Mean	$\overline{X} = \dfrac{\sum X}{N}$
Mean of grouped scores	$\overline{X} = \dfrac{\sum f \cdot X}{\sum f}$
Median	M_d : the value of the middle case (N odd) or the average of the values of the two middle cases (N even)
Mode	M_o : the most frequent X_i
Variance	$s^2 = \dfrac{\sum (X - \overline{X})^2}{N-1}$
Variance of grouped scores	$s^2 = \dfrac{\sum f\,(X - \overline{X})^2}{\sum f}$
Standard deviation	$s = \sqrt{s^2}$
Standard Score	$z = \dfrac{X - \overline{X}}{s}$
Regression model	$Y_i = b \cdot X_i + a + e_i = \hat{Y}_i + e_i$ $= Y_i = b_0 + b_1 * X_1 + e_i$
Covariance	$\text{Cov}(X, Y) = \dfrac{\sum (X - \overline{X})(Y - \overline{Y})}{N-1}$
Some important relations in regression and correlation :	$s_Y^2 = s_{\hat{Y}}^2 + s_e^2$

$$0 \leq s_{\hat{Y}}^2 \leq s_Y^2$$

$$r^2 = 1 - \frac{s_e^2}{s_Y^2} = \frac{s_{\hat{Y}}^2}{s_Y^2}$$

$$= \frac{\sum z_X z_Y}{N-1} = \frac{\text{Cov}(X, Y)}{s_X s_Y}$$

$$b_{YX} = r\,\frac{s_x}{s_y} = \frac{\text{Cov}(X, Y)}{s_X^2}$$

$$b_{XY} = b_{YX}\left(\frac{s_Y}{s_X}\right)^2$$

Enkelvoudige lineaire regressie (*simple lineair regression)*

Regressievergelijking	$E(y) = \alpha + \beta \cdot x$
Voorspellingsvergelijking *(predicion equation)*	$\hat{y} = a + b \cdot x$
Residuen, voorspellingsfouten *(residuals, errors)*	$e_i = y_i - \hat{y}_i$
Kleinste- kwadratenvoorspelling van β (*Least Squares Estimate*)	$b = \dfrac{Cov(x,y)}{s_x^2} = \dfrac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$
Kleinste-kwadratenvoorspelling van α	$a = \bar{y} - b \cdot \bar{x}$
Sum of Squared Errors of : Residual Sum of Squares (RSS)	$SSE = \sum (y_i - \hat{y}_i)^2$
Schatting van conditionele standaarddeviatie (*estimate of conditional standard deviation*)	$s = \sqrt{\dfrac{SSE}{n-2}}$ "standard error of the estimate" = RMSE
Pearson's correlatiecoëfficiënt	$r = \dfrac{s_x}{s_y} \cdot b$
Totale variatie in y (*Total Sum of Squares*)	$TSS = \sum (y_i - \bar{y})^2$
Verklaarde variatie in y (*Sum of Squares Regression/Model*)	$SSM = TSS - SSE = \sum (\hat{y}_i - \bar{y})^2$
Coefficient of Determination	$r^2 = R^2 = \dfrac{SSM}{TSS}$

Standaardfout voor $\bar{y}_d$	$se = \dfrac{s_d}{\sqrt{n}}$

Standaardfout voor $\bar{y}_2 - \bar{y}_1$	$se = \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$

Standaardfout voor $\hat{\pi}_2 - \hat{\pi}_1$	$se_0 = \sqrt{\hat{\pi}(1-\hat{\pi}) \cdot \left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}$ $\hat{\pi} = \dfrac{n_1\hat{\pi}_1 + n_2\hat{\pi}_2}{n_1 + n_2}$ = pooled estimate

Standaardfout voor $\hat{\pi}_2 - \hat{\pi}_1$	$se = \sqrt{\dfrac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \dfrac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}$

Standard Errors	
sample means	$\sigma_{\overline{X}} = \dfrac{\sigma_X}{\sqrt{N}}$, $s_{\overline{X}} = \dfrac{s_X}{\sqrt{N}}$
difference of sample means	$s_{\overline{D}} = \dfrac{s_D}{\sqrt{N}}$, $s_{\overline{X}_1 - \overline{X}_2} = \sqrt{\dfrac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2} \cdot \left(\dfrac{1}{N_1} + \dfrac{1}{N_2}\right)}$
correlation	$s_r = \sqrt{\dfrac{1 - r_{\text{obs}}^2}{N - 2}}$ $= 1/\sqrt{N-2}$ if H_0
Fisher's z – transformation of r	$s_{z_r} = \sqrt{\dfrac{1}{N - 3}}$
z – transformation of $r_1 - r_2$	$s_{z_{r_1} - z_{r_2}} = \sqrt{\dfrac{1}{N_1 - 3} + \dfrac{1}{N_2 - 3}}$

Table 4.1 *Standard errors for selected sampling distributions*

Sampling distribution	Standard error	Comments
Means	$\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$	Even if the population is not normal, the sampling distribution of means is nearly normal for $n \geq 30$. In all cases, $\mu_{\bar{X}} = \mu$.
Proportions	$\sigma_p = \sqrt{\dfrac{p(1-p)}{n}} = \sqrt{\dfrac{pq}{n}}$	See comments for the sampling distribution of the mean.
Medians	$\sigma_{\text{med}} = \sigma\sqrt{\dfrac{\pi}{2n}} = \dfrac{1.2533\sigma}{\sqrt{n}}$	The sampling distribution of the median is nearly normal for $n \geq 30$, as the population approaches normality. $\mu_{\text{med}} = \mu$.
Standard deviations	(1) $\sigma_s = \dfrac{\sigma}{\sqrt{2n}}$ (2) $\sigma_s = \sqrt{\dfrac{\mu_4 - \mu_2^2}{4n\mu_4}}$	The sampling distribution is nearly normal for $n \geq 100$. As population approaches normality, (1) applies; if population is not normal, (2) can be used. For $n \geq 100$, $\mu_s \approx \sigma$.
Variances	(1) $\sigma_{s^2} = \sigma^2\sqrt{\dfrac{2}{n}}$ (2) $\sigma_{s^2} = \sqrt{\dfrac{\mu_4 - \mu_2^2}{n}}$	See comments for standard deviations. $\mu_{s^2} = \sigma^2(n-1)/n$, which approaches σ^2 as n becomes large.
First and third quartiles	$\sigma_{Q_1} = \sigma_{Q_3} = \dfrac{1.3626\sigma}{\sqrt{n}}$	See comments for the sampling distribution of the median.
Semi-interquartile range	$\sigma_Q = \dfrac{0.7867\sigma}{\sqrt{n}}$	See comments for the median.

Note: See Table 3:2 for moment designations.

In the bivariate case, the standard error of the slope coefficient estimator can be calculated by:

$$s_b = \sqrt{\frac{\sum_{j=1}^{n} (Y_j - \hat{Y}_j)^2/(n-2)}{\sum_{j=1}^{n} (X_j - \bar{X})^2}} \qquad [1.11]$$

Extending this to the two variable case yields formulae for

$$s_{b_1} = \sqrt{\frac{\sum_{j=1}^{n} (Y_j - \hat{Y}_j)^2/(n-3)}{\sum_{j=1}^{n} (X_{1j} - \bar{X}_1)^2(1 - r^2_{X_1X_2})}}$$

$$s_{b_2} = \sqrt{\frac{\sum_{j=1}^{n} (Y_j - \hat{Y}_j)^2/(n-3)}{\sum_{j=1}^{n} (X_{2j} - \bar{X}_2)^2(1 - r^2_{X_1X_2})}} \qquad [1.12]$$

Finally, we can go one step further and derive a formula for the standard error of the partial slope coefficient estimator for a model with any number of independent variables:

$$s_{b_i} = \sqrt{\frac{\sum_{j=1}^{n} (Y_j - \hat{Y}_j)^2}{\sum_{j=1}^{n} (X_{ij} - \bar{X})^2(1 - R_i^2)\,(n-k-1)}} \qquad [1.13]$$