

Multipl. Regressionsmodell:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + e_i \quad \text{für } i = 1, \dots, n$$

z.B.

y Einkommen

$x_1$  Alter

$x_2$  Bildung

$x_3$  Geschlecht

...

→ man schätzt  $\hat{\beta}_0, \dots, \hat{\beta}_p$

z.B.  $\hat{\beta}_1$ : Welchen Einfluss hat das Alter auf das Einkommen?

mit den geschätzten  $\hat{\beta}$  und den Einflussgrößen lässt sich y vorhersagen

$$\Rightarrow \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_p x_{ip}$$

Streuungszerlegung:

$$\underbrace{\sum_{i=1}^n (y_i - \bar{y})^2}_{SQ_{Total}} = \underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)^2}_{SQ_{Res}} + \underbrace{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}_{SQ_{Reg}}$$

$SQ_{Total}$  gesamte Streuung in y (vgl. Varianz)

$SQ_{Reg}$  Streuung von y die durch Regression erklärt wird

$SQ_{Res}$  Streuung von y die nicht erklärt werden kann

⇒ Anteil erklärter Streuung

$$R^2 = \frac{SQ_{Reg}}{SQ_{Total}}$$