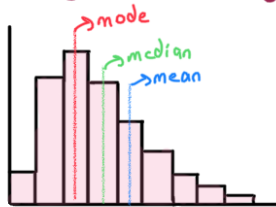


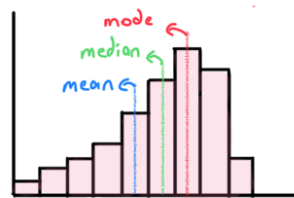
- ① a. nominal because colors can be categorized
 b. ratio because it is quantitative and it has natural zero
 c. ratio because the score is numerical and it has natural zero
 d. nominal because they are categorical
 e. if they are measured in celsius or fahrenheit it is interval because no natural zero, if they are measured in kelvin, it is ratio, because it has natural zero
- ② for the results from the sample to closely match the results obtained if the entire population was used

- ③ a. independent variable = regular 30-minute workout
 dependent variable = risk of catching a cold
 b. independent variable = meditation
 dependent variable = making rational decisions

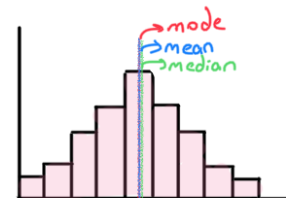
④ positively skewed histogram



negatively skewed histogram



symmetric histogram



- ⑤ a. mode, since it is most repeated one
 b. median, because exact values of mean and mode are unknown
 c. median, because it is not affected by the extreme values
 d. mode, since mean and median can be calculated and are meaningful only for numeric data
 e. mean, it is used in other statistical computations
- ⑥ There is no mode for the data

- ⑦ $\text{sum} = 272$ $n = 10$ $\text{mean} = 272/10 = 27.2$ $\text{median} = 19$ (5th and 6th biggest numbers)
 $\text{mode} = 17$ (repeated 3 times) $\text{midrange} = \frac{\text{max} + \text{min}}{2} = \frac{59 + 17}{2} = 38$

⑧ $\text{sum} = 51$ $n = 6$ $\text{mean} = 8.5$

$\text{sum of squares} = (3-8.5)^2 + (7-8.5)^2 + (8-8.5)^2 + (9-8.5)^2 + (10-8.5)^2 + (14-8.5)^2 = 65.5$

$\text{variance} = s^2 = 65.5 / (6-1) = 13.1$ $\text{standard deviation} = s = \sqrt{13.1} = 3.6194$

⑨ $\text{sum} = 1264$ $n = 12$ $\text{mean} = 105.333$

$\text{range} = 316 (\text{max}) - 10 (\text{min}) = 306$

$\text{sum of squares} = 79662.667$

↗ variance of population

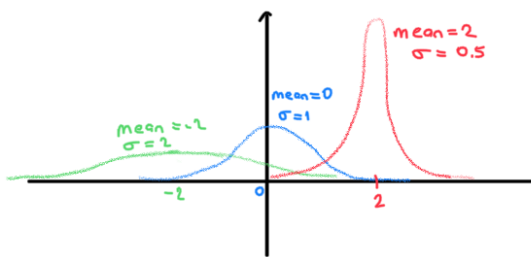
$\text{variance} = s^2 = 79662.667 / 12 = 6638.5556$

$\text{standard deviation} = s = 81.4773$

⑩ No, A and B are not independent. If they are mutually exclusive, they cannot occur at the same time. Since occurrence of one of them affects the occurrence of other event, they are dependent.

⑪ Normal distributions are symmetric; they have one mode; the mean, mode, and median are all equal. They have a bell shape, and they are asymptotic, area under curve is 1

Normal distribution formula = $\frac{1}{\sigma(2\pi)} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2}$, so mean (μ) and standard deviation affect its shape and position



- It is symmetric around its mean, so its position depends on mean.

- When the standard deviation is greater, shape is more wide bell shaped, so its shape depends on standard deviation

⑫ It is special normal distribution where mean = 0 and the standard deviation = 1.

- The total area under the curve is 1.

⑬ z-test = a statistical and significance test, It can be used when the data can be approximated by a normal distribution, has a sample size that is greater or equal to 30

t-test = a statistical and significance test, It can be used when the population mean and standard deviation is unknown, and sample size is less than 30

chi-square test = a statistical and significance test to compare observed results with expected results, to determine if there is a relationship between them. It is for comparing categorical variables

⑭ f-test is used, it calculates the ratio of variances

⑮ When one variable increases as the other increases the relationship is positive, when one decreases as the other increases it is negative relationship.

⑯ **explained variation** = the sum of the squared of the differences between each predicted y-value and the mean of y. $\sum (\hat{y} - \bar{y})^2$

↳ can be explained by the relationship between x and y

unexplained variation = the sum of the squared of the differences between the y-value of each ordered pair and each corresponding predicted y-value $\sum (y - \hat{y})^2$

↳ cannot be explained by the relationship between x and y, and is due to chance or other var.

total variation = sum of the squares of the differences between the y-value of each ordered pair and the mean of y $\sum (y - \bar{y})^2$

↳ total variation = explained variation + unexplained variation

⑰ Correlation coefficient is measure of the strenght of a linear relationship between two variables, range from -1 to 1. The value will be 0 if there is no relationship

⑱ Only mean value (a) will always change. For example if we change the max value to larger number the median remains same, if we change a variable value into the mode value, the mode remains same, but mean always changes

⑲ a. false, because it is square root of variance, cannot be negative

b. false, because a single outlier can drastically affect its value

c. true

d. false, we cannot find, because we have no information about the distribution of v

⑳ $MAE = \frac{1}{n} \sum |y - \hat{y}| = \frac{1}{5} (|165 - 151| + |170 - 161| + |175 - 168| + |180 - 171| + |185 - 179|) = \frac{1}{5} (42) = 8.4$