

Excel 12.0 Fehler im Beispiel nachgewiesen

Standardauswertungen in Kalibration, Trend, Haltbarkeit



► Dr. Stefan Schömer, Geschäftsführer, QM Balance

Forderungen zu Linearitätstests, beispielsweise aus Normen DIN 32645, DIN 38402 zur Kalibration ebenso geläufig wie in Richtlinien der ICH (hier z. B. Q2R), ist Ausgangspunkt unsere Beispielauswertung. Ein konkretes Zahlenbeispiel zeigt den in Excel 12.0 auftretenden, gefährlichen Fehler.

Tab. 1: Daten einer Kalibration

x [mg]	x [g]	Signal [arb.]
0,1	0,0001	0,168
0,2	0,0002	0,205
0,3	0,0003	0,334
0,4	0,0004	0,416
0,5	0,0005	0,519
0,6	0,0006	0,65
0,7	0,0007	0,739
0,8	0,0008	0,88
0,9	0,0009	0,968

Wir beschreiben erstmals einen erheblichen Fehler der Regressionsrechnung unter Excel 12.0. Der Fehler betrifft die zentrale Funktion „=RGP()“ bzw. „=Linest()“ und führt bereits mit bislang völlig unkritischen, ja unverdächtigen Daten zu unvorhergesehenen Falschergebnissen, also zu dem „größten anzunehmender Fehler“ (Lit. 1) einer Tabellenkalkulation. Das Feld der Betroffenen ist unabsehbar: Kalibrier- und Prüflabore, Produktions-, F&E-Abteilungen, Institute, Forschungseinrichtungen.

Grundlegendes Revalidieren firmeneigener Excel-Arbeitsblätter bis hin zum Festlegen neuer „kritischer Daten“ und Gültigkeitsgrenzen scheint branchenübergreifend unumgänglich.

Die Kalibrierproben mit Konzentrationsangaben in mg bzw. in g sind in Tabelle 1 identischen Signalwerten zugeordnet. Statt der in Normen

Tab. 2: Regressionskenndaten der Kalibration mit Excel 12.0 $y=a_1x+a_0$

x-Achse mit Angaben	in [mg]	in [g]
Steigung a_1	1,04483	1044,83333
Achsenabschnitt a_0	0,01969	0,01969
Bestimmtheitsmaß r^2	0,99348	0,99348
Reststandardabweichung s_{yx}	0,02479	0,02479
F-Statistik	1066,07262	1066,07262
Freiheitsgrad	7	7
QS der Regression	0,65501	0,65501
QS der Residuen	0,0043	0,0043

empfohlenen Anzahl begrenzen die Zahl der x,y-Datenpaare 9. Die äquidistanten x-Werte überstreichen so gerade eine Größenordnung.

Die Soll-Kenndaten der Kalibrationsgerade in Tabelle 2 werden auch mit Excel 12.0 ermittelt.

Selbstverständlich kann das Ergebnis – die Steigung ausgenommen – nicht davon abhängen, in welchen Einheiten die Konzentrationswerte angegeben sind. Der erste Koeffizient a_1 – also die Steigung – ist in Einheiten [arb./g] statt in [arb./mg] natürlich um den Faktor 1000 größer.

Der geforderte Linearitätstest basiert in jedem Fall auf dem Vergleich der Ergebnisse aus Tabelle 2 mit einer Anpassung 2. Ordnung (Parabel), deren Sollergebnisse Tabelle 3 nennt und den unter Excel 12.0 ermittelten Ergebnissen gegenüberstellt.

Auch diesmal spiegeln die Sollergebnisse nur den Wechsel der Einheit in der entsprechenden

Tab. 3: Regressionskenndaten der Kalibration mit Excel 12.0 $y=a_2x^2+a_1x+a_0$

	Sollergebnisse	Excel 12.0 Ergebnisse
	x in [mg]	x in [g]
Koeffizient x^2 , a_2	0,25076	2,5076E+05
Koeffizient x, a_1	0,79408	794,07576
Achsenabschnitt a_0	0,06567	0,01969
Bestimmtheitsmaß r^2	0,99641	0,99348
Reststandardabweichung s_{yx}	0,01985	0,02479
F-Statistik	833,61738	1066,07262
Freiheitsgrad	6,0	7,0
QS der Regression	0,65694	0,65501
QS der Residuen	0,00236	0,0043

Tab. 4: Vorgabe Parabel
 $y=a_2x^2+a_1x+a_0=1*x^2+2*x+3*a_0$

x_0 -Werte	$x_{0krit.}$ -Werte	y-Werte
1,0	0,0001	6,0
2,0	0,0002	11,0
3,0	0,0003	18,0
4,0	0,0004	27,0
5,0	0,0005	38,0
6,0	0,0006	51,0
7,0	0,0007	66,0
8,0	0,0008	83,0
9,0	0,0009	102,0

Tab. 5: Regressionskennenden der Parabel $y=a_2x^2+a_1x+a_0=1*x^2+2*x+3*a_0$

	Sollergebnisse		Excel 12.0 Ergebnisse
Größenordnung x-Werte:	1	10^{-4}	10^{-4}
Koeffizient x^2 , a_2	1,0	1,0000E+08	0,0000E+00
Koeffizient x , a_1	2,0	2,0E+4	1,2000E+05
Achsenabschnitt a_0	3,0		-1,5333E+01
Bestimmtheitsmaß r^2	1,0		0,96558
Reststandardabweichung s_{yx}	0		6,63325
Freiheitsgrad	6		7,0

Größenordnung wider. Die eigens aufgeführten Ergebnisse von Excel 12.0 für die Angabe der Konzentrationen in [g] sind unvorhersehbar falsch. Der Koeffizient der Parabelanpassung a_2 ist wohl das auffälligste Falschergebnis! Beim genaueren Hinsehen schaltet Excel 12.0 ohne Vorwarnung auf die Auswertung der ausdrücklich hier nicht relevanten und nicht zu erwartenden Geradengleichung. In diesem speziellen Fall ist die Excel-Fehlberechnung nicht allzu kritisch, denn das Ergebnis im Soll-Linearitätstest favorisiert ohnehin die Gerade.

Kritischer ist natürlich, wenn Excel dieses Verhalten auch mit einer tatsächlichen Parabel zeigt. Tatsächlich trifft dieser schlimmste anzunehmende Fall zu. Die auszuwertenden Vorgabedaten mit Soll-Koeffizienten nennt Tabelle 4.

Der bislang unkritische Wechsel der Größenordnung der x-Werte sollte allein – ihrer Dimension entsprechend, bei ansonsten identischem Numerus – den Unterschied der Koeffizienten a_2 , a_1 , a_0 bestimmen. Tabelle 5 stellt die Sollergebnisse den mit Excel 12.0 ermittelten Ergebnissen gegenüber – keine schlechte Überraschung.

Wieder wechselt Microsoft Excel 12.0 unvorhersehbar, ohne ersichtlichen Grund zu falschen Ergebnissen einer Geradenanpassung. Wir dürfen abschließend betonen, Microsoft Corporation über dieses Verhalten der Excel Version 12.0 bereits am 3.12.2009 detailliert informiert zu haben, bislang jedoch ohne Reaktion. Nach unserer aktuellsten Recherche Ende Februar 2009 hat Microsoft den Fehler weder korrigiert noch informiert (Lit. 2).

Fazit

Der hier nachgewiesene Excel-Fehler zeigt natürlich Auswirkungen von weit größerer Reichweite und ist keineswegs auf Linearitätstests begrenzt. Halbarkeitsstudien mit unerkannten oder gar falsch prognostizierten Trends und in der Folge ungültigen Spezifikationsangaben sind leicht vorstellbar. Die Notwendigkeit umfangreicher Revalidierungen für alle Anwen-

der von Excel 12.0 scheint uns offensichtlich und unumgänglich. Nach Auskunft des Bundesamtes für Datensicherheit vom 3.12.2008 sind hierfür tatsächlich die Firmen als Anwender – also wohl nicht der Hersteller – verantwortlich. Zugegeben, das hat uns auch verwundert. Gibt es doch ein Produkthaftungsgesetz?

Ein Ausweg für Anwender von Excel 12.0 ist das strikte Normieren der Eingabebereiche auf definierte, unkritische Größenordnungen (Lit. 3). Für durchaus übliche Auswertungen über mehr als 4 Größenordnungen (Lit. 4) ergibt sich naturgemäß ein besonderes, kaum lösbares Problem. Der „=RGP()“-Funktion verwandte Funktionen wie „=Trend()“, „=Steigung()“, „=Achsenabschnitt“ oder „=STFehlerYX()“ lassen ähnliche Schwächen vermuten und bieten keine Lösung, den hier aufgedeckten Fehler zuverlässig zu vermeiden [Lit 2]. Einzelheiten zum Auftreten dieses Fehlers, zum Ausloten der Gültigkeitsgrenzen und ein erfolgreiches Konzept, den hier berichteten Fehler unter Excel 12.0 zuverlässig auszuschalten, stellen wir auf den Hompages www.qmbalance.com im Flash-Film vor.

Literatur

- [1] Sawitzki, G.: Computational Statistics & Data Analysis 18 269–286 (1994)
- [2] Quelle: Microsoft Knowledge Base und Support, Online, z. B. Artikel-ID: 214230, Artikel-ID: 828801 u. a.
- [3] Dr. St. Schömer, Validierungsbericht ProControl 6.x – Angewandte Statistik für Prüflabors und Industrie
- [4] Melissa, M.; Kiser, J.; Dolan, W.: LCGC Europe 17(3), 138 143 (2004)

KONTAKT

Dr. Stefan Schömer
 QM Balance
 Saarbücken
 Tel.: 0681/9762-730
 Fax: 0681/9762-733
schoemer@qmbalance.com
www.qmbalance.com