we've drawn a box around the two middle levels

[3] symbols above or below the numbers indicate gatherings that make the (hand-)computation easier (unit digits add up to 10)

[4] don't ask why the wording excludes the 0 sales... so as to take it into account we decided to add the $0-5$ class (which has the same width as the others)

As for the median sales, since the total frequency is even (20) and is the double of $\boxed{10}$, it is the mean of the $\boxed{10}$th and 11th values, around which we've drawn a box above. Since they both equal 15, so does the median sales.

Last: *the mean sales.* Since the value 15 seems to be very frequent and quite around the center, let's take it as a new center to compute the mean sales. The shifted list (by -15) becomes

$$\begin{array}{cccccccc} -15 & \underline{-4} & \underline{-3} & \widehat{-2} & \widehat{-1} & \underline{\underline{-1}} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \underline{\underline{1}} & 1 & 1 & 2 & \widehat{3} & \underline{3 \quad 4} & 4. \end{array}$$

The values with graphics above/under add up to nil (same for the middle line of 0s), so the total sum boils down to $\overbrace{-15+1}+1+2\overbrace{+4} = -10+3 = -7$. Dividing by 20 (the total frequency) yields $\frac{-7}{20} = -\frac{3.5}{10} = -0.35$, then shifting back (by +15) yields the mean sales: 14.65 (which – comment – is slighty below the median).

Exercise 3. Let's compute, first, the total frequency, *i. e.* the total number of required tosses. We need only add the frequencies:

$$\begin{aligned} & \underline{384} + \underline{\underline{208}} + \overline{98} + \underline{56} + \widehat{28+12} + \underline{\underline{8}} + \overline{2} + \underline{\underline{3+1}} \\ = \; & \underline{440} + \underline{\underline{220}} + \overline{100} + \widehat{40} = 660 + 140 = 800. \end{aligned}$$

Since the latter sum is even and is the double of $\boxed{400}$, the median value is the mean of the $\boxed{400}$th and 401st *ordered* values. The first column deal with the (ordered) values up to the 384th and the first two columns with the first 384+208 ordered values: since we can compare

$$384 < \boxed{400} \qquad \text{and} \qquad 384 + 208 > 300 + 200 = 500 > \boxed{401},$$

wa can affirm both 400th and 401st values are dealt with by the 2d column and therefore equal 2. And so does their mean, hence the sought-after median.

As for the mean number of required tosses, it equals the sum of the values weighted by the relative frequencies. Since value 2 is central, let's compute the mean around this value. The sum of all (shifted) values becomes

$$\begin{aligned} & (-1)\,384 + 0 + 1\,(98) + 2\,(56) + 3\,(28) + 4\,(12) + 5\,(8) + 6\,(2) + 7\,(3) + 8\,(1) \\ = \; & \underline{-384} + \overline{98 + 112} + \underline{84} + \widehat{48} + 40 + \widehat{12} + 21 + 8 = \underline{-300} + \overline{210} + \widehat{60} + 69 \\ = \; & -90 + 129 = 39. \end{aligned}$$

Dividing by 800 (the total frequency) yields $\frac{39}{800} = 0.048\,75$, then shifting back (by +2) yields $2.04875 \simeq 2.05$ (which is slightly above the median value).