960-584– Biostatistics I– Fall, 2003

Homework 2 Solutions, 6 Oct 2003

1. The following are counts of deaths by falls in a certain cohort, by month, presented by Ryan and Joiner (1994) and originally published in the World Almanac and Book of Facts (1984).

Month	Falls	Days in Month	Ice Score
Jan	1150	31	1
Feb	1034	28	1
Mar	1080	31	2
Apr	1126	30	0
May	1142	31	0
Jun	1100	30	0
Jul	1112	31	0
Aug	1099	31	0
Sep	1114	30	0
Oct	1079	31	0
Nov	999	30	2
Dec	1181	31	2

This data may also be found at `http://www.stat.rutgers.edu/~kolassa/960-584/falls.dat`.

a. Investigators wish to test whether the rate of death by falls is the same in the iciest months as it is during the least icy months, and assume that the rate of deaths by falls per day is the same among all of the days in the most icy months, and the same among all the days in the least icy months. Perform this test.

There are $1080 + 999 + 1181 = 3260$ *falls in the iciest months, and* $1126 + 1142 + 1100 + 1112 + 1099 + 1114 + 1079 = 7772$ *falls in the least icy months. There are* 92 *days in the iciest months, and* $30 + 31 + 30 + 31 + 31 + 30 + 31 = 214$ *days in the least icy months. Under the null hypothesis, the proportion of falls in the iciest months to the number of deaths by fall in the icyest and least icy month is* $92/(92 + 214) = .3007$ *. We actually observe* $3260/(3260 + 7772) = 0.2955$ *of the deaths by fall during the icyest months. The* Z *statistic associated with the one–sample binomial test is* $(0.2955 - .3007)/\sqrt{.3007 \times .6993/11032} = -1.19$ *, corresponding to a two–sided p value of* 0.234 *. There is no evidence that a greater number of falls happen either in the icyest or the least icy months.*

Alternaltely, you might get a confidence interval for the difference in the number of deaths by fall per days in the month. This isn't quite the rate of deaths by falls, since you don't have the population size. The confidence interval is

$$(3260/92) - (7772/214) \pm 1.96\sqrt{3260/92^2 + 7772/214^2} = (-2.343, 0.577).$$

This confidence interval fails to exclude zero, and so do not reject the null hypothesis of equality.

b. Test the null hypothesis that the rate of deaths from falls is the same all year long, vs. the hypothesis that it may be different in different months. This is equivalent to testing equality of population SMRs, if the person–years at risk are proportional to the number of days in the month.

c. Test for an increasing dose–response relationship between ice and falls, against the null hypothesis of part (b).

I'll do the calculations for both parts (b) and (c) in one go.

```
data falls; infile 'falls.dat'; input falls days score; run;
proc means noprint data=falls; output out=tot sum=; run;
data tot; set tot; tfalls=falls; tdays=days;
 keep tfalls tdays; run;
data falls; set tot falls; retain ttfalls ttdays;
 if _n_=1 then do; ttdays=tdays; ttfalls=tfalls; end;
 exp=ttfalls*days/ttdays; stat+(falls-exp)**2/exp;
 pv=1-probchi(stat,_n_-2);
 if _n_=1 then delete;
 keep falls days score stat pv exp ttfalls; run;
data falls; set falls;
 keep falls days score exp stat pv stat2 pv2;
 v1+score**2*exp; v2+score*exp;
 ostat+score*(falls-exp);
 stat2=ostat/sqrt(v1-v2**2/ttfalls);
 pv2=2*(1-probnorm(abs(stat2))); run;
proc print data=falls noobs; run;
```

The output is

FALLS	DAYS	SCORE	EXP	STAT	PV	STAT2	PV2
1150	31	1	1122.45	.	.	.	.
1034	28	1	1013.83	.	.	.	.
1080	31	2	1122.45	.	.	.	.
1126	30	0	1086.25	.	.	.	.
1142	31	0	1122.45	.	.	.	.
1100	30	0	1086.25	.	.	.	.
1112	31	0	1122.45	.	.	.	.
1099	31	0	1122.45	.	.	.	.
1114	30	0	1086.25	.	.	.	.
1079	31	0	1122.45	.	.	.	.
999	30	2	1086.25	.	.	.	.
1181	31	2	1122.45	17.6924	0.088997	-0.96532	0.33439

Hence the p *-value for the test requested in part (b) of no association is* $.089$ *. From above, the* p *-value for the test requested in part (c) of no increase in falls as iciness increases is* $.334$ *. Also accept the one-sided test.*