Anlage zum AK-Bericht Nr. 2

Zusammenstellung der im Bericht verwendeten mathematischen Formalismen

1. **Korrelationskoeffizient** (Bartsch s. 573)

   Zusammenhang zwischen der Testreaktion x und anderen Zustandsparametern y

$$r_{xy} = \frac{m(xy) - m(x)m(y)}{\sqrt{(m(x^2) - m(x)^2)(m(y^2) - m(y)^2)}} \qquad \textbf{(1)}$$

   dabei ist m(x) = Mittelwert von x_i, m(y) = Mittelwert von y_i und
   m(xy) = Mittelwert von $x_i y_i$, also

$$m(x) = \tfrac{1}{n}\sum x_i \;;\; m(x^2) = \tfrac{1}{n}\sum x_i^2 \;;\; m(y) = \tfrac{1}{n}\sum y_i \;;\; m(y^2) = \tfrac{1}{n}\sum y_i^2 \text{ und}$$

$$m(xy) = \tfrac{1}{n}\sum x_i y_i \qquad \textbf{(2)}$$

2. **Regression** über die y_i-Werte, ausgehend vom Punkthaufen x_i, y_i mit der Anzahl n.
   Es wird nur die lineare Regression mit $y = a_0 + a_1 x$ betrachtet. **(3)**
   (siehe Brauch,Dreyer,Hacke s.665)

   Minimierung über die y_i-Werte mit der Quadratsumme

$$Q = \sum_{i=1}^{i=n} (y_i - a_0 - a_1 x_i)^2 = \sum v^2 \qquad \textbf{(4)}$$

   Bestimmung von a_0 und a_1 mit den 2 Gleichungen für das Q-Minimum

$$\frac{\partial Q}{\partial a_0} = -2\sum (y_i - a_0 - a_1 x_i) = 0 \text{ mit } 0 = -2n(m(y) - a_0 - a_1 m(x)) \qquad \textbf{(5a)}$$

$$\frac{\partial Q}{\partial a_1} = -2\sum (y_i - a_0 - a_1 x_i) x_i = 0 \text{ mit } 0 = -2n(m(xy) - a_0 m(x) - a_1 m(x^2)) \qquad \textbf{(5b)}$$

   Aus (5a) $a_0 = m(y) - a_1 m(x)$ eingesetzt in (5b), dann wird

$$a_1 = \frac{m(xy) - m(x)m(y)}{m(x^2) - m^2(x)} \qquad (6a)$$

$$a_0 = \frac{m(y)m(x^2) - m(x)m(xy)}{m(x^2) - m^2(x)} \qquad (6b)$$

Wird die Testreaktion x gesetzt mit Wertebereich 0 bis 1, dann kann die Regressionsgerade (3) mit $x=0$ und $y(0)=a_0$

$$x=1 \text{ und } y(1)=a_0+a_1 \tag{7}$$

gut gezeichnet werden.

3. Krümmung von Q im Minimum

Ausgehend von (4) wird

$$\begin{aligned} Q &= \sum\left[(y_i - a_0) - a_1 x_i\right]^2 = \sum\left[(y_i - a_0)^2 - 2a_1(y_i - a_0)x_i + a_1^2 x_i^2\right] \\ &= \sum\left[y_i^2 - 2a_0 y_i + a_0^2 - 2a_1 x_i y_i + 2a_0 a_1 x_i - a_1^2 x_1^2\right] \\ &= n\left[m(y^2) - 2a_0 m(y) + a_0^2 - 2a_1 m(xy) + 2a_0 a_1 m(x) + a_1^2 m(x^2)\right] \end{aligned} \tag{8}$$

Nach Bronstein S.205 ist mit (5a) und dem daraus folgenden

$$\frac{\partial^2 Q}{\partial a_0^2} = 2n$$

der Krümmungsradius im Minimum in a_0-Richtung

$$R_{a_0} = \frac{\left[1 + \left(\dfrac{\partial Q}{\partial a_0}\right)^2\right]^{3/2}}{\dfrac{\partial^2 Q}{\partial a_0^2}} = \frac{1}{2n} \tag{9a}$$

und mit (5b) und dem daraud folgenden

$$\frac{\partial^2 Q}{\partial a_1^2} = 2nm(x^2)$$

der Krümmungsradius im Minimum in a_1-Richtung

$$R_{a_1} = \frac{\left[1+\left(\frac{\partial Q}{\partial a_1}\right)^2\right]^{3/2}}{\frac{\partial^2 Q}{\partial a_1^2}} = \frac{1}{2nm(x^2)} \tag{9b}$$

Beide Krümmungsradien sind von der y-Streuung unabhängig, sind aber um so kleiner , je größer die Messzahl n ist. Sie zeigen, wie nah das Q dem Nullwert liegt. Q=0 würde ja größte Korrelation der x_i- und der y_i-Werte bedeuten.

4. Die Größe Q und der Zusammenhang mit der Korrelation r_{xy}.

(siehe auch Bartsch S.573).

Man geht von der entwickelten Gleichung (8) aus und ersetzt dort zunächst a_0 aus (5a) mit $a_0 = m(y) - a_1 m(x)$ und bekommt

$$\frac{Q}{n} = \left(m\left(y^2\right) - m^2(y)\right) - a_1^2\left(m\left(x^2\right) - m^2(x)\right) \tag{10a}$$

Nun wird mit dem Zähler von (6a) der Zähler von (1) kombiniert, so dass

$$a_1 = r_{xy}\sqrt{\frac{m\left(y^2\right) - m^2(y)}{m\left(x^2\right) - m^2(x)}} \tag{10b}$$

wird, das man wiederum in (10a) einsetzt und damit

$$\frac{Q}{n} = \left(1 - r_{xy}^2\right)\left(m\left(y^2\right) - m^2(y)\right) \tag{10c}$$

bekommt. Wählt man noch den rechten Klammerausdruck zu z, dann hat man das sogenannte Unabhängigkeitsmaß $M = \frac{Q}{nz} = 1 - r_{xy}^2$ (11)

eine noch schärfere Größe als der Korrelationskoeffizient r_{xy}.

Ist die Korrelation r_{xy}=1 (also 100%), dann wird M=Q=0.

Ist dagegen r_{xy}=0, also kein Zusammenhang, dann ist M=Q/nz = 1.

M=Q/nz liegt immer zwischen 0 und 1 wie r_{xy} auch, nur entgegengesetzt.

Für Korrelation r_{xy}=0 wird auch a_1=0 (siehe Gl.10b), d.h. die Regressionsgerade liegt horizontal und das quadratische Mittel Q ist

$$Q_{\min} = n\left[m\left(y^2\right) - m^2\left(y\right)\right]$$

Es mittelt also die y_i-Werte auf den Mittelwert von y_i , also m(y), die horizontale Gerade. a_0 ist dabei (siehe 5a) a_0 =m(y).

5. Die Zuverlässigkeit der Werte a_0 und a_1(siehe auch Brauch,Dreger,...S.665)

Die a_0 und a_1 aus der Regression sind Schätzwerte. Für sie lassen sich mittlere Fehler angeben, die von der gewünschten Aussagesicherheit abhängen, nämlich dem t-Faktor der Studentverteilung.

Die mittleren Fehler sind

$$\Delta a_0 = tS_y \sqrt{\frac{\sum x_i^2}{n\sum x_i^2 - \left(\sum x_i\right)^2}} \qquad (12a)$$

$$\Delta a_1 = tS_y \sqrt{\frac{n}{n\sum x_i^2 - \left(\sum x_i\right)^2}} \qquad (12b)$$

Dabei ist $$S_y^2 = \frac{\sum v_i^2}{n - \left(m+1\right)} = \frac{Q_{\min}}{n-2} \qquad (12c)$$

wobei nach (4) ja $Q_{min}=\sum v_i^2$, n die Zahl der Messungen und m der Grad des Ausgleichpolynoms ist. Wenn mit $y=a_0+a_1x$ ein lineares Polynom genommen wird, ist ja m=1.

Weiter werden die Gln.(12) mit $\sum x_i = nm(x)$ und

$$\begin{aligned}\sum x_i^2 &= \sum \left[x_i - m(x) + m(x)\right]^2 \\ &= \sum \left[(x_i - m(x))^2 + 2(x_i - m(x))m(x) + m^2(x)\right] \\ &= \sum (x_i - m(x))^2 + nm^2(x) = (n-1)S_x^2 + nm^2(x)\end{aligned}$$

weil die Streuung $S_x^2 = \frac{1}{n-1}\sum (x_i - m(x))^2 = \frac{n}{n-1}\left[m(x^2) - m^2(x)\right]$ ist.

Damit wird dann endgültig aus (12)

$$\Delta a_0 = tS_y \sqrt{\frac{(n-1)S_x^2 + nm^2(x)}{n(n-1)S_x^2 + n^2m^2(x) - n^2m^2(x)}} = \frac{tS_y}{\sqrt{n}}\sqrt{1 + \frac{n}{n-1}\frac{m^2(x)}{S_x^2}} \qquad (13a)$$

$$\Delta a_1 = tS_y \sqrt{\frac{n}{n(n-1)S_x^2 + n^2m^2(x) - n^2m^2(x)}} = \frac{tS_y}{\sqrt{n-1}S_x} \qquad (13b)$$

Die mittleren Fehler werden um so kleiner, je größer die Zahl der Messungen ist. Für die Unsicherheit der Regressionsgeraden nach unten oder oben ist das Δa_0 zuständig und damit in erster Linie die Schwankung von y_i, also S_y.

Werden die Messungen bei großem x_i und einer kleinen Schwankung S_x durchgeführt, dann wird

$$\Delta a_0 \approx \frac{tS_y}{\sqrt{n-1}}\frac{m(x)}{S_x}$$

also unverständlich groß. Dies rührt daher, dass große Bereiche von x_i nicht mit Messungen abgedeckt sind, sie hier berücksichtigt werden sollten.

Im Fall von Messungen großer x_i-Werte in einem kleinen Bereich, auf den man sich beschränken will oder muß, ist es zweckmäßig, die x-Koordinatenachse um m(x) zu verschieben. Dann liegt nämlich der Nullpunkt der neuen x-Koordinatenachse im alten m(x) und das neue m(x) verschwindet. Dann wird aus (13a)

$$\Delta a_0 \approx \frac{tS_y}{\sqrt{n}} \tag{13c}$$

wie es auch sofort verständlich ist.

T ist der Studentparameter und wird Tabellen (z.B. Brauch...) entnommen, z.B. für 95%-ige Sicherheit. Der Nenner in (12c) ist der Freiheitsgrad $f = n - 2$ (Zahl der überschießenden Gleichungen bei der Bestimmung von a_o und a_1).

Es folgen dazu 3 Beispiele, die in den Bilder des anderen Anhangs dargestellt sind.

A) Diagramm Luftdruck/Testreaktion

$a_0 = 766{,}28135$ $\qquad a_1 = 6{,}32315113$

t = 2,306 bei f = 8 und 95%-ige Sicherheit, n=10

$$S_x = \sqrt{\frac{10}{9}(0{,}336 - 0{,}2116)} = 0{,}37 \;, S_y = \sqrt{\frac{Q_{\min}}{8}} = \sqrt{\frac{149{,}31109}{8}} = 4{,}32$$

$$\Delta a_0 = \frac{2{,}306x4{,}32}{\sqrt{10}} = 3{,}1525 \;; \qquad \Delta a_1 = \frac{2{,}306x4{,}32}{\sqrt{9}x0{,}37} = 8{,}909$$

Das Ergebnis ist in das zuständige Diagramm gestrichelt eingezeichnet. Die Verschiebung der Regressionsgeraden um Δa_0 umfasst praktisch die Streupunkte. Die Versteilerung der Geraden um Δa_1 liegt sehr einleuchtend in der Punkteschar.

B) Um das Verfahren mit einer 100%-igen Korrelation zu prüfen, wird die Testreaktion mit sich selber korreliert.

Dabei ist $S_y = \sqrt{\frac{Q_{\min}}{8}} = 0$

Und damit auch $\Delta a_0 = \Delta a_1 = 0$

C) Zur Überprüfung der Wetterdaten von den offiziellen Stationen und den eigenen Messungen wurde auch hier eine Regression vorgenommen. Es ergibt sich aus den Daten

-Luftdruck off./Luftdruck eig.-folgende Sicht:

$$a_0 = -37{,}9173168$$
$$a_1 = 1{,}28722073$$
$$S_y = \sqrt{\frac{Q^o_{\min}}{8}} = \sqrt{\frac{9{,}7883034}{8}} = 1{,}11$$
$$S_x = \sqrt{\frac{Q^e_{\min}}{8}} = \sqrt{\frac{149{,}311}{8}} = 4{,}32$$

Hier ist angebracht, $\Delta a_0 \approx \frac{tS_y}{\sqrt{n}}$, also Gl.(13c) zu benutzen.

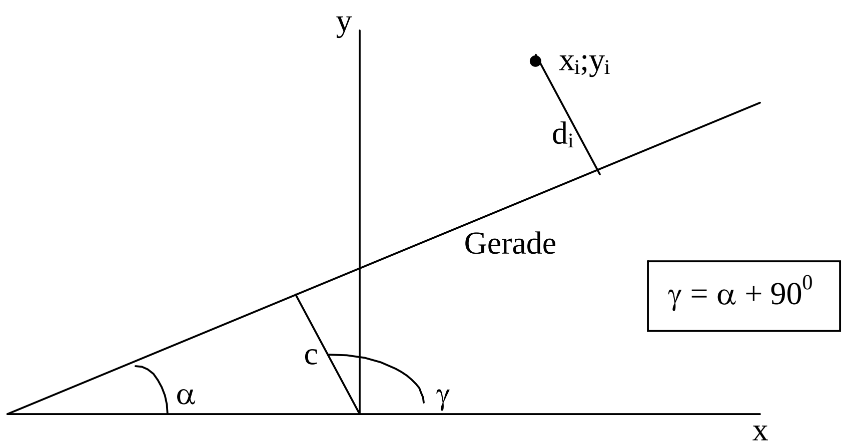
Dann ergibt sich

$$\Delta a_0 = \frac{2{,}306 x 1{,}11}{\sqrt{10}} = 0{,}81$$

$$\Delta a_1 = \frac{2{,}306 x 1{,}11}{\sqrt{9 x 4{,}32}} = 0{,}197$$

Die Unsicherheit für die Regressionsgeraden liegt sowohl für Ihre Lage als auch ihre Richtung in einem sehr kleinen Intervall, d.h. die Übereinstimmung der beiden Luftdruckmessungen ist recht gut.

6. Regression über die Abstandsminimierung der Messpunkte



Man geht aus von der Hesse'schen Normalform einer Geraden

$$x \cos\gamma + y \sin\gamma - c = 0$$

oder $$-x \sin\alpha + y \cos\alpha - c = 0 \qquad (14)$$

Für den Punktabstand d_i von der Geraden gilt dann

$$d_i = -x_i \sin\alpha + y_i \cos\alpha - c$$

Die Abstände d_i sind zu minimisieren mit

$$Q = \sum \left(y_i \cos\alpha - x_i \sin\alpha - c \right)^2 \Rightarrow Minimum \quad \text{zu}$$

(1)

$$\frac{\partial Q}{\partial \alpha} = \sum \left(y_i \cos\alpha - x_i \sin\alpha - c \right)\left(- y_i \sin\alpha - x_i \cos\alpha \right) = 0$$

$$m(y^2)\sin\alpha \cos\alpha - m(xy)\cos^2\alpha + m(xy)\sin^2\alpha + m(x^2)\sin\alpha \cos\alpha +$$
$$cm(y)\sin\alpha + cm(x)\cos\alpha = 0$$

$$- m(xy)\cos 2\alpha + \frac{1}{2}\sin 2\alpha \left(m\left(x^2\right) - m\left(y^2\right) \right) + c\left(m(y)\sin\alpha + m(x)\cos\alpha \right) = 0$$

(2)

$$\frac{\partial Q}{\partial c} = m(y)\cos\alpha - m(x)\sin\alpha - c = 0$$
$$c = m(y)\cos\alpha - m(x)\sin\alpha$$

Aus (1) und (2) ergibt sich

$$m(xy)\cos 2\alpha - \frac{1}{2}\left(m\left(x^2\right) - m\left(y^2\right) \right)\sin 2\alpha =$$

$$\left(m(y)\cos\alpha - m(x)\sin\alpha \right)\left(m(y)\sin\alpha + m(x)\cos\alpha \right) = \frac{1}{2}\left(m^2(y) - m^2(x) \right)\sin 2\alpha +$$
$$m(x)m(y)\cos 2\alpha$$

Zusammengefasst wird daraus

$$\tan 2\alpha = 2\frac{m(xy) - m(x)m(y)}{\left[m(x^2) - m^2(x) \right] - \left[m(y^2) - m^2(y) \right]} \qquad (15)$$
$$c = m(y)\cos\alpha - m(x)\sin\alpha$$

Die Gleichung der Regressionsgeraden ist nach (14)

$$y = \frac{c}{\cos\alpha} + x\tan\alpha \qquad (16)$$

Wird die Testreaktion wieder x gesetzt mit Wertebereich 0 bis 1, dann kann die Regressionsgerade (16) mit

$$x = 0 \text{ und } y(0) = c/\cos\alpha$$

$$x = 1 \text{ und } y(1) = c/\cos\alpha + x \tan\alpha$$ gezeichnet werden.

7. Eine Bemerkung zur $\tan 2\alpha$-Formel

Ist volle Korrelation mit $r_{xy} = 1$ gegeben, dann wird nach (4)

$$m(xy) - m(x)m(y) = \sqrt{\left[m(x^2) - m^2(x)\right]\left[m(y^2) - m^2(y)\right]}$$

und damit(14a) $\tan 2\alpha = 2\dfrac{ab}{a^2 - b^2}$ mit $\begin{array}{l} a^2 = m(x^2) - m^2(x) \\ b^2 = m(y^2) - m\ (y) \end{array}$

bzw. dann auch $\tan\alpha = \dfrac{b}{a} = a_1$ (siehe Gl.(10a)), wie es auch nach der anderen Art der Fall ist.

Ist aber keine Korrelation $r_{xy}=0$ gegeben, dann wird $m(xy) - m(x)m(y) = 0$

und $$\tan 2\alpha = 2\frac{0}{\left[m(x^2) - m^2(x)\right] - \left[m(y^2) - m^2(y)\right]} = 0$$

2α ist demnach 0° oder 180°, also α=0° oder 90°.

Bei α=0° ist die Punktwolke horizontal gestreckt, bei α=90° dagegen vertikal.

Ist dazu auch noch a=b, also in x-Richtung und y-Richtung kein Unterschied in der Punkteverteilung (runde Punktwolke). Dann wird

$$\tan 2\alpha = 0/0 = 0$$ also unbestimmt

und α kann jede Richtung gleichberechtigt annehmen.