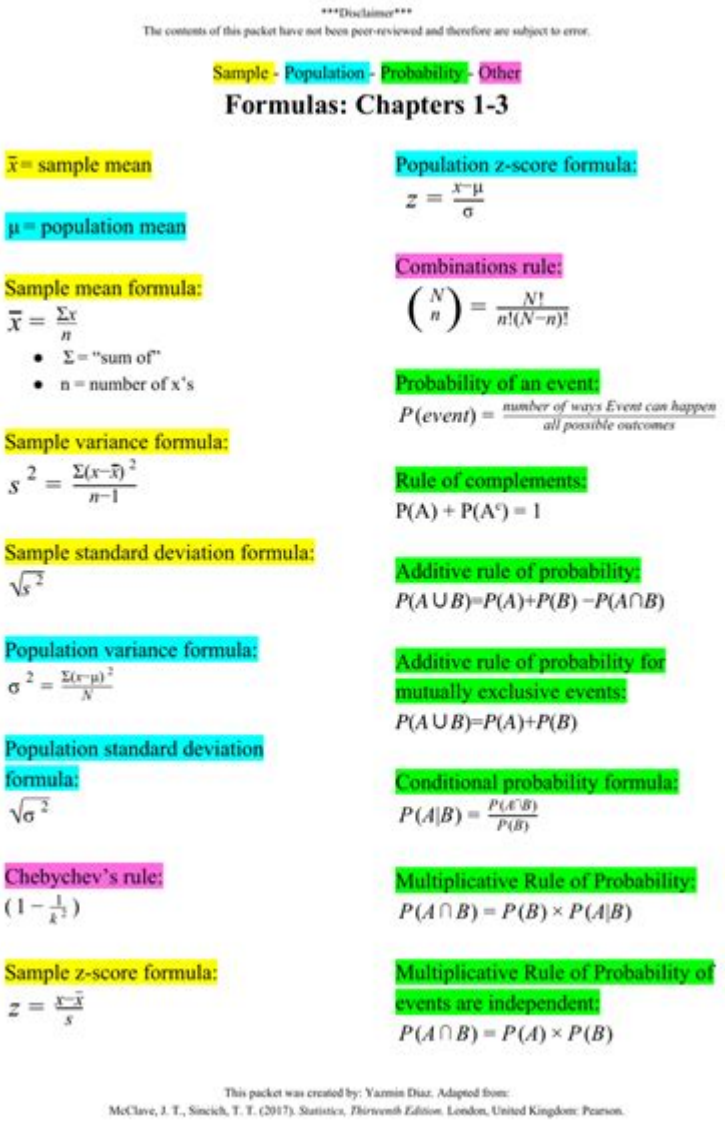
Statistics Formulas Cheat Sheet



Disclaimer

The contents of this packet have not been peer-reviewed and therefore are subject to error.

Sample - Population - Probability - Other

Formulas: Chapters 1-3

$\bar{x}$ = sample mean

μ = population mean

Sample mean formula:

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

- Σ = "sum of"
- n = number of x's

Sample variance formula:

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

Sample standard deviation formula:

$$\sqrt{s^2}$$

Population variance formula:

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

Population standard deviation formula:

$$\sqrt{\sigma^2}$$

Chebychev's rule:

$$(1 - \frac{1}{k^2})$$

Sample z-score formula:

$$z = \frac{x-\bar{x}}{s}$$

Population z-score formula:

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

Combinations rule:

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

Probability of an event:

$$P(event) = \frac{number\ of\ ways\ Event\ can\ happen}{all\ possible\ outcomes}$$

Rule of complements:

$$P(A) + P(A^c) = 1$$

Additive rule of probability:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Additive rule of probability for mutually exclusive events:

$$P(A \cup B) = P(A) + P(B)$$

Conditional probability formula:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Multiplicative Rule of Probability:

$$P(A \cap B) = P(B) \times P(A|B)$$

Multiplicative Rule of Probability of events are independent:

$$P(A \cap B) = P(A) \times P(B)$$

This packet was created by: Yazmin Diaz. Adapted from:

McClave, J. T., Sincich, T. T. (2017). *Statistics, Thirteenth Edition.* London, United Kingdom: Pearson.

statistics formulas cheat sheet

In the realm of data analysis, research, and decision-making, understanding the fundamental statistical formulas is essential. Whether you're a student, data analyst, or researcher, having a solid grasp of key statistical formulas can greatly enhance your ability to interpret data accurately and efficiently. This **statistics formulas cheat sheet** aims to provide a comprehensive overview of the most important statistical formulas, organized for easy reference and quick understanding.

Basic Statistical Measures

Understanding basic statistical measures is the foundation of data analysis. These include measures of central tendency and measures of dispersion, which describe the distribution of data points.

Measures of Central Tendency

- **Mean (Average):**

  $$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$
  where x_i are data points and n is the total number of data points.

- **Median:**

  The middle value when data is ordered.

  For odd n, median is the $\frac{n+1}{2}$th data point.

  For even n, median is the average of $\frac{n}{2}$ and $\frac{n}{2}+1$ data points.

- **Mode:**

  The most frequently occurring data point(s).

  Can be bimodal or multimodal if multiple values share the highest frequency.

Measures of Dispersion

- **Range:**

  Difference between maximum and minimum values:
  $$\text{Range} = x_{max} - x_{min}$$

- **Variance:**

  Measures the average squared deviation from the mean.

  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:**

  The square root of variance, providing dispersion in original units:
  $$\sigma = \sqrt{\sigma^2} \quad \text{(population)}$$

$$ s = \sqrt{s^2} \quad \text{(sample)} $$

- **Interquartile Range (IQR):**

  Difference between third quartile (Q3) and first quartile (Q1):
  $$ \text{IQR} = Q_3 - Q_1 $$

Probability and Distributions

Probability concepts and distribution formulas are vital for inferential statistics.

Basic Probability Formulas

- **Probability of an event A:**
  $$ P(A) = \frac{\text{Number of favorable outcomes}}{\text{Total number of outcomes}} $$
- **Complement rule:**
  $$ P(A') = 1 - P(A) $$
- **Addition rule:**
  $$ P(A \cup B) = P(A) + P(B) - P(A \cap B) $$
- **Multiplication rule (independent events):**
  $$ P(A \cap B) = P(A) \times P(B) $$

Common Probability Distributions

- **Binomial Distribution:**
  $$ P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} $$
  where n is the number of trials, k is the number of successes, and p is the probability of success in each trial.
- **Normal Distribution:**
  $$ f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{ -\frac{(x - \mu)^2}{2\sigma^2} } $$
  where μ is the mean, σ is the standard deviation.
- **Poisson Distribution:**
  $$ P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!} $$
  where λ is the average rate (mean number of events in interval).

Inferential Statistics Formulas

These formulas are used to make inferences about a population based on sample data.

Confidence Interval

- **For the population mean (known σ):**
  $$ \bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} $$
- **For the population mean (unknown σ, using t-distribution):**
  $$ \bar{x} \pm t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}} $$

Hypothesis Testing

- **Z-test statistic:**
  $$ Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} $$
- **t-test statistic:**
  $$ t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} $$

Correlation and Regression

- **Pearson correlation coefficient (r):**
  $$ r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{ \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} } $$
- **Linear regression equation:**
  $$ y = a + bx $$
  where
  $$ b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} $$
  and
  $$ a = \bar{y} - b \bar{x} $$

Additional Important Formulas

Chi-Square Test

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$
where O = observed frequency, E = expected frequency.

ANOVA (Analysis of Variance)

- F-statistic:
$$F = \frac{\text{Between-group variance}}{\text{Within-group variance}}$$
Calculations involve sums of squares (SS) and degrees of freedom (df), but the core formula for the F-value is crucial for hypothesis testing.

Summary of Key Formulas

Concept	Formula	Description
Mean	$\bar{x} = \frac{\sum x_i}{n}$	Average of data points
Variance (sample)	$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$	Measure of data spread
Standard deviation	$s = \sqrt{s^2}$	Dispersion in original units
Z-score	$Z = \frac{x - \mu}{\sigma}$	Distance from mean in SD units
Confidence interval (mean)	$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$	Range estimate for population mean
Correlation coefficient	r	Measure of linear relationship between two variables

---

Conclusion

Mastering these statistical formulas is vital for accurate data analysis, interpretation, and reporting. This **statistics formulas cheat sheet** serves as a quick reference guide to essential formulas,

Frequently Asked Questions

What is the formula for calculating the mean in statistics?

The mean is calculated by summing all data points and dividing by the number of data points: Mean = (Σx) / n.

How do you compute the variance of a data set?

Variance is calculated as the average of the squared differences from the mean: Variance = $\Sigma(x_i - \mu)^2 / n$ for population, or $\Sigma(x_i - \bar{x})^2 / (n - 1)$ for a sample.

What is the formula for the standard deviation?

Standard deviation is the square root of variance: SD = $\sqrt{\text{Variance}}$.

How is the z-score calculated?

Z-score = $(X - \mu) / \sigma$, where X is the data point, μ is the mean, and σ is the standard deviation.

What is the formula for the coefficient of correlation (Pearson's r)?

$r = \Sigma[(x_i - \bar{x})(y_i - \bar{y})] / \sqrt{[\Sigma(x_i - \bar{x})^2 \Sigma(y_i - \bar{y})^2]}$, measuring the strength and direction of the linear relationship between two variables.

Statistics Formulas Cheat Sheet

Statistics Formulas Cheat Sheet

statistics formulas cheat sheet In the realm of data analysis, research, and decision-making, understanding the fundamental statistical formulas is essential. Whether you're a student, data analyst, or researcher, having a solid grasp of key statistical formulas can greatly enhance your ability to interpret data accurately and efficiently. This statistics

Back to Home