## VERTEILUNG DER GESCHLECHTER IN FAMILIEN

von HEINZ KLAUS STRICK

Um 1860 gab es in Sachsen 10690 Familien mit 12Kindern. Von den insgesamt 128.280 Kindern waren 66299 Jungen (51,68%) und 61981 Mädchen (48,32%). Die folgende Tabelle enthält die theoretische und die empirische Geschlechterverteilung dieser Familien.

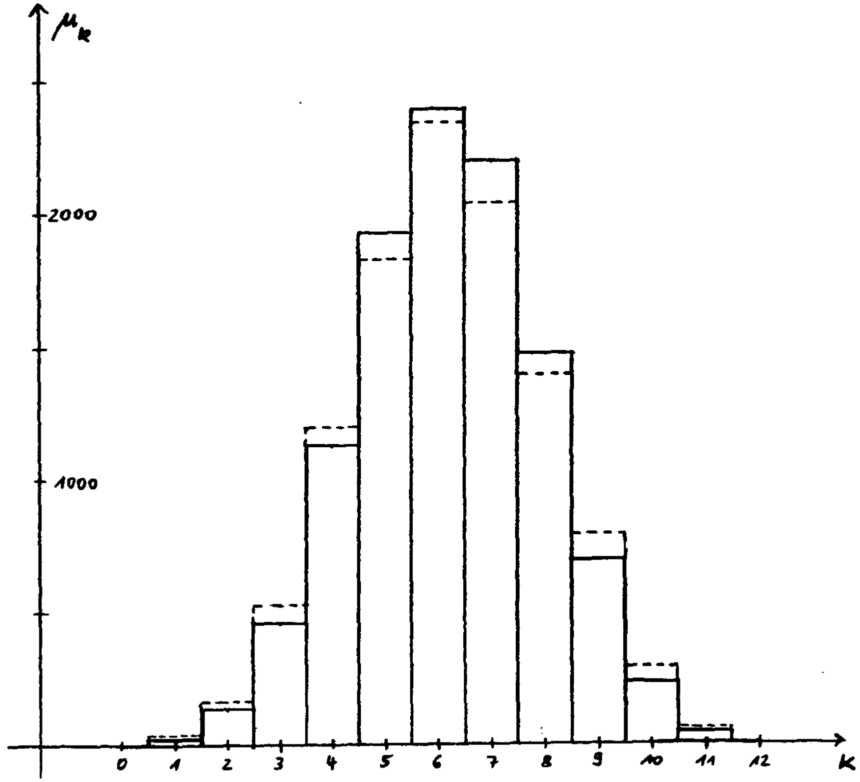
| Anzahl der Jungen | Anzahl der Mädchen | Wahrschein-lichkeit | theoretische Verteilung | empirische Verteilung |
|---|---|---|---|---|
| 0 | 12 | 0,00016 | 1,7 | 6 |
| 1 | 11 | 0,0021 | 22,2 | 29 |
| 2 | 10 | 0,0122 | 130,7 | 160 |
| 3 | 9 | 0,0436 | 466,1 | 521 |
| 4 | 8 | 0,1049 | 1121,7 | 1198 |
| 5 | 7 | 0,1796 | 1919,6 | 1821 |
| 6 | 6 | 0,2241 | 2395,2 | 2360 |
| 7 | 5 | 0,2054 | 2195,8 | 2033 |
| 8 | 4 | 0,1373 | 1467,8 | 1398 |
| 9 | 3 | 0,0653 | 697,7 | 799 |
| 10 | 2 | 0,0209 | 223,9 | 298 |
| 11 | 1 | 0,0041 | 43,5 | 60 |
| 12 | 0 | 0,00036 | 3,9 | 7 |

Hierbei wurde mit dem Binomial-Ansatz gerechnet:
p=0,5168 (Wahrscheinlichkeit einer Jungengeburt), q=0,4832 (Wahrscheinlichkeit für eine Mädchengeburt),
X: Anzahl der Jungen in einer Familie mit 12 Kindern,

$P(X=k) = \binom{12}{k} p^k q^{12-k}$ Wahrscheinlichkeit für k Jungen in einer Familie mit 12 Kindern

und $\mu_k = 10690 \cdot P(X=k)$ Erwartungswert der Anzahl der Familien mit k Jungen und 12-k Mädchen.






Zum Vergleich : Empirische (---) und theoretische Verteilung

Wie man aus der letzten Spalte und der Abbildung entnimmt, ist die Abweichung der theoretischen Verteilung von der der Erhebung z.T. beträchtlich: Eine Untersuchung mit Hilfe des $\chi^2$-Anpassungstest zeigt, daß die Abweichung signifikant auf sehr hohem Niveau ist ($\chi^2_{emp} \approx 99$ ; $P(\chi^2 \leq 32,91) = 0,999$).

Der Genetiker STERN bietet für die Verformung der empirischen Verteilung (außen zu hoch, innen zu flach) eine Erklärung an: Man könne sich die Verteilung entstanden denken als 'Summe' von 2 (oder mehr)

Binomialverteilungen. In der Bevölkerung gebe es 2 (oder mehr) gleich große Gruppen von Familien mit 12 Kindern. In der einen Gruppe sei z.B. die Wahrscheinlichkeit für eine Jungengeburt $p_1 = 0{,}6168$, in der anderen $p_2 = 0{,}4168$ ; dann liegt im Mittel eine Wahrscheinlichkeit von $p = 0{,}5168$ für eine Jungengeburt vor.

In der Tat stellt man an Beispielen fest, daß die so konstruierte Überlagerung zweier Binomialverteilungen gegenüber der Verteilung der zugehörigen 'mittleren' Erfolgswahrscheinlichkeit $p = \frac{1}{2}(p_1+p_2)$ eine ähnlich geartete Verformung aufweist wie die des obigen Beispiels.

Man betrachte etwa das Beispiel mit
$p_1 = 0{,}4$ ; $p_2 = 0{,}6$ ; $p = \frac{1}{2}(0{,}4 + 0{,}6) = 0{,}5$ und $n = 12$:

| k | $P(X_1=k)$ | $P(X_2=k)$ | $\frac{1}{2}(P(X_1=k)+P(X_2=k))$ | $P(X=k)$ |
|---|---|---|---|---|
| 0 | 0,0022 | 0,0000 | 0,0011 | 0,0002 |
| 1 | 0,0174 | 0,0003 | 0,0089 | 0,0029 |
| 2 | 0,0639 | 0,0025 | 0,0332 | 0,0161 |
| 3 | 0,1419 | 0,0125 | 0,0772 | 0,0537 |
| 4 | 0,2128 | 0,0420 | 0,1274 | 0,1208 |
| 5 | 0,2270 | 0,1009 | 0,1640 | 0,1934 |
| 6 | 0,1766 | 0,1766 | 0,1766 | 0,2256 |
| 7 | 0,1009 | 0,2270 | 0,1640 | 0,1934 |
| 8 | 0,0420 | 0,2128 | 0,1274 | 0,1208 |
| 9 | 0,0125 | 0,1419 | 0,0772 | 0,0537 |
| 10 | 0,0025 | 0,0639 | 0,0332 | 0,0161 |
| 11 | 0,0003 | 0,0174 | 0,0089 | 0,0029 |
| 12 | 0,0000 | 0,0022 | 0,0011 | 0,0002 |

Mit Hilfe eines Computers wurden nun solche Paare $(p_1, p_2)$ gesucht, daß die Anpassung der theoretischen Verteilung an die empirische Geschlechterverteilung des Eingangsbeispiels besser wird.

Die Anpassung wurde mit Hilfe der Größe $\chi^2$ gemessen (vgl. die nachfolgende Ergänzung). Einige Paare $(p_1, p_2)$ mit den zugehörigen Werten von $\chi^2$ sind in der folgenden Tabelle abgedruckt:



| $p_1$ | $p_2$ | $\chi^2$ |
|---|---|---|
| 0,53 | 0,5036 | 68,4 |
| 0,54 | 0,4936 | 68,2 |
| 0,55 | 0,4836 | 44,6 |
| 0,56 | 0,4736 | 25,7 |
| 0,57 | 0,4636 | 20,3 |
| 0,58 | 0,4536 | 36,6 |
| 0,575 | 0,4586 | 25,2 |
| 0,565 | 0,4686 | 20,8 |
| 0,568 | 0,4656 | 19,9 |

Der $\chi^2$-Wert der besten Anpassung liegt knapp unterhalb des kritischen Werts zum 95%-Niveau ( $P(\chi^2 \leq 21{,}03) = 0{,}95$ ) !
Eine Untersuchung der Anpassung zu einer 'mittleren' Verteilung von 3 Binomialverteilungen zeigt, daß der kleinste zugehörige $\chi^2$-Wert ebenfalls noch zwischen 19 und 20 liegt.

Aufgrund der gegebenen Daten ist keine Bewertung des STERN-schen Modells möglich. Es wäre nützlich, wenn weitere empirische Daten über die Geschlechterverteilung vorlägen.

WELCHER LESER KANN WEITERHELFEN ?

Literatur:

STERN : Grundlagen der Humangenetik, Gustav-Fischer-Verlag, Stuttgart, 1968

STRICK: Einführung in die Beurteilende Statistik, Schroedel-Verlag, Hannover, 1980

Ergänzung zu: Verteilung der Geschlechter in Familien

Zur Benutzung des $\chi^2$-Anpassungstests

Hat man einen n-stufigen Zufallsversuch vorliegen, bei dem auf jeder Stufe r verschiedene Ausgänge möglich sind, die mit den festen Wahrscheinlichkeiten $p_1, p_2, \ldots, p_r$ auftreten, dann kann man nach n Stufen $\mu_1 = n \cdot p_1$-mal mit Ausgang Nr.1 , $\mu_2 = n \cdot p_2$-mal mit Ausgang Nr.2 ... , $\mu_r = n \cdot p_r$-mal mit Ausgang Nr.r rechnen.

Abweichungen von diesen Erwartungswerten kann man mit Hilfe der Größe $\chi^2$ messen:

$$\chi^2 = \frac{(\mu_1 - k_1)^2}{\mu_1} + \frac{(\mu_2 - k_2)^2}{\mu_2} + \ldots + \frac{(\mu_r - k_r)^2}{\mu_r}$$

wenn bei einem konkreten Zufallsversuch $k_1$-mal Ausgang Nr.1, ... , $k_r$-mal Ausgang Nr.r auftritt.

Liegt für einen Zufallsversuch eine Hypothese über die zugrundeliegenden $p_1$, ..., $p_r$ vor, dann deuten kleine Werte von $\chi^2$ auf eine gute Anpassung und große Werte auf eine schlechte Anpassung des Modells (der Hypothese) hin.

Es mag überraschend erscheinen, daß es unter bestimmten Voraussetzungen für nahezu beliebige $p_1,\ldots,p_r$,n gemeinsame kritische Werte für $\chi^2$ gibt, die nur von r abhängen, dh. daß man für diese Werte von $p_1,\ldots,p_r$ und n trotzdem dieselbe Tabelle von $\chi^2$ verwenden kann !

Liegen in unserem Beispiel dem 10690-stufigen Zufallsversuch für die 13 möglichen Ausgänge (0 Jungen,...,12 Jungen) tatsächlich die Wahrscheinlichkeiten 0,00016; 0,0021; ... ; 0,00036 zugrunde (vgl. erste Tabelle), dann erhält man mit der Wahrscheinlichkeit 0,95 eine Realisierung mit einem $\chi^2$-Wert, der kleiner oder gleich 21,03 ist. Lediglich mit der Wahrscheinlichkeit 0,05 erhält man einen größeren Wert für $\chi^2$. (In 99,9% der Stichproben treten $\chi^2$-Werte kleiner/gleich 32,91 auf und nur in 0,1% der Stichproben treten zufällig größere $\chi^2$-Werte auf.)

(Im vorliegenden Beispiel wurde der Wert für p (=0,5168) aus der Stichprobe geschätzt - dadurch reduziert sich die Anzahl der Freiheitsgrade auf f=11. Die zugehörigen kritischen Werte von $\chi^2$ sind: zum 95%-Niveau: 19,68 ; zum 99,9%-Niveau: 31,26.)

Literaturhinweis:

STRICK: Der Chiquadrat-Anpassungstest im Mathematik- und im Biologieunterricht der Sekundarstufe II, MNU 1981, Heft 3 (S. 139-147)