

Paper-ID: VGI_196903



Ausreißerkriterien

K. Pauly ¹

¹ *Centro de Geodesia de la Universidad de Chile, Santiago de Chile, Casilla 2777*

Österreichische Zeitschrift für Vermessungswesen **57** (1), S. 13–21

1969

Bib_TE_X:

```
@ARTICLE{Pauly_VGI_196903,  
  Title = {Ausreißerkriterien},  
  Author = {Pauly, K.},  
  Journal = {{\u}sterreichische Zeitschrift f{{\u}r Vermessungswesen},  
  Pages = {13--21},  
  Number = {1},  
  Year = {1969},  
  Volume = {57}  
}
```



passung an die Wirklichkeit ist hervorragend, besonders wenn man bedenkt, daß ja nur von den Bodenwerten ausgegangen wurde. Man darf im übrigen hoffen, daß über längere Strecken bzw. größere Höhenunterschiede, wie sie besonders bei Shoranmessungen auftreten, sich die Anomalien weitgehend kompensieren. Leider ist noch kaum Beobachtungsmaterial vorhanden, das die Frage zu beantworten erlaubt, wie groß die horizontalen Erstreckungen von Dampfdruckanomalien sind. Dazu wären gleichzeitige Sondenaufstiege an nicht zu weit entfernten Stationen nötig.

Literatur:

- [1] *Jordan-Eggert-Kneißl*: Handbuch der Vermessungskunde, Band VI.
- [2] *Robitzsch, M.*: Die mittlere Abnahme des Dampfdruckes mit der Höhe. Meteorologische Zeitschrift, Band 61 (1944), Seite 273.
- [3] *Löser, H.-G.*: Untersuchungen zur praktischen Berechnung von Refraktionsanomalien . . . , DGK, Reihe B, Heft Nr. 36.
- [4] *Cehak, K.*: Zehnjährige Mittelwerte der meteorologischen Elemente in der freien Atmosphäre bis 30 km über Wien, Archiv für Meteorologie, Geophysik und Bioklimatologie, Serie A, Band 15, Heft 2 (1966).
- [5] *Aerologische Berichte, Radiosondenaufstiege und Höhenwindmessungen*. Zentralanstalt für Meteorologie und Geodynamik, Publikation Nr. 169.

Ausreißerkriterien

Von *K. Pauly*, Santiago de Chile

A) Allgemeines

Jeder messende Ingenieur wird einmal vor dem Problem gestanden haben, daß in einer Serie von Messungen der gleichen Größe ein oder mehrere Ergebnisse auftraten, von denen er vorgezogen hätte, daß sie nicht vorhanden seien.

Diese sogenannten „Ausreißer“ sind, falls es sich nicht um ganz extreme Werte handelt, meistens nicht erklärbar. Außerdem ist es natürlich nicht zulässig, diese Werte rein gefühlsmäßig auszusondern, denn die daraus resultierende scheinbare Genauigkeitssteigerung des Restsystems kann zu falschen Schlüssen über die Systemgenauigkeit führen.

Durch die zunehmende Einführung von physikalischem Gerät in die Geodäsie, wie Kreisel, Gravimeter, elektromagnetische Entfernungsmesser, ist dieses Problem in den letzten Jahrzehnten wieder besonders aktuell geworden. Es handelt sich nämlich um Instrumente, bei denen der Genauigkeitsgrad nicht mehr allein von der Geschicklichkeit des Beobachters, und der Fertigungsgenauigkeit abhängt; denn infolge der komplizierten Technologie können Teile oder die Gesamtheit einer oft sprunghaften Veränderung unterliegen, die die Meßgenauigkeit beeinflußt*).

*) In diesem Zusammenhang sei auf die interessante Arbeit von Baumeister, Die Zuverlässigkeit von Flugreglern, in *Flugwelt* 10/65, S. 858–859, hingewiesen.

Das Problem der „Ausreißer“ besteht also grundsätzlich darin, innerhalb einer Meßserie einen Grenzwert zu bestimmen, von dem man behaupten kann, daß alle Werte, die jenseits dieses Maximalwerts liegen, auszuschließen sind.

Diesen kritischen Wert hat man als Vielfaches eines der bekannten Genauigkeitsmaße, z. B. des n -fachen wahrscheinlichen Fehlers oder des n -fachen mittleren Fehlers auszudrücken. Im allgemeinen muß man zwei Fälle der Grenzwertbestimmung unterscheiden:

1. Man wünscht bei einer bereits vorliegenden Meßreihe die Ausreißer zu ermitteln und damit zu eliminieren.

2. Für künftige Messungen ist der größte zu erwartende Fehler einer Messung wieder als Funktion des wahrscheinlichen oder mittleren Fehlers zu bestimmen.

Dazu müßte natürlich zuerst eine Untersuchung nach 1) stattgefunden haben, da nach Elimination der Ausreißer der mittlere Fehler des Restsystems kleiner und damit das System wahrscheinlicher wird.

Es soll hier nur der Fall 1. dargestellt werden, da bekanntlich die Auswertung der Wahrscheinlichkeitsfunktion

$$P_0 = \frac{2}{\sqrt{\pi}} \int_0^{h\bar{m}} e^{-u^2} du \quad (**) \quad \dots (1)$$

einen 3- bis 4-fachen mittleren Fehler als Grenzfehler akzeptieren läßt. Hiernach wird für $3\bar{m}$ bereits eine Wahrscheinlichkeit von 0.997 erreicht, d. h. bei 1000 Fehlern liegen (absolut genommen) 997 innerhalb der Grenzen 0 und $3\bar{m}$.

Es wurde oben mit Absicht gesagt, daß die Untersuchung über Ausreißer wieder aktuell geworden ist, denn einige grundsätzliche Untersuchungen hierüber sind bereits um die Jahrhundertwende veröffentlicht [1] und auch in der neuen deutschen geodätischen Literatur wieder angewandt worden [2]. Es besteht jedoch die begründete Vermutung, daß diese Untersuchungen einem großen Teil der Leser nicht bekannt sind.

Diese in Vergessenheit geratenen Dinge der Fachwelt wieder näherzubringen, ist die Absicht des Verfassers.

B) Kriterien für Ausreißer

I. Kriterium von Chauvenet

$$P_0^{nr} = \frac{2}{\sqrt{\pi}} \int_0^{hnr} e^{-u^2} du = \frac{2}{\sqrt{\pi}} \left[0,4769 n - \frac{(0,4769 n)^3}{3 \cdot 1!} + \frac{(0,4769 n)^5}{5 \cdot 2!} - \frac{(0,4769 n)^7}{7 \cdot 3!} + \dots \right] \quad \dots (2)$$

ist die Wahrscheinlichkeit, daß ein Fehler zwischen Null und den n -fachen wahrscheinlichen Fehler fällt. [3] [4]

** \bar{m} sei im folgenden der mittlere quadratische Fehler des Systems.

Ist m die Anzahl der Beobachtungen, so fallen $m \cdot P_0^{nr}$ Fehler innerhalb dieses Bereiches und

$$m - m \cdot P_0^{nr} = m(1 - P_0^{nr}) \quad \dots (3)$$

außerhalb desselben.

Nehmen wir jetzt m als Bezugspunkt an, so folgt daraus, daß, wenn die Differenz $< 0,5$ ist, ein Fehler von der Größe $\geq n \cdot r$ eine geringere Wahrscheinlichkeit hat, und deshalb auszuschneiden ist.

Man kann also setzen

$$\frac{1}{2} = m - m P_0^{nr} \quad \text{und} \quad P_0^{nr} = \frac{2m - 1}{2m} \quad \dots (4)$$

Mit P_0^{nr} und einer Tabelle für das Integral (2) wird der zugehörige Wert n_r gefunden. Da $\frac{r}{m} = 0,6745$, wird $\frac{n_r \cdot r}{m} = n_r \cdot 0,6745 = k \quad \dots (5)^*$

Dann ist der Grenzfehler $n_r \cdot r$, oder wie ihn Chauvenet angibt, mit

$$k = n_r \cdot 0,6745 \text{ gleich } k \cdot \bar{m}.$$

Dieser Weg ist noch geringfügig zu vereinfachen, was aber von Bedeutung ist,

wenn keine Tafel für $P_0^{n \cdot r}$ sondern nur für

$$P_0^{\bar{m}} = \frac{2}{\sqrt{\pi}} \int_0^{h\bar{m}} e^{-u^2} du = \sqrt{\frac{2}{\pi}} \left(n - \frac{1}{2} \cdot \frac{n^3}{3 \cdot 1!} + \frac{1}{2^2} \cdot \frac{n^5}{5 \cdot 2!} - \frac{1}{2^3} \cdot \frac{n^7}{7 \cdot 3!} + \dots \right)$$

zur Verfügung steht. \dots (6)

Aus einer Tafel für P_0^{nr} erhalten wir n_r , aus einer Tafel für $P_0^{\bar{m}}$ erhalten wir $n_{\bar{m}}$.

Wir müssen also $n_r \cdot 0,6745 \cdot \bar{m}$ bilden, um den Grenzfehler zu erhalten, oder, was dasselbe ist, $n_r \cdot r$.

Der Wert $n_{\bar{m}}$ ist also identisch mit obigem k , also

$$P_0^{n_r \cdot r} = P_0^{\bar{m}} = \frac{2m - 1}{2m}$$

$$n_r \cdot 0,6745 = n_{\bar{m}} = k \quad \dots (7)$$

Grenzfehler $k \cdot \bar{m}$

Häufig wird auch nur eine Tabelle für das Wahrscheinlichkeitsintegral

$$P_{-a}^{+a} = \frac{2}{\sqrt{\pi}} \left(ah - \frac{(ah)^3}{3 \cdot 1!} + \frac{(ah)^5}{5 \cdot 2!} - \frac{(ah)^7}{7 \cdot 3!} + \dots \right) \quad \dots (8)$$

angegeben, wobei P_{-a}^{+a} die Wahrscheinlichkeit bedeutet, daß ein Fehler zwischen den

Grenzen $-a$ und $+a$ liegt, und zwar als Funktion von $(a \cdot h)$. [6]

In diesem Fall ist der Tafel für P_{-a}^{+a} der zugehörige Wert (ah) zu entnehmen; man eliminiert dann h mit der Beziehung

$$h = \frac{1}{\bar{m} \sqrt{2}}$$

und erhält bei bekanntem m in diesem Fall direkt den Grenzfehler

$$(ah) \cdot \sqrt{2} \cdot \bar{m} = a$$

Es läßt sich also zusammenfassend für den Grenzfehler aufstellen:

$$n_r \cdot r = n_{\bar{m}} \cdot \bar{m} = a$$

$$\frac{r}{\bar{m}} = \frac{n_{\bar{m}}}{n_r} = 0,6745$$

$$n_{\bar{m}} = (ah) \sqrt{2}$$

Wahrscheinlichkeit und Grenzfehler:

$$P_0^{n\bar{m}} = \int_0^{h n \bar{m}} n m$$

$$P_0^{nr} = \int_0^{h n r} n \cdot r$$

$$P_0^a = \int_0^{ah} a$$

... (9)

II. Kriterium nach Peirce

Im folgenden soll gelten:

u = Anzahl der Unbekannten

m = die Zahl aller Beobachtungen

n = die Anzahl der auszuschließenden Beobachtungen

n' = $m - n$ die Anzahl der verbleibenden Beobachtungen

m_1, m_2 = die mittleren Fehler des 1. bzw. des 2. Systems

y = die Wahrscheinlichkeit für einen Ausreißer

y' = $1 - y$ die Wahrscheinlichkeit, daß eine Beobachtung kein Ausreißer ist

k = Faktor, der den Grenzfehler bestimmt

$\varepsilon_1, \varepsilon_2$, Fehler im 1. bzw. 2. System, wobei System I = ohne Ausschluß und System II = mit Ausschluß bedeuten soll.

Die Wahrscheinlichkeit eines Fehlers ε ist

$$\varphi(\varepsilon) = \frac{h}{\sqrt{\pi}} e^{-h^2 \varepsilon^2} \quad \dots (10)$$

$P = \varphi(\varepsilon) \cdot \varphi(\varepsilon') \cdot \varphi(\varepsilon'')$ die Gesamtwahrscheinlichkeit
oder bei n Beobachtungen

$$P = \frac{h^n}{(\sqrt{\pi})^n} e^{-h^2 [\varepsilon\varepsilon]}$$

Die Wahrscheinlichkeit, daß ein Fehler größer als $k m_1$ ist, beträgt

$$2 \int_{\varepsilon = k m_1}^{\varepsilon = \infty} \varphi(\varepsilon) d\varepsilon$$

oder durch entsprechende Substitution mit (10) und $h = \frac{1}{m \sqrt{2}}$

$$\psi k = \frac{2}{m_1 \sqrt{2\pi}} \int_{k m_1}^{\infty} e^{-\frac{\varepsilon\varepsilon}{2m_1^2 d\varepsilon}} \dots (10a)$$

Substituieren wir jetzt in (10a)

$$\varepsilon = k \cdot m_1 \text{ durch } t = \frac{\varepsilon}{m_1 \sqrt{2}}$$

so wird mit $t m_1 \sqrt{2} = k m_1$, $t = \frac{k}{\sqrt{2}}$

$$\psi k = \frac{2}{\sqrt{\pi}} \int_{t = \frac{k}{\sqrt{2}}}^{t = \infty} e^{-t^2} dt \dots (11)$$

das ist die Wahrscheinlichkeit für einen Fehler, der über $\frac{k}{\sqrt{2}}$ liegt, ein Integral, das sich aus einer Tabelle für Formel (9) berechnen läßt.

Nach Chauvenet-Peirce ist die Wahrscheinlichkeit für das System I, einschließlich der Bedingung, daß n Beobachtungen jenseits des Grenzfehlers $k m_1$ liegen [1], S. 560.

$$P = [\varphi(\varepsilon)]^m \cdot \left[\frac{\psi k}{\varphi(k m_1)} \right]^n$$

Daraus läßt sich mit der Beziehung $n' = m - n$ herleiten

$$P = \frac{\left(\frac{h}{\sqrt{\pi}}\right)^m e^{-\frac{\varepsilon\varepsilon}{2m_1^2}}}{\left(\frac{h}{\sqrt{\pi}}\right)^n e^{-n\left(\frac{k}{\sqrt{2}}\right)^2}} \cdot (\psi k)^n = \frac{1}{m_1 (2\pi)^{\frac{1}{2}n'}} e^{-\frac{[\varepsilon\varepsilon] - n k^2 m_1^2}{2m_1^2}} \cdot (\psi k)^n$$

Ist n' die Zahl der Fehler, die nach Aussonderung übrig sind, so wird

$P^* = \frac{1}{m_1 n' (\sqrt{2\pi})^{n'}} e^{-\frac{[\varepsilon\varepsilon] - n k^2 m_1^2}{2 m_1^2}}$ das ist die Wahrscheinlichkeit aller Fehler innerhalb der Grenzen, was durch $n k m_1$ charakterisiert ist,

$P^{**} = (\psi k)^n$ die Wahrscheinlichkeit, daß n Fehler über die Grenzen $k m_1$ hinausgehen,

$P = P^* P^{**}$ ist die Wahrscheinlichkeit des Gesamt-Systems.

Da nun $[\varepsilon\varepsilon] = (m - u) m_1^2$ ist, . . . (12)

wird daraus

$$P_1 = \frac{1}{m_1 n' (2\pi)^{\frac{1}{2} n'}} \cdot e^{\frac{1}{2} (-m+u+nk^2)} \cdot (\psi k)^n \quad \dots (13)$$

Peirce postuliert nun [1]: Die Wahrscheinlichkeit des Fehlersystems einschließlich der Ausreißer (P_1) ist geringer als das ohne Ausreißer multipliziert mit der Wahrscheinlichkeit, daß die und nur die Ausreißer vorkommen (P_2), wobei

$$P_2 = \frac{1}{m_2 n' (2\pi)^{\frac{1}{2} n'}} \cdot e^{-\frac{[\varepsilon_2 \varepsilon_2]}{2m_2^2}} \cdot y^n \cdot y'^n \quad \dots (14)$$

wir haben also als Kriterium

$P_1 < P_2$ oder ausführlicher

$$\left(\frac{m_2}{m_1}\right)^{n'} e^{-\frac{1}{2} m + \frac{1}{2} u + \frac{1}{2} nk^2 + \frac{1}{2} n' - \frac{1}{2} n} \cdot (\psi k)^n < y^n y'^n$$

$$\left(\frac{m_2}{m_1}\right)^{n'} e^{\frac{1}{2} n(k^2-1)} (\psi k)^n < y^n y'^n \text{ und mit } m - n = n' \quad \dots (15)$$

Die Wahrscheinlichkeit ist für $y = \frac{n}{m}$ und $y' = \frac{n'}{m}$

deren Summe Gewißheit sein muß, also $y + y' = 1$.

Somit erhält man:

$$y^n \cdot y'^n = \left(\frac{n}{m}\right)^n \cdot \left(\frac{n'}{m}\right)^{n'} = \frac{n^n \cdot n'^{n'}}{m^{n+n'}}$$

$$y + y' = \frac{n + n'}{m} = 1$$

$n + n' = m$ und damit

$$y^n \cdot y'^n = \frac{n^n \cdot n'^{n'}}{m^n} = T^n$$

$$e^{\frac{1}{2} (k^2-1)} \cdot (\psi k) = R$$

$$\boxed{\left(\frac{m_2}{m_1}\right)^{n'} = R^n = T^n} \quad \dots (16)$$

Dieser Ausdruck muß im engeren Sinne als Kriterium von Peirce bezeichnet werden.

Jetzt wird als Approximation, um das Verhältnis $\frac{m_2}{m_1}$ zu bestimmen, angenommen, daß die Differenz der Fehlerquadratsummen nicht größer sei als das n -fache Fehlerquadrat $(k m_1)^2$, also

$$\begin{aligned} [\varepsilon_1 \varepsilon_1] - [\varepsilon_2 \varepsilon_2] &= n K^2 m_1^2 & \dots (17) \\ (m - u) m_1^2 - (n' - u) m_2^2 &= n k^2 m_1^2 \\ (m - u) m_1^2 - (m - u - n) m_2^2 &= n k^2 m_1^2 \\ \frac{m - u - n k^2}{m - u - n} &= \left(\frac{m_2}{m_1}\right)^2 \end{aligned}$$

$$\left(\frac{m_2}{m_1}\right)^2 = \frac{m - u - n k^2}{m - u - n} = \left(\frac{T}{R}\right)^{\frac{2n}{m-u-n}} = \lambda^2 \text{ und damit}$$

$$\begin{aligned} m - u - n k^2 &= (m - u - n) \lambda^2 \\ 0 &= m - u - n k^2 - \lambda^2 m + \lambda^2 u + \lambda^2 n \end{aligned}$$

von beiden Seiten wird n subtrahiert

$$\begin{aligned} n k^2 - n &= m - u - n - \lambda^2 m + \lambda^2 u + \lambda^2 n \\ (k^2 - 1) &= \frac{m - u - n}{n} (1 - \lambda^2) & \dots (18) \end{aligned}$$

In [1] findet man Tafeln für k^2 als Funktion von u , m und n wie auch $\log T$ und $\log R$.

C) Beispiele

Im folgenden sollen die abgeleiteten Gleichungen an einigen Beispielen erläutert werden:

Aus einer Serie von 40 Kreismessungen, die 1965 mit einem TK 4 (Fennel) vom Verfasser ausgeführt wurden, ergab sich ein mittlerer Fehler von

$$m = \pm \sqrt{\frac{39,33}{39}} = \pm 1,0$$

Folgende Restfehler sind gegeben (in $^{\circ}$):

+ 0,68	+ 1,05	+ 0,10	+ 0,02
+ 0,41	+ 0,84	+ 0,20	- 0,54
+ 3,17	- 0,16	+ 0,14	+ 0,80
+ 0,28	- 1,14	- 0,48	+ 0,35
+ 0,48	- 1,93	+ 0,56	- 0,87

- 0,28	+ 0,40	- 0,22	- 0,68
+ 1,32	- 1,88	+ 0,01	- 0,23
+ 1,80	+ 0,31	- 0,70	+ 0,06
+ 1,26	- 0,61	- 1,05	- 1,40
+ 0,91	- 0,79	- 0,83	- 1,41

Eine Untersuchung, ob systematische Anteile vorhanden waren, ergab nach Cornu

$$m:t:r = 1:0,76:0,58$$

nach Wolf

$$\left| \frac{\pi}{2} \cdot \frac{t^2}{\bar{m}^2} - 1 \right| < G_{\Delta} \text{ mit } G_{\Delta 40} \approx 0,24$$

$$0,15 < 0,24$$

wodurch also bestätigt war, daß keine systematischen Fehler vorkommen [5].
Ausreißer:

a) *Nach Chauvenet*

$$P_0''' = \frac{79}{80} = 0,9875$$

$$n_r = 3,7, \quad n_{\bar{m}} = 2,5, \quad (\alpha \cdot h) \sqrt{2} = 2,5;$$

$$\frac{n_{\bar{m}}}{n_r} = 0,6745, \quad K = 2,5$$

$$K \cdot \bar{m} = 2,5 \cdot 1 = 2,5$$

es ist also eine Beobachtung auszuschließen.

Jetzt

$$P_0''' = \frac{77}{78} = 0,9872, \quad n_r = 3,69, \quad K = 2,48$$

$$K \cdot \bar{m} = 2,48 \cdot 0,88 = 2,2$$

es ist keine weitere Beobachtung auszuschließen.

b) *Nach Peirce*

$$K_{40}^2 = 6,270 \quad K_{40} = 2,5$$

$$K_{39}^2 = 4,925 \quad K_{39} = 2,2 \quad \text{Ergebnis wie unter a)}$$

Es muß hier erläutert werden, daß die n -te Beobachtung nicht ausgeschlossen werden kann, bevor die $(n - 1)$ te Beobachtung nicht ausgeschlossen wurde.

c) *Fall mit mehreren Unbekannten:*

Am 3. 3. 1967 wurden auf dem Astro-Pfeiler der TH der Universidad de Chile übungsmäßige Breitenbestimmungen mit einem T4-Wild nach Sterneck durchge-

führt, wobei eine Ausgleichung mit den Unbekannten φ und Z_0 (Zenitpunkt) erfolgte, die als wahrscheinlichsten Wert für

$$\varphi = -33^{\circ}27'29'',30$$

ergab.

$$\bar{m} = \pm \sqrt{\frac{44,5}{13-2}} = \pm 2''$$

$$M = \pm 0'',5$$

Restfehler (v) (in''):

1,05	0,17
0,18	0,62
3,62	0,65
2,07	0,58
2,21	1,57
2,28	2,75
1,40	

Die asymmetrische Anordnung soll hier nicht stören, da ja bei vollkommener Symmetrie und evtl. Ausschluß von Beobachtungen die Serie asymmetrisch werden kann.

Nach der in [1] enthaltenen Tafel wird k für $u = 2$ ermittelt und man erhält:

$$k^2 = 3,782$$

$$k = 1,95, \quad k \cdot \bar{m} = 3'',9$$

Als Ergebnis erhalten wir also, daß keine Beobachtung ausgeschlossen werden darf.

Abschließend können wir feststellen, daß der Vorteil der ziemlich verwickelten Theorie von Peirce über das Princip von Chauvenet nicht erwiesen ist, selbst nicht mit der Begründung, daß sie die Anwendung auf mehr als eine Unbekannte erlaubt.

Im übrigen verweisen wir auf [4], S. 365 und 366, wo darauf aufmerksam gemacht wird, „vorsichtig zu sein, vor allem bei geringer Anzahl der Beobachtungen“.

Literatur:

- [1] *Chauvenet*: A Manual of Spherical Astronomy, Vol II, Dover 1960 (reprint).
- [2] *Schrück, K. W.*: Astronomisch-geodätische Messungen, DGK, Reihe 8, Nr. 85.
- [3] *Jordan*: Handbuch der Vermessungskunde 1. Bd., 5. Aufl. 1904.
- [4] *Helmert*: Die Ausgleichsrechnung nach der Methode der kleinsten Quadrate, 2. Aufl. B. G. Teubner, 1907.
- [5] *Grossmann*: Grundzüge der Ausgleichsrechnung, 2. Aufl. Springer-Verlag 1961.
- [6] *Rinehart*, Mathematical Tables, Formulas and Curves, Holt, Rinehart & Winston, New York 1960, p. 162 ff.