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 ▸ 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 Duchfallrisiko

Das ist ein einseitiger Test, da man bei Fleiß wohl kaummehr 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.

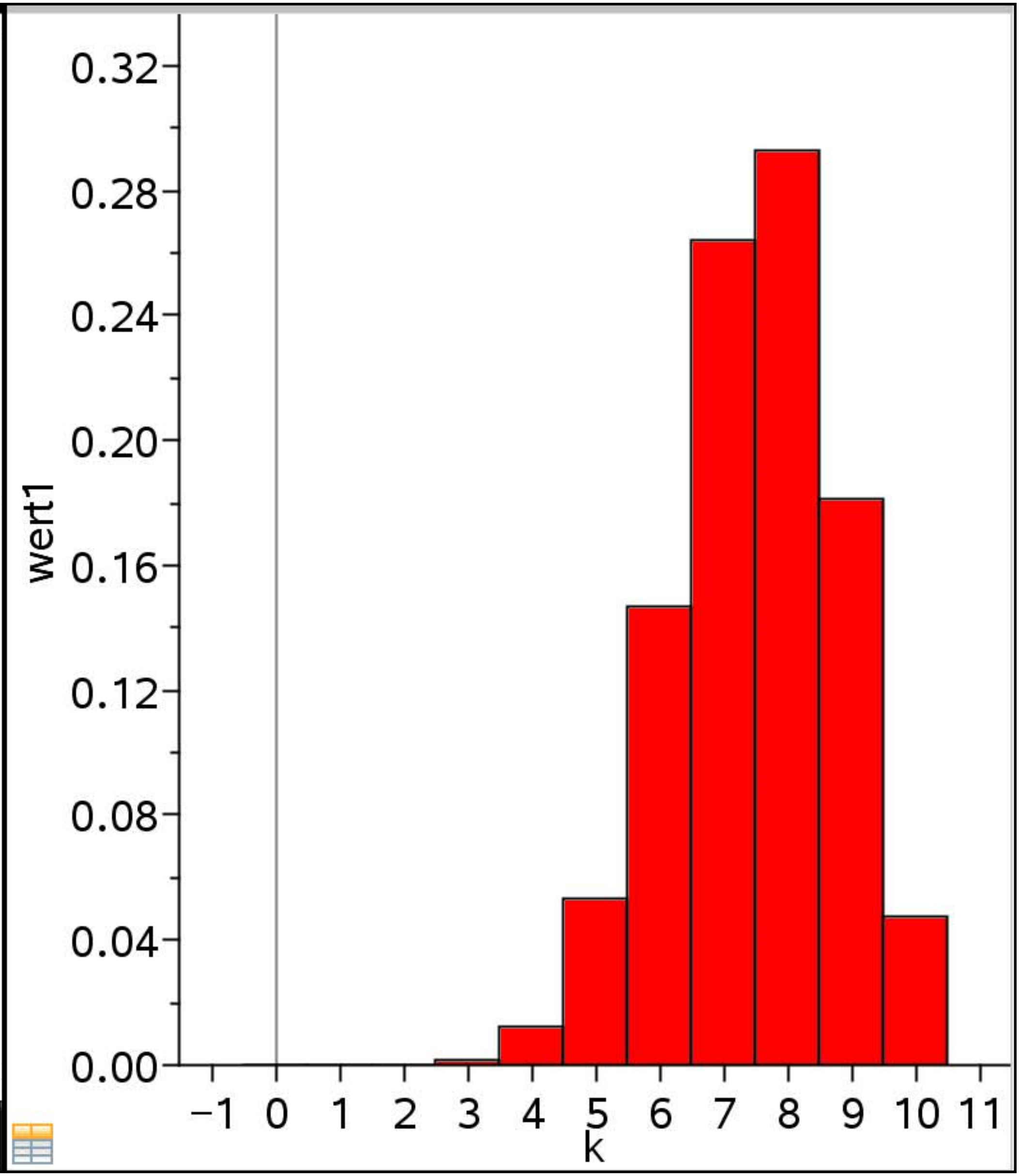
$\mathbf{t}(x):=\frac{\mathrm{nCr}(10,x)\cdot\mathrm{nCr}(90,25-x)}{\mathrm{nCr}(100,25)}$ ▸ *Fertig* Also $\mathbf{t}(1)$ ▸ 0.181388 errechnet sich aus der
Zahl der Möglichkeiten, den einen Erfolglosen unter den 10 Fleißigen anzukeuzen **mal** die Zahl der Möglichkeiten die 24 Erfolglosen unter den 90 Faulen auszuwählen dividiert durch die Gesamtzahl überhaupt 25 Erfolglose 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. Ebeso $\mathbf{t}(0)$ ▸ 0.047887 Zusammen tritt also so eine oder günstigere Tabelle mit $\mathbf{t}(1)+\mathbf{t}(0)$ ▸ 0.229275 23% Wahrscheinlichkeit auf, ein so großes alpha ist nicht akzeptierbar. Darum ist Fleiß nicht signifikant besser als Fauheit.

Auf Seite 1.3 ist $\mathbf{f}(x):=\frac{\mathrm{nCr}(10,x)\cdot\mathrm{nCr}(90,75-x)}{\mathrm{nCr}(100,75)}$ ▸ *Fertig* 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.

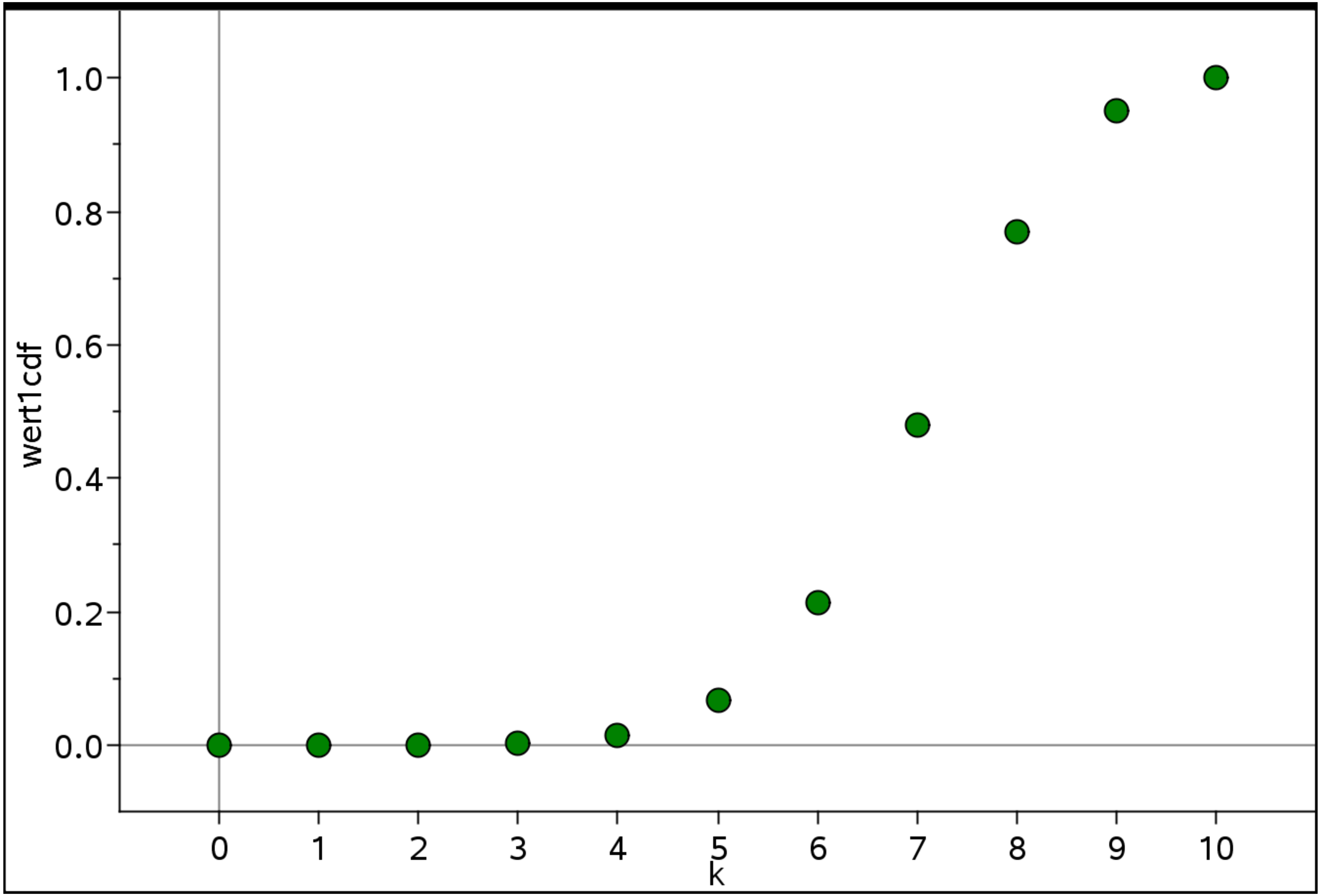
1.2

	A k	B wert1	C wert1cdf
◆	=seq(j,j,0,'	=approx(f(	
1	0	1.88833...	1.88833...
2	1	0.000009	0.000009
3	2	0.000173	0.000182
4	3	0.001875	0.002058
5	4	0.012435	0.014493
6	5	0.052974	0.067466
7	6	0.147149	0.214616
8	7	0.263722	0.478338
9	8	0.292387	0.770725
10	9	0.181388	0.952113
11	10	0.047887	1.
12			
13			
14			

A **k**:=seq(j,j,0,'**a**)



1.3



1.4

Großes Besipiel

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

	B	$\overline{B}$	
A	95	5	100
$\overline{A}$	655	245	900
	750	250	1000

Anzahl Fleißige (machne Aufgaben, A) $\textbf{ag}:=100$ ▸ 100

$\textbf{fg}(x):=\dfrac{\mathrm{nCr}(100,x)\cdot\mathrm{nCr}(900,750-x)}{\mathrm{nCr}(1000,750)}$ ▸ *Fertig* $\textbf{fg}(95)$ ▸ $3.18494\mathrm{E}{-8}$ winzig!!

$\textbf{fg}(83)$ ▸ 0.014258 Schon mit nur 83 Bestehenden unter den 100 Fleißigen wäre das Ergebnis signifikant gewesen.

$\textbf{li}:=\mathrm{seq}(\textbf{fg}(x),x,83,100)$

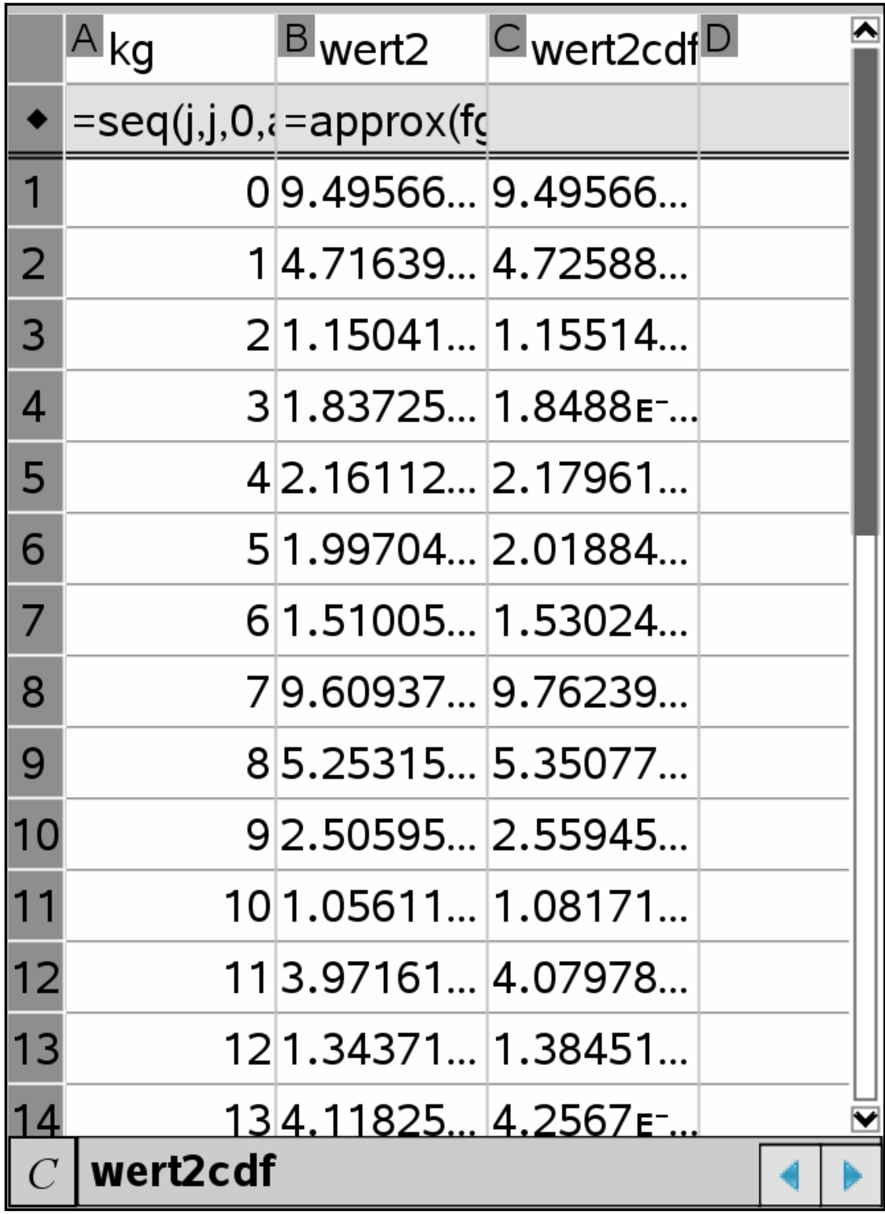
▸ $\{0.014258,0.008225,0.004388,0.002156,0.000972,0.0004,0.000149,0.00005,0.000015,0.$

$\mathrm{sum}(\textbf{li})$ ▸ 0.030619 und zwar auf dem 3%−Niveau.

Das bekommt man interaktiv mit der kumulierten Verteilungsfunktion Seite 2.3 heraus.

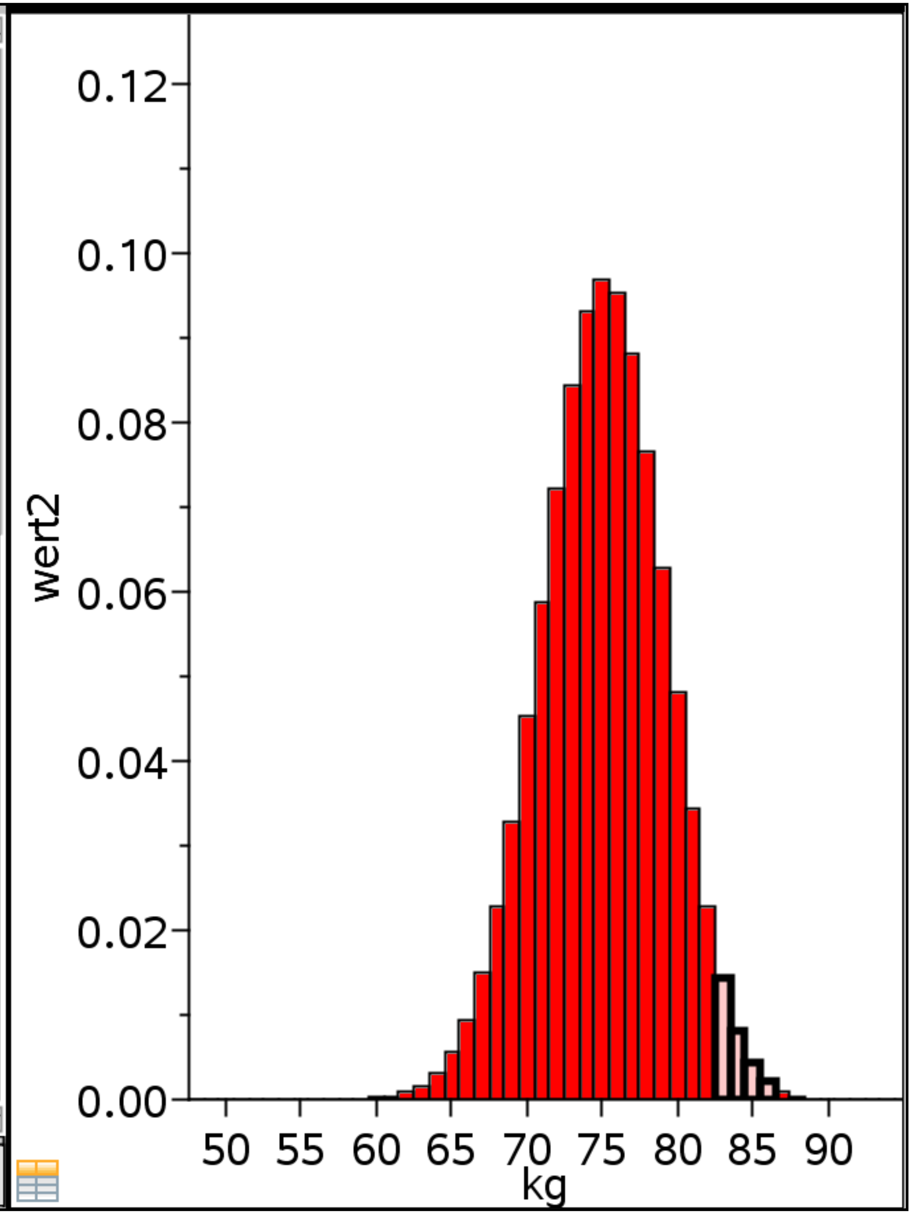
Es wird deutlich, dass mit mit dem 4−Felder−Test von Fisher ohne Computer bei größeren Zahlen keine Chance hat.

2.1

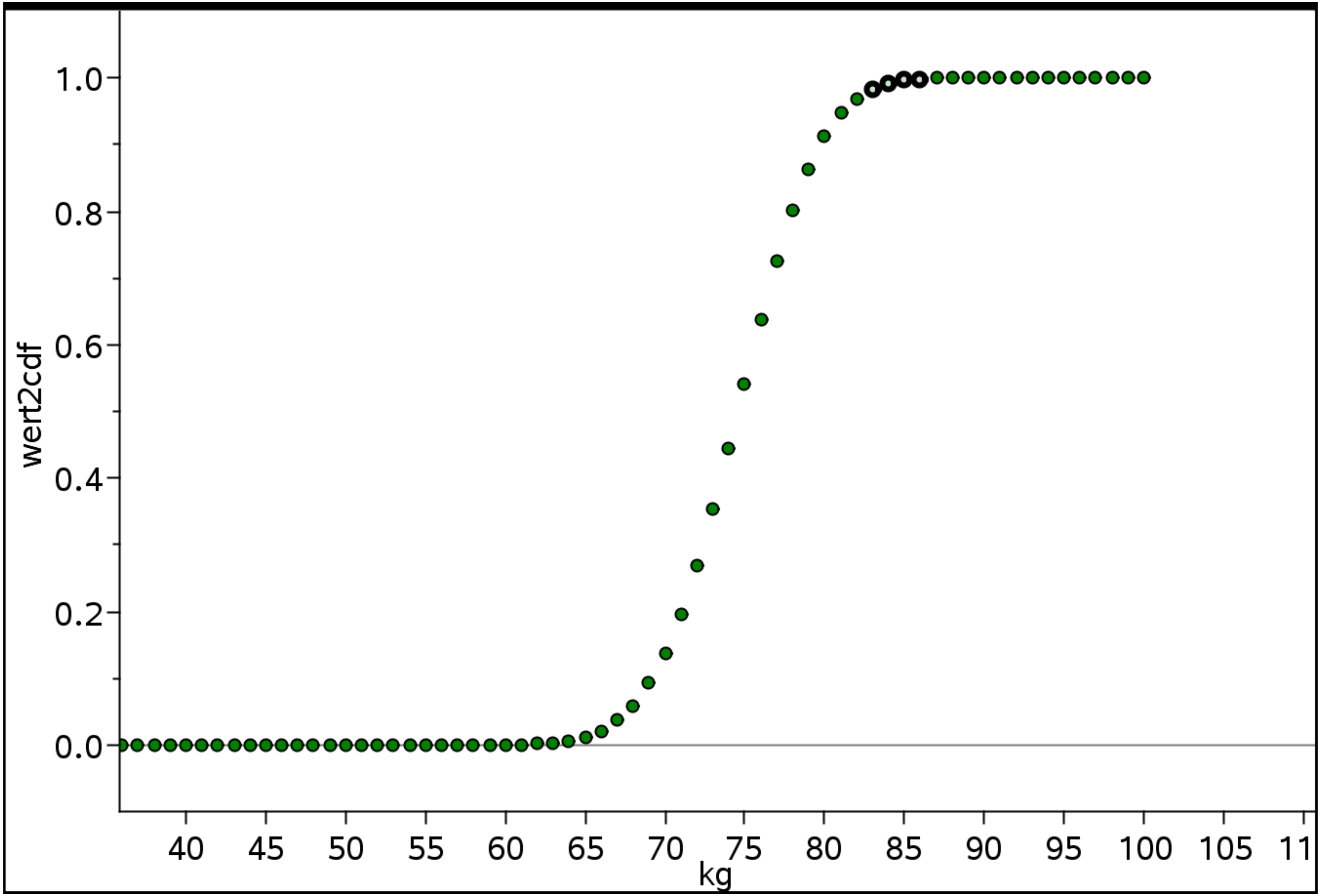


	A kg	B wert2	C wert2cdf	D
◆	=seq(j,j,0,ε	=approx(fc		
1	0	9.49566...	9.49566...	
2	1	4.71639...	4.72588...	
3	2	1.15041...	1.15514...	
4	3	1.83725...	1.8488ᴇ⁻...	
5	4	2.16112...	2.17961...	
6	5	1.99704...	2.01884...	
7	6	1.51005...	1.53024...	
8	7	9.60937...	9.76239...	
9	8	5.25315...	5.35077...	
10	9	2.50595...	2.55945...	
11	10	1.05611...	1.08171...	
12	11	3.97161...	4.07978...	
13	12	1.34371...	1.38451...	
14	13	4.11825...	4.2567ᴇ⁻...	

C **wert2cdf**



2.2


1.0
0.8
0.6
0.4
0.2
0.0
wert2cdf
40
45
50
55
60
65
70
75
80
85
90
95
100
105
11
kg

2.3

Bemerkung zur Anordnung der Tafel:

Es ist egal, ob man sich auf die Fleißigen oder die Faulen bezieht.

Es ist sogar egal, ober man überhaupt stattdessen die Bestehenden und die Durchfallenden betrachtet.

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

So eine Tafel tritt zufallig mit der Wahrscheinlichkeit

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

Der Term ist symmetrisch in Zeilen bzw. Spalten.

Man braucht sich also bezüglich der Anordnung keinen Kopf zu machen.

2.4