dieses Problems, die vielleicht noch untersucht werden sollten, wären, welche Maschinen über einen größeren Zeitraum öfters gewartet werden müssen, ob die Ausfallraten von Schicht zu Schicht oder Woche zu Woche sich ändern, oder ob weniger Ausfälle ein Ergebnis einer besseren Wartung sein können.

Anmerkungen

1) Als Quantil der Ordnung p (Quantil p-ter Ordnung) einer Zufallsgröße X mit der Verteilungsfunktion F_X bezeichnet man jede Zahl Q, für die $F_X(Q) \leq p \leq F_X(Q+0)$ gilt. Für p=0,5 ergibt sich der Median.

2) Es sei X eine Zufallsveränderliche, deren mögliche Werte $x_1, x_2, \ldots, x_n$ mit den Wahrscheinlichkeiten $p(x_1), p(x_2), \ldots, p(x_n)$ eintreten. Dann heißt die Funktion $F_X: R \to [0;1]$ mit $F_X(x) = \sum_{x_s<x} p(x_s)$ Verteilungsfunktion der diskreten Zufallsvariablen X. Sie gibt die Wahrscheinlichkeit dafür an, daß die Zufallsvariable X Werte annimmt, die x nicht überschreiten

$$F_X(x) = \sum_{x_s<x} p(x_s) = \sum_{s=1}^{r} f(x_s) = P(X=x_1)+P(X=x_2)+\ldots+P(X=x_r)$$
$$= P(X=x_1 \vee X=x_2 \vee \ldots \vee X=x_r)$$

3) Das Geburtstagsproblem läßt sich genauer ausgeführt in den meisten Büchern über Wahrscheinlichkeitsrechnung finden. Es sei hier jedoch noch auf den Aufsatz hingewiesen von SPENCER, N.: Celebrating the birthday problem. In: The Mathematics Teacher v. 70(4), S.348-353 (April 1977).

4) Das Geburtstagsparadoxon wird z.B. aufgelöst bei ENGEL, A.: Wahrscheinlichkeitsrechnung und Statistik, Bd. 1, Stuttgart, Klett 1973, S. 50-51.

Statistiken über verschiedene Stile
oder
Es war keine stilvolle Heirat

von J.Swift, übersetzt von M.Nuske

Nachdem ich kürzlich in meiner Unterrichtsklasse im Fach Statistik die Stile einiger Autoren untersuchte, muß meine Klasse Eliza große Sympathien entgegengebracht haben. Sie mußten also nicht nur im Englischunterricht mit Wörtern und Ausdrücken umgehen, sondern hatten sich damit auch im Statistik-Unterricht zu befassen.
Das Experiment wurde dadurch stimuliert, daß wir mit Jack Hodgins einen erfolgreichen Autor in unserem Kollegium besaßen. Unser Ziel war es herauszufinden, ob wir alleine unter dem Einsatz mathematischer Methoden Jack's Stil von dem anderer Autoren unterscheiden konnten. In "Probability und Statistics" behandelt John Durran Statistiken, die sich mit der Satzlänge und den unterschiedlich oft gebrauchten (=Variabilität) Substantiven befassen. Für die erste Untersuchung wurden deswegen diese beiden Gebiete gewählt.

Eine Statistik über den Gebrauch von Substantiven

Eine der Übungen in "Probability and Statistics" baut auf Material von G.U.Yule's "Statistical Study of Literary vocabulary (CUP 1944) auf. In der Aufgabe wird eine Statistik entwickelt, die den Gebrauch von Substantiven "mißt". Dies geschieht folgendermaßen:

a) Durch zufällige Auswahl wird ein Abschnitt herausgesucht.
b) Für den ausgewählten Abschnitt werden nun die darin erscheinenden Substantive aufgeschrieben.
c) Unter Berücksichtigung der Wiederholungen eines Substantives sammelt man so die ersten 100 verschiedenen Substantive. Dabei wird man viele Substantive haben, die nur einmal vorkommen; einige werden zweimal oder dreimal und einige noch öfters in diesem ausgezählten Teil enthalten sein.
d) Das Ergebnis wird nun in einer Häufigkeitstabelle angeordnet, wie dies für einen Abschnitt aus Hemingways Buch "For Whom the Bell Tolls" (dt.: Wem die Stunde schlägt) gezeigt wird:

x ist die Häufigkeit, mit der ein Substantiv erscheint
f ist die Anzahl der Substantive, die x-mal erscheinen

x	1	2	3	4	5	6	7	8	13	21
f	55	25	6	6	3	1	2	0	1	1

Berechne Σfx und Σfx^2

x	1	2	3	4	5	6	7	8	13	21
f	55	25	6	6	3	1	2	0	1	1
fx	55	50	18	24	15	6	14	0	13	21
fx^2	55	100	54	96	75	36	91	0	169	441

$\Sigma fx = 216$, $\Sigma fx^2 = 1117$

e) Die von Yule verwendete Statistik lautet:

$$m = 10^4 \frac{\Sigma fx^2 - \Sigma fx}{(\Sigma fx)^2} = 193.1$$

Yule gibt Gründe an, warum diese Statistik mit der Variabilität eines Substantives in einem Abschnitt variiert. Zwei Beispiele mögen dies verdeutlichen. Der Extremfall der größten Variabilität tritt dann ein, wenn ein Abschnitt 100 Substantive enthält, die je einmal auftreten. Dann gilt:

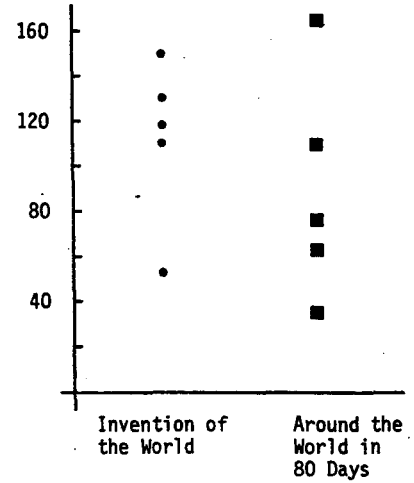
$$\Sigma fx^2 = \Sigma fx = 1 \quad \text{und} \quad m = 0.$$

Ein Abschnitt des Buches "Children of Dune" von Frank Herbert kommt diesem Extrem sehr nahe: es wurden 94 Substantive gezählt, die einmal vorkamen und 6, die vom Autor zweimal verwendet wurden.

$$\Sigma fx = 106 \; \Sigma fx^2 = 118 \quad \text{und} \quad m = 10.7$$

Der andere Extremfall tritt etwa in einem Lesebuch der 1. Grundschulklasse auf. Einige Substantive werden sehr oft gebraucht, bevor man die 100 verschiedenen Substantive gesammelt hat. Das heißt: Σfx^2 ist viel größer als Σfx und m ist daher sehr groß. Dadurch, daß der Faktor 10^4 in Yules Formel verwendet wird, liegen die Werte für m normalerweise zwischen 0 und 200. Der weiter oben behandelte Hemingway-Abschnitt ergibt etwa den Wert $m \approx 193$. Jeder Schüler wählte einen Autor aus und errechnete m für mindestens 5 Abschnitte in einem Werk dieses Autors. Für "The Invention of the Word" von Jack Hodgins und "Around the World in 80 days" von Jules Verne ergab sich folgendes Bild:

$$m = \frac{10^4(\Sigma fx^2 - \Sigma fx)}{(\Sigma fx)^2}$$



Es ist klar, daß mit dieser einen Variablen, die den Gebrauch von Substantiven mißt, noch nicht einwandfrei zwischen diesen beiden Stücken unterschieden werden kann. Dies ändert sich aber, wenn noch eine zweite Variable betrachtet wird. In unserem Fall wird es die Satzlänge sein.

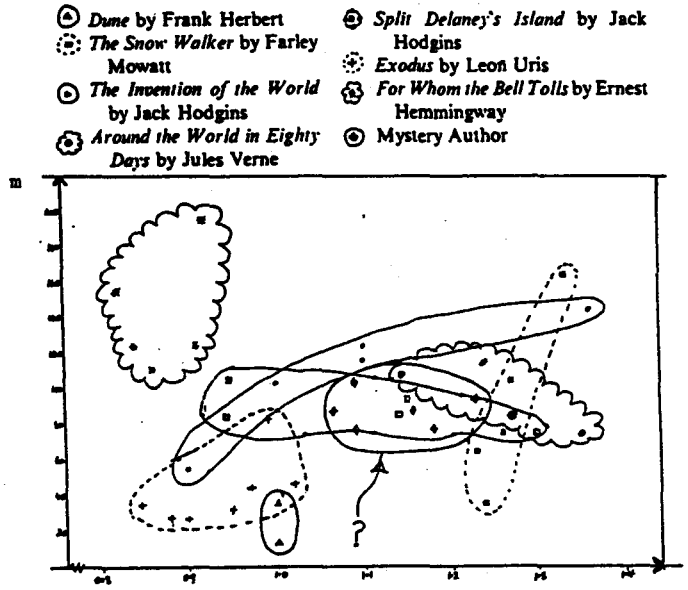
Eine Statistik für die Satzlänge

In "Probability and Statistics" erwähnt John Durran weiterhin eine Studie von C.B. Williams, die zeigt, daß bei vielen Autoren der Logarithmus der Satzlänge (gemessen in Anzahl der Worte) angenähert normalverteilt ist. Deshalb wurden aus jedem Abschnitt 20 Sätze zufällig ausgewählt, von der Satzlänge $\log_{10}$ gebildet und davon das arithmetische Mittel $\bar{x}$ errechnet. $\bar{x}$ lag zwischen 0.8 und 1.4.
Die Schüler errechneten $\bar{x}$ für jeden Abschnitt, den sie schon für die Substantivstatistik untersucht hatten.

Unterschiede zwischen den Autoren

Für jede Stichprobe aus einem Buch hatten nun die Schüler ein Paar von Statistiken vorliegen. Eine zweidimensionale Darstellung der Ergebnisse erleichtert die weitere Analyse. Einige ziemlich unterschiedliche Muster kann man erkennen und die Daten für "The Invention of the World" und "Around the World in 80 Days" zeigen nun entschieden größere Unterschiede auf, als dies bei der Substantivstatistik allein der Fall war. Am Ende der Untersuchung sollte der

Schüler dann anhand einer Kurzgeschichte den jeweiligen Autor erkennen. Die Daten des englischen Originalartikels sind auch in dieser Darstellung angegeben. Wer macht sich die Mühe den "J.Swift"-Punkt herauszusuchen?



Weitere Entwicklung

Eine Vielzahl weiterer Statistiken über den Stil eines Autors können erstellt werden. Unser Projekt soll weiterverfolgt werden; wir wollen dabei auf die folgenden beiden Statistiken Wert legen: Maurice Kendall beschreibt einen "Verschleierungs-Index", der mit der Satzlänge und langen Worten korreliert. Ein Verschleierungs-Index von 10 oder weniger läßt ihn das Buch als gut lesbar empfehlen, während bei einem Verschleierungsindex von 25 und mehr Vorsicht bezüglich der Lesbarkeit angebracht sein soll. Jack Hodgins ist der Meinung, daß der Anteil von Adverben und Adjektiven in Sätzen ebenfalls sehr viel über den Autor aussagt. Welche Auswirkung dieses Projekt auf die gängige Meinung, daß Mathematiker "nicht literatisch veranlagt" seien, hat, mag die Zukunft zeigen.

Einige Vorstellungen der Studenten über den Median und den Modalwert

von G. V. Barr
frei übersetzt von B. Stumpf

Der Zweck dieses Berichts ist es, in einem Versuch darzulegen, wie Studenten die statistischen Begriffe "Median und Modalwert" verstehen.
Dieser Artikel bezieht sich auf eine Musterstudie, die der Autor durchgeführt hat. Sie versucht folgende Fragen zu beantworten, die mit den studentischen Fehlern zusammenhängen:

- Welche Schwierig-keiten gibt es ?
- Wieviele Studenten haben solche Schwierigkeiten ?
- Wie wird das statistische Vokabular bezüglich dieser Begriffe bekannt ? (entwickelt)

Die Studie betrifft 95 Studenten im Alter zwischen 17 und 21 Jahren, welche alle die mathematischen Grundlagen für ein Studium der Technik mitbrachten. 69 % studierten Ingenieurwissenschaften, die restlichen 31 % Naturwissenschaft und Technik Von beiden Gruppen wurde erwartet, daß sie Lageparameter und Streuungsparameter aus numerischen Daten berechnen können.
Die vorgestellten Beispiele stammen aus den Tests, die dafür entwickelt wurden. Es sind Multiple-Choice-Tests, wobei aus vier Antworten a) b) c) d) dierichtige ausgewählt werden soll.
e) kann geantwortet werden, wenn das errechnete Ergebnis mit keiner der vier Auswahlantworten übereinstimmt.
f) Hier werden diejenigen zusammengefaßt, die keine Antwort geben konnten.

Der Median

Tabelle A gibt die Anzahl der zerbrochenen Stücke in einer Packung Schokolade an.

Tabelle A

Anzahl	Häufigkeit
2	5
4	6
6	7
9	2
11	1

Frage: Was ist der Median dieser Verteilung ?
Antworten: a) 4 (20%) b) 5 (33%) c) 7 (14%)
d) 8 (2 %) e) (24 %) f) (6 %)
Bemerkung: 21 % unter den 24 %, die sich für e) entschieden, haben 6 als Wert angegeben.
Das unterstrichene Ergebnis ist das richtige.