Tutorial 2

CB2200 Business Statistics, YANG Yihang

September 15, 2020

Intended Learning Outcomes

After this tutorial, you may know how to

- differentiate sample with population
- compute several statistics featuring on central tendency, variation, and shape.
- form a box-and-whisker plot.

Recap

- Central tendency
  - population mean: $\mu = \frac{1}{N}(\sum_{i=1}^{N} x_i)$
  - sample mean: $\bar{x} = \frac{1}{n}(\sum_{i=1}^{n} x_i)$
  - Median $=$ the middle value in the ordered array
  - Mode $=$ the most frequently observed value
- Variation
  - Range $= X_{largest} - X_{smallest}$
  - IQR (Interquartile Range)$=$ Q3-Q1
  - Variance
    - Population variance: $\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$
    - Sample variance: $S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$
  - Standard deviation
    - Population standard deviation: $\sigma = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}}$
    - Sample standard deviation: $S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$

Recap

- Shape
  - Skewness:
    - Left-skewed: mean$<$ median
    - Symmetrical: mean $=$ median
    - Right-skewed: mean $>$ median
  - Boxplot

Question 7

- The following is a set of data for a Population of size $N = 10$:
  7 5 11 8 3 5 2 1 10 8
  a) Compute the population mean.
  b) Compute the population standard deviation

Question 7

- a) Compute the population mean.
  $\mu = \frac{1}{N}(\sum_{i=1}^{N} x_i) = 6$
- b) Compute the population standard deviation
  $\sigma = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}} = 3.1937$

Question 8

- A food inspector, examining 10 bottles of a certain brand of honey, obtained the following percentages of impurities:
  23.5 19.8 21.3 22.6 19.4 18.2 24.7 21.9 20.0 21.1
  What are the mean and standard deviation of this sample?

Question 8

Q8

$$\text{Mean, } \bar{X} = \frac{23.5+19.8+21.3+22.6+19.4+18.2+24.7+21.9+20.0+21.1}{10} = 21.25$$

$$\text{Standard Deviation, S} = \sqrt{\frac{(23.5-21.25)^2+...+(21.1-21.25)^2}{10-1}} = 1.9896$$

Question 9

- Q9
  The data contain the price for two tickets with online service charges, large popcorn, and two medium soft drinks at a sample of six theater chains:
  \$36.15 \$31.00 \$35.05 \$40.25 \$33.75 \$43.00
  a) Compute the mean and median
  b) Compute the variance, standard deviation and range
  c) Are the data skewed? If so, how?
  d) Based on the results of (a) through (c), describe the data.

Question 9:Quantile calculation

a)Steps:

- 1. Write the data points in order.
- 2. Using the formulas below to calculate
  Q1 position $= \frac{n+1}{4}$
  Q2 position $= \frac{2(n+1)}{4}$
  Q3 position $= \frac{3(n+1)}{4}$
- 3. Identify the categories that the result belongs to.
  (i)Integers: We end up with it.
  (ii) A fractional half(e.g. 2.5, 8.5): We take the average of the two corresponding data values.
  (iii) Neither (i) nor (ii): We round the result to the nearest integer to find the ranked position.

Question 9

Q9

a) Mean, $\overline{X} = 36.53$
Median = 35.6

b) Variance, $S^2 = 19.27$
Standard Deviation, S = 4.39
Range = 43 – 31 = 12

c) Since the mean is only slightly greater than the median, the data are slightly right-skewed

d) There is a \$12 difference between the most expensive and the least expensive outlet. The prices vary around \$36.53 with half of the outlets being more expensive than \$35.6. The middle half of the prices fall between \$33.75 and \$40.25.

- The data contains the total fat, in grams per serving, for a sample of 20 chicken sandwiches from fast-food chains. The data are as follows:
  7 8 4 5 16 20 20 24 19 30
  23 30 25 19 29 29 30 30 50 56
  a) Compute the first quartile (Q1), the third quartile (Q3), and the interquartile range.
  b) List the five-number summary.
  c) Construct a boxplot and describe the shape.

Question 10

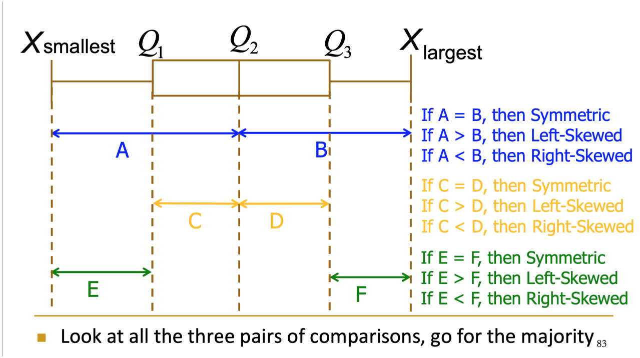
b)

- $X_{Smallest}$
- $Q1$
- $Median$
- $Q3$
- $X_{Largest}$

c) Box-and-whisker plot
Steps:

- 1. Make sure that five numbers are ready.
- 2. Make the x-axis consistent with the maximum.
- 3. Mark those five numbers on the x-axis
- 4. Draw it formally.

Skewness

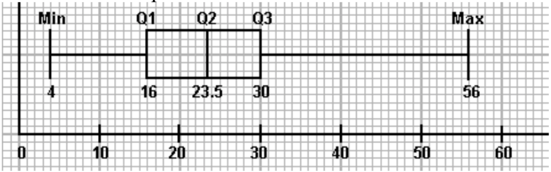

$X_{smallest}$
Q_1
Q_2
Q_3
$X_{largest}$
A
B
C
D
E
F
If A = B, then Symmetric
If A > B, then Left-Skewed
If A < B, then Right-Skewed
If C = D, then Symmetric
If C > D, then Left-Skewed
If C < D, then Right-Skewed
If E = F, then Symmetric
If E > F, then Left-Skewed
If E < F, then Right-Skewed
Look at all the three pairs of comparisons, go for the majority
83

Q10

a) $Q_1 = 5.25$ th obs ~5 th obs = 16
$Q_3 = 15.75$ th obs ~16 th obs = 30
Interquartile range = $Q_3 - Q_1 = 30 - 16 = 14$

b) Minimum = 4 | 1st Quartile = 16 | Median = 23.5
3rd Quartile = 30 | Maximum = 56

c) Box-and-whisker plot



The amount of fat is right-skewed.

Question 11

- The following data is the number of vitamin supplements sold by a health food store in a sample of 11 days:
  19 19 20 20 20 22 23 25 26 27 30
  a) What are the average and standard deviation of daily sale of vitamin supplements of the health food store?
  b) Work out a five-number summary of the data in the sample. Comment on the distribution of the sample data.

Question 11

Q11

a) sample mean, $\bar{X}$ = (19+19+20+20+20+ 22+ 23+ 25+ 26+ 27+ 30) /10 = 22.8182

sample standard deviation, S = 3.7099

b) Min. value = 19
Q1 = value of ((11+1)/4)-th obs = 3rd obs = 20
Q2 = value of (2*(11+1)/4)-th obs = 6th obs = 22
Q3 = value of (3*(11+1)/4)-th obs = 9th obs = 26
Max. value = 30

Five-number summary: 19, 20, 22, 26, 30
Since (Q3-Q2) > (Q2-Q1), the distribution is right-skewed.

Question 12

Q12

A bank branch located in a commercial district of a city has developed an improved process for serving customers during the 12:00 to 1 p.m. peak lunch period. The waiting time in minutes (operationally defined as the time the customer enters the line to the time he or she is served) of all customers during this hour is recorded over a period of a week. A random sample of 15 customers is selected, and the results are as follows:

4.21	5.55	3.02	5.13	4.77	2.34	3.54	3.20
4.50	6.10	0.38	5.12	6.46	6.19	3.79	

Another branch located in a residential area is most concerned with the Friday evening hours from 5 to 7 p.m. The waiting time in minutes (operationally defined as the time the customer enters the line to the time he or she is served) of all customers during these hours is recorded over a period of a week. A random sample of 15 customers is selected, and the results are as follows:

9.66	5.90	8.02	5.79	8.73	3.82	8.01	8.35
10.49	6.68	5.64	4.08	6.17	9.91	5.47	

a) For each bank branch, compute the mean, median and interquartile range.
b) Form the box-and-whisker plot and describe the shape of the distribution of waiting time at the two bank branches.
c) Compare and contrast the distributions of the waiting time at the two bank branches.

Question 12

Q12

a) Bank branch in commercial district:

Mean, $\bar{x} = \dfrac{\sum X}{n} = 4.287$ mins

Median, $Q_2 = \dfrac{n+1}{2}$ th obs. $= \dfrac{15+1}{2}$ th obs $=8$ th obs $= 4.5$ mins

$\because Q_1 = \dfrac{n+1}{4}$ th obs$=$ 4th obs $= 3.2$ mins $\quad Q_3 = \dfrac{3(n+1)}{4} = 12$ th obs $= 5.55$ mins

$\therefore$ interquartile range $= Q_3 - Q_1 = 5.55 - 3.20 = 2.35$ mins

Bank branch in residential area:

Mean, $\bar{x} = \dfrac{\sum X}{n} = 7.115$ mins

Median, $Q_2 = \dfrac{n+1}{2}$ th obs.$= \dfrac{15+1}{2}$ th obs $=8$ th obs $= 6.68$ mins

$\because Q_1 = \dfrac{n+1}{4}$ th obs$=$ 4th obs $= 5.64$ mins $\quad Q_3 = \dfrac{3(n+1)}{4} = 12$ th obs $= 8.73$ mins

$\therefore$ interquartile range $= Q_3 - Q_1 = 8.73 - 5.64 = 3.09$ mins

b) Commercial: min=0.38, Q_1=3.20, Q_2=4.50, Q_3=5.55, max=6.46, left-skewed

Residential: min=3.82, Q_1=5.64, Q_2=6.68, Q_3=8.73, max=10.49, right-skewed

c) The central tendency of the waiting time for the bank branch located in the commercial district is lower than that of the branch located in residential area. Also, the normal waiting time for residential area is longer than that of commercial area.