Regressionsgerade

Gegeben sind n Punkte $P_1(x_1|y_1), P_2(x_2|y_2) \dots P_n(x_n|y_n)$, die eine *Punktwolke* darstellen. Durch diese Punktwolke soll eine Gerade g mit $g(x) = mx + b$ so gelegt werden, dass die Summe der quadrierten Abweichungen der Punkte von der Geraden g in y-Richtung minimal wird. Die Gerade mit dieser Eigenschaft heißt *Regressions- oder Ausgleichsgerade*.

1. Betrachtet man die beiden arithmetischen Mittel $\bar{x}$ für die x- und $\bar{y}$ für die y-Werte mit

   $$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ und } \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

   so liegt es nahe, die Gerade durch den *Schwerpunkt* der Punktwolke $M(\bar{x}|\bar{y})$ gehen zu lassen. Dies ergibt folgende Beziehung:

   $$g(\bar{x}) = \bar{y} \iff m\bar{x} + b = \bar{y} \iff b = \bar{y} - m\bar{x}.$$

   Hat man das Problem gelöst, die Steigung m zu bestimmen, dann ergibt sich aus der letzten Gleichung der Achsenabschnitt b.

2. Die Beziehung für den Achsenabschnitt b wird in die Gleichung der Geraden eingesetzt:

   $$g(x) = mx + b = mx + \bar{y} - m\bar{x} \iff$$

   $$g(x) = m(x - \bar{x}) + \bar{y}.$$

3. An einer Stelle x_i ist die Abweichung des Punktes $P(x_i|y_i)$ von der Geraden in y-Richtung quadriert folgender Term:

   $$\begin{aligned} (g(x_i) - y_i)^2 &= [m(x_i - \bar{x}) + \bar{y} - y_i]^2 \\ &= [m(x_i - \bar{x}) - (y_i - \bar{y})]^2 \\ &= m^2(x_i - \bar{x})^2 - 2m(x_i - \bar{x})(y_i - \bar{y}) + (y_i - \bar{y})^2 \end{aligned}$$

4. Die Summe S aller Abweichungen erfordert eine Summation aller Abweichungen der Punkte P_1 bis P_n. Mit dem vorher gefundenen Ergebnis ergibt sich dann:

   $$S = \sum_{i=1}^{n}\left[m^2(x_i - \bar{x})^2 - 2m(x_i - \bar{x})(y_i - \bar{y}) + (y_i - \bar{y})^2\right]$$

   Klammert man die Konstanten m und $2m$ aus und zerlegt man die große Summe, so ergibt sich

   $$\begin{aligned} S &= m^2 \underbrace{\sum_{i=1}^{n}(x_i - \bar{x})^2}_{S_x} - 2m \underbrace{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}_{S_{xy}} + \underbrace{\sum_{i=1}^{n}(y_i - \bar{y})^2}_{S_y} \\ &= S_x m^2 - 2S_{xy}m + S_y \end{aligned}$$

   mit den Abkürzungen S_x, S_{xy} und S_y.

5. Die Summe S kann als quadratische Funktion von m aufgefasst werden. Graphisch dargestellt ergäbe sie eine nach oben geöffnete Parabel, weil $S_x > 0$. Eine nach oben geöffnete Parabel nimmt ihren minimalen Funktionswert in ihrem Scheitelpunkt an. Also muss nur noch die m-Koordinate des Scheitelpunktes durch Umformung der Parabelgleichung in die Scheitelpunktsform bestimmt werden (quadratische Ergänzung):

$$\begin{aligned}
S(m) &= S_x m^2 - 2S_{xy}m + S_y \\
&= S_x \left[m^2 - 2\frac{S_{xy}}{S_x}m + \frac{S_y}{S_x}\right] \\
&= S_x \left[m^2 - 2\frac{S_{xy}}{S_x}m + \left(\frac{S_{xy}}{S_x}\right)^2 - \left(\frac{S_{xy}}{S_x}\right)^2 + \frac{S_y}{S_x}\right] \\
&= S_x \left[\left(m - \frac{S_{xy}}{S_x}\right)^2 - \left(\frac{S_{xy}}{S_x}\right)^2 + \frac{S_y}{S_x}\right] \\
&= S_x \left(m - \frac{S_{xy}}{S_x}\right)^2 + \left(S_y - \frac{S_{xy}^2}{S_x}\right)
\end{aligned}$$

Die m-Koordinate des Scheitelpunktes ist $m = \frac{S_{xy}}{S_x}$. Damit ist die Steigung der gesuchten Geraden gefunden.

6. Für die praktische Anwendung der gefundenen Formel macht man noch einige Vereinfachungen:

$$\begin{aligned}
m &= \frac{S_{xy}}{S_x} \\
&= \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2} \\
&= \frac{\sum\limits_{i=1}^{n}(x_i y_i - x_i\bar{y} - \bar{x}y_i + \bar{x}\bar{y})}{\sum\limits_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)} \\
&= \frac{\sum\limits_{i=1}^{n} x_i y_i - \bar{y}\sum\limits_{i=1}^{n} x_i - \bar{x}\sum\limits_{i=1}^{n} y_i + \bar{x}\bar{y}\sum\limits_{i=1}^{n} 1}{\sum\limits_{i=1}^{n} x_i^2 - 2\bar{x}\sum\limits_{i=1}^{n} x_i + \bar{x}^2\sum\limits_{i=1}^{n} 1} \\
&= \frac{\sum\limits_{i=1}^{n} x_i y_i - \bar{y}n\bar{x} - \bar{x}n\bar{y} + \bar{x}\bar{y}n}{\sum\limits_{i=1}^{n} x_i^2 - 2\bar{x}n\bar{x} + \bar{x}^2 n}
\end{aligned}$$

$$= \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

7. Zusammenfassung:

$$m = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \qquad \text{und} \qquad b = \bar{y} - m\bar{x}$$

8. **Mathematisches Beispiel:**

   Gegeben seien die 5 Punkte $P_1(1|4)$, $P_2(2|6)$, $P_3(3|7)$, $P_4(5|8)$ und $P_5(7|7)$. Zeichne die Punkte in ein geeignet skaliertes Koordinatensystem, bestimme m und b und zeichne die gefundene Gerade ebenso in das Koordinatensystem.

x_i	y_i	$x_i y_i$	x_i^2
1	4		
2	6		
3	7		
5	8		
7	7		
$\sum_{i=1}^{5} x_i =$	$\sum_{i=1}^{5} y_i =$	$\sum_{i=1}^{5} x_i y_i =$	$\sum_{i=1}^{5} x_i^2 =$

$\bar{x} =$ $\qquad$ $\bar{y} =$

9. **Physikalisches Beispiel:**

Bei seinen Experimenten zum Photoeffekt erhielt MILLIKAN 1916 die folgenden Messwerte:

λ_i in nm	f_i in 10^{14} Hz	W_i in eV	$f_i W_i$	f_i^2
254		3,04		
312,5		2,14		
365		1,61		
546		0,48		
	$\sum_{i=1}^{4} f_i =$	$\sum_{i=1}^{4} W_i =$	$\sum_{i=1}^{4} f_i W_i =$	$\sum_{i=1}^{4} f_i^2 =$

$\bar{f} =$ $\bar{W} =$

Bestimme daraus die Werte für

(a) das Plancksche Wirkungsquantum in eVs.
(b) das Plancksche Wirkungsquantum in Js.
(c) die Ablösearbeit W_A in eV.
(d) die Ablösearbeit W_A in J.
(e) die Grenzfrequenz f_{grenz}.
(f) die Grenzwellenlänge λ_{grenz}.
(g) die prozentuale Abweichung des experimentellen Wertes des Planckschen Wirkungsquantums vom Literaturwert $h = 6{,}626176 \cdot 10^{-34}$ Js.