



RGP-Fehler in Excel 14.0 korrigiert

Rundungsfehler der Regression scheint behoben

Dr. Stefan Schömer

Gute Nachricht für alle Excel-Anwender. Ein Fehler der RGP-Funktion [1] wurde in Excel 2013 (Version 14.0.6129.5000) behoben. Prüflabore und Produktionsabteilungen können aufatmen. Ein kurzer Rückblick stellt mit kritischen Daten die richtigen Ergebnisse den Falschergebnissen in Excel 12.0 nochmals gegenüber.

Ein Rundungsfehler, inzwischen von Microsoft bestätigt [2], ist Anlass, die Grenzen numerischer Genauigkeit zu beleuchten. Die prinzipielle Eignung der RGP-Funktion zur Lösung linearer Regressionen wird nachgewiesen. Begriffe zur Regression zu klären greift Fragen und Rückmeldungen von Lesern auf.

Der Fehler: Funktionen RGP() und LINEST() unter Excel 12.0

Das Auswerten einer linearen Regression 2. Ordnung mit augenscheinlich unkritischen Daten (Tab. 1) endet unter Excel 12.0 unversehens in den Kenndaten einer Geraden.

Prüflaboratorien droht z.B. in Kalibrationen das Risiko einer unerkannten Nicht-Linearität wegen unzutreffender Rest-Standardabweichungen.

Daten einer tatsächlichen Parabel (Tab. 2) werden unter Excel 12.0 mit der RGP-Funktion als Gerade angepasst. Der Koeffizient a_2 in der Parabelgleichung ist ohne weitere Information für den Anwender auf „0“ gesetzt.

Die Messwerte, die korrekte Parabel und die mit Excel 12.0 erhaltene Gerade zeigt Abb. 1. Anwenderfehler, wie sie J. Bourquain unterstellt [4], sind ausgeschlossen, denn die identische Auswertung (nicht nur die gleiche Auswertung) liefert für Werte $>10^{-3}$ [willkürliche Einheiten] korrekte Ergebnisse.

Selbst in validierten und bisher fehlerfreien Auswertebättern drohen so unversehens falsche Ergebnisse. Risiken einer Falschbewertung sind offensichtlich, ob zu Kalibrationen in Prüflaboren oder zu Haltbarkeit und Spezifikationsfindung in Produktion und Entwicklung.

Ursache ist ein Rundungsfehler der Kollinearitätsprüfung. Der exakt gleiche Knowledge Base Eintrag [2] war noch im Jahr 2009 ohne einen solchen Hinweis [1]. Die Kollinearitätsprüfung ist mit Excel 11.0 etabliert.

Somit sind Trendauswertungen seit 2003 betroffen, laut Hersteller auch das Auswerten exponentieller Trends mit der RKP-Funktion. Diese Funktion erfreut sich in kaufmännischen Bereichen großer Beliebtheit.

Trendfunktion RGP() ohne Alternative

Kenndaten linearer Regressionen zu berechnen, wird in „einfachen Formeln“ mit steigender Ordnung sehr schnell komplex und unübersichtlich. Gerade aus Gründen der Validierung von Excel Arbeitsblättern, ist die RGP-Funktion vorzuziehen. Direkter Zugang zu Kenndaten eines Polynoms und zusätzlicher Schutz der „Bereichsformel“ vor unbeabsichtigten Änderungen sind echte Vorteile. „Einfachere Berechnungen“ zu favorisieren bleibt ein frommer Wunsch [4].

Verlockende Alternative und von Rundungsfehlern nicht betroffen ist die Trendlinie für Polynome im xy-Punktdiagramm. Kenndaten sind allerdings auf die Koeffizienten a_n und das Bestimmtheitsmaß r^2 beschränkt. Ein numerischer Linearitätstest

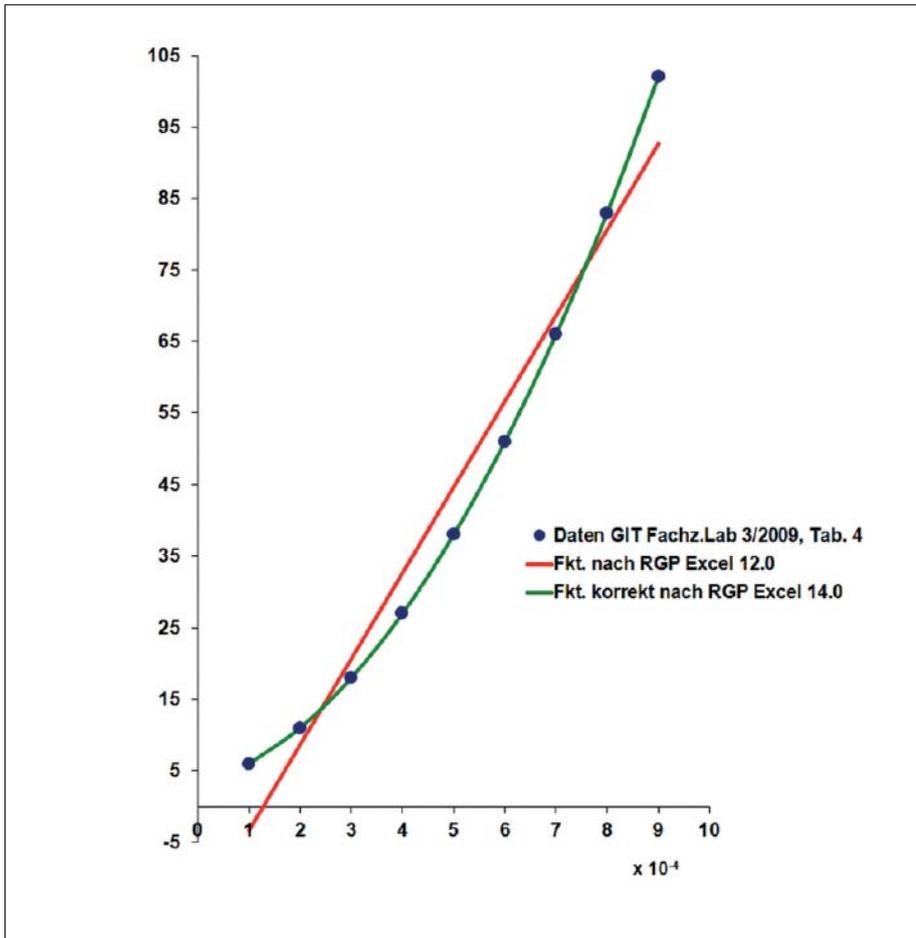


Abb. 1: Soll/Ist-Vergleich – RGP-Trendergebnis aus Daten der Tabelle 2

ist wegen der fehlenden Rest-Standardabweichung nicht unterstützt. Die Beschriftungen von Grafik-Elementen in einer Tabellenkalkulation sind zudem der Gefahr von Aktualisierungsfehlern stärker ausgesetzt. Diese Erfahrung ist am Beispiel eines einfachen Kreisdiagramms nachgewiesen [5]. Moderne ungleich leistungsstärkere Rechner beheben das Problem nicht, solange ein Überfluss an CPU-Ressourcen nur die Performance auf Nebenschauplätzen erhöht. Das Verwenden einer Diagramm-Beschriftung als Berechnungswerkzeug entspricht zudem nicht dem Verwendungszweck einer Tabellenkalkulation, namentlich also das Rechnen in Tabellen, eben nicht in Grafiken. Angesichts einer notwendigen Validierung von Excel-Arbeitsblättern sollte diese Möglichkeit als Alternative also ausscheiden.

Die Behauptung, die RGP-Funktion sei kaum verbreitet, geradezu exotisch und allein schon deshalb kaum jemand betroffen [4] verkehrt sich so ins Gegenteil. Eine fundierte Trendauswertung ohne die RGP-Funktion ist kaum denkbar oder nur durch komplexe eigene Berechnungen zu ersetzen [3].

Rundungsfehler für kleine Änderungen großer Zahlen

Fehleranfällig in numerischen Rechnungen sind die im „nasty dataset“ vorgestellten Datenreihen „big“ und „little“ [3,7,8]: „große“ Werte (10^8) mit Änderungen im einstelligen Bereich sowie „kleine“ Werte mit Änderungen nach 8 weiteren Stellen (10^{-8}). Absolut kleine oder große Werte, ja sogar „winzige“ und „riesige“ Werte – „tiny“ und „huge“ im „nasty dataset“ – sind meist vollkommen unkritisch. Ein richtiges Ergebnis oder aber eine korrekte Fehlermeldung ist zu erwarten [6].

Diese Eigenschaften sind in der binären Architektur der zentralen Rechereinheit begründet. Zur Vertiefung sei die Originalliteratur [7,8] empfohlen.

„Klein“ oder „groß“ ist immer ein relativer Begriff

Kleiner/Wachter [6] warnen zu Recht vor dem Problem kleiner Zahlen – genauer „kleine Differenzen großer Zahlen“. Ihrem Appell, sich der begrenzten Genauigkeit nu-

merischer Rechnungen stets bewusst zu sein, schließt sich der Autor des vorliegenden Artikels an.

Deren Beispiel, ein Multiplizieren 10^{-307} mit 0,1 ergibt „0“, verfehlt die Fragestellung. Ein aufmerksamer Anwender wird Ergebnisse jenseits offen mitgeteilter Genauigkeitsgrenzen nicht verwenden. Das Beispiel und die beiläufige Bewertung, das Ergebnis „0“ sei lediglich „aus mathematischer Sicht“ falsch, lässt in den Augen des Lesers ein tatsächliches Risiko in gefährlich weite Ferne, ins „Reich der Theorie“ entschwinden.

Gut zu wissen: Die Praxis mit Rundungsfehlern beginnt schon ab Werten von 0,009 (Tab. 2). Das Beispiel mit x-Werten um 10^{-4} erreicht mit Eingangsgrößen x^2 die kritische Größenordnung 10^{-8} . Ebenso die Abweichungsquadrate im Kollinearitätstest mit 10^{-8} bzw. 10^{-16} . Es entsteht also in beiden Eingangsgrößen eine kritische Größenordnung 10^{-8} . Microsoft selbst nennt 10^{-12} als kritisch: „...‘almost equal’ means a very small sum of squared deviations ...of c1, c2,...; ,very small’ might be less than 10^{-12} ...“ [2]. Die Eingangsgrößen c1 und c2 entsprechen übrigens den vorher genannten x und x^2 . Die Größenordnung der Werte trifft mit $9 \cdot 10^{-4}$ in den Abweichungsquadraten von x^2 den von Microsoft als kritisch genannten Punkt. Besonders kritisch ist das Überstreichen dieser Grenze

X [g]	Signal [willkürliche Einheiten]
0,0001	0,168
0,0002	0,205
0,0003	0,334
0,0004	0,416
0,0005	0,519
0,0006	0,65
0,0007	0,739
0,0008	0,88
0,0009	0,968

Tab. 1: Daten einer Kalibration [1,3]

x_{krit} -Werte	y-Werte
0,0001	6,0
0,0002	11,0
0,0003	18,0
0,0004	27,0
0,0005	38,0
0,0006	51,0
0,0007	66,0
0,0008	83,0
0,0009	102,0

Tab. 2: Vorgabe Parabel

$$y = a_2 x^2 + a_1 x + a_0 = 1 \cdot x^2 + 2 \cdot x + 3$$

innerhalb des hier gewählten Datensatzes. Die Funktion Trend(), auf die sich Microsoft bezieht, kann allerdings kaum betroffen sein. Die Eingabe mehrerer Eingangsgrößen, die einem Kollinearitätstest unterzogen werden könnten, ist für die Geraden-Anpassung in der Funktion Trend() nicht möglich.

Risiko nicht auf Excel beschränkt

Numerische Rechnungen und Computer mögen keine „kleinen Änderungen großer Zahlen“. Eine naturgemäß begrenzte Genauigkeit ist die Ursache solcher Rundungsfehler. Einfaches Summieren kann schon kritisch werden. Das Risiko ist nicht auf statistische Rechnungen und auch keineswegs auf Excel beschränkt. Es gilt vom Taschenrechner bis zu Giga-Herz getakteten CPUs für jegliche Soft- und Hardware.

Die ausgesprochene Warnung [1] war für viele Kunden hilfreich, zumal sich „ganz normale Daten“ ([mg] in [g] angegeben) schon als kritisch erweisen. Der Eindruck, dem Autor jenes Artikels ginge es um einen Programmierfehler [6] oder gar um Panikmache [4], sei damit ausgeräumt.

Begriffe und Lösungen zu Trendanalyse und Regressionsauswertungen

Die RGP Funktion ist Basis für Funktionen wie Trend() oder RKP() [2]. Zur Frage einiger Leser, ob es zulässig sei, eine Polynomanpassung mit der RGP-Funktion, also einer „mehrfachen Regression“ zu lösen, nennt Hartung [9] zur multiplen Regression: „Die Polynomanpassung geht demnach durch Substituieren der Matrix ihrer Eingangsgrößen ($x^0, x^1, x^2 \dots x^n$) durch die Matrix der Eingangsgrößen ($x_0, x_1, x_2, \dots, x_n$) in eine mehrfache Regression über“.

Die RGP Funktion ist damit geeignet, eine Polynomanpassung vorzunehmen. Exakt so, wie es die Excel-Online-Hilfe ausdrücklich beschreibt. Das Anpassen einer Geraden oder Parabel ist, exakt bezeichnet als lineare Regression 1. Ordnung bzw. 2. Ordnung zwanglos als eine mehrfache Regression lösbar. Es bedarf keineswegs eines „Kunstgriffes“ und einer Zweckentfremdung der RGP-Funktion [6].

Auch die notwendige Randbedingung linear unabhängiger Eingangsgrößen ist erfüllt. Die Eingangsgrößen x und x^2 sind niemals linear abhängig.

Die Bedeutsamkeit der Warnung des Autors wird durch Kleiner/Wachter eindrucksvoll bestätigt [6]. Deren Appell, Randbedingungen einzuhalten impliziert, wir hätten mit linear abhängigen Eingangsgrößen genau diese Randbedingungen missachtet. Diese Fehleinschätzung ist – Ironie des Schicksals – genau auf den Rundungsfehler zurückzuführen, vor dem gewarnt wurde.

Nach Kleiner/Wachter die multivariate Regression mit der RGP Funktion auszuwerten ist nicht möglich [6]. Sorry, das ist einfach „eine andere Baustelle“ [9].

Fazit und Ausblick

Frei nach Leonardo da Vinci: „Und es bewegt sich doch“, auch wenn es im Falle Microsoft mehr als drei Jahre dauerte. Die Beispieldaten erlauben ein einfaches Prüfen, ob Microsoft auch in Excel 15.0 bzw. Excel 365 der Kollinearitätsprüfung eine so erfreulich robuste Prozedur spendiert hat.

Literatur

- [1] Schömer S.: GIT Labor-Fachzeitschrift 3, 160 (2009)
- [2] Microsoft Knowledge Base und Support, Online, Artikel-ID: 828801, Rev. 18.9.2011
- [3] Schömer S.: QMBalance - Formeldokumentation & Validierungsbericht ProControl® 6.x - Angewandte Statistik für Prüflabors und Industrie, kritische Daten zum freien Download, www.qmbalance.com/pages/level1/procontrol_demo_dt.html
- [4] Bourquain J.: Keine Panik mit Excel, www.chemgineering.com
- [5] Sawitzki G.: Numerischer Test von Statistik Programmen, PC-Professionell 9, 34 (1994)
- [6] Kleiner J., Wachter G.: GIT Labor-Fachzeitschrift 5, 338 (2011)
- [7] Sawitzki G.: Computational Statistics & Data Analysis 18, 269-286 (1994)
- [8] Wilkinson L.: Statistics Quiz, Systat, Evanston, IL, 1985
- [9] Hartung J.: Lehr- und Handbuch der angewandten Statistik, 7. Auflage, 1989

KONTAKT |

Dr. Stefan Schömer
 QMBalance
 Saarbrücken
 Tel.: 0681/9762-730
schoemer@qmbalance.com
www.qmbalance.com