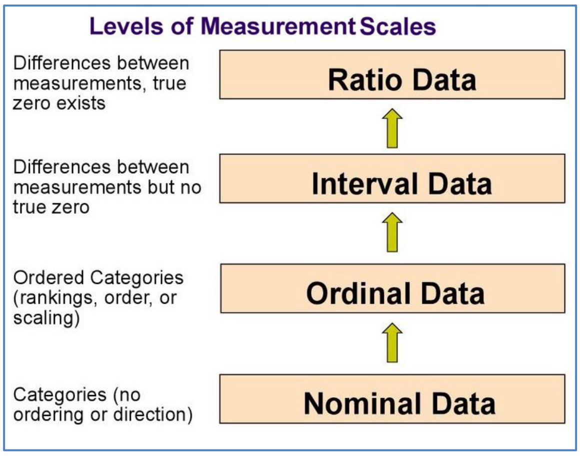
CE/CZ 2100 Prob & Stat

Solution to Tutorial #1

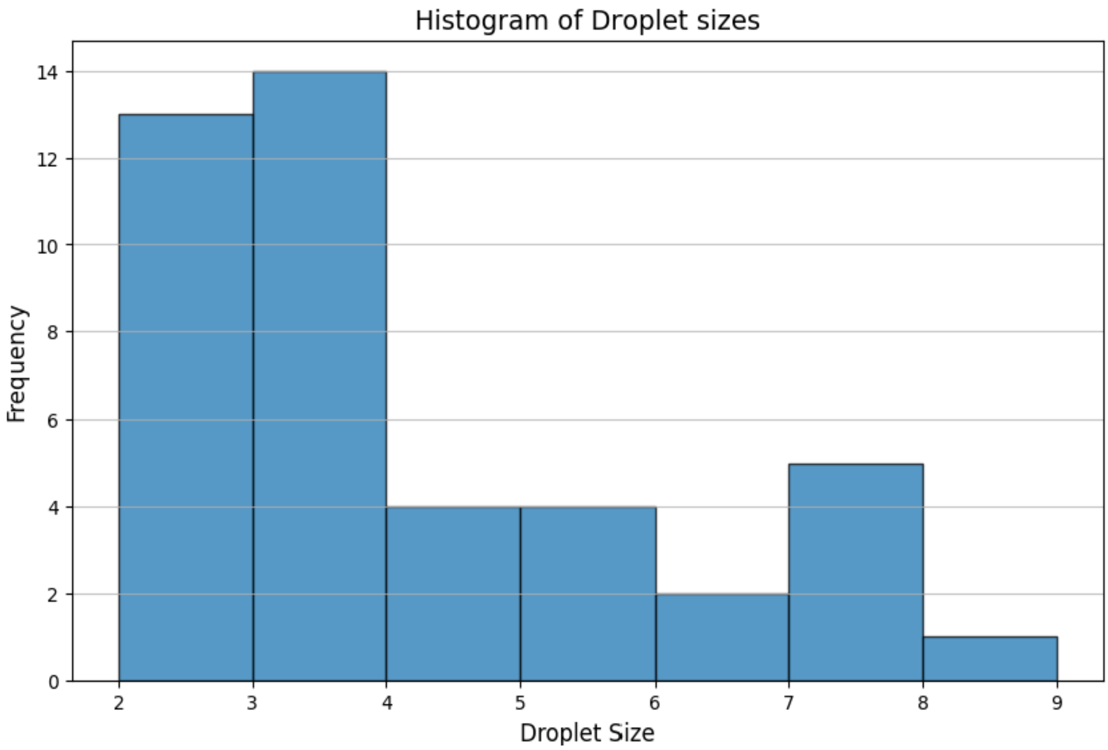
Descriptive Statistics



1. (a) qualitative, nominal
   (b) quantitative, ratio
   (c) qualitative, ordinal
   (d) quantitative, interval
   (no true 0 for intelligence)
   (e) qualitative, nominal

2. (a)

Class	[2,3)	[3,4)	[4,5)	[5,6)	[6,7)	[7,8)	[8,9)
Frequency	13	14	4	4	2	5	1



This distribution has a positive skew, i.e. skewed to the right.

(b) Stem-and-Leaf:

8 |9
7 |18999
6 |01

5 |1337
4 |0259
3 |01123334566677
2 |1223345558899

(c) mean = 178.4/43 = 4.15

For percentile, the first step is to compute the rank (R) of the P^{th} percentile using the formula:

$$R = (P/100) \times (N - 1) + 1$$

where P is the desired percentile (eg: 25th) and N is the number of data (i.e. 41). Therefore,

$$R = (25/100) \times (43 - 1) + 1 = 11.5$$

Name	Formula	Value

If R is an integer, then P^{th} percentile is the data with rank R. When R is not an integer, we compute the P^{th} percentile by interpolation as follows:

1. Define I_R as the integer portion of R (i.e. largest integer smaller than R). For this eg, $I_R = 11$.
2. Define F_R as the fractional portion of R (i.e. $F_R = R - I_R$). For this example, $F_R = 0.5$.
3. Find the data with Rank I_R and with Rank $I_R + 1$. For this example, the data with Rank 11 = 2.8 and the data with Rank 12 = 2.9.
4. Interpolate by multiplying the difference between the data by F_R and add the result to the lower data. For these data, this is (0.5)(2.9 – 2.8) + 2.8 = 2.85 (i.e. 25^{th} percentile).

Similarly, for 50^{th} and 75^{th} percentile, we have the following:

50th percentile = 3.5 + (3.6 − 3.5) × 0 = 3.5,
since R = 50 /100 × (43 - 1)+1 = 22,
$I_R = 22$, $F_R = 0$ and values are arranged in ascending order.

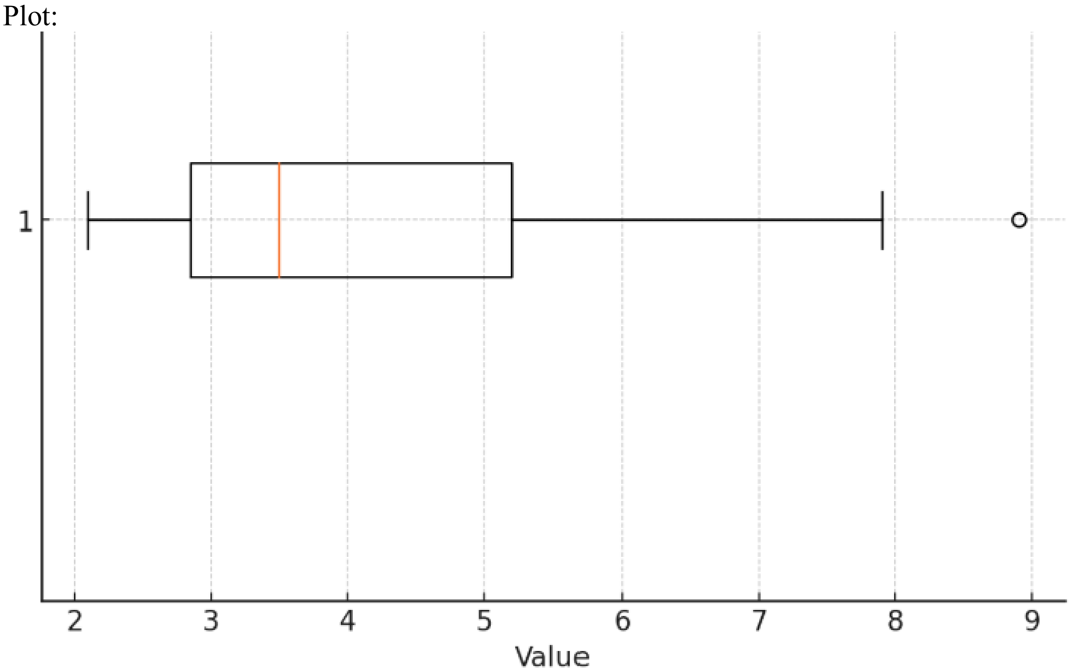
75th percentile = 5.1 + (5.3 − 5.1) × 0.5 = 5.2,
since R = 75/100 × (43 - 1) + 1 = 32.5, $I_R = 32$, $F_R = 0.5$ and values are arranged in ascending order.

(d) Construction of Box Plot according to the following formulas:

Upper Hinge	75th Percentile	5.2
Lower Hinge	25th Percentile	2.85
H-Spread	Upper Hinge - Lower Hinge	2.35
Step	1.5 x H-Spread	3.525
Upper Inner Fence	Upper Hinge + 1 Step	8.725
Lower Inner Fence	Lower Hinge - 1 Step	0*
Upper Outer Fence	Upper Hinge + 2 Steps	12.25
Lower Outer Fence	Lower Hinge - 2 Steps	0
Upper Adjacent	Largest value below Upper Inner Fence	7.9
Lower Adjacent	Smallest value above Lower Inner Fence	2.1
Outside Value	A value beyond an Inner Fence but not beyond an Outer Fence	8.9
Far Out Value	A value beyond an Outer Fence	None

* the minimum value for the droplet size is 0, i.e., it can not be a negative number.

Box Plot:



3.

Class interval	Middle value x_i	frequency f_i
[0, 2)	1	10
[2, 4)	3	25
[4, 6)	5	20
[6, 8)	7	40
[8, 10)	9	5
Totals		100

mean $\bar{x} = \frac{1}{n}\sum x_i f_i = \frac{510}{100} = 5.1$ (min)

mode = 7 (with the highest frequency = 40)

4. $$s^2 = \frac{1}{n-1}\left[\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right] = \frac{1}{9}\left[500 - \frac{(50)^2}{10}\right] = 27.78; \quad \text{so, } s = \sqrt{27.78} = 5.27$$

If $\sum x^2 = 100$, then $s^2 = \frac{1}{9}\left[100 - \frac{(50)^2}{10}\right] = -16.67$; Impossible to have $s^2 < 0$

5. Using the formula: $\rho = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\sum X^2 - \frac{(\sum X)^2}{N}}\sqrt{\sum Y^2 - \frac{(\sum Y)^2}{N}}}$,

we get 0.955176

Strong positive linear correlation between X and Y.

6. Given $\bar{x} = \frac{\sum x_i}{10} = 2.4$, so $\sum x_i = 10 \times 2.4 = 24$

$\bar{y} = \frac{\sum y_i}{5} = 2.0$, so $\sum y_i = 5 \times 2.0 = 10$

therefore, new mean $\bar{z} = \frac{\sum x_i + \sum y_i}{n_x + n_y} = \frac{24+10}{10+5} = 2.267$

$$s_x^2 = \frac{1}{9}\left[\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{10}\right] = 0.8^2, \text{ so } \sum x_i^2 = 9 \times 0.8^2 + \frac{24^2}{10} = 63.36$$

$$s_y^2 = \frac{1}{4}\left[\sum y_i^2 - \frac{\left(\sum y_i\right)^2}{5}\right] = 1.2^2, \text{ so } \sum y_i^2 = 4 \times 1.2^2 + \frac{10^2}{5} = 25.76$$

therefore, new variance $s_z^2 = \frac{1}{10+5-1}\left[\sum x_i^2 + \sum y_i^2 - \frac{\left(\sum x_i + \sum y_i\right)^2}{10+5}\right]$

$$= \frac{1}{14}\left[63.36 + 25.76 - \frac{(24+10)^2}{15}\right] = 0.86$$

$$s_z = \sqrt{s_z^2} = \sqrt{0.86} = 0.93$$