Manfred BOROVCNIK, Klagenfurt

Abkürzungen zur Beurteilenden Statistik

Abkürzungen zur Beurteilenden Statistik – wie sie hier dargestellt und in EXCEL eingebettet werden – ermöglichen ein verständiges Anwenden formaler Ergebnisse auf ein Sachproblem.

0. Vorbemerkungen

Beurteilende Statistik liegt im Brennpunkt von *Kontext*, aus dem ein Problem stammt, *Logik des induktiven Schließens*, einer Hülle für „statistisches Restrisiko“, sowie von *mathematischen Methoden*, die zur Herleitung der Verfahren in „irgendeinem“ optimalen Sinne dienen. Dementsprechend gilt Beurteilende Statistik als schwierig. Im Folgenden sollen Abkürzungen zur Beurteilenden Statistik, oder besser zu einer statistischen Beurteilung angesprochen werden, die allesamt ein vertieftes Verständnis der Methoden wie auch der damit in konkreten Problemen ermöglichten Lösungsansätze anbahnen können.

Im ersten Abschnitt zur Modellierung eines Kontexts wird eine hautnahe Interpretation des Korrelationskoeffizienten wiedergegeben, die eine sachgemäße Einschätzung von dessen Größe weit über das üblicherweise angestrebte Ziel „der Korrelationskoeffizient ist signifikant von Null verschieden, das bedeutet die in den Daten gefundenen Zusammenhänge sind statistisch abgesichert“ hinausweisen. Und das mit einfachsten, konkreten Mitteln – also eine Abkürzung, wie angedeutet.

Im zweiten Abschnitt wird die Möglichkeit der Simulation von stochastischen Annahmen über einen Kontext ausgenützt, um die Folgen des Zufalls – in der vorgegebenen Form – zu illustrieren. Zur Kontrastierung wird dies hier „Sampling“ genannt, d.h. wiederholte Realisierung der Zufallssituation („Stichprobenziehen“) unter genau definierten Bedingungen. Die Ergebnisse werden einerseits genutzt, um die eigentliche Variabilität des Zufalls zu illustrieren, andererseits sollen „Gesetze“ des Zufalls als stabiles Nebenprodukt des Zufalls sichtbar werden. Auf diesen Gesetzen basierend wird der Begriff „Restrisiko“ einsehbar.

Im dritten Abschnitt werden die sogenannten Resampling-Methoden eingesetzt, um den Zufallsfehler zu erforschen. Unter Zufallsfehler versteht man die Fluktuation irgendeiner – aus den Daten abgeleiteten – Kenngröße, die für sich genommen einen Sachverhalt messen bzw. beurteilen soll. Wenn z.B. Daten über die Wirkung eines Medikaments vorliegen – eine Gruppe mit dem Medikament behandelt, eine andere Gruppe mit Placebo – dann dient der ausgewiesene Mittelwertsunterschied im Zielmerkmal (Zeit bis zur Heilung, Anteil der Patienten, die vollständig geheilt werden, o.ä.) zur Beurteilung der Wirksamkeit des Medikaments. Könnte man wiederholt weitere Daten bekommen, so hätte man dieselbe Situation wie beim „Sampling“ und könnte die Auswirkungen des Zufalls bewerten und damit ein Vertrauensintervall für die Behandlungsunterschiede in den Gruppen angeben, oder bestimmen, ob das Medikament signifikant besser ist als Placebo. Man kann man in der Regel nicht den gesamten Versuch wiederholen. Hier setzt die Beurteilende Statistik ein: Man modelliert die Daten durch dahinterstehende Verteilungen und so weiter .. , und leitet daraus die „gewünschten“ Schlüsse ab.

„Re-Sampling“ dient nun dazu, ohne Theorie, also nur durch Simulation von wiederholten Teilstichproben aus einer schon *vorhandenen* Datenmenge die Genauigkeit von Schätzgrößen (hier der Behandlungsunterschied) zu beurteilen. Damit kann man Vertrauensintervalle angeben oder statistische Tests auf Signifikanz durchführen.

1. Modellierung

Ziel ist es, u.a. eine Zielgröße zu erklären, etwa die Frage zu beantworten, was Körpergewicht (einer vergleichbaren Gruppe) von Personen erklärt. Man sucht in der Phase der Systemanalyse nach erklärenden Variablen, wie Körpergröße, Geschlecht, Körpertyp, oder Ernährung im frühkindlichen Alter. Das resultierende Modell daraus könnte – im einfachsten Fall – eine (lineare) Strukturgleichung

$$\text{Gewicht} = a + b \times \text{Körpergröße}$$

sein. Anhand von Daten soll überprüft werden, ob dieses Modell eine brauchbare Beschreibung liefert und zu Prognosen für das Gewicht verwendet werden kann, wenn man die Größe einer Person kennt. Der Einfluss der Körpergröße gilt als gesichert, wenn der Koeffizient b signifikant von 0 verschieden ist. Damit hängt zusammen, ob der

Korrelationskoeffizient zwischen Körpergröße und Gewicht signifikant verschieden von 0 ist. Daten selbst werden aus vielerlei Gründen vom Modell abweichen:

Daten = Modell + Residuen = Signal + Rauschen

In EXCEL kann man bequem zu einem Datensatz die Punktwolke mit Trendlinie (nicht nur ein lineares Modell wie oben) zeichnen sowie R^2, das Quadrat des Korrelationskoeffizienten berechnen lassen – das erfolgt nach leicht einsehbaren Optimalitätskriterien. Für einen Datensatz von 15jährigen erhält man z.B.

Gewicht = 0,697 × Größe – 67,299

Anstatt jetzt das Modell zu verfeinern, etwa indem man getrennte Modelle für Buben und Mädchen aufstellt, sei jetzt direkt die Bedeutung des Korrelationskoeffizienten R erschlossen. Die Modellgleichung lässt in EXCEL einfach Modelldaten und die Residuen, d.s. die Abweichungen der echten Daten von denen auf der Ausgleichsgeraden, berechnen. Wertet man nun die Spalten Daten, Modelldaten sowie Residuen statistisch aus, so ergibt sich z.B. Mittel der Residuen = 0. Es ergibt sich auch empirisch, dass gilt

Varianz der Daten × (1-R^2) = Varianz der Residuen

Wenn man nun bei Kenntnis der Größe eine Voraussage auf das Gewicht macht, so kann man die ursprüngliche Voraussage verbessern:

Mittel Gew. ± 2×Stabw Gew. ⟶ Modell–Gew. für bekannte Größe ± 2×Stabw Gew.×√(1-R^2)

Der Faktor ×√(1-R^2) entspricht der Verkürzung der Voraussageintervalle, spiegelt also die Verbesserung an Präzision der Kenntnisse unter Ausnutzung des Modells wider. Die folgende Tabelle zeigt u.a., dass ein R von 0,83 nur zu einer Verbesserung von 44% (Faktor 0,56) führt. Wenn man weiß, dass aus Unkenntnis solcher methodischer Eigenschaften Korrelationskoeffizienten von 0,2 und darunter, weil in einer Studie als signifikant erwiesen, hochgejubelt werden, wird man die Abkürzung hier nicht unterschätzen.

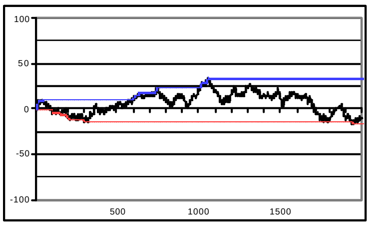
Nr	Geschlecht	Größe	Gewicht	L. Modell alle	Residuen
5	0	153	38	39,33	-1,33
18	1	166	50	48,39	1,61
			data	fit	res
Mittel			45,15	45,15	0,00
Varianz			34,34	23,58	10,76
R			0,83		
			Varianz	´ R^2	´ (1- R^2)
Verkleinerungsfaktor				0,69	0,31
Standardabw.			5,86	4,86	3,28
					´ Ö(1- R^2)
Verkleinerungsfaktor					0,56

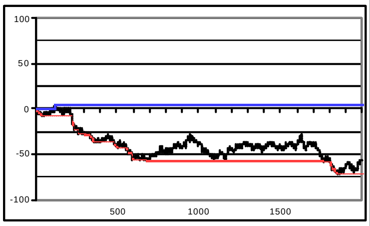
2. Sampling

Mit Simulation kann man die Auswirkungen des Zufalls bzw. von Zufallsmodellen studieren und damit Näherungen für Modellrechnungen bekommen – im Prinzip kann man weite Teile der Wahrscheinlichkeitstheorie auf diese Weise empirisch vorwegnehmen. Man könnte etwa illustrieren, wie variabel der Zufall ist, oder, *was* man unter der Variabilität des Zufalls verstehen soll. Oder man könnte trotz der Fluktuation des Zufalls Invarianten herausschälen, wie etwa ein Restrisiko von Entscheidungen. Man könnte die Folgen des Modells, des „was wäre, wenn“ materiell anhand simulierter Daten untersuchen. Etwa könnte man der Frage nachgehen, wie stark der Korrelationskoeffizient R fluktuiert, obwohl vom Modell her klar ist, das er eigentlich 0 ist – damit wird es möglich, Schwellenwerte für R zu bestimmen, bei deren Überschreitung wir nicht mehr davon ausgehen, dass R doch noch 0 sein könnte., bei deren Überschreitung wir dann sagen, R ist statistisch signifikant (gesichert, von 0 verschieden). An dieser Stelle seien nur zwei kommentierte Bildfolgen wiedergegeben.

Roulette – Was ist eigentlich unter Variabilität des Zufalls aufzufassen:

Die durchgezogene Kurve zeigt den Saldo, die obere den maximalen Gewinn, die untere den maximalen Verlust. Während die linke Serie lange in der Gewinnzone bleibt, geht die rechte Serie ganz dramatisch in den Verlust.



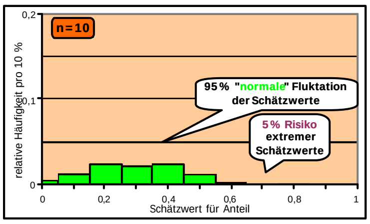


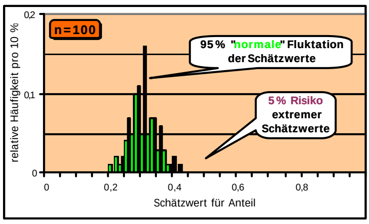
Das Gesetz der großen Zahlen bezieht sich wohl nicht auf die Summenkurve (= Saldo des Gewinns), sondern auf die relative Häufigkeit des Gewinns. Die Einpendelung der Kurve der relativen Häufigkeiten auf eine waagrechte Linie ist bei üblicher Betrachtung (auch im Unterricht) im Zentrum des Interesses; die *echte* Variabilität zeigt jedoch erst die Wiederholung der *gesamten* Serie. Und darauf bezieht sich die Aussage des Gesetzes: Das (Rest-)Risiko, dass die gesamte Kurve zu einem beliebigen Zeitpunkt außerhalb festgelegter Grenzen zu liegen kommt, ist kalkulierbar. Bis dahin jedoch ist ihr Verlauf sehr starken Schwankungen unterworfen. Der Wechsel von den relativen Häufigkeiten des Gewinns zum bisherigen Saldo wirkt hier wie ein Vergrößerungsglas und lässt das Ausmaß der Fluktuation (des Zufalls) trotz Einpendelung (also einer statistischen Gesetzmäßigkeit) viel besser erkennen.

Die Verteilung des „Anteils der Einsen" („Erfolge") im klassischen Bernoulli-Versuch:

Aus einer Bernoulli-Kette schätzt man mittels des Stichprobenanteils den unbekannten Wert p für die Wahrscheinlichkeit eines „Erfolgs". Die Schätzung wird mit zunehmender Zahl der Daten präziser; genauer hat man eine Abnahme des Schätzfehlers mit dem Faktor 1/Wurzel (n). Hier haben schon viele Autoren eine Elementarisierung versucht, so etwa Riemer. Mit EXCEL lassen sich die Daten effizient simulieren und auswerten. Die folgenden beiden Abbildungen zeigen qualitativ den Sachverhalt der Verbesserung und ließen den Begriff des Restrisikos wiederum thematisieren.

Bei p=0,25 und n=10 ist ein Anteil der „Erfolge" zwischen 0 und 0,65 *normal*; setzt man das als Maßstab für Prognosen für die nächste 10er Serie, so beträgt das Risiko 5%. Bei 100 Daten gilt: *normale* Serien liegen zwischen 0,2 und 0,4: bei gleichem Risiko sind Voraussagen genauer.





3. Re-sampling

Man hat Daten – eine Stichprobe – für ein Merkmal, etwa Zeiten der Belastung für ein „Projekt", mit Mittelwert 8,9 sowie einer Standardabweichung von 10,39. Wie genau kennt man den Mittelwert für *alle* (=Population)? Das Vertrauensintervall berechnet sich zu $8{,}9 \pm 1{,}96 \times 10{,}39$ (statt 1,96 das entsprechende Quantil der t-Verteilung).

Stattdessen geht man im Re-sampling so vor: Man entnimmt der vorhandenen Stichprobe – deren Verteilungsfunktion ja eine Schätzung der theoretischen Verteilungsfunktion ist – eine weitere Stichprobe (mit Wiederholung). Dieses 1. Resampling ergibt im Tabellenblatt unten ein Mittel von 5,4 – ein Hinweis über die Genauigkeit der Schätzung von 8,9 aus den Urdaten. Wie man in EXCEL die Stichprobe „zieht", lese man in Christie nach. Mit F9 kann man nun bequem den Zufall „erneuern", die Zahl 5,4 ändert entsprechend der neuen Zufallsauswahl. Man erhält konkret vor Augen geführt, wie genau der Ausgangswert von 8,9 ist. Man könnte nun beliebig oft F9 drücken, die Zwischenergebnisse speichern und somit die resampelte Verteilung der wiederholten Mittelwerte gewinnen. Diese Verteilung kann man mit Methoden der beschreibenden Statistik untersuchen, so kann man ihren 2,5%- bzw. 97,5%-Punkt ermitteln – 4,38 bzw. 17,58, womit man ein zentrales 95%-Intervall für die Fluktuation des Mittelwerts bei resampelten Stichproben erhält, das R(esampling)-Vertrauensintervall. (Oder man berechnet ein Intervall mit 2 Standardabweichungen um den Mittelwert dieser resampelten Mittelwerte.)

Das Resampling der Stichprobe verkürzt man nun effizient mit einer sogenannten Mehrfachoperation aus dem Menü Daten>Tabelle; für Details siehe wieder Christie. Resampling funktioniert für jedes Merkmal, so etwa auch für die Differenz von Mittelwerten aus zwei verschiedenen Gruppen. Man kann das resampelte Intervall auch mit Hypothesen vorweg vergleichen, kommt also auch ans Testen heran.

Urdaten

Nr	Zeiten
1	12
2	2
3	6
4	2
5	19
6	5
7	34
8	4
9	1
10	4
Mittel	8,90
St-abw.	10,39

1. Resampling

Nr	Zeiten
9	1
5	19
2	2
10	4
1	12
4	2
2	2
2	2
6	5
6	5
Mittel	5,40

Wiederholte Mittelwerte

Daten>Tabelle	
	5,40
	8,60
	18,20
	10,50
	8,60
	8,50
	12,90
	4,60
	9,10
	5,70
	9,60
	6,90
	...

Verteilung wiederholter Mittelwerte

Kennziffern	
Mittel	St-abw.
9,80	3,20

2-s R-Intervall	
3,52	16,08

R-Vertrauensintervall	
unten	oben
4,38	17,58

95% Vertrauensintervall		
unten	oben	Präzision
2,33	15,47	6,57

4. Re-sumé

Mit Studierenden aus verschiedensten Studienrichtungen, speziell aus BWL konnte der Autor ausgesprochen gute Erfahrungen machen. Das angestrebte tiefere Verständnis für die Möglichkeiten und Grenzen, auch konkret das Verstehen der Standardprozeduren von Vertrauensintervallen wie statistischen Tests – insbesondere des Tests auf Unterschiede in verschiedenen Gruppen – war da. Es wurde nicht mehr durch die nicht erreichbare Mathematik behindert. Die Schwierigkeiten mit der induktiven Logik mit dem „was wäre, wenn" – Szenario-Charakter waren weg. Die künstlichen Szenarien waren keine Gedankenexperimente mehr, sondern materiell fassbar, es waren konkrete Daten über die Auswirkungen des „wenn" – man konnte „was wäre" angreifen und untersuchen.

Das Re-Sampling aus der schon vorhandenen Stichprobe statt des Ziehen neuer Stichproben machte kein Problem. Auch das Argument, dass der Fehler einer schon vorhandenen Stichprobe – wenn sie nicht-repräsentativ ist – sich dadurch „hochschaukelt", ist in Wirklichkeit kein Nachteil. Denn, wenn das Design der Untersuchung schon einmal zu einer so schlechten Stichprobe geführt hat, dann wird es diese „Neigung" auch bei einer echten Wiederholung der Stichprobe haben. Und: Der Fehler der vorhandenen Stichproben multipliziert sich dann auch mit der klassischen Auswertung der Daten auf Basis der üblichen Modelle. Im übrigen sind die Re-Sampling-Methoden in der Angewandten *und* der Theoretischen Statistik heute oft das Mittel der Wahl. Resampling-Methoden revolutionieren den Zugang zur gesamten Beurteilenden Statistik. Man kann die erforderlichen Simulationen in einer Spezialsprache realisieren. Für den Hausgebrauch reicht jedoch EXCEL.

Bleibt noch, auf das Werkzeug EXCEL – oder eine andere Tabellenkalkulation – einzugehen: Nicht alles wird lösbar, nicht alle Problemstellungen lassen sich einfach in ein Re-Sampling-Design umformulieren. Zeit geht auch mit trivialen Computer-Fragen verloren. Der Arbeitsstil ist empirisch-explorativ, was sehr vielen Lernenden entgegenkommt. Man probiert spielerisch und sieht die Konsequenzen direkt und nicht in abstrakten Größen verklausuliert. Das hängt auch mit der begrifflichen Einfachheit der Verfahren zusammen, die zudem noch auf dem Computer materiell repräsentiert und damit nachvollziehbar werden.

Literatur und EXCEL-Files: Siehe http://www.uni-klu.ac.at/stochastik.schule/ unter Links.