

DALITE Q1 - Histograms, Medians and Means Solutions.

EPIB607 - Inferential Statistics^a

^aFall 2018, McGill University

This version was compiled on September 12, 2018

This DALITE quiz will cover the basic concepts of histograms, means and medians.

Histogram | Density Plot | Mean | Median | Mode

1. Measures of Center Q2

The following link <https://www150.statcan.gc.ca/n1/pub/84-537-x/2013005/tbl/tbl1b-eng.htm> contains the Complete life table for females in Canada from 2009 to 2011. For a given age x , column l_x gives the number alive (the cohort starts with 100,000 people) and column d_x gives the number of deaths (this column adds up to 100,000 people). The mean age of death is 83.60. What percentage of people live longer than the mean?

- a) 50%
- b) **More than 50% (Correct)**
- c) Less than 50%

1.1. Correct rationales.

- At 84 years of age there are 58,833 people alive which is more than 50% of the population as the cohort started with 100,000 people.
- At the mean age of death of 83 years old, more than half the population (50,000) is still alive.
- Looking at the number of people alive, at 84 we can see that it is roughly 58,000. Considering we started with a population of 100,000 this is more than 50%.

1.2. Incorrect rationales.

- Because that is the definition of the mean
- The mean is the center of the distribution of deaths. Therefore exactly 50% of people live longer than the mean.
- Usually people live longer than the mean
- Mean represents average value, so if on average people die at 83.6, less than 50% will live after

2. Measures of Center Q3

When a group of statisticians from Concordia moved to McGill, they raised the average intelligence in both universities. Is this possible? (answer YES or NO and explain).

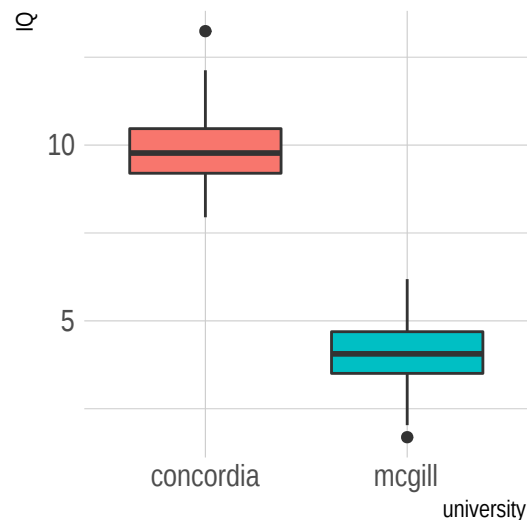
- a) **YES (Correct)**
- b) NO

2.1. Correct rationales.

- Technically this is possible, if the statisticians have lower intelligence than the overall concordia population, and higher intelligence than the overall McGill population, they would raise the average at both universities.
- If the least smartest statisticians from Concordia became the smartest statisticians at another school like McGill, then yes, this is possible.
- Yes because of knowledge sharing between the groups from Concordia and from McGill. They can then share their new knowledge with their respective universities resulting in raising the average intelligence in both universities.

2.2. Incorrect rationales.

- To raise the intelligence at McGill means that the Concordia statisticians must be really smart. If they are really smart, they are likely at the top of the distribution of intelligence at Concordia. If they leave, they will surely decrease the average.
- The average(mean) is pulled to the lower tail, meaning it is affected by the extreme numbers in the dataset. If the average needs to be increased, people with relatively high IQ needed to be brought to each university.
- Since the number of observations are decreasing in Concordia, average intelligence cannot increase in both schools



3. Histograms Q3

Six sketches are shown below and labelled (i) to (vi) going from left to right, top to bottom. Four of them are histograms for the following variables (in a study of a small town):

- a) heights of all members of households where both parents are less than 24 years old
- b) heights of married couples
- c) heights of all people
- d) heights of all automobiles

Match each one of the four variables (a)-(d) with the histograms shown in the figure below.

- a) (a) – (iv), (b) – (ii), (c) – (iv), (d) – (i)
- b) (a) – (ii), (b) – (iv), (c) – (i), (d) – (iv)
- c) (a) – (ii), (b) – (iii), (c) – (iv), (d) – (i) **(Correct)**
- d) (a) – (ii), (b) – (vi), (c) – (v), (d) – (i)

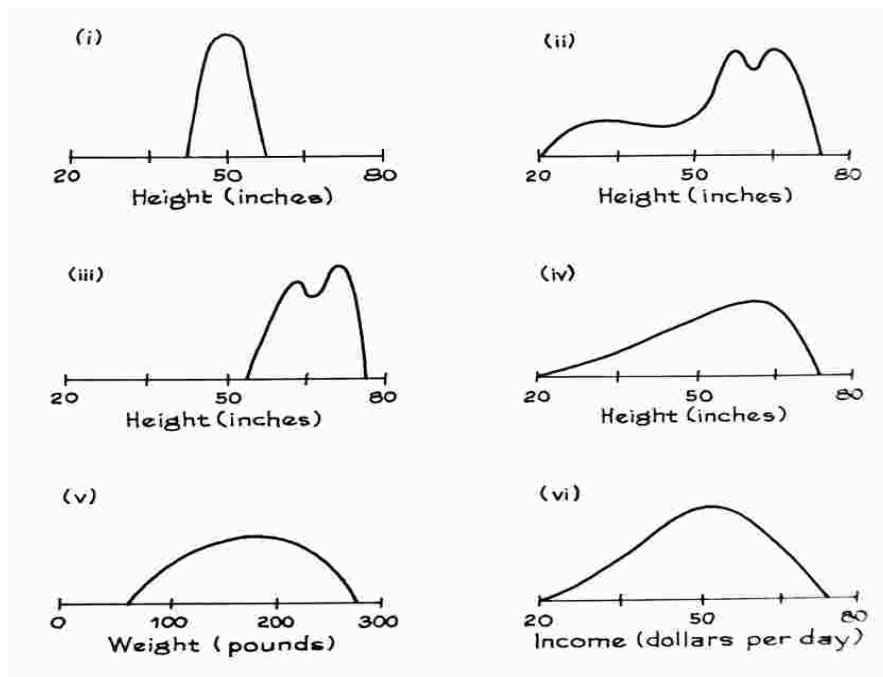


Fig. 1. Histograms Q3

3.1. Correct rationales.

- Histograms v and vi are irrelevant because of their x-axis variable. a(ii) because a there are many different heights which is expected for a family with people from different ages including children. b(iii) married couples are adults and on this histogram you only have heights above 50 inches. c(iv) is plausible because height in the general population should be more or less normally distributed. d(i) would be for automobiles because of the narrow range of heights.

3.2. Incorrect rationales.

- Heights of all people → (iv) because it shows the incremental growth of the population.

4. Histogram Q2

Below are sketches of histograms for test scores in three different classes. The scores range from 0 to 100; a passing score was 50. For each class, was the percent who passes about 50%, well over 50%, or well under 50% ?

- (a) well over 50%, (b) about 50%, (c) about 50%
- (a) about 50%, (b) well under 50%, (c) about 50%
- (a) **well over 50%**, (b) **well under 50%**, (c) **about 50%** (Correct)
- (a) well over 50%, (b) about 50%, (c) about 50%

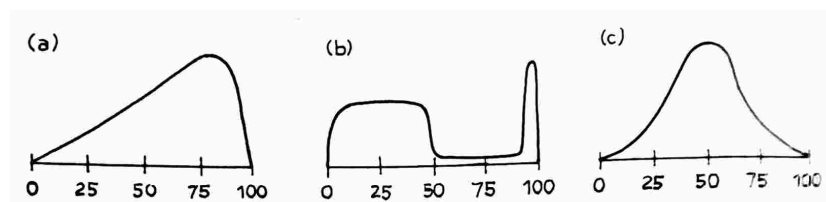


Fig. 2. Histogram Q2

4.1. Correct rationales.

- (a) the density of the plot is greatest above 50, (b) the density of the plot is greatest below 50, (c) the density centres around 50
- a has an area under the curve where more people get higher than 50% b has an area under the curve where more people get lower than 50 % c has an area under the curve where about half of the people get above 50% and about half of people get 50%

4.2. Incorrect rationales.

- (b) high number of people with high scores, so may be about 50
- (b) about 50% there seems to be a high spike of people that got almost 100 which would bring the % of people that passed closer to 50%

5. Histogram Q1

An investigator collects data on hourly wage rates for three groups of workers. Workers in group B earn about twice as much as those in group A; workers in group C earn about \$10 an hour more than those in group A. Which histogram belongs to which group?

- a) A(ii), B(i), C(iii) (Correct)
- b) A(ii), B(iii), C(i)
- c) A(i), B(ii), C(iii)
- d) A(iii), B(i), C(ii)

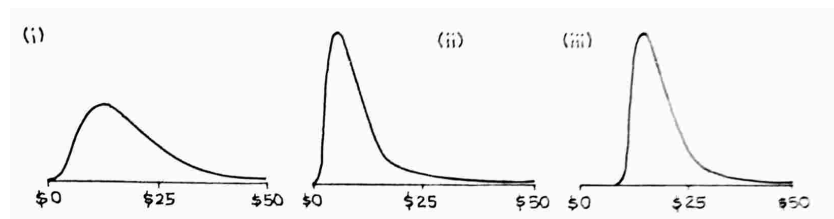


Fig. 3. Histogram Q1

5.1. Correct rationales.

- workers in group C earn about \$10 more than A which means that they will have the same histogram except it will be shifted \$10 to the right
- i) is group B because the density of the graph is more spread out indicating a wider spread of wage. ii) is group A because the largest amount of people make the least amount of money. iii) is group C because the graph is shifted over to the left indicating the the most amount of people make a higher wage than those of group A.

5.2. Incorrect rationales.

- (iii) is shifted to the right therefore workers receive double that of the pay of workers in (ii). Using \$25 as a reference point, we find that workers in (i) receive more (but not double) that of workers in (ii). (i) is also distributed more widely as opposed to having a peak.
- The right shift in the overall histogram (from ii to iii) indicates that group B earns twice as much as A.
- Twice as much = histogram higher 10 dollars more = histogram shifted away from zero-point

6. Measures of center Q1

In a salary dispute involving professional hockey players, the owners claimed that the average player's salary was \$1 million, while the players association (NHLPA) said it was 'only' \$500k. How can this be? (assume that these calculations were from the same database)

- a) This is impossible
- b) The owners were actually talking about the median salary while the NHLPA was referring to the mean salary.
- c) **The owners were actually talking about the mean salary while the NHLPA was referring to the median salary. (Correct)**

6.1. Correct rationales.

- Some players earn a very large salary, raising the mean, but most players earn a more modest salary resulting in the lower median.
- The owners are most likely referring to the mean because the mean includes outliers and with a few players in the NHL collecting considerably more money, the owners are most likely going to include this in their interpretation of the average player salary, which will skew the average towards 1 million. On the other hand, the players are most likely referring to the median because they probably looked at the middle salary entry in the database.
- This is possible as the mean is sensitive to very high/extreme values. It is likely that in a sport such as hockey, a small number of players are making well above the mean or median which would skew the mean value to higher. However because the median simply reports the value at which 50% of players earn more and less, it is not affected by what outliers at the top would make. This means that both could be right if they used different measures of central tendency

6.2. Incorrect rationales.

- The distribution is skewed to the left since hockey players would be making top salaries, so it would make sense the the mean is lower than the median.
- Since this is a high paying profession, the graph will be skewed to the left. This means that the mean is smaller than the median. Therefore, it is possible that the players association has the correct mean of \$500k, but that the owners who want their players to be paid more and would use the median of \$1 million.
- NHLPA was talking about mean/average salary which is lower than the median salary (referred by the owners) since few players had significantly lower salaries than others which had led to the dispute.