Sample Standard Deviation

$$S = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n - 1}}$$

Population Standard Deviation

$$\sigma = \sqrt{\frac{\Sigma (x - \mu)^2}{N}}$$

Sample Standard Deviation without calculating the mean

$$s = \sqrt{\frac{n(\Sigma x^2) - (\Sigma x)^2}{n(n-1)}}$$

Mean of grouped Data

$$\bar{x} = \frac{\Sigma f x}{n}$$

NOTE

Variance = $(\sigma)^2$ OR $(s)^2$

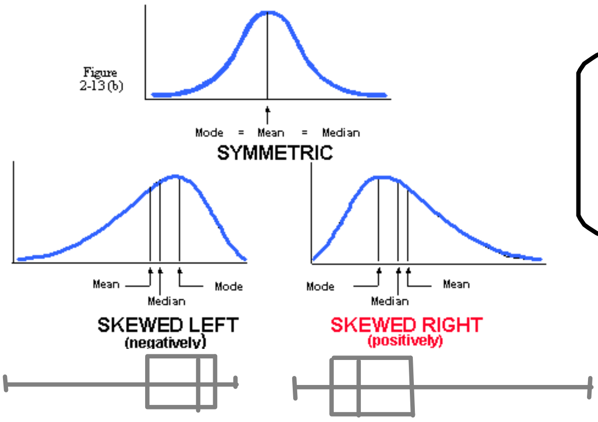
Standard Deviation of frequency table or grouped data

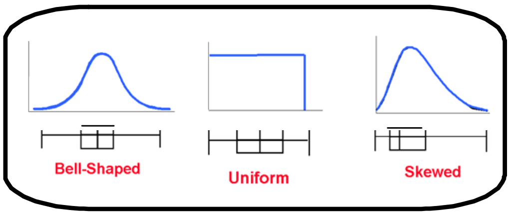
$$S = \sqrt{\frac{n[\Sigma(f \cdot x^2)] - [\Sigma(f \cdot x)]^2}{n(n-1)}}$$

Standard Deviation of frequency table or grouped data where x_m is the midpoint of the class interval

$$S = \sqrt{\frac{n[\Sigma(f \cdot x_m^2)] - [\Sigma(f \cdot x_m)]^2}{n(n-1)}}$$

Skewness





Percentile score $= \dfrac{\text{number of scores below data value} + 0.5}{\text{total number of data values}} \times 100$

Finding a Data Value Corresponding to a given Percentile

$$c = \frac{np}{100}$$

where c = data value position
n = total number of values
p = percentile

-if c is a whole number count go to the number between c and c+1

-if c is not a whole number round up

Outliers

Interquartile Range (or IQR): $Q_3 - Q_1$

To Check for Outliers: $Q_3 + 1.5\ IQR$

$Q_1 - 1.5\ IQR$

RANGE: Highest -Lowest

MIDRANGE: $\dfrac{\text{Highest + Lowest}}{2}$

BOXPLOT

5 - number summary

- Minimum
- first quartile Q1
- Median (Q2)
- third quartile Q3
- Maximum

Z-scores

Sample z-score: $z = \dfrac{x - \bar{x}}{s}$

Population z-score $z = \dfrac{x - \mu}{\sigma}$

To find the x-value: $x = z\sigma + \mu$

Probability Distribution

mean of a probability distribution $\mu = \Sigma[x \cdot P(x)]$

variance of probability distribution $\sigma^2 = [\Sigma x^2 \cdot P(x)] - \mu^2$

standard deviation of probability distribution $\sigma = \sqrt{[\Sigma x^2 \cdot P(x)] - \mu^2}$

Expected Value $E = \mu = \Sigma [x \cdot P(x)]$
The average value of outcomes

Binomial Distribution

Mean of a binomial distribution
$\mu = n \cdot p$

Standard deviation of a binomial distribution
$\sigma = \sqrt{n \cdot p \cdot q}$

Binomial Probability Formula $P(X) = {}_nC_x \cdot p^x \cdot q^{n-x}$

Central Limit Theorem

the mean of the sampling distribution $\mu_{\bar{x}} = \mu$

the standard deviation of the sampling distribution (standard error) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

central limit theorem formula for z-score $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

Population z-score

$$z = \frac{x - \mu}{\sigma}$$

To find the x-value:

$$x = z\sigma + \mu$$

Confidence Interval

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Confidence Interval for t-distribution

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Confidence Interval for Proportion

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Margin of error (E) or Error Estimate:

$$z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Sample size

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 \quad \text{or} \quad n = \frac{(z_{\alpha/2})^2 \hat{p}\hat{q}}{E^2}$$

Note: Point estimate is the mean

Hypothesis Testing: null always contains the equality
p-value method and critical value method

Test Statistic

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

Test Statistic for a proportion

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

Critical Value Method

If the test statistic falls in the tail, reject the null

P-Value Method

P-value is the area in the tail of the test statistic

if: $P \leq \alpha$ reject Ho

$P > \alpha$ fail to reject Ho

Testing The difference Between Two Sample Means Large Samples } **Test for Independence**

$$\mu_{\bar{x}_1 - \bar{x}_2} = 0 \qquad \sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Test Statistic

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

	H_0 True Innocent	**H_0 False** Not Innocent Guilty
Reject H_0 Convict	Type 1 Error	correct decision
Fail to Reject H_0 Not Convict	correct decision	Type 2 Error

Contingency Tables
Chi Square test of Independence

Expected (E) Value: $\frac{\text{(row total) x (column total)}}{\text{grand total}}$

Test of Independence

Test Statistic

$$X^2 = \sum \frac{(O - E)^2}{E}$$

Ho: row variable is independent of column variable

Ha: row variable is dependent on column variable

Critical Values

1. Found in Table G using degrees of freedom = (r - 1)(c - 1)
   r is the number of rows
   c is the number of columns
2. Tests of Independence are always right-tailed.

Formula for the Correlation Coefficient *r*

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n(\Sigma x^2) - (\Sigma x)^2][n(\Sigma y^2) - (\Sigma y)^2]}}$$

where n is the number of data pairs.

Compute *r* on the calculator

- When calculating a correlation coefficient, an obvious question arises: *Is the implied relationship statistically significant, or due to random chance?*
- We can perform a hypothesis test testing whether there is significant evidence against the correlation coefficient being zero

$$H_0 : \rho = 0$$
$$H_A : \rho \neq 0$$

Formula for t Test Value for the correlation Coefficient

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

Use T table and n-2 degrees of freedom for critical value