Zur Studie der Epi.Consult GmbH vom 01.09.2009

1. Einleitung

Im Folgenden wird unsere Rechnung zur Epi.Consult-Studie vom 01.09.2009 /Link/ vorgestellt.

Zunächst werden in Abschnitt 2 die Formeln angegeben, mit denen wir die Tabellen in der Studie ausgewertet haben. Dann werden wir in Abschnitt 3 unsere Ergebnisse darstellen und schließlich in Abschnitt 4 noch auf methodische Fehler in o. g. Studie eingehen.
Im Anhang werden die hier verwendeten Formeln hergeleitet, da diese in der Literatur leider nicht einheitlich sind.

2. Benutzte Formeln

Zur Auswertung werden folgende Formeln benutzt:

Gepoolter Mittelwert x_m

$$x_m = \frac{1}{n}\sum_{j=1}^{r}(n^j y_m^j) \quad \text{Gl. (5)}$$

Gepoolte Standardabweichung

$$s^2 = \frac{1}{n-1}\left[\sum_{j=1}^{r}(n^j - 1)(s^j)^2 + \sum_{j=1}^{r} n^j (y_m^j - x_m)^2\right] \quad Gl.(7)$$

mit

$$n = \sum_{j=1}^{r} n^j \quad Gl.(6)$$

Dabei bedeuten:
r = Anzahl der Kernkraftwerke (AKW)
n^j = Anzahl der erwarteten Fälle bei einem AKW
(Der Index j kennzeichnet das jeweilige AKW, er wurde für eine bessere Übersichtlichkeit hochgestellt.)
n = Anzahl aller erwarteten Fälle nach Gl. (6)
y_m^j = SIR-Wert
s^j = Standardabweichung bei einem AKW

Wir möchten darauf hinweisen, dass für n^j die Anzahl der erwarteten Fälle einzusetzen ist. Da die Zahlen für n^j wiederum aus einer Statistik für das jeweilige AKW stammen, ergeben sich dafür im Allgemeinen Dezimalzahlen und keine ganze Zahlen.

Der SIR-Wert ist gemäß der Studie die Anzahl der tatsächlichen Fälle z^j dividiert durch die Anzahl der erwarteten Fälle n^j in der betrachteten Region:

$$y_m^j = \frac{z^j}{n^j} \quad \text{Gl. (8)}$$

Setzen wir Gl. (8) in Gl. (5) ein, so erhalten wir:

$$x_m = \frac{1}{n}\sum_{j=1}^{r}\left(n^j \frac{z^j}{n^j}\right) = \frac{1}{n}\sum_{j=1}^{r} z^j \quad \text{Gl. (9)}$$

Die letzte Formel besagt: Der gepoolte Mittelwert ist gleich der Summe der Anzahl der tatsächlichen Fälle z^j dividiert durch die Anzahl n aller erwarteten Fälle, was man auch erwarten würde.

3. Ergebnisse

Es wurden die Tabellen 4 bis 9 der Epi.Consult-Studie mit den o. g. Formeln ausgewertet. Das Ergebnis ist in der folgenden Tabelle zusammengestellt. Außerdem enthält die Tabelle zum Vergleich die entsprechenden Zahlen aus der Studie:

Tab.-Nr.	**Unsere Rechnung**		**Epi.Consult-Studie**	
	SIR	**95%-Vertrauensbereich**	**SIR**	**95%-Vertrauensbereich**
4*	**1.06**	**(0.17 - 1.95)**	1.19	(1.13 - 1.25)
5	**1.02**	**(0.09 - 1.95)**	1.14	(1.05 - 1.25)
6	**1.09**	**(0.02 - 2.15)**	1.24	(1.12 - 1.37)
7	**1.03**	**(0.** - 2.09)**	1.20	(1.08 - 1.33)
8	**0.99**	**(0.** - 2.10)**	1.22	(1.08 - 1.36)
9	**1.03**	**(0.22 - 1.84)**	1.13	(1.10 - 1.17)
9a	**0.98**	**(0.58 - 1.39)**	--	--

Tab. Nr. 9a: Ergebnis nur für die deutschen AKWs aus Tabelle 9
*Bei Tab. Nr. 4 wurde die dritte Zeile (Fall DE) nicht berücksichtigt, da hier wohl das Ergebnis der KiKK-Studie eingeflossen ist.
**Es ergab sich rechnerisch ein negativer Wert

Wir verwendeten für die Berechnung die Gl. (7). Die Gl. (7) besteht aus zwei Summanden, die beide immer positiv sind. Wir hatten aus Gründen der Arbeitsersparnis und vor allem wegen der Unsicherheit der Berechnung von s^j nur den zweiten Teil benutzt. Durch das Weglassen des ersten Teils von Gl. (7) wird der Vertrauensbereich kleiner.

Man erkennt, dass die hier berechneten (gepoolten) SIR-Werte wesentlich kleiner sind als die in der Studie ausgewiesenen Werte. Für die deutschen AKWs (s. Tab. Nr. 9a) liegt der Wert sogar unterhalb von 1.00, d. h. es sieht so aus, als ob das Risiko für Leukämie in der Nähe eines deutschen AKWs im Durchschnitt kleiner wäre als an einem Ort weiter weg. Außerdem sind in allen Fällen die Vertrauensbereiche trotz o. g. Vernachlässigung des ersten Teils der Gl. (7) wesentlich größer als die in der Studie ausgewiesenen, wobei die untere Grenze des Vertrauensbereichs immer deutlich unterhalb von SIR = 1 liegt. Dies bedeutet, aus den Tabellen 4 bis 9 der o. g. Studie kann man nicht entnehmen, dass in der Nähe eines AKWs das Risiko, an Leukämie zu erkranken, größer ist.

Wir sind der Meinung, dass die Vertrauensbereiche so groß sind, dass auch mit einer anderen Studie, bei der **nur** die statische Methode (etwa wie bei der KiKK Studie) anders ist, eine statistisch gesicherte Aussage über das Leukämierisiko nicht möglich ist. Dies kann u. E. - wenn überhaupt - nur mit einer verbesserten Datenbasis gelingen.

Wir sind der Meinung, dass die Ergebnisse in der Studie falsch sind. Weiterhin sind wir der Meinung, dass die Studie der Epi.Consult insgesamt unseriös ist, weil sie keine Angaben darüber enthält, wie die gepoolten Mittelwerte und die zugehörigen gepoolten Vertrauensbereiche berechnet worden sind (außer der Angabe, dass „0“ durch „0.01“ ersetzt worden ist).

4. Systematische Fehler

Die Professoren Walter Krämer, Dortmund, und Gerhard Arminger, Wuppertal, weisen darauf hin, dass die Studie unabhängig von der Berechnung – wie viele Studien - zusätzliche systematische Fehler enthält /Link/. Diese Fehler betreffen die Auswahl der Datensätze, die Behauptung einer Signifikanz, die in Wirklichkeit gar nicht gegeben ist, und das Nichtbeachten anderer Ursachen für die Leukämie. So ergab z. B. eine Studie in den USA, dass das Risiko für die Erkrankung an Leukämie bei Kinder mit dem Wohlstand der Eltern ansteigt, und eine andere, dass in der Nähe von Baumschulen dieses Risiko ansteigt. Ebenso spielt die Abstammung eine Rolle. Man kann sich selbst leicht ausmalen, dass AKWs nicht von ganz armen Menschen erbaut und betrieben werden. Oder allgemeiner: Selbst wenn in der Nähe eines AKWs ein erhöhtes Leukämierisiko festgestellt würde, muss dies nicht unbedingt etwas mit dem AKW zu tun haben.

Weiterhin wiesen die beiden Professoren darauf hin, dass über Studien, aus denen sich an AKWs (angeblich) ein höheres Risiko in deren Nähe ergibt, in der Presse viel häufiger berichtet wird als über solche, in denen kein erhöhtes Risiko nachgewiesen wurde. Dies führt dazu, dass in der Öffentlichkeit der Eindruck entsteht, in der Nähe von AKWs bestehe ein erhöhtes Risiko.

Zusammenfassend lässt sich feststellen, dass es bisher keine Studie gibt, die ein erhöhtes Leukämierisiko in der Nähe von AKWs seriös nachgewiesen hat. (Während dieser Bericht entstand, „geistert" schon wieder eine neue Studie, diesmal aus Italien, durch die Medien.)

Anhang: Ableitung der Formeln

Das Mittelwert x_m von n Zahlen x_i wird wie folgt berechnet:

$$x_m = \frac{1}{n}\,(x_1 + x_2 + \cdots + x_n)$$

Gl. (1a)

oder die gleiche Formel mit Summenzeichen geschrieben

$$x_m = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Gl.(1b)

Beispiel:
Der Mittelwert der 5 Zahlen 1.4, 1.6, 1.8, 2.0 und 2.2 beträgt
x_m = (1.4 + 1.6 +1.8 + 2.0 + 2.2) / 5 = 1.8
(Wir haben hier für die Zahlen statt eines Kommas einen Punkt wie im englischen Sprachraum benutzt, um eine Verwechslung mit dem Komma zum Abtrennen der Zahlen zu vermeiden.)

Die Streuung s (quadratische Abweichung) von n Zahlen x_i wird wie folgt berechnet:

$$s^2 = \frac{1}{n-1}\left(\,(x_1 - x_m)^2 + \ (x_2 - x_m)^2 + \ \ldots + \ (x_n - x_m)^2\right) \quad \text{Gl. (2a)}$$

oder mit Summenzeichen geschrieben

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - x_m)^2 \quad \text{Gl. (2b)}$$

s^2 bedeutet, dass die Formel zunächst nur einen Wert für das Quadrat von s liefert und man deshalb noch die Wurzel aus s^2 berechnen muss, um den Wert für s zu erhalten.
(siehe z. B. „Taschenbuch der Mathematik" von Bronstein · Semendjajew, Verlag: Harri Deutsch 1980, ab Seite 717)

Beispiel:
Der Streuung der o. g. fünf Zahlen 1.4, 1.6, 1.8, 2.0 und 2.2 beträgt:

$$s^2 = \frac{1}{5-1}\,(\,(1.4-1.8)^2 + \,(1.6-1.8)^2 + \,(1.8-1.8)^2 + (2.0-1.8)^2 + (2.2-1.8)^2)$$

$$s^2 = \frac{1}{4}\,(\,(-0.4)^2 + \,(-0.2)^2 + \,(0)^2 + (0.2)^2 + (0.4)^2)$$

$$s^2 = \frac{1}{4}\,(\,0.16 + 0.04 + 0. + 0.04 + 0.16) = \,0.10$$

s = 0.316

Die Streuung ist ein Maß dafür, wie weit die Zahlen auseinanderliegen.

In vielen Fällen liegen die Zahlen in Gruppen unterteilt vor.
Beispiel:
Eine (sehr kleine) Schule hat r=3 Klassen mit n^1=4, n^2=5, n^3=2 Schüler (bzgl. hochgestellter Indizes s. obige Bem. zu j). Jeder Schüler hat die Durchschnittsnote x_i:

Erste Klasse:	x_1, x_2, x_3, x_4	= 2.3, 3.4, 3.4, 4.2
Zweite Klasse:	x_5, x_6, x_7, x_8, x_9	= 1.3, 2.1, 3.4, 4.2, 5.3
Dritte Klasse:	x_{10}, x_{11}	= 1.9, 2.5

Man kann nun die elf Zahlen x_1, x_2, .. x_{11} in Gl. (1a) und Gl. (2a) einsetzen, um den Mittelwert x_m und die Streuung s für die ganze Schule zu berechnen. Oft möchte man aber auch die Klassen untereinander vergleichen und zunächst für jede Klasse x_m und s bestimmen und daraus dann die Werte für die gesamte Schule.

Um die Formeln besser formulieren zu können, führen wir für die Durchschnittsnote y_i^j ein. Es sind dieselben Zahlen wie die x_i, nur dass wir für jede Klasse die Nummerierung der Zahlen jedes Mal mit 1 beginnen und hochgestellt die Klassennummer schreiben.
In unserem Beispiel also:

Erste Klasse:	$y_1^1, y_2^1, y_3^1, y_4^1$	= 2.3, 3.4, 3.4, 4.2
Zweite Klasse:	$y_1^2, y_2^2, y_3^2, y_4^2, y_5^2$	= 1.3, 2.1, 3.4, 4.2, 5.3
Dritte Klasse:	y_1^3, y_2^3	= 1.9, 2.5

Der Mittelwert y_m^j für die Schulklasse j berechnet sich gemäß Gl. (2a) wie folgt:

$$y_m^j = \frac{1}{n_j}\sum_{i=1}^{n_j} y_i^j \quad Gl.(3)$$

und die Streuung

$$(s^j)^2 = \frac{1}{n^j - 1} \sum_{i=1}^{n^j} \left(y_i^j - y_m^j\right)^2 \quad Gl. (4)$$

Wir überlegen uns nun, wie wir y_m^j in Gl. (1b) einführen können. Gl. (1b) kann wie folgt geschrieben werden:

$$x_m = \frac{1}{n} \sum_{j=1}^{r} \sum_{i=1}^{n^j} y_i^j$$

r = Anzahl der Gruppen von Zahlen

Es wird Gl. (3) verwendet und wir erhalten als Endergebnis für x_m:

$$x_m = \frac{1}{n} \sum_{j=1}^{r} (n^j y_m^j) \quad \text{Gl. (5)}$$

Dabei ist n die Summe der n^j:

$$n = \sum_{j=1}^{r} n^j \quad \text{Gl. (6)}$$

Ähnlich wird Gl. (2b) bearbeitet:

$$s^2 = \frac{1}{n-1} \sum_{j=1}^{r} \sum_{i=1}^{n^j} \left(y_i^j - y_m^j + y_m^j - x_m\right)^2$$

$$s^2 = \frac{1}{n-1} \sum_{j=1}^{r} \sum_{i=1}^{n^j} \left[(y_i^j - y_m^j)^2 + 2\left(y_i^j - y_m^j\right)(y_m^j - x_m) + \left(y_m^j - x_m\right)^2\right]$$

Der zweite Summand verschwindet, wie im Folgenden gezeigt wird:

$$\sum_{j=1}^{r} \sum_{i=1}^{n^j} 2(y_i^j - y_m^j)\left(y_m^j - x_m\right) = 2 \sum_{j=1}^{r} \left[\left(y_m^j - x_m\right) \sum_{i=1}^{n^j} (y_i^j - y_m^j)\right]$$

Man beachte, dass die erste Summe auch die zweite Summe einschließt. Die zweite Summe kann wie folgt umgeformt werden:

$$\sum_{i=1}^{n^j} (y_i^j - y_m^j) = \left[\sum_{i=1}^{n^j} y_i^j \right] - n^j y_m^j = 0 \text{ wegen Gl. (3)}$$

Zur Berechnung von s^2 brauchen wir nur den ersten und letzten Summanden zu berücksichtigen:

$$s^2 = \frac{1}{n-1} \left[\sum_{j=1}^{r} \sum_{i=1}^{n^j} (y_i^j - y_m^j)^2 + \sum_{j=1}^{r} \sum_{i=1}^{n^j} (y_m^j - x_m)^2 \right]$$

In den ersten Summand wird Gl. (4) eingesetzt. Beim zweiten Summand kann die zweite Summe sofort ausgeführt werden.

$$s^2 = \frac{1}{n-1} \left[\sum_{j=1}^{r} (n^j - 1)(s^j)^2 + \sum_{j=1}^{r} n^j (y_m^j - x_m)^2 \right] \quad \text{Gl. (7)}$$

Die Gl. (5) wird zur Berechnung des gepoolten Mittelwertes benutzt und Gl. (7) für die gepoolte Standardabweichung.

Während Gl. (5) für die Berechnung des gepoolten Mittelwertes wohl allgemein üblich ist, findet man für die gepoolte Standardabweichung auch andere Gleichungen wie z. B. die folgende

$$s^2 = \frac{1}{n-r} \sum_{j=1}^{r} (n-j)(s^j)^2 \quad \text{Gl. (8)}$$

Die Standardabweichung wird zur Berechnung des Vertrauensbereichs benutzt. Wir benutzen dafür die t-Verteilung (Student-Verteilung). Welche Verteilung in der Epi.Consult-Studie benutzt wird, steht dort nicht. In der Epi.Consult-Studie wird von einem 95% Vertrauensbereich ausgegangen. Die Zahlenwerte für den 95% Vertrauensbereich werden folgender **Tabelle A** entnommen:

m	1	2	3	4	5	6	7	8	9	10	12	14
t	12.706	4.303	3.182	2.776	2.571	2.447	2.365	2.306	2.262	2.228	2.179	2.145

m	16	18	20	22	24	26	28	30	unendlich
t	2.120	2.101	2.086	2.074	2.064	2.056	2.048	2.042	1.960

m = Anzahl der Freiheitsgrade, t=Abstand vom Maximum der Verteilung
(Tabelle aus „Taschenbuch der Mathematik“ von Bronstein · Semendjajew, Verlag: Harri Deutsch 1980, Seite 74 für α=0,05)

Der Vertrauensbereich für das gepoolte Ergebnis zwischen vb1 (untere Grenze) und vb2 (obere Grenze) wird folgt berechnet:

$vb1 = x_m – t*s$
$vb2 = x_m + t*s$ Gl.(9)

Für t werden folgende Werte (s. Tabelle A) benutzt:
Für die Auswertung der Tabelle 5, 6, 7, 8 t =2.101 (m=18)
Für die Auswertung der Tabelle 4 und 9 t= 1.960 (m=unendlich)
Für die Auswertung der Tabelle 9a t= 2.145 (m=14)

26.03.2011 gr