The Greek Alphabet

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Upper | Lower | English | Upper | Lower | English |
| Α | α | Alpha | Ν | ν | Nu |
| Β | β | Beta | Ξ | ξ | Xi |
| Γ | γ | Gamma | Ο | ο | Omicron |
| Δ | δ | Delta | Π | π | Pi |
| Ε | ε | Epsilon | Ρ | ρ | Rho |
| Ζ | ζ | Zeta | Σ | σ | Sigma |
| Η | η | Eta | Τ | τ | Tau |
| Θ | θ | Theta | Υ | υ | Upsilon |
| Ι | ι | Iota | Φ | φ | Phi |
| Κ | κ | Kappa | Χ | χ | Chi |
| Λ | λ | Lambda | Ψ | ψ | Psi |
| Μ | μ | Mu | Ω | ω | Omega |

Statistical Symbols

|  |  |
| --- | --- |
| Symbol | Meaning |
| S | Sample space (probability theory) |
| E | Event (probability theory) |
| ∪ | Union of sets |
| ∩ | Intersection of sets |
| *P(A)* | Probability of event A |
| *P(A|B)* | Probability of event A, given event B |
| *P(~A)* | Probability of A complement (not A) |
| ln | Natural logarithm (base *e*) |
| logx | Logarithm to base x |
| *x*i | The *i*th member of sample x |
| x̄ | Sample mean |
| μ | Population mean |
| *s* | Sample standard deviation |
| σ | Population standard deviation |
| *s*2 | Sample variance |
| *s2p* | Pooled sample variance |
| σ 2 | Population variance |
| *n* | Sample size |
| *N* | Population size |
| *r* | Sample correlation |
| *r*pb | Point-biserial correlation |
| *r*s | Spearman’s rho (Spearman rank-order correlation) |
| γ | Goodman and Kruskal’s gamma |
| Ta, TB, TC | Kendall’s tau-A, tau-B, tau-C |
| Φ | Phi (measure of association between binary variables) |
| P | Concordant pairs (ordinal measures of association) |
| Q | Discordant pairs (ordinal measures of association) |
| ρ | Pearson population correlation coefficient |
| *r* | Pearson sample correlation coefficient |
| *r*2 | Coefficient of determination (% of variance explained) |
| χ2 | Chi-square |
| *O* | Observed value (chi-square) |
| *E* | Expected value (chi-square) |
| *R×C* | A table with R rows and C columns |
| *H*0 | Null hypothesis |
| *H*A, *H*1 | Alternate hypothesis |
| α | Alpha, the probability of a Type I error |
| β | Beta, the probability of a Type II error |
| Σ | Summation |
| *n*P*k* | Permutation |
| *n*C*k*(*nk*) | Combination |
| *n*! | *n* factorial |
| *t* | Student’s *t* |
| *df* | Degrees of freedom |
| *Z* | Standard normal score/distribution |
| *SS* | Sums of squares |
| *MS* | Mean square |
| *I*t | Index for time *t* (business statistics) |
| *Qit* | Quantity of product *I* for time *t* (business statistics) |
| *Pit* | Price of product *I* for time *t* (business statistics) |
| *Tt* | Secular trend (time series) |
| *Ct* | Cyclical trend (time series) |
| *St* | Seasonal effect (time series) |
| *Rt* | Residual or error effect trend (time series) |
| *RR* | Risk ratio (relative risk) |
| *OR* | Odds ratio |
| *OR*MH | Mantel-Haenszel odds ratio |
| *D*+, *D*- | Disease, no disease (epidemiology) |
| *E*+, *E*- | Exposure, no exposure (epidemiology) |

Distributions

|  |  |  |
| --- | --- | --- |
| Symbol | Meaning | Parameters |
| Normal | 68% within σ, 95% within 2σ, 99% within 3σ | μ, σ |
| Binomial | Probability of *k* successes in *n* trials with independent probability, *p* = *pk*(1-*p*)*n*-*k* | *p*, *n* |
| Student’s t | Alternative to Normal distribution for samples with an unknown population σ | *df* |
| Poisson |  |  |
| Chi-square (χ2) |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

Definitions

|  |  |  |
| --- | --- | --- |
| Symbol | Meaning | SINS page |
| Z-scores | The probability of a particular observation measured in standard deviations, Z = (*x*– μ) / σ | 49 |
| p-value | “The probability that results at least as extreme as those obtained in an analysis of sample data are due to chance.” | 68 |
| Z-Statistic | The probability of a particular sample mean, Z = (x̄ - μ) / (σ / √n) | 70 |
| Kolmogorov-Smirnov (K-S) | Tests if data are normally distributed | 72 |
| Anderson-Darling | Revision of Kolmogorov-Smirnov | - |
| Cohen’s kappa | Measure of agreement between categorical series | 123 |
| Chi-square test (independence) | Tests that two categorical variables are independent | 127 |
| Chi-square test (equality of proportions) | Tests whether proportions (rates) are the same across samples from independent populations | 130 |
| Chi-square test (goodness of fit) | Test if the distribution of a categorical variable follows expected pattern of proportions | 130 |
| Fisher’s Exact Test | Alternative to the chi-square test better suited to small or sparsely distributed datasets | 132 |
| McNemar’s test | Variation on chi-square for use with matched samples (e.g. before and after some event) | 134 |
| Phi (Φ) | Measure of association between binary variables | 139 |
| Cramer’s V | Variation on Phi for non-binary categorical variables |  |
| Point-biserial correlation (*r*pb) | Measure of association between a binary variable and a continuous variable | 141 |
| Spearman’s rho (*r*s) | Measure of association between ordinal variables | 142 |
| Goodman and Kruskal’s gamma (γ) | Alternative to Spearman’s rho based on number of concordant and discordant pairs | 143 |
| Kendall’s tau (Ta, TB, TC) | Additional alternatives to Spearman’s rho | 144 |
| Somer’s d | An asymmetrical version of Goodman and Kruskal’s gamma | 145 |
| t-test (one-sample) | Tests if there is no significant difference between sample mean and population mean | 157 |
| t-test (two-samples) | Tests if the means of the populations from which two samples were drawn are the same | 160 |
| t-test (repeated measures) | Tests if the mean difference between samples is 0 | 164 |
| Levene statistic | Tests that two samples have approximately equal variance (homogeneity of variance) | 167 |
| Pearson correlation coefficient (*r)* | Measure of association between interval or ratio variables | 182 |
| Coefficient of determination (*r*2) | Amount of variation in one variable that can be explained by variation in another | 187 |
| Linear regression | Derives the line which best predicts a continuous outcome variable from one or more predictor variables | 195 |
| ANOVA (Analysis of Variance) | Tests whether independent groups have the same mean | 206 |
| F ratio | The ratio of the “between group” and “within group” means in an ANOVA test | 206 |
| Scheffe test | Tests for significant differences among multiple group means (ANOVA post hoc) | 210 |
| Factorial ANOVA | ANOVA with multiple grouping variables | 223 |
| ANCOVA | ANOVA with controls for one or more continuous covariates | 233 |
| Logistic regression | Regression to a binary outcome variable | 273 |
| logit (log odds) | The outcome variable of a logistic regression—a measure of probability | 273 |
| Wald test | Tests if a predictor variable adds explanatory power to a model (a type of chi-square test) | 275 |
| -2 log likelihood | A measure of goodness of fit for a logistic regression model | 277 |
| Cox and Snell R2 | A pseudo-R2 variable that measures the explanatory power of a logistic regression | 277 |
| Nagelkerke R2 | A version of Cox and Snell R2 adjusted so its maximum value is 1 | 277 |
| Multinomial logistic regression | Regression to a categorical outcome variable | 279 |
| Polynomial regression | Regression with exponential predictors | 282 |