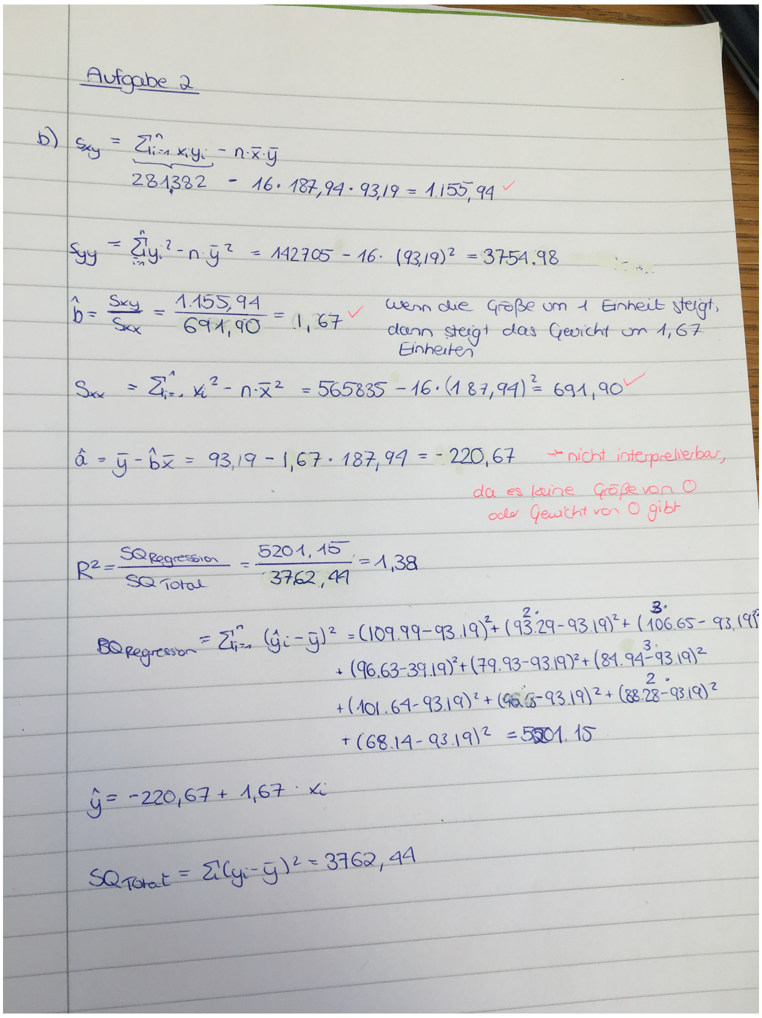
Frage:

Ich rechne gerade das ÜB 9 und die Ergebnisse ergeben keinen Sinn und von der Lösung die ich habe stimmt das auch nicht.



Aufgabe 2

b) $S_{xy} = \underbrace{\sum_{i=1}^{n} x_i y_i}_{281.382} - n \cdot \bar{x} \cdot \bar{y}$

$281.382 - 16 \cdot 187,94 \cdot 93,19 = 1.155,94$

$S_{yy} = \sum_{i=1}^{n} y_i^2 - n \cdot \bar{y}^2 = 142705 - 16 \cdot (93,19)^2 = 3754,98$

$\hat{b} = \frac{S_{xy}}{S_{xx}} = \frac{1.155,94}{691,90} = 1,67$ Wenn die Größe um 1 Einheit steigt, dann steigt das Gewicht um 1,67 Einheiten

$S_{xx} = \sum_{i=1}^{n} x_i^2 - n \cdot \bar{x}^2 = 565835 - 16 \cdot (187,94)^2 = 691,90$

$\hat{a} = \bar{y} - \hat{b}\bar{x} = 93,19 - 1,67 \cdot 187,94 = -220,67$ → nicht interpretierbar, da es keine Größe von 0 oder Gewicht von 0 gibt

$R^2 = \frac{SQ_{Regression}}{SQ_{Total}} = \frac{5201,15}{3762,44} = 1,38$

$SQ_{Regression} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = (109.99-93.19)^2 + 2\cdot(93.29-93.19)^2 + 3\cdot(106.65-93.19)^2 + (96.63-39.19)^2 + (79.93-93.19)^2 + 3\cdot(84.94-93.19)^2 + (101.64-93.19)^2 + (96.5-93.19)^2 + 2\cdot(88.28-93.19)^2 + (68.14-93.19)^2 = 5201.15$

$\hat{y} = -220,67 + 1,67 \cdot x_i$

$SQ_{Total} = \sum (y_i - \bar{y})^2 = 3762,44$

Um das Bestimmtheitsmaß, zu bestimmen kann man laut Formelsammlung $R^2 = \dfrac{SQ_{Regression}}{SQ_{Total}}$ rechnen, aber da kommt bei mir ein anderes Ergebnis raus, als wenn man $R^2 = \dfrac{S_{xy}}{S_{xx} * S_{yy}}$ wie in der Übung rechnen würde.

Wieso ist das so? Ich mache doch die Rechnung nach der Formelsammlung? Und wie kommt man drauf anstatt meiner Formel einfach den Korrelationskoeffizienten nach Bravais- Pearson zu quadrieren?? Wann erkennt man, wann man welche dieser Rechenschritte machen sollte?

Antwort:

In der Übung wurde zuerst der Korrelationskoeffizient nach Bravais-Pearson berechnet:

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}},$$

danach wurde dieser quadriert,

$$r^2 = \left(\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}\right)^2 = \frac{(S_{xy})^2}{S_{xx}S_{yy}}$$

Bei der Berechnung hast du in der Fragestellung vergessen, den Term S_{xy} im Zähler zu quadrieren.

In Zahlen

$$r^2 = \frac{1155,94}{\sqrt{691,90 * 3754,98}} = \frac{1155,94^2}{691,90 * 3754,98} = 0,51$$

Bei dieser Übung hängt das Gewicht nur von der Größe des Spielers ab, also von einer Einflußgröße.
Bei nur einer Einflußgröße gilt:

$$R^2 = r^2$$

In diesem Fall ist es schlicht schneller, R^2 über r^2 zu berechnen. Wie man sieht, kann man aber beide Wege nutzen.

Bei der Berechnung von $SQ_{Regression}$ musst du dich verrechnet haben:

Richtig ist (wenn man alle $\hat{y}_i$ vorhergesagt hat):
z.B $\hat{y}_1 = -220,67 + 198 \cdot 1,67 = 109,99$

$$\begin{aligned}
SQ_{Regression} &= (109,99 - 93,19)^2 + (93,29 - 93,19)^2 + (106,65 - 93,19)^2 \\
&+ (96,63 - 93,19)^2 + (79,93 - 93,19)^2 + (84,94 - 93,19)^2 \\
&+ (106,65 - 93,19)^2 + (106,65 - 93,19)^2 + (101,64 - 93,19)^2 \\
&+ (84,94 - 93,19)^2 + (84,94 - 93,19)^2 + (96,63 - 93,19)^2 \\
&+ (88,28 - 93,19)^2 + (93,29 - 93,19)^2 + (88,28 - 93,19)^2 \\
&+ (68,24 - 93,19)^2 \\
&= 1971,64
\end{aligned}$$

In R^2 eingesetzt:

$$R^2 = \frac{1971,64}{3762,40} = 0,52$$

Der Unterschied zwischen den Lösungen $0,51$ und $0,52$ ist schlichtweg auf die Rundung der Parameter $\hat{a}$ aund $\hat{b}$ zurückzuführen, die dann ja für jede Beobachtung zur Vorhersage verwendet wurden.