

Kleine 4-Felder-Tafel

**Vierfeldertafeln-Fisher-Test** Ha Dez 2011

Angelehnt an das Beispiel aus dem Skript Seite 21b. (Auf der Site ist es bayes.pdf)  
 A sind die Studis, die die Aufgaben machen, B diejenigen, die die Klausur bestehen.  
 Hier sind zunächst die Zahlen um Faktor 10 kleiner.

	B	$\bar{B}$	
A	9	1	10
$\bar{A}$	66	24	90
	75	25	100

Ersichtlich fällt bei der Fleißigen (A) ein geringerer Anteil durch.  $a=10 \cdot 10$  Aber ist das signifikant?????

H0: Fleiß lohnt sich nicht, die Durchfallquote ist gleich der allgemeinen Quote  
 H1: Fleiß lohnt sich, man hat ein geringeres Durchfallrisiko  
 Das ist ein einseitiger Test, da man bei Fleiß wohl kaum mehr durchfällt.  
 Die Tabelle gibt die Beobachtung wieder. Noch günstiger für H1 wäre 0 bei A mit nicht-B. |

1.1

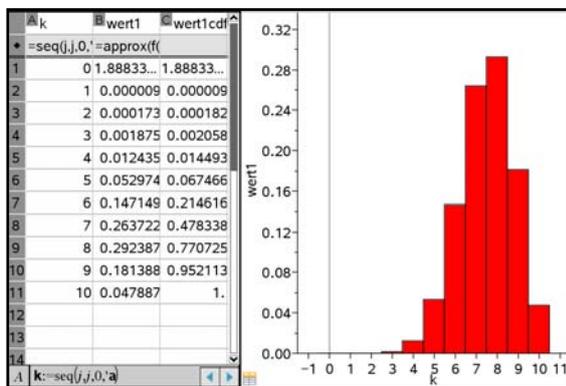
Darum ist die Irrtumswahrscheinlichkeit für die fälschliche Annahme von H1 die Wahrscheinlichkeit, dass diese Tafel oder eine günstigere unter H0 zufällig eintritt.

$$f(x) = \frac{nCr(10,x) \cdot nCr(90,25-x)}{nCr(100,25)} \cdot \text{Fertig}$$

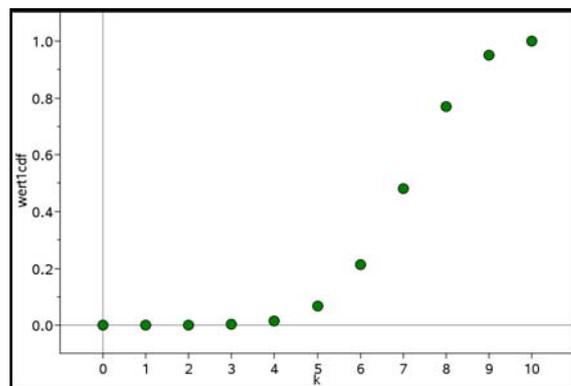
Also  $f(1) = 0,181388$  errechnet sich aus der Zahl der Möglichkeiten, den einen Erfolgrlosen unter den 10 Fleißigen anzukreuzen mal die Zahl der Möglichkeiten die 24 Erfolgrlosen unter den 90 Faulen auszuwählen dividiert durch die Gesamtzahl überhaupt 25 Erfolgrlose von 100 Studis anzukreuzen. Das ist die Wahrscheinlichkeit, dass genau ein Fleißiger unten den 10 Fleißigen durchfällt und 24 Faule unten den 90 Faulen durchfallen, wenn sowieso 25 von 100 Studis durchfallen. Ebenso  $f(0) = 0,047887$  Zusammen tritt also eine oder günstigere Tabelle mit  $f(1) - f(0) = 0,229275$  23% Wahrscheinlichkeit auf, ein so großes alpha ist nicht akzeptierbar. Darum ist Fleiß nicht signifikant besser als Faulheit.

Auf Seite 1.3 ist  $f(x) = \frac{nCr(10,x) \cdot nCr(90,75-x)}{nCr(100,75)}$  bezogen auf die bestehenden Studis mit alle Werten berechnet und dargestellt. Die beiden letzten Balken entsprechen den eben berechneten Zahlen. Seite 1.4 zeigt die kumulierte Verteilungsfunktion.

1.2



1.3



1.4

Großes Beispiel

Beispiel mit den 10-fach größeren Zahlen aus der Bayes-Seite 21c

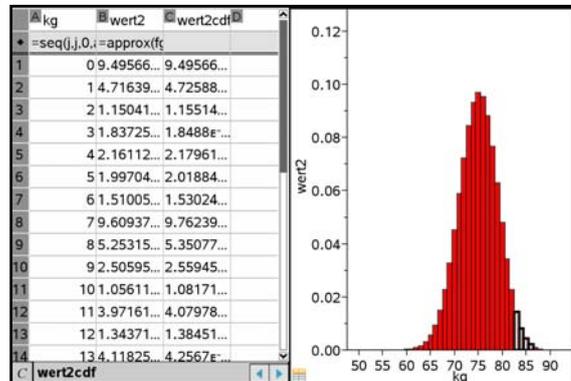
	B	$\bar{B}$	
A	95	5	100
$\bar{A}$	955	245	1000
	1050	250	1300

Anzahl Fleißige (machne Aufgaben, A)  $ag=100 \cdot 100$

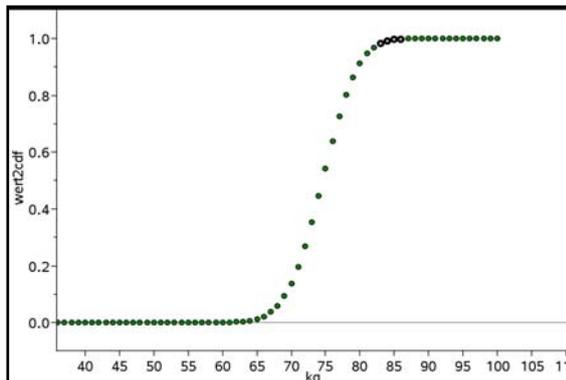
$$fg(x) = \frac{nCr(100,x) \cdot nCr(900,750-x)}{nCr(1000,750)} \cdot \text{Fertig}$$

$fg(95) = 3,18494e 8$  winzig!!  
 $fg(83) = 0,014258$  Schon mit nur 83 Bestehenden unter den 100 Fleißigen wäre das Ergebnis signifikant gewesen.  
 $ll:=seq(fg(x),x,83,100)$   
 $\sum(ll) = 0,030619$  und zwar auf dem 3%-Niveau.  
 Das bekommt man interaktiv mit der kumulierten Verteilungsfunktion Seite 2.3 heraus.  
 Es wird deutlich, dass mit dem 4-Felder-Test von Fisher ohne Computer bei größeren Zahlen keine Chance hat.

2.1



2.2



2.3

**Bemerkung zur Anordnung der Tafel:**  
 Es ist egal, ob man sich auf die Fleißigen oder die Faulen bezieht.  
 Es ist sogar egal, ob man überhaupt stattdessen die Bestehenden und die Durchfallenden betrachtet.

	E	$nE$	
A	ea	na	a
B	eb	nb	b
	e	n	n

So eine Tafel tritt zufällig mit der Wahrscheinlichkeit

$$\frac{a! \cdot b! \cdot e! \cdot n!}{(ea! \cdot na! \cdot eb! \cdot nb!) \cdot n!}$$

ein.  
 Der Term ist symmetrisch in Zeilen bzw. Spalten.  
 Man braucht sich also bezüglich der Anordnung keinen Kopf zu machen.

2.4