Comprehensive Statistics Formula Sheet

StatsCalculators.com

10/29/2024

Contents

1 Descriptive Statistics

1.1 Measures of Central Tendency

Basic Statistics

Let $\{x_1, x_2, \ldots, x_n\}$ be a sample of size n.

$$\text{Mean:} \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\text{Median:} \quad \begin{cases} x_{(\frac{n+1}{2})} & \text{if } n \text{ is odd} \\ \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} & \text{if } n \text{ is even} \end{cases}$$

Mode: Most frequent value

1.2 Measures of Spread

Variance and Standard Deviation

$$\text{Sample Variance:} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$\text{Sample SD:} \quad s = \sqrt{s^2}$$

$$\text{Population Variance:} \quad \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

$$\text{Population SD:} \quad \sigma = \sqrt{\sigma^2}$$

Quartiles and Percentiles

$$\text{Range:} \quad R = \max(x_i) - \min(x_i)$$

$$\text{IQR:} \quad \text{IQR} = Q_3 - Q_1$$

$$\text{Percentile k:} \quad P_k = \text{value at position } \left\lceil \frac{k}{100} \cdot n \right\rceil$$

Outlier Boundaries:

Lower fence: $Q_1 - 1.5 \cdot \text{IQR}$

Upper fence: $Q_3 + 1.5 \cdot \text{IQR}$

2 Probability Theory

2.1 Basic Probability Rules

Fundamental Probability Laws

For events A and B in sample space S:

$$\text{Addition Rule:} \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$\text{Multiplication Rule:} \quad P(A \cap B) = P(A) \cdot P(B|A)$$

$$\text{Complement Rule:} \quad P(A^c) = 1 - P(A)$$

$$\text{Conditional Probability:} \quad P(A|B) = \frac{P(A \cap B)}{P(B)}$$

2.2 Counting Principles

Combinatorics

$$\text{Permutations:} \quad P(n, r) = \frac{n!}{(n-r)!}$$

$$\text{Combinations:} \quad C(n, r) = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

3 Probability Distributions

3.1 Discrete Distributions

Binomial Distribution

$X \sim \text{Bin}(n, p)$

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\text{Mean:} \quad \mu = np$$
$$\text{Variance:} \quad \sigma^2 = np(1-p)$$

Poisson Distribution

$X \sim \text{Pois}(\lambda)$

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

$$\text{Mean:} \quad \mu = \lambda$$
$$\text{Variance:} \quad \sigma^2 = \lambda$$

3.2 Continuous Distributions

Normal Distribution

$X \sim N(\mu, \sigma^2)$

$$\text{Z-score:} \quad Z = \frac{X - \mu}{\sigma}$$

$$\text{Probability Density:} \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Empirical Rule:

- $\mu \pm 1\sigma :\approx 68\%$
- $\mu \pm 2\sigma :\approx 95\%$
- $\mu \pm 3\sigma :\approx 99.7\%$

Exponential Distribution

$X \sim \text{Exp}(\lambda)$

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

$$\text{Mean:} \quad \mu = \frac{1}{\lambda}$$

$$\text{Variance:} \quad \sigma^2 = \frac{1}{\lambda^2}$$

4 Statistical Inference

4.1 Sampling Distributions

Properties of Sample Mean

For a random sample of size n:

$$\text{Mean:} \quad \mu_{\bar{X}} = \mu$$

$$\text{Standard Error:} \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

$$\text{Central Limit Theorem:} \quad \bar{X} \xrightarrow{n \to \infty} N\left(\mu, \frac{\sigma^2}{n}\right)$$

Sample Proportion Distribution

For $\hat{p}$ (sample proportion):

$$\text{Mean:} \quad \mu_{\hat{p}} = p$$

$$\text{Standard Error:} \quad \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Conditions:

- $np \geq 10$
- $n(1-p) \geq 10$

4.2 Confidence Intervals

Common Confidence Intervals

For population mean μ:

$$\text{Known } \sigma: \quad \bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\text{Unknown } \sigma: \quad \bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

Common Confidence Levels:

Confidence Level	α	$z_{\alpha/2}$
90%	0.10	1.645
95%	0.05	1.960
99%	0.01	2.576

5 Regression Analysis

5.1 Simple Linear Regression

Linear Regression Equations

Model: $\hat{y} = b_0 + b_1 x$

$$\text{Slope:} \quad b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$\text{Intercept:} \quad b_0 = \bar{y} - b_1\bar{x}$$

$$\text{Correlation:} \quad r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

Regression Analysis

$$\text{R-squared:} \quad R^2 = r^2$$

$$\text{Standard Error:} \quad s_e = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}}$$

6 Hypothesis Testing

6.1 Test Statistics

Common Test Statistics

$$\text{Z-test:} \quad Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

$$\text{T-test:} \quad t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

$$\text{Paired t-test:} \quad t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

Two-Sample Tests:

$$\text{Z-test:} \quad Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$\text{T-test:} \quad t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$

Types of Errors

Decision	H_0 True	H_0 False
Reject H_0	Type I Error (α)	Correct Decision
Fail to Reject H_0	Correct Decision	Type II Error (β)

Power Analysis:

$$\text{Power} = 1 - \beta$$
$$\text{Sample Size} = f(\alpha, \beta, \text{effect size})$$

6.2 Effect Size Measures

Cohen's d Effect Size

$$d = \frac{|\mu_1 - \mu_2|}{s_p}$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Effect Size Interpretation:

- Small effect: $d \approx 0.2$
- Medium effect: $d \approx 0.5$
- Large effect: $d \approx 0.8$

7 Additional Tests

7.1 Chi-Square Tests

Chi-Square Statistics

$$\text{Goodness of Fit:} \quad \chi^2 = \sum \frac{(O - E)^2}{E}$$

$$\text{Independence:} \quad \chi^2 = \sum\sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Degrees of Freedom:

- Goodness of Fit: $df = k - 1$ (k = number of categories)
- Independence: $df = (r - 1)(c - 1)$ (r = rows, c = columns)

Expected Frequencies (Independence):

$$E_{ij} = \frac{\text{(row total)(column total)}}{\text{grand total}}$$

7.2 ANOVA

One-Way ANOVA

$$F = \frac{\text{MS}_{\text{between}}}{\text{MS}_{\text{within}}}$$

$$\text{MS}_{\text{between}} = \frac{\text{SS}_{\text{between}}}{k-1}$$

$$\text{MS}_{\text{within}} = \frac{\text{SS}_{\text{within}}}{N-k}$$

where k is the number of groups and N is total sample size.

8 Common Critical Values

Z-scores for Common Percentiles

Percentile	Z-score
90%	±1.645
95%	±1.96
99%	±2.576

Critical Values for t-distribution

df	90%	95%	99%
10	1.812	2.228	3.169
20	1.725	2.086	2.845
30	1.697	2.042	2.750
∞	1.645	1.960	2.576

9 Effect Size Measures

Cohen's d Effect Size

$$d = \frac{|\mu_1 - \mu_2|}{s_p}$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Effect Size Interpretation:

- Small effect: $d \approx 0.2$
- Medium effect: $d \approx 0.5$
- Large effect: $d \approx 0.8$

10 Additional Probability Distributions

Exponential Distribution

$X \sim \text{Exp}(\lambda)$

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

$$\text{Mean:} \quad \mu = \frac{1}{\lambda}$$

$$\text{Variance:} \quad \sigma^2 = \frac{1}{\lambda^2}$$

Geometric Distribution

$X \sim \text{Geom}(p)$

$$P(X = k) = p(1 - p)^{k-1}, \quad k \geq 1$$

$$\text{Mean:} \quad \mu = \frac{1}{p}$$

$$\text{Variance:} \quad \sigma^2 = \frac{1 - p}{p^2}$$

11 Sampling Distribution Properties

Sample Proportion Distribution

For $\hat{p}$ (sample proportion):

$$\text{Mean:} \quad \mu_{\hat{p}} = p$$

$$\text{Standard Error:} \quad \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Conditions for Normal Approximation:

- $np \geq 10$
- $n(1-p) \geq 10$

Central Limit Theorem Applications

For $\bar{X}$ to be approximately normal:

- Random sample
- Independent observations
- $n \geq 30$ or population is normal

Standard Error Forms:

$$\text{For mean:} \quad SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

$$\text{For proportion:} \quad SE(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$$

$$\text{For difference of means:} \quad SE(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$