Wie kommt man auf die Formel der Streuungszerlegung in der linearen Regression?

**Ansatz**

vgl. Folie 7.28

$$\begin{aligned}
\hat{e}_i &= y_i - \hat{y}_i = (y_i - \bar{y}) - (\hat{y}_i - \bar{y}) \\
\sum_{i=1}^{n} e_i^2 &= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}((y_i - \bar{y}) - (\hat{y}_i - \bar{y}))^2 \\
&= \sum_{i=1}^{n}(y_i - \bar{y})^2 - 2\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{y}) + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \qquad (1)
\end{aligned}$$

Jetzt wirds kompliziert. Unsere Schätzer sind

$$\begin{aligned}
\hat{b} &= \frac{S_{xy}}{S_{xx}} \\
\hat{a} &= \bar{y} - \hat{b}\bar{x}
\end{aligned}$$

Unsere Vorhersagen sind

$$\hat{y}_i = \hat{a} + \hat{b}x_i$$

Also können wir statt $\bar{y}$ auch schreiben $\bar{y} = \hat{a} + \hat{b}\bar{x}$ und in $\hat{y}_i - \bar{y}$ einsetzen:

$$\hat{y}_i - \bar{y} = \hat{a} + \hat{b}x_i - (\hat{a} + \hat{b}\bar{x}) = \hat{b}(x_i - \bar{x}) \qquad (2)$$

Setzt man nun (2) in den Term $\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{y})$ aus (1) ein erhält man:

$$\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{y}) = \sum_{i=1}^{n}(y_i - \bar{y})\hat{b}(x_i - \bar{x}) = \hat{b}\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) = \hat{b}S_{xy} \qquad (3)$$

Aus der Formel $\hat{b} = \frac{S_{xy}}{S_{xx}}$ lässt sich wiederum folgern dass

$$S_{xy} = \hat{b}S_{xx}$$

sodass man (3) mit Hilfe von (2) nun weiter umformen kann zu

$$\hat{b}S_{xy} = \hat{b}^2 S_{xx} = \hat{b}^2\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}\hat{b}^2(x_i - \bar{x})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

Wir haben damit also gezeigt dass

$$\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{y}) = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

Damit lässt sich dieser Term wieder in (1) einsetzen:

$$\begin{aligned}
\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 &= \sum_{i=1}^{n}(y_i - \bar{y})^2 - 2\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{y}) + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \\
&= \sum_{i=1}^{n}(y_i - \bar{y})^2 - 2\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \\
&= \sum_{i=1}^{n}(y_i - \bar{y})^2 - \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2
\end{aligned}$$

Also gilt:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$