# CORRELATIONS

The user-specified treatment for missing values is used for computation of all statistics except, under certain conditions, the means and standard deviations.

## Notation

The following notation is used throughout this chapter unless otherwise specified:

| | |
|---|---|
| $N$ | Number of cases |
| $X_{kl}$ | Value of the variable *k* for case *l* |
| $w_l$ | Weight for case *l* |
| $W_k$ | Sum of weights of cases used in computation of statistics for variable *k* |
| $W_{kj}$ | Sum of weights of cases used in computation of statistics for variables *k* and *j* |

## Statistics

### Means and Standard Deviations

$$\bar{X}_k = \frac{\sum_{l=1}^{N} w_l X_{kl}}{W_k}$$

$$S_k = \sqrt{\left( \sum_{l=1}^{N} w_l X_{kl}^2 - \bar{X}_k^2 W_k \right) \Big/ \left( W_k - 1 \right)}$$

*Note:* If no treatment for missing values is specified (default is pairwise), means and standard deviations are computed based on all nonmissing values for each variable. If missing values are to be included or listwise is chosen, that option is used for means and standard deviations as well.

### Cross-product Deviations and Covariances

The cross-product deviation for variables *i* and *j* is

$$C_{ij} = \sum_{l=1}^{N} w_l X_{il} X_{jl} - \left( \sum_{l=1}^{N} w_l X_{il} \right) \left( \sum_{l=1}^{N} w_l X_{jl} \right) \Big/ W_{ij}$$

The covariance is

$$S_{ij} = \frac{C_{ij}}{W_{ij} - 1}$$

### Pearson Correlation

$$r_{ij} = \frac{C_{ij}}{\sqrt{C_{ii} C_{jj}}}$$

### Significance Level of *r*

The significance level for $r_{ij}$ is based on

$$t = r_{ij} \sqrt{\frac{W_{ij} - 2}{1 - r_{ij}^2}}$$

which, under the null hypothesis, is distributed as a $t$ with $W_{ij} - 2$ degrees of freedom. By default, the significance level is two-tailed.

## References

Blalock, H. M. 1972. *Social statistics*. New York: McGraw-Hill.