Problem Set 1 Solutions

ECON 340: Economic Research Methods — Instructor: Div Bhagia

1. (1 pt) The statement is <u>true</u> because

$$\begin{aligned}\sum_{i=1}^{n}(3X_i^2 + 2k) &= \sum_{i=1}^{n} 3X_i^2 + 2nk \\ &= 3\sum_{i=1}^{n} X_i^2 + 2nk \\ &= 3\sum_{j=1}^{n-1} X_j^2 + 3X_n^2 + 2nk\end{aligned}$$

   The last step follows from the fact that

$$\sum_{i=1}^{n} X_i^2 = \underbrace{X_1^2 + X_2^2 + \ldots + X_{n-1}^2}_{\sum_{i=1}^{n-1} X_i^2} + X_n^2$$

   So we can re-write it as

$$\sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n-1} X_i^2 + X_n^2$$

   Also, since i and j are just indexes for each observation, it doesn't matter what letter we use.

2. (1 pt) Fill in the following frequency distribution table.

X_i	Relative Frequency	Cumulative Frequency
1	**0.1**	0.1
2	**0.2**	0.3
3	0.4	**0.7**
4	**0.3**	**1**
Total	**1**	X

3. (2 pts) Data on calorie consumption of five adults is given:

$$1400, 1800, 2000, 2200, 2600$$

(a) The mean calorie consumption is given by

$$\bar{x} = \frac{\sum_{i=1}^{5} X_i}{5} = \frac{1400 + 1800 + 2000 + 2200 + 2600}{5}$$
$$= \frac{10000}{5} = 2000$$

(b) Since $n = 5$ is odd, the median calorie consumption for this group is given by $\frac{n+1}{2}$ th term which is 2000.

We added one more individual to the data set whose calorie consumption is 4100.

(c) Since we added another individual to the data

$$\bar{x}_{new} = \frac{\sum_{i=1}^{6} X_i}{6} = \frac{10000 + 4100}{6} = \frac{14100}{6} = 2350$$

Aside: If we were not given the calorie consumption of the first five individuals but just their mean $\bar{x}_{old} = 2000$, we could still find the new mean with the addition of one more person. How?

$$\bar{x}_{new} = \frac{\sum_{i=1}^{6} X_i}{6} = \frac{\sum_{i=1}^{5} X_i + 4100}{6} = \frac{5\bar{x}_{old} + 4100}{6}$$

(d) Now $n = 6$ is even so the median calorie consumption is average of $\frac{n}{2}$th and $\frac{n}{2} + 1$th term. So the median is

$$\frac{2000 + 2200}{2} = 2100$$

(e) Mean is more susceptible to outliers than the median. This is because every observation contributes to the mean with a weight of $\frac{1}{n}$. So if an observation is disproportionately larger (smaller) than others, it can pull the mean upwards (downwards). Whereas the median only depends on observations in the center of the data.

4. (1 pt) Here is the amount (in $) that I spent on groceries in the last three weeks:

$$100, 120, 80$$

To calculate the variance, we first need the average spending on groceries in the last three weeks:

$$\mu = \frac{\Sigma_{i=1}^{3} X_i}{3} = \frac{100 + 120 + 80}{3} = 100$$

Now to calculate the variance:

$$\begin{aligned}\sigma^2 &= \frac{\Sigma_{i=1}^{3}(X_i - \mu)^2}{3} \\ &= \frac{(100-100)^2 + (120-100)^2 + (80-100)^2}{3} \\ &= \frac{0 + 400 + 400}{3} = 266.67\end{aligned}$$

5. (3 pts) The following table is constructed from a sample of 6 students. X_i represents the number of hours an individual usually sleeps and Y_i represents the number of hours the individual typically exercises per week.

Obs	X_i	Y_i	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
1	8	3	0.5	-1	0.25	1	-0.5
2	7	4	-0.5	0	0.25	0	0
3	6.5	2	-1	-2	1	4	2
4	7.5	4	0	0	0	0	0
5	9	6	1.5	2	2.25	4	3
6	7	5	-0.5	1	0.25	1	-0.5
Total	**45**	**24**	**0**	**0**	**4**	**10**	**4**

(a) What is the variance of X and Y?

We can use the formula for the sample variance:

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

From the above table, we already have $\sum_{i=1}^{n}(X_i - \bar{X})^2 = 4$. Here $n = 6$, so $n - 1 = 5$. Plugging this in the formula we get:

$$S_X^2 = \frac{4}{5} = 0.8 \qquad S_Y^2 = \frac{10}{5} = 2$$

(b) What is the standard deviation of X and Y?

Standard deviation is given by the square root of the variance. So we have:

$$S_X = \sqrt{S_X^2} = \sqrt{0.8} = 0.89$$

$$S_Y = \sqrt{S_Y^2} = \sqrt{2} = 1.41$$

(c) How many standard deviations is the fifth observation away from the average hours of sleep?

The average hours of sleep is given by:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{45}{6} = 7.5$$

The fifth observation is $9 - 7.5 = 1.5$ hours away from the average. The standard deviation for hours of sleep is 0.89. So the fifth observation is $1.5/0.89 = 1.68$ standard deviations away from the mean.

(d) What is the covariance between hours of sleep per night and hours of exercise per week?

Formula for sample covariance:

$$S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

Since we have $\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = 4$ from the table.

$$S_{XY} = \frac{4}{5} = 0.8$$

(e) What is the correlation between X and Y?

$$r_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{0.8}{0.89 \times 1.41} = 0.64$$

6. (2 pts) For this question, we can use the formulas to calculate mean and variance for grouped data.

X_k	n_k	f_k	$f_k X_k$	$(X_k - \bar{X})^2$	$f_k(X_k - \bar{X})^2$
1	200	0.2	0.2	0.64	0.128
0	800	0.8	0	0.04	0.032
Total	1000	1	0.2	NA	0.16

So we can calculate the mean and variance of X as follows:

$$\bar{X} = \sum_{k=1}^{K} f_k X_k = 0.2$$

$$S_X^2 = \frac{n}{n-1} \sum_{k=1}^{K} f_k (X_k - \bar{X})^2 = \frac{1000}{999} \cdot 0.16 \approx 0.16$$

The variance without sample correction would be given by:

$$\sigma_X^2 = \sum_{k=1}^{K} f_k (X_k - \mu)^2 = 0.16$$