Manfred Borovcnik, Klagenfurt

Ideen zur Unterweisung in Statistik im Gewand von Tabellenkalkulation

Zusammenfassung:

In der Beschreibenden Statistik werden viele Begriffe für die Wahrscheinlichkeitsrechnung und Beurteilende Statistik vorbereitet. Man kann die graphischen Möglichkeiten, die heutzutage etwa durch den Einsatz von Tabellenkalkulation offen stehen, dafür einsetzen, um Lernenden ein vertieftes Verständnis zu ermöglichen. Dazu zählt etwa, was eine Kennziffer der Lage (Mittelwert oder Median etwa) über einen Datensatz aussagen kann. Oder, wie man die Breite eines Datensatzes durch Kennzahlen erfasst und was man aus der Standardabweichung eigentlich ablesen kann. Oder, welche Vorteile etwa Boxplots gegenüber den üblichen Histogrammen oder Stabdiagrammen bieten. Die vorliegenden Ausführungen werden in EXCEL präsentiert.

Einleitung

Der Mittelwert von Daten hat eine Reihe bekannter Eigenschaften. Hier sollen nicht alle wiederholt werden. Lediglich einige sollen herausgegriffen werden, die von einer graphischen Ergänzung besonders profitieren. Gleiches gilt für die Standardabweichung. Durch dynamische Graphiken sollen Vorteile und Nachteile spezieller Kennziffern (etwa im Vergleich Standardabweichung gegenüber dem Interquartil-Abstand) illustriert werden. Die Stabilität der Prozentsätze an Daten, die innerhalb von einer (zwei) Standardabweichungen vom Mittelwert liegen, kann durch Simulation geeigneter Daten schön illustriert werden – speziell die Möglichkeit, das gesamte Szenario der Simulation mit einem Knopfdruck zu erneuern, illustriert den Spielraum des Zufalls dynamisch. Diese so genannten s-Intervalle entsprechen in gewisser Weise zentralen Darstellungselementen des Boxplots, sodass Boxplot und Histogramm (mit diesen Ergänzungen) gar nicht so verschieden sind – oder doch?

1. Mittelwert und Median

Der Mittelwert kann als Ausgleichswert interpretiert werden: wenn alle gleich viel haben, haben alle so viel, wie der Mittelwert angibt. Der Median ist einfach die Mitte der (geordneten) Daten. Beide Interpretationen sind einfach. Auffallend jedoch ist, dass der Median nicht eindeutig bestimmt ist – entsprechend ist die Formel eigentlich kompliziert und die Auswertung in verschiedener Software unterschiedlich gelöst. Hier geht es jedoch um die Demonstration eines wesentlichen Unterschieds zwischen diesen beiden Kennziffern für die *Lage* der Verteilung (siehe Abb. 1): Während der Mittelwert auf Bewegungen des vierten Punktes (gesteuert über den Schieberegler) empfindlich reagiert, lässt das den Median unberührt, es sei denn, der Punkt überstreicht die gegenwärtige Lage des Medians.

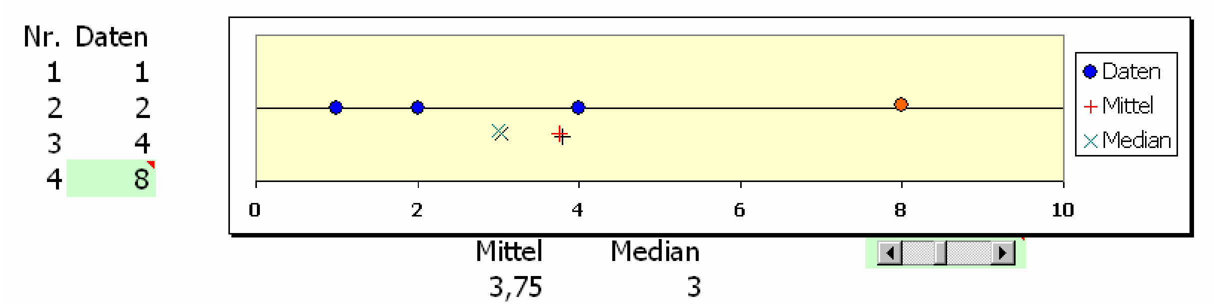


Abb. 1: Mittelwert und Median repräsentieren in unterschiedlicher Weise die Lage der Daten – das macht der bewegliche rechte Punkt deutlich

Der Mittelwert fasst alle Daten zusammen und reagiert empfindlich, wenn die Qualität der Daten unterschiedlich ist: kann sein, die Daten stammen nicht alle aus derselben Grundgesamtheit; oder es gibt Ausreißer unter den Daten. In beiden Fällen gibt es einzelne Daten, die eigentlich nicht „dazu“ gehören. In diesem Falle liefert der Mittelwert keine geeignete Zahl, welche die Lage der Daten anzeigt.

Der Median reagiert kaum auf Veränderung einzelner Daten, er ist robust dagegen; insbesondere ist er unempfindlich bezüglich Ausreißern. Wenn sich nur wenige Daten aus einer anderen Gesamtheit unter die Daten gemischt haben, werden sie nur wenig Einfluss auf seinen Wert haben. Der Wert des Medians zeigt also das generelle Muster in den Daten, unbeeinflusst von Besonderheiten.

2. Standardweichung als Maß für die Breite einer Verteilung und wozu das Maß noch dient

Anstelle einer Formel für die Standardabweichung

Die Standardabweichung gilt als das Maß für die Kennzeichnung der Breite einer Verteilung. Es wird jedoch kaum motiviert, wieso Abweichungen *von einem Bezugspunkt* (hier dem Mittelwert) und nicht Abweichungen der Daten untereinander; wieso nicht absolute Abweichungen etc. Mit einem ‚Schätzspiel’ für die Vorhersage einzelner Werte desselben Datensatzes kann man motivieren, dass Abweichungen von einem festen Wert zu nehmen sind: Kennt man Daten und ihren Mittelwert so ist – vorausgesetzt wir benutzen als Vorhersagewert das Mittel – die Abweichung vom Mittel unser Vorhersagefehler. Die Summe dieser Fehler ist Null, wie man schnell in einer Tabellenkalkulation sieht. Bleibt noch zu motivieren, warum die Quadrate (und nicht die Absolutbeträge) der Fehler summiert werden. Die Prozedur ähnelt (bis auf die Division mit n-1) der Festlegung der Länge von Vektoren, das ist ein Grund dafür. Ein weiterer kann aus einer Quadratzerlegung in Zusammenhang mit linearer Regression gewonnen werden (siehe Borovcnik 2006b; SQ Summe der Quadrate):

SQ Daten	=	SQ Regression	+	SQ Fehler
Totale Variation		Durch Modell *erklärte* Variation		*Nicht-Erklärte* Rest-Variation

Auch hier hilft die *Anordnung* der Daten in einem Tabellenblatt, damit man die *Beziehungen übersichtlich* darstellen und erkennen kann. In Tab. 1 wird statt der Formel für die Standardabweichung ein Blatt zu ihrer übersichtlichen Berechnung gezeigt.

	Daten	Mittel	ABW vom Mittel	Quadrierte ABW	ABW = Abweichung
	x_i	m	x_i - m	$(x_i - m)^2$	
	1	3,5	-2,5	6,25	
	2	3,5	-1,5	2,25	
	4	3,5	0,5	0,25	
	7	3,5	3,5	12,25	
Summe	14	14	0	21,00	
		Division durch n-1		7,00	s^2 Varianz
		Wurzel		2,65	s Standardabweichung

Tab. 1: Ausschnitt aus einem Tabellenblatt zur Berechnung der Standardabweichung – die übersichtliche Anordnung macht eine Formel überflüssig.

Vergleich mit Interquartil-Abstand – wie gut repräsentiert die Standardweichung die Daten

Viel wichtiger als eine Begründung, warum man ein Maß für die Breite einer Verteilung wie eben die Standardabweichung heranzieht, ist zu klären, wozu dieses Maß dienen kann, was man aus der Standardabweichung ablesen kann. Hier bietet sich als Alternative der so genannte Interquartil-Abstand an: Wie der Median in der Mitte der (geordneten) Daten steht, also salopp gesagt sind 50% der Daten darüber und 50% darunter (das ist nicht ganz richtig; die genaue Definition ist deswegen auch so kompliziert; darauf soll hier nicht eingegangen werden), kann man andere Punkte der Verteilung auszeichnen, die 100p% nach unten bzw. 100(1-p)% nach oben markieren. Solche Punkte nennt man p-Quantile; speziell sind die 25% und 75%-Quantile zu nennen, innerhalb derer (wieder salopp formuliert) die zentralen 50% der Daten liegen. Die Länge dieses Quartil-Intervalles (Quartil von Viertel; das Intervall reicht vom „1. Viertel" bis zum „3. Viertel"), der Interquartil-Abstand, ist ebenso geeignet, die Breite einer Verteilung zu kennzeichnen. Ein anderes Intervall wäre (2,5%-Punkt; 97,5%-Punkt), welches die inneren 95% der Daten zusammenfasst.

Wie gut kennzeichnen die Intervalle (Mittel – s, Mittel + s) bzw. (1. Viertel; 3. Viertel) die Breite einer Verteilung? Wieder kann man in einer Tabellenkalkulation einen Datensatz graphisch darstellen lassen und etwa einen Punkt beweglich gestalten (Abb. 2). Mit der Bewegung des einen Punktes nach außen wird schnell klar, dass das „1s-Intervall" ganz auseinander gezogen wird, während das Quartil-Intervall nicht reagiert. Der Quartil-Abstand repräsentiert die Breite einer Verteilung viel „robuster", d.h. er reagiert auf Änderungen einzelner Daten wenig bis gar nicht. Er zeigt auch viel besser auf, dass der *eine* Wert so weit draußen vielleicht gar nicht zu den Daten gehört (ist ein Ausreißer, gehört zu einer anderen Population etc. und sollte vielleicht besser eliminiert werden.) Allerdings fallen die beiden Intervalle bei Fehlen von solchen ausreißerverdächtigen Werten (mittlere Lage des beweglichen Punktes) ganz gut zusammen (abgesehen davon, dass das Quartil-Intervall dann etwas schmäler ausfällt).

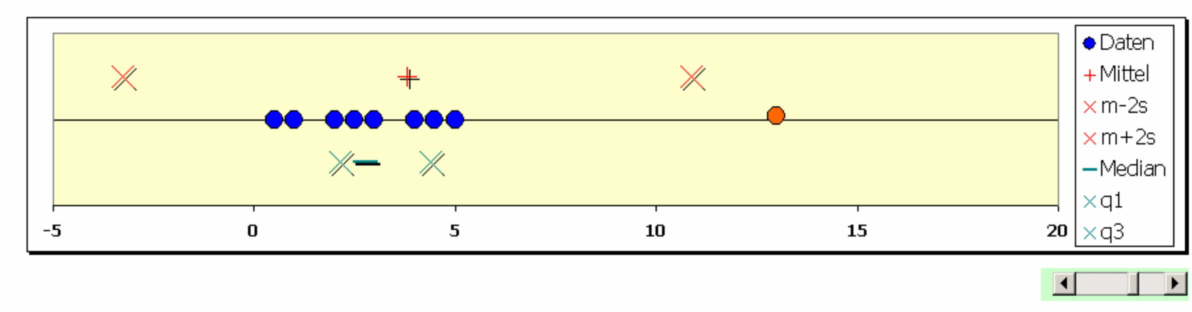


Abb. 2: Die Breite einer Verteilung wird durch Standardabweichung und Quartile ganz unterschiedlich erfasst – wie der bewegliche rechte Punkt deutlich macht. Das Quartil-Intervall bleibt starr, das 1s-Intervall reagiert voll auf jede Veränderung.

Die Möglichkeit, in einer Tabellenkalkulation relativ einfach dynamische Graphiken zu erzeugen, unterstützt den Vergleich enorm. Die Standardabweichung erscheint hier als optimiert für homogene Verteilungen – i.e. Verteilungen, die keine Ausreißer aufweisen, die gut ausgeprägt eingipfelig sind, die im Wesentlichen der Normalverteilung ähnlich sind. Durch Datenpflege wird man nicht nur zu diesem Zweck die Rohdaten bereinigen: Ausreißer entfernen, Daten trennen, wenn sie verschiedenen Gruppen zugehören und die Gruppen getrennt darstellen und analysieren. Die Beseitigung von solchen Störgrößen hilft auch bei der späteren klaren Interpretation der Ergebnisse enorm.

Die s-Regeln zur Interpretation der Standardabweichung

Wozu ist die Standardabweichung noch gut? Man kann aus ihr die so genannten s-Intervalle berechnen: (Mittel – *k**s, Mittel + *k**s). Nach der Tschebyschew-Ungleichung sind mindestens ¾ der Daten im 2*s-Intervall und mindestens 8/9 im 3*s-Intervall. Diese Abschätzungen sind jedoch nur theoretisch von Belang. (Aus der Tschebyscheff-Ungleichung kann man elementar das Schwache Gesetz der Großen Zahlen ableiten, wonach die Zufallsvariable, welche für Stichproben die jeweilige relative Häufigkeit eines fixen Ereignisses E als Wert annimmt, eine Verteilung hat, die sich um P(E) zusammenzieht.) Für praktische Zwecke kann man davon ausgehen, dass die *k**s-Intervalle etwa folgende Prozentsätze des gesamten Datensatzes „beherbergen“ (siehe Tab. 2; man muss schon lange nach Datensätzen suchen, in denen die angegebene Zahl unterschritten wird) – man nennt das die *s-Regeln*:

*k**s-Intervall	Enthält ca. so viele der Daten	Mindestanteile nach Tschebyschew-Ungleichung
m – s, m + s	2/3	—
m – 2*s, m + 2*s	95%	75%
m – 3*s, m + 3*s	99,7%	89%

Tab. 2: Ungefähre Prozentsätze an Daten, die innerhalb von symmetrischen Intervallen um den Mittelwert liegen – zum Vergleich die Mindestanteile dieser Intervalle nach der Tschebyschew-Ungleichung

Diese s-Regeln kann man an einzelnen Datensätzen überprüfen. In einer Tabellenkalkulation bietet sich hierzu die Simulation von Daten an – wie in folgendem Tabellenblatt (Abb. 3) wiedergegeben: Es werden 20 Münzwürfe simuliert, die Wahrscheinlichkeit für Kopf wird mit 0,4 unterstellt. Man kann andere Wahrscheinlichkeiten durch Änderung der 1-Einträge in der links stehenden Matrix für die Münze erreichen; man kann sich stattdessen auch einen Schieberegler basteln, der die Anzahl der 1-Einträge in der Matrix für die Münze steuert.

In der zweiten Spalte der rechten Tabelle wählt man zufällig eine Zahl aus der Menge {1, 2, .., 10} über die Funktion **ZUFALLSBEREICH** . Danach wählt man aus der Matrix für die Münze das dem Zufallsergebnis entsprechende Ergebnis des Münzwurfs aus; der Befehl hierzu lautet: **INDEX** bzw. **SVERWEIS** . Anschließend summiert man die Zahl der Köpfe über die gesamte Serie von 20 Würfen. Damit das Tabellenblatt immer übersichtlich bleibt, blendet man überzählige Zeilen mit Vorteil aus. Damit hat man aber erst eine einzige Serie mit 20 Münzwürfen realisiert. Durch eine geschickte Ausnutzung von Mehrfachoperationen erhält man hierauf etwa die 20 Würfe (mit Anzahl der Köpfe) gleich 25-mal wiederholt und das dann gleich in 3facher Ausfertigung, siehe die Ergebnistabelle in Abb. 3.

Zufall	Münze	
1	0	Z
2	0	Z
3	0	Z
4	0	Z
5	0	Z
6	0	Z
7	**1**	K
8	**1**	K
9	**1**	K
10	**1**	K

Die Zahl der 1'en bestimmt **p**

Nr.	Zufall ergibt	Münzwurf lautet
1	8	1
2	10	1
3	6	0
4	4	0
5	7	1
6	5	0
7	4	0
8	8	1
9	8	1
10	3	0
11	8	1
20	2	0
Anzahl Köpfe		12

Abb. 3: Links: Schematische Darstellung der Grundgesamtheit für den Münzwurf mit Wahrscheinlichkeit 0,4 für Kopf – Rechts: Simulation der 1. Serie von 20 Würfen mit Anzahl der Köpfe

Serie von 20 Würfen:			
	#1	#2	#3
	12	12	12
1	7	9	6
2	6	7	9
22	11	11	5
23	6	8	7
24	7	6	9
25	8	7	11

Tab. 3: Erzeugung der Daten durch eine trickreiche Anwendung von Mehrfachoperationen in EXCEL – Die Serie von 20 Münzwürfen wird 25mal wiederholt (jedes Mal Anzahl der Köpfe angegeben); insgesamt werden 3 solche Serien erzeugt

Um die ganze Wurfserie beliebig (hier 25mal) zu wiederholen, fügt man hierzu (in Tab. 3)

- in die oberste Zelle (die unterlegte 12) den Zellbezug auf das Ergebnis der ersten 20er Serie ein,
- lässt die Zelle links daneben leer (!),
- markiert die leere Zelle und die 12 sowie nach unten einen Bereich nach unten (dort werden die Wiederholungen der gesamten 20er Serie eingeschrieben) und
- wählt hernach **Daten >Tabelle > Werte aus Spalten** und
- fügt im Pop-up-Menü **Werte aus Spalten** den Bezug zu einer leeren (!) Zelle, etwa **Z100** ein.

Wie in der Ergebnisliste (Tab. 4) entstehen dann der Reihe nach die Daten; hier 7, 6, usw. (es wurden gleich drei Serien von Wiederholungen in Einem erzeugt). Man kann sich das so vorstellen:

- EXCEL sucht in der 2. Zeile in der leeren Zelle die Input-Größe; diese ist leer, provoziert aber eine Neuberechnung des Zufalls (die 12 als Anzahl der Köpfe wird neu berechnet als 7);
- die vom Input abhängige Größe wird berechnet; in diesem Fall ist dies gerade das Ergebnis dieses Zufalls (der Wert 7 für die Anzahl der Köpfe in der neuen Serie);
- dieser Wert wird in die Ergebnisspalte (=rechte Spalte) eingetragen;
- dann geht das Programm in die 3. Zeile und sucht sich den neuen Wert für den Input usw.

Von den so simulierten Daten kann man sich (für die 3 Serien getrennt, siehe Tab. 4) jeweils Mittel und Standardabweichung berechnen lassen und anschließend über einen geschachtelten **Wenn** – Befehl eine Indikatorvariable erzeugen lassen, die zählt, ob ein spezielles Datum innerhalb des z.B. 2*s-Intervalls liegt (=1; oder nicht =0). Von hier ist es nicht weit, die Anteile der Daten innerhalb der *k**s-Intervalle zu bestimmen.

Daten	7	8	6
...	1. Serie	2. Serie	3. Serie
Mittel	7,92	8,28	7,40
STABW	2,4	1,9	2,0
s-Intervall	5,5	6,3	5,4
	10,3	10,2	9,4
% innerhalb	76	68	68
2-s-Intervall	3,1	4,4	3,4
	12,7	12,2	11,4
% innerhalb	92	92	96
3-s-Intervall	0,7	2,4	1,5
	15,1	14,1	13,3
% innerhalb	100	100	100

Tab. 4: Berechnung der s-Intervalle für alle 3 Serien und Auszählung, wie viel Prozent der Daten innerhalb dieser Intervalle liegen

Im Beispiel etwa sind innerhalb des 2*s-Intervalls 92%, 92% bzw. 96% (siehe Tab. 4). Lässt man dann den Zufall über die gesamte Datensimulation erneuern (mittels Funktionstaste **F9**), so kann man rasch sehen, wie stabil eigentlich die s-Regeln sind. Unterstützt wird die „dynamische Stabilität“ der s-Regeln durch geeignete Stabdiagramme in Abb. 4.

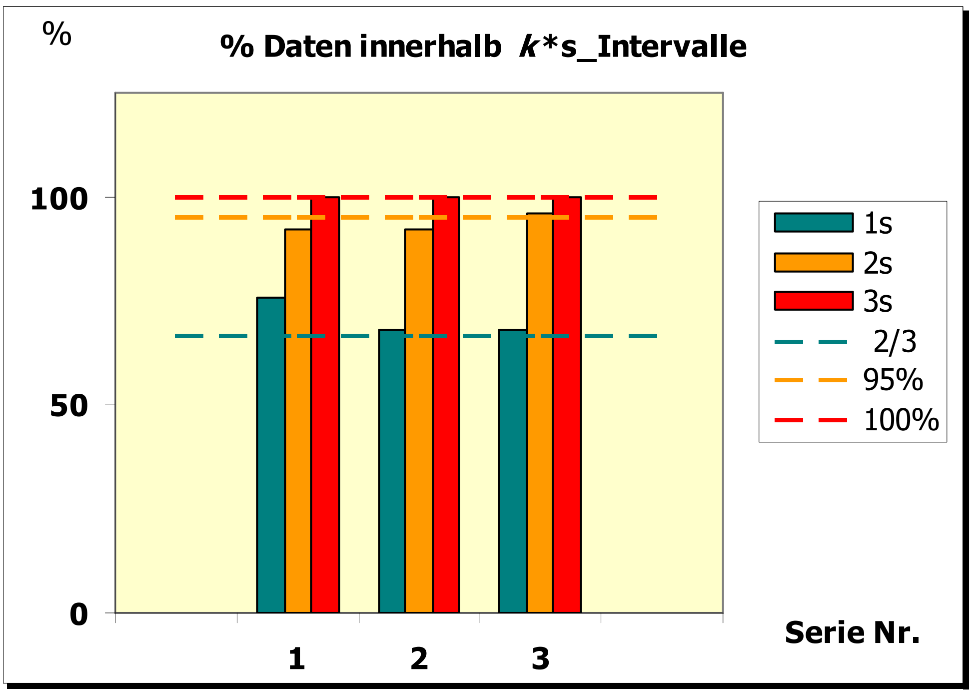


Abb. 4: Stabilität der s-Regeln – der Anteil an Daten innerhalb der entsprechenden Intervalle schwankt nur wenig um die gestrichelten Vorgabelinien

3. Boxplots und Histogramme zur Darstellung der Verteilung von Daten

Boxplots

In EXCEL ist unter den vorgegebenen Diagrammen der Boxplot nicht enthalten. Man kann sich ein Template herstellen (oder vom Autor anfordern), in welchem man allenfalls die Länge der Ausläufer an die gegebenen Daten anpassen muss. Wie immer funktioniert das Ganze mit Tricks. Was ist ein Boxplot? Eine mögliche Antwort darauf: Eine geschickte graphische Anordnung von bestimmten Quantilen der Verteilung. Am besten erläutert man dies an einer Darstellung: Für jene Daten, die innerhalb der Quartile sind, zeichnet man eine Box, die Lage des Medians kennzeichnet man durch eine Marke. Hernach zeichnet man im einfachen Boxplot Ausläufer bis zu Minimum und Maximum der Daten. Im erweiterten Boxplot zeichnet man auch Marken für Minimum und Maximum; die Ausläufer zeichnet man jedoch nur bis zu den 2,5%- und 97,5%-Quantilen (oder auch bis zum 5%- und 95%-Punkt der Verteilung). Das Ergebnis kann man in EXCEL durch Nachbearbeiten etwa in die Form von Abb. 5 bringen.

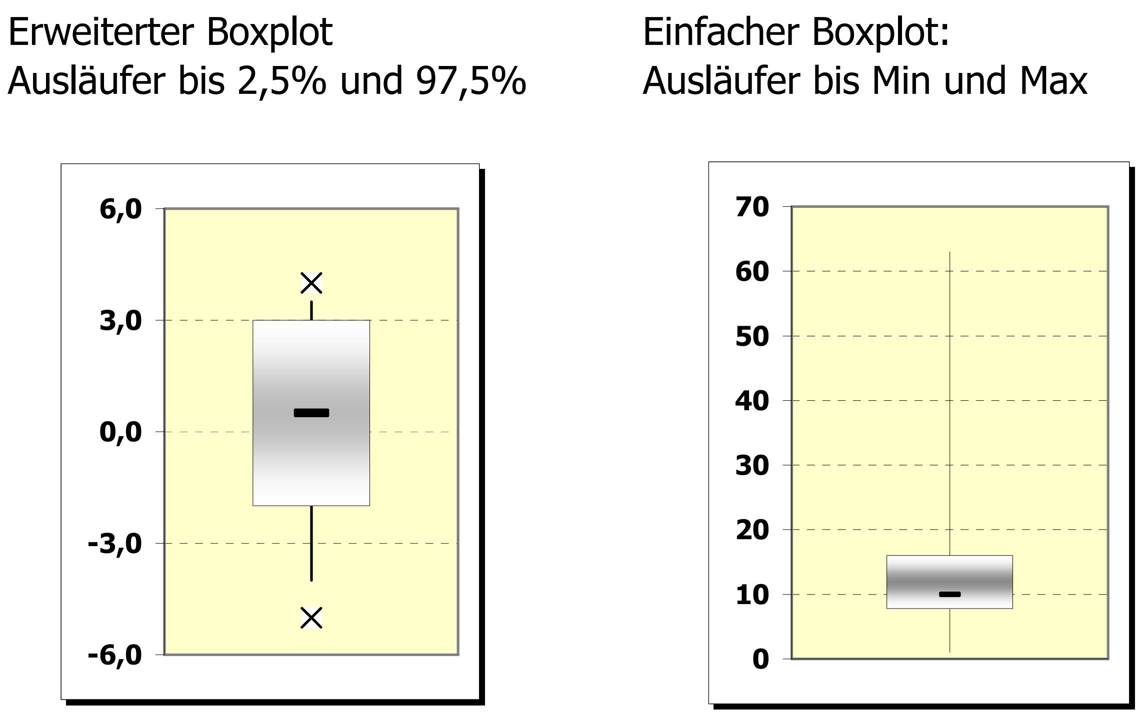


Abb. 5: Templates für einfachen und erweiterten Boxplot – mit ein paar Tricks in EXCEL erstellt

Die eigentlichen Tricks zur Erstellung von Boxplots sind nur für EXCEL-Freaks interessant (siehe die Links zu den EXCEL-Files des Autors). In den folgenden Tabellen 5-7 wird nur erläutert, wie man das Template ladet und anwendet und zwar gleich für den komplizierteren Fall des erweiterten Boxplots.

Tab. 5: Auszug aus dem Tabellenblatt für die Erstellung eines Templates mit der Beschreibung für die Vorgangsweise in EXCEL.

Template erstellen

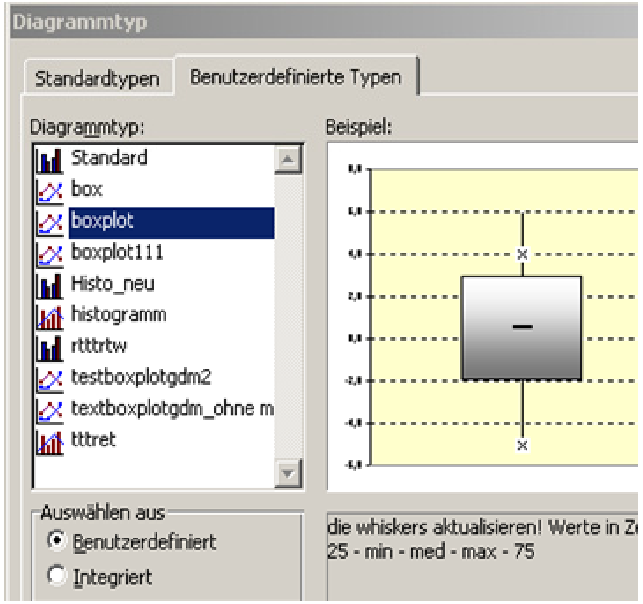
Vorlagen für Boxplot   Template: erstellen   anwenden   anpassen   Konstruktion   Inhalt

Rechts-klicken auf die Vorlage für den Boxplot
dann auf Diagrammtyp
die Karte "Benutzerdefiniert" wählen
auf Schaltfläche "Hinzufügen" klicken

Namen wie "boxplot" wählen

Beschreibung eingeben wie:
Reihenfolge beachter 25, min, med, max, 75
Daten in Zeilen eingeben
Whiskers ergänzen: Differenz 25-2,5; 97,5-75

Beim einfachen Boxplot entfällt
die Angabe für die Whiskers



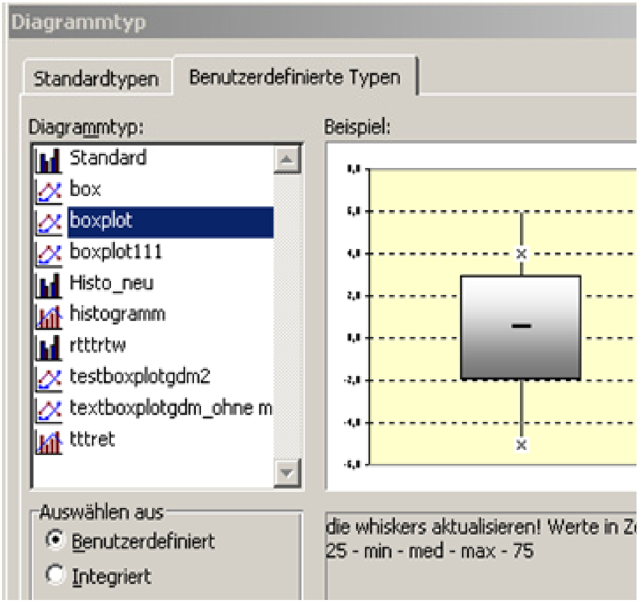
Tab. 6: Beschreibung der Anwendung eines Templates für den Boxplot. Da Tricks ausgenützt werden, ist die Angabe der dargestellten Quantile in einer eigenartigen Reihenfolge erforderlich. Man könnte dafür ein Makro schreiben, aber Makros machen eventuell bei der Übergabe an Fremde Probleme.

Template anwenden

Vorlagen für Boxplot   Template: erstellen anwenden anpassen Konstruktion Inhalt

In dieser Reihenfolge müssen die Quantile angegeben werden

Quantil	Wert	
97,5-75	0,50	Länge des oberen Ausläufers
25	-2,00	unteres Ende der Box
min	-5,00	Minimum
med	0,50	Median - Mitte der Box
max	4,00	Maximum
75	3,00	oberes Ende der Box
25-2,5	2	Länge des unteren Ausläufers



markiert wird nur der gestrichelte Bereich

die Ausläufer müssen extra aktualisiert werden Template anpassen

Tab. 7: Unangenehm, dass im Template für den Boxplot die Länge der Ausläufer nicht automatisch übergeben werden kann. Diese müssen „händisch“ ergänzt werden. Das entfällt bei der Erstellung eines einfachen Boxplots mit Ausläufern bis zu Minimum und Maximum. Allerdings sind diese Boxplots aus vielerlei Hinsicht zu wenig informativ.

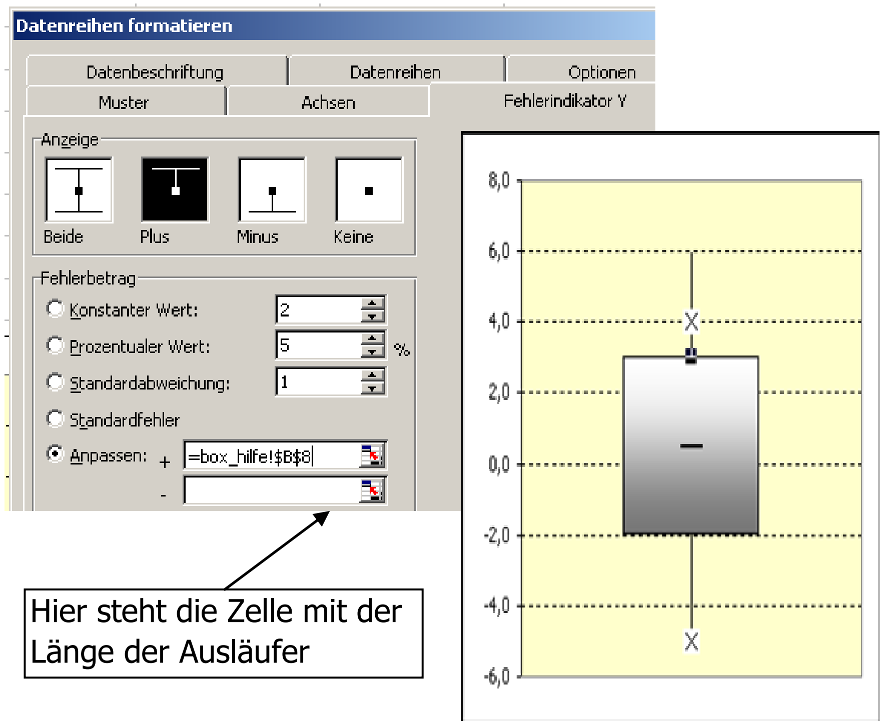
Template anpassen - Ausläufer für gegebene Daten aktualisieren

Vorlagen für Boxplot   Template: erstellen   anwenden   anpassen   Konstruktion   Inhalt

3. Quartil rechts-klicken
  Datenreihen formatieren wählen
  Option "Fehlerindikator" wählen

unter "Anpassen"
  mit der Maus die Adresse mit
  Differenz zwischen Quantilen 97,5 und 75
  = Länge des oberen Ausläufers eintragen

analog 1. Quartil…
  unteren Ausläufer aktualisieren



Vergleich von Boxplots mit Histogrammen – Kernbereich, normaler Bereich und Ausreißer

Für die Beurteilung einer Verteilung und einzelner Werte könnte man in Histogramme die *k**s-Intervalle einzeichnen und das entstehende Diagramm mit Boxplots vergleichen. Der *Kernbereich*, das 1s-Intervall entspricht hierbei der zentralen Box; der *Bereich des noch üblichen* (mit Ausnahme des Kernbereichs) – was durch das 2s-Intervall neu hinzukommt, entspricht den Ausläufern („whiskers") des Boxplots (der hier u.A. bis zu den 2.5%-Quantilen gezeichnet wurde); der extreme Bereich liegt außerhalb des 2s-Intervalls im Histogramm – es ist der außerhalb der Ausläufer befindliche Bereich im Boxplot, dessen Daten üblicherweise extra eingezeichnet und durch Bezeichnungen aus der statistischen Anonymität herausgenommen werden; diese Werte bedürfen einer eigenen Interpretation, sie sind jedenfalls Kandidaten dafür, als Ausreißer aus dem Kollektiv der Daten entfernt zu werden (weil sie nicht dazu gehören oder weil sie simpel die Folge von Fehlern sind) – siehe Abb. 6 und 7 und Tabelle 8.

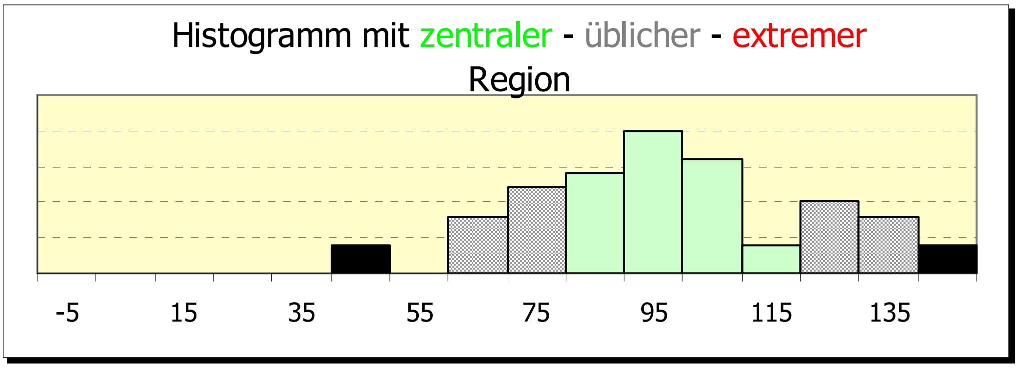


Abb. 6: Histogramm – hier farblich in Regionen unterteilt:
Kernbereich, Normalregion und ausreißerverdächtige Daten

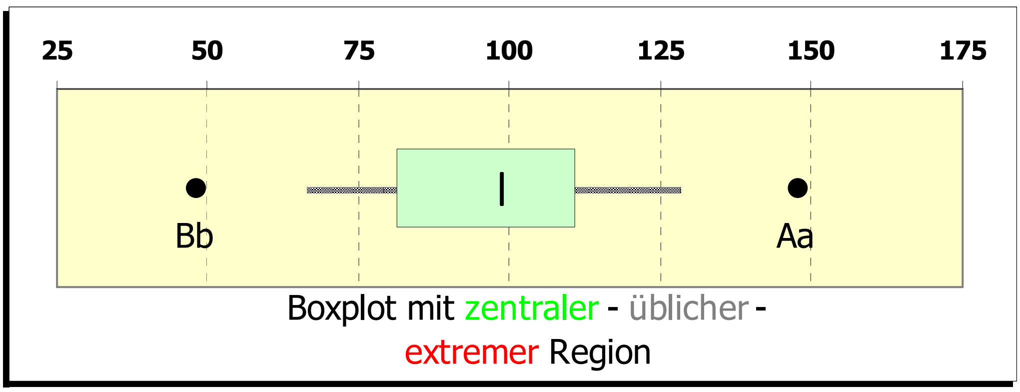


Abb. 7: Boxplot mit indirekter Darstellung von Kernbereich, Normalregion und ausreißerverdächtigen Daten

Im Übrigen ist es mit ein bisschen Mühe durchaus möglich, in EXCEL ordentliche Histogramme zu zeichnen, ja sogar mit ungleich breiten Klassen. Damit sich das Unterfangen lohnt, sollte man sich ehest angewöhnen, die eigenen Layout-Vorgaben und Vorlieben in Templates zu fassen und speichern, weil sonst allein die Nachbearbeitung des Layouts zu viel an Zeit wegnimmt. Hier sei wiederum auf die in der Literatur angegebenen EXCEL-Files des Autors verwiesen.

Tab. 8: Vergleich einiger Bereiche in Histogramm und Boxplot

	Histogramm			Boxplot	
	Mittelwert	Mittel		Mitte	Median
Innerhalb	1s-Intervall	Kernbereich: ca. 2/3	Grün	Zentral: 50%	Zentrale Box: $(x_{0,25}, x_{0,75})$
Im Intervall, aber nicht im darüber liegenden	m–2*s, m+2*s	Noch üblich: ca. 95%-2/3	Grau	Noch üblich: 95%-50%	Ausläufer $(x_{0,025}, x_{0,975})$
Im Intervall, aber nicht im darüber liegenden	m–3*s, m+3*s	Außergewöhnlich: ca. 5%	Rot	Außergewöhnlich: 5%	Marken bei Extrema (min, max)
Außerhalb	3s-Intervall	Extrem: 3‰	Violett	0	

4. Zusammenfassung

Zur professionellen Datenanalyse für einen echten Anwender mag EXCEL zweite Wahl sein, nicht jedoch für die Datenanalyse im Unterricht. Spielt der didaktische Blickwinkel eine Rolle, so gewinnen die klare Übersicht in den Tabellen und die dynamische Kapazität an Bedeutung. Viele Einsichten, die zum Verständnis der Begriffe unumgänglich sind, werden leicht zugänglich. Die Erneuerung des Zufalls mittels Tastendrucks lässt, speziell in den Graphiken, den Zufall, dessen Auswirkungen und dessen Spielraum live erleben. Studierende der Mathematik, Informatik sowie der Angewandten Betriebswirtschaft an der Universität geben viel versprechendes Feedback. Für den voruniversitären Unterricht sollte das didaktische Potential einer Tabellenkalkulation noch stärker an Gewicht gewinnen.

Literatur

Borovcnik, M.: EXCEL für den Unterricht in Stochastik. http://www.uni-klu.ac.at/stochastik.schule unter Links Manfred Borovcnik.

Borovcnik, M.: Nützliche Gesetze über den Zufall – Experimente mit EXCEL. Schriftenreihe Didaktik der Mathematik der Österr. Math. Ges. 32 (2001), 1-22.

Borovcnik, M.: EXCELlent Statistics – Statistik mit Unterstützung von Tabellenkalkulation. Schriftenreihe Didaktik der Mathematik der Österr. Math. Ges. 33 (2002).

Borovcnik, M.: Neue Wege zur statistischen Beurteilung. Schriftenreihe Didaktik der Mathematik der Österr. Math. Ges. 38 (2006a), 1-12.

Borovcnik, M.: Daten – Zufall – Resampling. J. Meyer (Hsg.): Anregungen zum Stochastikunterricht Bd. 3, Fanzbecker, Hildesheim 2006b, 143-158.

Borovcnik, M., Neuwirth, E.: Ein rekursiver Zugang zur Binomialverteilung. http://www.osg.or.at/ (2006)

Christie, D.: Resampling mit Excel. Stochastik in der Schule 24 (2004), Heft 3, 22-27.

Hunt, N.: Piktogramme mit Microsoft Excel. Stochastik in der Schule 21 (2001), Heft 2, 18-20.

Hunt, N.: Punkte-Diagramme und Stengel-Diagramme mit EXCEL. Stochastik in der Schule 22 (2002), Heft 3, 8-9.

Meyer, J.: EXCEL-Files in der Lehrerfortbildung. Auf Anfrage beim Autor: J.M.Meyer@t-online.de

Neuwirth, E.: http://sunsite.univie.ac.at/mailman/listinfo/Improve-excel

Neuwirth, E., Arganbright, D.: The Active Modeler: Mathematical Modeling with Microsoft Excel. Brooks/Cole 2004.

Reckelkamm, B.: Der Tanz der Residuen - Erarbeitung statistischer Grundbegriffe mit Hilfe von EXCEL. Stochastik in der Schule 24 (2004), Heft 3, 14-21.

Spreadsheets in Education: http://www.sie.bond.edu.au/

Velleman, P.F.: ActivStats for Excel 2003-2004 Release. Boston: Addison-Wesley 2004.

Whingham, D.: Normalverteilungen mit EXCEL. Stochastik in der Schule 19 (1999), Heft 2, 28-31.