(Due Tuesday 01/29/2019 right before the class)

&

(Your homework shall be stapled if it contains multiple pages.)

SPRING/2019/MA526: HOMEWORK 1 SOLUTION

Instructor: Guangqu Zheng[1]; Grader: Chessa Mccalla[2]

Total points: 20

Q1 (4 points)

(1) Exercise 1.16 in textbook.

Proof: This follows from the definition of sample mean:

$$\sum_{i=1}^{n}(x_i - \overline{x}) = \left(\sum_{i=1}^{n} x_i\right) - \sum_{i=1}^{n}\overline{x} = n\overline{x} - n\overline{x} = 0\,.$$

(2) Verify

$$n\bar{x}^2 + \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i)^2$$

where $\bar{x}$ is the sample average of $x_1, \ldots, x_n$.

Proof: This follows from the definition of sample mean:

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n}(x_i^2 - 2x_i\overline{x} + \overline{x}^2)^2 = \left(\sum_{i=1}^{n} x_i^2\right) - \left(2\overline{x}\sum_{i=1}^{n} x_i\right) + \sum_{i=1}^{n}\overline{x}^2 = \left(\sum_{i=1}^{n} x_i^2\right) - \big(2n\overline{x}^2\big) + n\overline{x}^2$$

which equals

$$\left(\sum_{i=1}^{n} x_i^2\right) - n\overline{x}^2\,.$$

So the desired equality can be seen from an easy rearrangement of the terms.

Q2 ($4+3$ points) In a study conducted by the Department of Mechanical Engineering at Virginia Tech, the steel rods supplied by two different companies were compared. Ten sample springs were made out of the steel rods supplied by each company, and a measure of flexibility was recorded for each. The data are as follows:

Company A:	9.3	8.8	6.8	8.7	8.5
	6.7	8.0	6.5	9.2	7.0
Company B:	11.0	9.8	9.9	10.2	10.1
	9.7	11.0	11.1	10.2	9.6

(a) Calculate the sample mean and median for the data for the two companies.

(b) Compute the sample variance for the two companies. You may use the formula in Q1-(2).

Students need to show enough details in order to convince the grader that they understand the definition of sample average, median. For example, if student A gets the step (0.1) below,

[1] gzheng90@ku.edu; Office hours: TuTh 11:00-11:50; Office = 641 Snow Hall

[2] chessa_m@ku.edu

`but fails to find the right value` (0.2)`, student A would only get half of the point.`

Answer: It follows from the definition that the sample average of the data from company A, denoted by $\bar{x}_A$ is equal to

{step1}
$$\frac{1}{10}\big(9.3+8.8+6.8+8.7+8.5+6.7+8.0+6.5+9.2+7.0\big) \tag{0.1}$$
{value1}
$$= 7.95\,; \tag{0.2}$$

in the same way, the sample average of the data from company B, denoted by $\bar{x}_B$ is equal to

$$\frac{1}{10}\big(11.0+9.8+9.9+10.2+10.1+9.7+11.0+11.1+10.2+9.6\big)$$
$$= 10.26\,.$$

`In order to find the sample median, it will be helpful to first re-order the numbers in a monotone manner`[3].

reordering Company A	6.5	6.7	6.8	7.0	8.0	8.5	8.7	8.8	9.2	9.3
reordering Company B	9.6	9.7	9.8	9.9	10.1	10.2	10.2	11.0	11.0	11.1

so the median for A is $(8.0+8.5)/2 = 8.25$ and the median for B is $(10.1+10.2)/2 = 10.15$.

(b) Now we compute the sample variance for A and B. Recall first the formula:

$$\texttt{sample variance } s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\ .$$

In our question, $n = 10$ and we already got the sample averages $\bar{x}_A$, $\bar{x}_B$. Although we can simply plug our data into the above formula, let us rewrite the above formula as follows: since

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum_{i=1}^{n}\Big[(x_i)^2 - 2\bar{x}x_i + \bar{x}^2\Big] = \left(\sum_{i=1}^{n}(x_i)^2\right) - \left[2\bar{x}\sum_{i=1}^{n}x_i\right] + \sum_{i=1}^{n}\bar{x}^2$$

and $\sum_{i=1}^{n} x_i = n\bar{x}$ by the definition of sample average, then

{use1}
$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}(x_i)^2\right) - \frac{n}{n-1}\bar{x}^2\,. \tag{0.3}$$

So the sample variance for A, denoted by s_A^2, is equal to

$$\frac{1}{9}\big(9.3^2+8.8^2+6.8^2+8.7^2+8.5^2+6.7^2+8.0^2+6.5^2+9.2^2+7.0^2\big)-\frac{10}{9}\times 7.95^2 = \frac{642.89}{9}-\frac{632.025}{9} = \frac{10.865}{9} \approx 1.20722$$

(`Anything close to the above number can be seen as a ‘‘true" answer if the arguments towards it are ok.`) So the sample variance for B, denoted by s_B^2, is equal to

$$\frac{1}{9}\big(11.0^2+9.8^2+9.9^2+10.2^2+10.1^2+9.7^2+11.0^2+11.1^2+10.2^2+9.6^2\big)-\frac{10}{9}\times 10.26^2 = \frac{1055.6}{9}-\frac{1052.676}{9} = \frac{2.924}{9} \approx 0.3249$$

[3] in a non-decreasing order or in a non-increasing order, as you prefer!

Q3 (1 + 2 + 2 points) Fix an integer $n \geq 2$ and consider the sample x_1, ..., x_n.

(a) Recall the definition of sample average and median. Is sample median always bigger than sample average?

Ans: No. You can state any evidence to support your answer. For example, the previous question already gave us an example. Do you see this?

`Only writing down ‘‘ No ’’ without any explanation will not get the 1 point for this question (a).`

(b) If $f(x) = mx + c$ with $m \neq 0$ and c any given real number, then we call f a linear function. Putting $y_i = f(x_i)$, we get a new sample $y_1, \ldots, y_n$. What is the relation between the sample average of x_1, ..., x_n and the sample average of $y_1, \ldots, y_n$?

Ans: We know that $y_i = mx_i + c$ for $i = 1, \ldots, n$. So the sample average for the new data, denoted by $\bar{x}_{\text{new}}$ is equal to

$$\bar{x}_{\text{new}} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n}\big(mx_i + c\big) = c + m\frac{1}{n}\sum_{i=1}^{n} x_i = c + m\bar{x}$$

where $\bar{x}$ stands for the sample average associated to the old data $x_1, \ldots, x_n$.

Meanwhile, the sample variance for the new data, denoted by s^2_{new} is equal to

$$s^2_{\text{new}} = \frac{1}{n-1}\sum_{i=1}^{n}\big(y_i - \bar{x}_{\text{new}}\big)^2 = \frac{1}{n-1}\sum_{i=1}^{n}\Big(mx_i + c - \big[c + m\bar{x}\big]\Big)^2 = \frac{1}{n-1}\sum_{i=1}^{n}\Big(m\big[x_i - \bar{x}\big]\Big)^2 = m^2 s^2$$

where s^2 stands for the sample variance associated to the old data $x_1, \ldots, x_n$.

(c) Following question (b): What is the relation between the sample variance of x_1, ..., x_n and the sample variance of $y_1, \ldots, y_n$?

Ans: The sample variance for the new data, denoted by s^2_{new} is equal to

$$s^2_{\text{new}} = \frac{1}{n-1}\sum_{i=1}^{n}\big(y_i - \bar{x}_{\text{new}}\big)^2 = \frac{1}{n-1}\sum_{i=1}^{n}\Big(mx_i + c - \big[c + m\bar{x}\big]\Big)^2 = \frac{1}{n-1}\sum_{i=1}^{n}\Big(m\big[x_i - \bar{x}\big]\Big)^2 = m^2 s^2$$

where s^2 stands for the sample variance associated to the old data $x_1, \ldots, x_n$.

Q4 (4 points) Let Ω be our sample space in this question. It models all possible outcomes from our (imaginary) experiment, so it is a nonempty set by default. Let A, B, C be subsets of Ω, written as $A \subset \Omega$, $B \subset \Omega$ and $C \subset \Omega$. Are the following statements true or false? Justify your response carefully.

(i) If $A \subset B$, then $A^c \subset B^c$. (A^c stands for the complement of the set A.)

False. If $\Omega = \{1, 2\}$, $A = \{1\}$, $B = \Omega$, then $A^c = \{2\}$ and $B = \emptyset$. This gives us a counterexample.

(ii) If $A \subset B$, then $A \cap C \subset B \cap C$.

True: If $A \cap C$ is empty, then the above statement is true; if it is not empty and consider any element x inside this set, we have $x \in A$ and $x \in C$. Since A is a subset of B, so x also belongs to B. Then we have $x \in B \cap C$. So $A \cap C \subset B \cap C$.

(iii) If $A \cup C \subset B \cup C$, then $A \subset B$.

False, see the following counterexample: $\Omega = \{1, 2, 3\}$, $A = \Omega$, $B = \{1\}$ **and** $C = \{2, 3\}$.

(iv) If $A \cap C \subset B \cap C$, then $A \subset B$.

False, see the following counterexample: $\Omega = \{1, 2, 3\}$, $A = \{1, 2\}$, $B = \{1\}$ **and** $C = \{3\}$.