Examples of Poisson, ‘Extra-Poisson’ and ‘Less-than-Poisson’ Variation

- One definition of epidemiology:

Examples of Poisson, ‘Extra-Poisson’ and ‘Less-than-Poisson’ Variation

- One definition of epidemiology:

  “Disease is not distributed at random”

- One definition of epidemiology:

  "Disease is not distributed at random"

- When might we expect (just) Poisson variation?

Examples of Poisson, 'Extra-Poisson' and 'Less-than-Poisson' Variation

- One definition of epidemiology:

  "Disease is not distributed at random"

- When might we expect (just) Poisson variation?

- When might we expect more than (i.e., extra-) Poisson variation?

  And what do we do about it?

Examples of Poisson, 'Extra-Poisson' and 'Less-than-Poisson' Variation

- One definition of epidemiology:

  "Disease is not distributed at random"

- When might we expect (just) Poisson variation?

- When might we expect more than (i.e., extra-) Poisson variation?

  And what do we do about it?

- Can you think of a context where counts show less-than-Poisson variation?

Deaths from Horsekicks

Deaths from Horsekicks

“The chance of a man being killed by horsekick on any one day is exceedingly small, but if an army corps of men are exposed to this risk for a year, often one or more of them will be killed in this way" [R. A. Fisher, 1925, using just 10 of 14 corps used by Bortkiewicz in his 1898 work, Das Gesetz der kleinen Zahlen.]

Deaths from Horsekicks

"The chance of a man being killed by horsekick on any one day is exceedingly small, but if an army corps of men are exposed to this risk for a year, often one or more of them will be killed in this way" [R. A. Fisher, 1925, using just 10 of 14 corps used by Bortkiewicz in his 1898 work, Das Gesetz der kleinen Zahlen.]

10 Prussian army-corps for 20 years, i.e., 200 observations, 1 per "corps-year" (CY).

Deaths from Horsekicks

"The chance of a man being killed by horsekick on any one day is exceedingly small, but if an army corps of men are exposed to this risk for a year, often one or more of them will be killed in this way" [R. A. Fisher, 1925, using just 10 of 14 corps used by Bortkiewicz in his 1898 work, Das Gesetz der kleinen Zahlen.]

10 Prussian army-corps for 20 years, i.e., 200 observations, 1 per "corps-year" (CY).

No. Deaths in corps-year	Frequencies) (No. "corps-years" with y deaths)		
y	Observed	Expected$	$y \times$ *Obs. Freq.*
0	109	108.7	0
1	65	66.3	65
2	22	20.2	44
3	3	4.1	9
4	1	0.6	4
5	-	0.1	-
Sum:	200	200	No. Deaths: 122

Deaths from Horsekicks

"The chance of a man being killed by horsekick on any one day is exceedingly small, but if an army corps of men are exposed to this risk for a year, often one or more of them will be killed in this way" [R. A. Fisher, 1925, using just 10 of 14 corps used by Bortkiewicz in his 1898 work, Das Gesetz der kleinen Zahlen.]

10 Prussian army-corps for 20 years, i.e., 200 observations, 1 per "corps-year" (CY).

No. Deaths in corps-year	Frequencies) (No. "corps-years" with y deaths)		
y	Observed	Expected[$]	$y \times$ *Obs. Freq.*
0	109	108.7	0
1	65	66.3	65
2	22	20.2	44
3	3	4.1	9
4	1	0.6	4
5	-	0.1	-
Sum:	200	200	No. Deaths: 122

$\hat{\lambda} = \bar{y} = \frac{122 \ deaths}{200CY} = 0.61/CY$;

Deaths from Horsekicks

"The chance of a man being killed by horsekick on any one day is exceedingly small, but if an army corps of men are exposed to this risk for a year, often one or more of them will be killed in this way" [R. A. Fisher, 1925, using just 10 of 14 corps used by Bortkiewicz in his 1898 work, Das Gesetz der kleinen Zahlen.]

10 Prussian army-corps for 20 years, i.e., 200 observations, 1 per "corps-year" (CY).

No. Deaths in corps-year	Frequencies) (No. "corps-years" with y deaths)		
y	Observed	`Expected`$^{\$}$	$y \times$ *Obs. Freq.*
0	109	`108.7`	0
1	65	`66.3`	65
2	22	`20.2`	44
3	3	`4.1`	9
4	1	`0.6`	4
5	-	`0.1`	-
Sum:	200	`200`	No. Deaths: 122

$\hat{\lambda} = \bar{y} = \frac{122\ deaths}{200CY} = 0.61/CY$; $^{\$}$`200 * dpois(0:6,lambda=0.61)`

Deaths from Horsekicks

"The chance of a man being killed by horsekick on any one day is exceedingly small, but if an army corps of men are exposed to this risk for a year, often one or more of them will be killed in this way" [R. A. Fisher, 1925, using just 10 of 14 corps used by Bortkiewicz in his 1898 work, Das Gesetz der kleinen Zahlen.]

10 Prussian army-corps for 20 years, i.e., 200 observations, 1 per "corps-year" (CY).

No. Deaths in corps-year	Frequencies) (No. "corps-years" with y deaths)		
y	Observed	`Expected`$^{\$}$	$y \times$ *Obs. Freq.*
0	109	108.7	0
1	65	66.3	65
2	22	20.2	44
3	3	4.1	9
4	1	0.6	4
5	-	0.1	-
Sum:	200	200	No. Deaths: 122

$\hat{\lambda} = \bar{y} = \frac{122\ deaths}{200CY} = 0.61/CY$; $^{\$}$`200 * dpois(0:6,lambda=0.61)`

- Poisson Distribution $\Rightarrow$ SD = $\sqrt{mean}$.

Deaths from Horsekicks

“The chance of a man being killed by horsekick on any one day is exceedingly small, but if an army corps of men are exposed to this risk for a year, often one or more of them will be killed in this way" [R. A. Fisher, 1925, using just 10 of 14 corps used by Bortkiewicz in his 1898 work, Das Gesetz der kleinen Zahlen.]

10 Prussian army-corps for 20 years, i.e., 200 observations, 1 per “corps-year” (CY).

No. Deaths in corps-year	Frequencies) (No. “corps-years” with y deaths)		
y	Observed	Expected$	$y \times$ *Obs. Freq.*
0	109	108.7	0
1	65	66.3	65
2	22	20.2	44
3	3	4.1	9
4	1	0.6	4
5	-	0.1	-
Sum:	200	200	No. Deaths: 122

$\hat{\lambda} = \bar{y} = \frac{122\ deaths}{200CY} = 0.61/CY$; $ `200 * dpois(0:6,lambda=0.61)`

- Poisson Distribution $\Rightarrow$ SD = $\sqrt{mean}$.
- In this series, the SD of the 200 y's is 0.78, which happens to equal $\sqrt{0.61}$.

Deaths from Horsekicks

"The chance of a man being killed by horsekick on any one day is exceedingly small, but if an army corps of men are exposed to this risk for a year, often one or more of them will be killed in this way" [R. A. Fisher, 1925, using just 10 of 14 corps used by Bortkiewicz in his 1898 work, Das Gesetz der kleinen Zahlen.]

10 Prussian army-corps for 20 years, i.e., 200 observations, 1 per "corps-year" (CY).

No. Deaths in corps-year	Frequencies) (No. "corps-years" with y deaths)		
y	Observed	Expected$	$y \times$ *Obs. Freq.*
0	109	108.7	0
1	65	66.3	65
2	22	20.2	44
3	3	4.1	9
4	1	0.6	4
5	-	0.1	-
Sum:	200	200	No. Deaths: 122

$\hat{\lambda} = \bar{y} = \frac{122\ deaths}{200CY} = 0.61/CY$; $ `200 * dpois(0:6,lambda=0.61)`

- Poisson Distribution $\Rightarrow$ SD = $\sqrt{mean}$.
- In this series, the SD of the 200 y's is 0.78, which happens to equal $\sqrt{0.61}$.
- " SD = $\sqrt{mean}$ " $\not\Rightarrow$ Poisson Distribution, but there is close agreement ('fit') between the observed and expected/theoretical distributions.

Deaths from Horsekicks

"The chance of a man being killed by horsekick on any one day is exceedingly small, but if an army corps of men are exposed to this risk for a year, often one or more of them will be killed in this way" [R. A. Fisher, 1925, using just 10 of 14 corps used by Bortkiewicz in his 1898 work, Das Gesetz der kleinen Zahlen.]

10 Prussian army-corps for 20 years, i.e., 200 observations, 1 per "corps-year" (CY).

No. Deaths in corps-year	Frequencies) (No. "corps-years" with y deaths)		
y	Observed	`Expected`$	$y \times$ *Obs. Freq.*
0	109	`108.7`	0
1	65	`66.3`	65
2	22	`20.2`	44
3	3	`4.1`	9
4	1	`0.6`	4
5	-	`0.1`	-
Sum:	200	`200`	No. Deaths: 122

$\hat{\lambda} = \bar{y} = \frac{122\ deaths}{200CY} = 0.61/CY$; $ `200 * dpois(0:6,lambda=0.61)`

- Poisson Distribution $\Rightarrow$ SD = $\sqrt{mean}$.
- In this series, the SD of the 200 y's is 0.78, which happens to equal $\sqrt{0.61}$.
- " SD = $\sqrt{mean}$ " $\not\Rightarrow$ Poisson Distribution, but there is close agreement ('fit') between the observed and expected/theoretical distributions.

http://www.epi.mcgill.ca//hanley/bios601/Intensity-Rate/horsekicks_anthrax_poisson.pdf
http://www.epi.mcgill.ca//hanley/bios601/Intensity-Rate/horsekicks.txt

What about decade-to-decade variations in U.S. Hurricane Strikes?

Number of hurricanes by Saffir-Simpson Category to strike the mainland U.S. each decade.

Decade	Saffir-Simpson Category[1]					All 1,2,3,4,5	Major 3,4,5
	1	2	3	4	5		
1851-1860	8	5	5	1	0	19	6
1861-1870	8	6	1	0	0	15	1
1871-1880	7	6	7	0	0	20	7
1881-1890	8	9	4	1	0	22	5
1891-1900	8	5	5	3	0	21	8
1901-1910	10	4	4	0	0	18	4
1911-1920	10	4	4	3	0	21	7
1921-1930	5	3	3	2	0	13	5
1931-1940	4	7	6	1	1	19	8
1941-1950	8	6	9	1	0	24	10
1951-1960	8	1	5	3	0	17	8
1961-1970	3	5	4	1	1	14	6
1971-1980	6	2	4	0	0	12	4
1981-1990	9	1	4	1	0	15	5
1991-2000	3	6	4	0	1	14	5
2001-2004	4	2	2	1	0	9	3
1851-2004	109	72	71	18	3	273	92
Average Per Decade	7.1	4.7	4.6	1.2	0.2	17.7	6.0

Number of hurricanes by Saffir-Simpson Category to strike the mainland U.S. each decade.

Decade	Saffir-Simpson Category[1]					All 1,2,3,4,5	Major 3,4,5
	1	2	3	4	5		
1851-1860	8	5	5	1	0	19	6
1861-1870	8	6	1	0	0	15	1
1871-1880	7	6	7	0	0	20	7
1881-1890	8	9	4	1	0	22	5
1891-1900	8	5	5	3	0	21	8
1901-1910	10	4	4	0	0	18	4
1911-1920	10	4	4	3	0	21	7
1921-1930	5	3	3	2	0	13	5
1931-1940	4	7	6	1	1	19	8
1941-1950	8	6	9	1	0	24	10
1951-1960	8	1	5	3	0	17	8
1961-1970	3	5	4	1	1	14	6
1971-1980	6	2	4	0	0	12	4
1981-1990	9	1	4	1	0	15	5
1991-2000	3	6	4	0	1	14	5
2001-2004	4	2	2	1	0	9	3
1851-2004	109	72	71	18	3	273	92
Average Per Decade	7.1	4.7	4.6	1.2	0.2	17.7	6.0

15 FULL DECADES

mean(sd)	7.0(2.3)	4.7(2.2)	4.6(1.8)	1.1(1.1)	0.2(0.4)	17.6(3.2)	5.9(2.3)
$\sqrt{mean}$:	2.6	2.2	2.1	1.1	0.4	4.2	2.4 !!

Number of hurricanes by Saffir-Simpson Category to strike the mainland U.S. each decade.

Decade	Saffir-Simpson Category[1]					All 1,2,3,4,5	Major 3,4,5
	1	2	3	4	5		
1851-1860	8	5	5	1	0	19	6
1861-1870	8	6	1	0	0	15	1
1871-1880	7	6	7	0	0	20	7
1881-1890	8	9	4	1	0	22	5
1891-1900	8	5	5	3	0	21	8
1901-1910	10	4	4	0	0	18	4
1911-1920	10	4	4	3	0	21	7
1921-1930	5	3	3	2	0	13	5
1931-1940	4	7	6	1	1	19	8
1941-1950	8	6	9	1	0	24	10
1951-1960	8	1	5	3	0	17	8
1961-1970	3	5	4	1	1	14	6
1971-1980	6	2	4	0	0	12	4
1981-1990	9	1	4	1	0	15	5
1991-2000	3	6	4	0	1	14	5
2001-2004	4	2	2	1	0	9	3
1851-2004	109	72	71	18	3	273	92
Average Per Decade	7.1	4.7	4.6	1.2	0.2	17.7	6.0

15 FULL DECADES

mean(sd)	7.0(2.3)	4.7(2.2)	4.6(1.8)	1.1(1.1)	0.2(0.4)	17.6(3.2)	5.9(2.3)
$\sqrt{mean}$:	2.6	2.2	2.1	1.1	0.4	4.2	2.4 !!

Can use regression models to fit temporal trends to these ('noisy') data.

Simulate 15 decades $\pm$ a time pattern: `summary(rpois(15, lambda= ..  ))`

Number of hurricanes by Saffir-Simpson Category to strike the mainland U.S. each decade.

Decade	Saffir-Simpson Category[1]					All 1,2,3,4,5	Major 3,4,5
	1	2	3	4	5		
1851-1860	8	5	5	1	0	19	6
1861-1870	8	6	1	0	0	15	1
1871-1880	7	6	7	0	0	20	7
1881-1890	8	9	4	1	0	22	5
1891-1900	8	5	5	3	0	21	8
1901-1910	10	4	4	0	0	18	4
1911-1920	10	4	4	3	0	21	7
1921-1930	5	3	3	2	0	13	5
1931-1940	4	7	6	1	1	19	8
1941-1950	8	6	9	1	0	24	10
1951-1960	8	1	5	3	0	17	8
1961-1970	3	5	4	1	1	14	6
1971-1980	6	2	4	0	0	12	4
1981-1990	9	1	4	1	0	15	5
1991-2000	3	6	4	0	1	14	5
2001-2004	4	2	2	1	0	9	3
1851-2004	109	72	71	18	3	273	92
Average Per Decade	7.1	4.7	4.6	1.2	0.2	17.7	6.0

15 FULL DECADES

mean(sd)	7.0(2.3)	4.7(2.2)	4.6(1.8)	1.1(1.1)	0.2(0.4)	17.6(3.2)	5.9(2.3)
$\sqrt{mean}$:	2.6	2.2	2.1	1.1	0.4	4.2	2.4 !!

Can use regression models to fit temporal trends to these ('noisy') data.
Simulate 15 decades $\pm$ a time pattern: `summary(rpois(15, lambda= .. ))`

Source: `https://www.nhc.noaa.gov/pastdec.shtml`

Rate of *de novo* mutations and the importance of father's age to disease risk

http://www.epi.mcgill.ca/hanley/bios601/FathersAgeMutations.pdf

Rate of *de novo* mutations and the importance of father's age to disease risk

http://www.epi.mcgill.ca/hanley/bios601/FathersAgeMutations.pdf

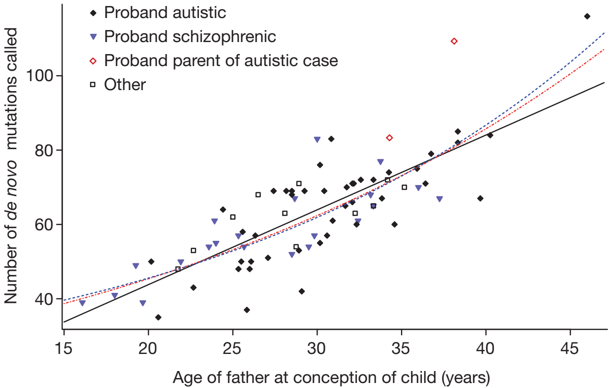

Figure 2 | Father's age and number of *de novo* mutations. The number of *de novo* mutations called is plotted against father's age at conception of child for the 78 trios. The solid black line denotes the linear fit. The dashed red curve is based on an exponential model fitted to the combined mutation counts. The dashed blue curve corresponds to a model in which maternal mutations are assumed to have a constant rate of 14.2 and paternal mutations are assumed to increase exponentially with father's age.

Rate of *de novo* mutations and the importance of father's age to disease risk

`http://www.epi.mcgill.ca/hanley/bios601/FathersAgeMutations.pdf`

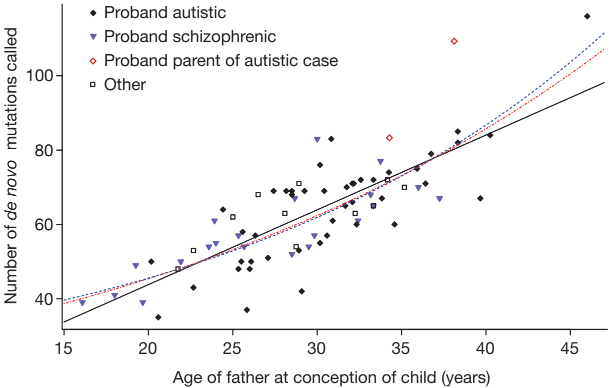

Figure 2 | Father's age and number of *de novo* mutations. The number of *de novo* mutations called is plotted against father's age at conception of child for the 78 trios. The solid black line denotes the linear fit. The dashed red curve is based on an exponential model fitted to the combined mutation counts. The dashed blue curve corresponds to a model in which maternal mutations are assumed to have a constant rate of 14.2 and paternal mutations are assumed to increase exponentially with father's age.

$\mu[y|age] \approx 2 \times age$; $\approx$ Poisson variation of y's around each $\mu[y|age]$; i.e. $\sigma[y|x] \approx \sqrt{\mu[y|age]}$

What events are these?

http://www.epi.mcgill.ca/hanley/mysteryData/

What about Daily No.s of BIRTHS 9 MONTHS AFTER BLACKOUT ?

What about Daily No.s of BIRTHS 9 MONTHS AFTER BLACKOUT ?

On November 9, 1965, the power went out in New York City, and it stayed out for a day – the **Great Blackout**. Nine months later, the newspapers suggested that New York was experiencing a baby boom. The columns below show the numbers of babies born every day during a 25-day period, Aug1-Aug25, centered nine months and ten days after the Great Blackout. These numbers average out to 436. This turns out not to be unusually high for New York.

What about Daily No.s of BIRTHS 9 MONTHS AFTER BLACKOUT ?

On November 9, 1965, the power went out in New York City, and it stayed out for a day – the **Great Blackout**. Nine months later, the newspapers suggested that New York was experiencing a baby boom. The columns below show the numbers of babies born every day during a 25-day period, Aug1-Aug25, centered nine months and ten days after the Great Blackout. These numbers average out to 436. This turns out not to be unusually high for New York.

Aug.	Births
1	451
2	468
3	429
4	448
5	466
6	377
7	344
8	448
9	438
10	455
11	468
12	462
13	405
14	377
15	451
16	497
17	458
18	429
19	434
20	410
21	351
22	467
23	508
24	432
25	426

What about Daily No.s of BIRTHS 9 MONTHS AFTER BLACKOUT ?

On November 9, 1965, the power went out in New York City, and it stayed out for a day – the **Great Blackout**. Nine months later, the newspapers suggested that New York was experiencing a baby boom. The columns below show the numbers of babies born every day during a 25-day period, Aug1-Aug25, centered nine months and ten days after the Great Blackout. These numbers average out to 436. This turns out not to be unusually high for New York.

Aug.	Births	Day
1	451	Mon
2	468	Tue
3	429	Wed
4	448	Thu
5	466	Fri
6	377	Sat
7	344	Sun
8	448	Mon
9	438	Tue
10	455	Wed
11	468	Thu
12	462	Fri
13	405	Sat
14	377	Sun
15	451	Mon
16	497	Tue
17	458	Wed
18	429	Thu
19	434	Fri
20	410	Sat
21	351	Sun
22	467	Mon
23	508	Tue
24	432	Wed
25	426	Thu

What about Daily No.s of BIRTHS 9 MONTHS AFTER BLACKOUT ?

On November 9, 1965, the power went out in New York City, and it stayed out for a day – the **Great Blackout**. Nine months later, the newspapers suggested that New York was experiencing a baby boom. The columns below show the numbers of babies born every day during a 25-day period, Aug1-Aug25, centered nine months and ten days after the Great Blackout. These numbers average out to 436. This turns out not to be unusually high for New York.

Statistics (2nd Ed) by D Freedman, R Pisani et al

Aug.	Births	**Day**
1	451	Mon
2	468	Tue
3	429	Wed
4	448	Thu
5	466	Fri
6	377	Sat
7	344	Sun
8	448	Mon
9	438	Tue
10	455	Wed
11	468	Thu
12	462	Fri
13	405	Sat
14	377	Sun
15	451	Mon
16	497	Tue
17	458	Wed
18	429	Thu
19	434	Fri
20	410	Sat
21	351	Sun
22	467	Mon
23	508	Tue
24	432	Wed
25	426	Thu

What about Daily No.s of BIRTHS 9 MONTHS AFTER BLACKOUT ?

On November 9, 1965, the power went out in New York City, and it stayed out for a day – the **Great Blackout**. Nine months later, the newspapers suggested that New York was experiencing a baby boom. The columns below show the numbers of babies born every day during a 25-day period, Aug1-Aug25, centered nine months and ten days after the Great Blackout. These numbers average out to 436. This turns out not to be unusually high for New York.

Statistics (2nd Ed) by D Freedman, R Pisani et al

- "Apparently, the *New York Times, NYT* sent a reporter around to a few hospitals on Monday, August 8, and Tuesday, August 9, nine months after the blackout.

Aug.	Births	**Day**
1	451	Mon
2	468	Tue
3	429	Wed
4	448	Thu
5	466	Fri
6	377	Sat
7	344	Sun
8	448	Mon
9	438	Tue
10	455	Wed
11	468	Thu
12	462	Fri
13	405	Sat
14	377	Sun
15	451	Mon
16	497	Tue
17	458	Wed
18	429	Thu
19	434	Fri
20	410	Sat
21	351	Sun
22	467	Mon
23	508	Tue
24	432	Wed
25	426	Thu

What about Daily No.s of BIRTHS 9 MONTHS AFTER BLACKOUT ?

On November 9, 1965, the power went out in New York City, and it stayed out for a day – the **Great Blackout**. Nine months later, the newspapers suggested that New York was experiencing a baby boom. The columns below show the numbers of babies born every day during a 25-day period, Aug1-Aug25, centered nine months and ten days after the Great Blackout. These numbers average out to 436. This turns out not to be unusually high for New York.

Aug.	Births	**Day**
1	451	Mon
2	468	Tue
3	429	Wed
4	448	Thu
5	466	Fri
6	377	Sat
7	344	Sun
8	448	Mon
9	438	Tue
10	455	Wed
11	468	Thu
12	462	Fri
13	405	Sat
14	377	Sun
15	451	Mon
16	497	Tue
17	458	Wed
18	429	Thu
19	434	Fri
20	410	Sat
21	351	Sun
22	467	Mon
23	508	Tue
24	432	Wed
25	426	Thu

Statistics (2nd Ed) by D Freedman, R Pisani et al

- "Apparently, the *New York Times, NYT* sent a reporter around to a few hospitals on Monday, August 8, and Tuesday, August 9, nine months after the blackout.
- The hospitals reported that their obstetrics wards were busier than usual probably because of the general pattern that weekends are slow, Mondays and Tuesdays are busy.

What about Daily No.s of BIRTHS 9 MONTHS AFTER BLACKOUT ?

On November 9, 1965, the power went out in New York City, and it stayed out for a day – the **Great Blackout**. Nine months later, the newspapers suggested that New York was experiencing a baby boom. The columns below show the numbers of babies born every day during a 25-day period, Aug1-Aug25, centered nine months and ten days after the Great Blackout. These numbers average out to 436. This turns out not to be unusually high for New York.

Aug.	Births	**Day**
1	451	Mon
2	468	Tue
3	429	Wed
4	448	Thu
5	466	Fri
6	377	Sat
7	344	Sun
8	448	Mon
9	438	Tue
10	455	Wed
11	468	Thu
12	462	Fri
13	405	Sat
14	377	Sun
15	451	Mon
16	497	Tue
17	458	Wed
18	429	Thu
19	434	Fri
20	410	Sat
21	351	Sun
22	467	Mon
23	508	Tue
24	432	Wed
25	426	Thu

Statistics (2nd Ed) by D Freedman, R Pisani et al

- “Apparently, the *New York Times, NYT* sent a reporter around to a few hospitals on Monday, August 8, and Tuesday, August 9, nine months after the blackout.
- The hospitals reported that their obstetrics wards were busier than usual probably because of the general pattern that weekends are slow, Mondays and Tuesdays are busy.
- Headline, p.1, Wed, Aug10, ’66: “**Births Up 9 Months After the Blackout**’ `http://www.epi.mcgill.ca//hanley/bios601/Intensity-Rate/BirthsUp9MonthsAfterBlackout.pdf`

What about Daily No.s of BIRTHS 9 MONTHS AFTER BLACKOUT ?

On November 9, 1965, the power went out in New York City, and it stayed out for a day – the **Great Blackout**. Nine months later, the newspapers suggested that New York was experiencing a baby boom. The columns below show the numbers of babies born every day during a 25-day period, Aug1-Aug25, centered nine months and ten days after the Great Blackout. These numbers average out to 436. This turns out not to be unusually high for New York.

Aug.	Births	**Day**
1	451	Mon
2	468	Tue
3	429	Wed
4	448	Thu
5	466	Fri
6	377	Sat
7	344	Sun
8	448	Mon
9	438	Tue
10	455	Wed
11	468	Thu
12	462	Fri
13	405	Sat
14	377	Sun
15	451	Mon
16	497	Tue
17	458	Wed
18	429	Thu
19	434	Fri
20	410	Sat
21	351	Sun
22	467	Mon
23	508	Tue
24	432	Wed
25	426	Thu

Statistics (2nd Ed) by D Freedman, R Pisani et al

- "Apparently, the *New York Times, NYT* sent a reporter around to a few hospitals on Monday, August 8, and Tuesday, August 9, nine months after the blackout.
- The hospitals reported that their obstetrics wards were busier than usual probably because of the general pattern that weekends are slow, Mondays and Tuesdays are busy.
- Headline, p.1, Wed, Aug10, '66: "**Births Up 9 Months After the Blackout**' `http://www.epi.mcgill.ca//hanley/bios601/Intensity-Rate/BirthsUp9MonthsAfterBlackout.pdf`
- This [overlooked source of variation] seems to be the origin of the baby-boom myth."

What about Daily No.s of BIRTHS 9 MONTHS AFTER BLACKOUT ?

On November 9, 1965, the power went out in New York City, and it stayed out for a day – the **Great Blackout**. Nine months later, the newspapers suggested that New York was experiencing a baby boom. The columns below show the numbers of babies born every day during a 25-day period, Aug1-Aug25, centered nine months and ten days after the Great Blackout. These numbers average out to 436. This turns out not to be unusually high for New York.

Aug.	Births	**Day**
1	451	Mon
2	468	Tue
3	429	Wed
4	448	Thu
5	466	Fri
6	377	Sat
7	344	Sun
8	448	Mon
9	438	Tue
10	455	Wed
11	468	Thu
12	462	Fri
13	405	Sat
14	377	Sun
15	451	Mon
16	497	Tue
17	458	Wed
18	429	Thu
19	434	Fri
20	410	Sat
21	351	Sun
22	467	Mon
23	508	Tue
24	432	Wed
25	426	Thu

Statistics (2nd Ed) by D Freedman, R Pisani et al

- "Apparently, the *New York Times, NYT* sent a reporter around to a few hospitals on Monday, August 8, and Tuesday, August 9, nine months after the blackout.
- The hospitals reported that their obstetrics wards were busier than usual probably because of the general pattern that weekends are slow, Mondays and Tuesdays are busy.
- Headline, p.1, Wed, Aug10, '66: "**Births Up 9 Months After the Blackout**' `http://www.epi.mcgill.ca//hanley/bios601/Intensity-Rate/BirthsUp9MonthsAfterBlackout.pdf`
- This [overlooked source of variation] seems to be the origin of the baby-boom myth."

Closer to Poisson variation if use week as unit, or if one 'conditions on' day-of-week; Over the year, some variation in weekly numbers.

Freedman cites Izenman & Zabell, "Babies and the blackout: the genesis of a misconception," Soc. Sci. Res.,1981,282- 99. available here `http://www.epi.mcgill.ca/hanley/bios601/Intensity-Rate/GenesisOfAMisconception.pdf`

Daily (and hourly!) variations in numbers of births `https://www.significancemagazine.com/585` and `https://rss-onlinelibrary-wiley-com.proxy3.library.mcgill.ca/doi/full/10.1111/j.1740-9713.2017.01026.x` or `http://www.epi.mcgill.ca/hanley/mysteryData/`

Are “accidents” distributed “randomly” over bus drivers?

Are "accidents" distributed "randomly" over bus drivers?

Observed & "expected" numbers of accidents during a 3-year period among 708 Northern Ireland Transport Authority bus drivers. [Table 2.5]

Are "accidents" distributed "randomly" over bus drivers?

Observed & "expected" numbers of accidents during a 3-year period among 708 Northern Ireland Transport Authority bus drivers. [Table 2.5]

No. accidents in 3-year period	Number of drivers with y accidents	
(y)	Observed	Expected[$]
0	117	71.5
1	157	164.0
2	158	187.9
3	115	143.6
4	78	82.3
5	44	37.7
6	21	14.4
7	7	4.7
8	6	1.4
9	1	0.3
10	3	0.1
11	1	0.0
	708	708

Are “accidents” distributed “randomly” over bus drivers?

Observed & “expected” numbers of accidents during a 3-year period among 708 Northern Ireland Transport Authority bus drivers. [Table 2.5]

No. accidents in 3-year period	Number of drivers with y accidents	
(y)	Observed	Expected$
0	117	71.5
1	157	164.0
2	158	187.9
3	115	143.6
4	78	82.3
5	44	37.7
6	21	14.4
7	7	4.7
8	6	1.4
9	1	0.3
10	3	0.1
11	1	0.0
	708	708

$\hat{\lambda} = \bar{y} = \frac{1623 \ \textit{accidents}}{708 \ \textit{drivers}} = \frac{2.29}{\textit{driver}}.$

Are “accidents” distributed “randomly” over bus drivers?

Observed & “expected” numbers of accidents during a 3-year period among 708 Northern Ireland Transport Authority bus drivers. [Table 2.5]

No. accidents in 3-year period	Number of drivers with y accidents	
(y)	Observed	Expected[$]
0	117	71.5
1	157	164.0
2	158	187.9
3	115	143.6
4	78	82.3
5	44	37.7
6	21	14.4
7	7	4.7
8	6	1.4
9	1	0.3
10	3	0.1
11	1	0.0
	708	708

$\hat{\lambda} = \bar{y} = \frac{1623 \ \textit{accidents}}{708 \ \textit{drivers}} = \frac{2.29}{\textit{driver}}$.

SD of the 708 y's :1.86; $\sqrt{2.29} = 1.51$.

[$] `708 * dpois(0:11,lambda=2.29)`

Are “accidents” distributed “randomly” over bus drivers?

Observed & “expected” numbers of accidents during a 3-year period among 708 Northern Ireland Transport Authority bus drivers. [Table 2.5]

No. accidents in 3-year period (y)	Number of drivers with y accidents: Observed	Expected[$]
0	117	71.5
1	157	164.0
2	158	187.9
3	115	143.6
4	78	82.3
5	44	37.7
6	21	14.4
7	7	4.7
8	6	1.4
9	1	0.3
10	3	0.1
11	1	0.0
	708	708

$\hat{\lambda} = \bar{y} = \frac{1623 \ accidents}{708 \ drivers} = \frac{2.29}{driver}$.

SD of the 708 y's :1.86; $\sqrt{2.29} = 1.51$.

[$] `708 * dpois(0:11,lambda=2.29)`

Colton:

"Comparison of observed and expected frequencies:

Are “accidents” distributed “randomly” over bus drivers?

Observed & “expected” numbers of accidents during a 3-year period among 708 Northern Ireland Transport Authority bus drivers. [Table 2.5]

No. accidents in 3-year period (y)	Number of drivers with y accidents: Observed	Expected[$]
0	117	71.5
1	157	164.0
2	158	187.9
3	115	143.6
4	78	82.3
5	44	37.7
6	21	14.4
7	7	4.7
8	6	1.4
9	1	0.3
10	3	0.1
11	1	0.0
	708	708

$\hat{\lambda} = \bar{y} = \frac{1623 \ accidents}{708 \ drivers} = \frac{2.29}{driver}$.

SD of the 708 y's :1.86; $\sqrt{2.29} = 1.51$.

[$] `708 * dpois(0:11,lambda=2.29)`

Colton:

"Comparison of observed and expected frequencies:

- More than the expected number of drivers with no accidents

Are “accidents” distributed “randomly” over bus drivers?

Observed & “expected” numbers of accidents during a 3-year period among 708 Northern Ireland Transport Authority bus drivers. [Table 2.5]

No. accidents in 3-year period (y)	Number of drivers with y accidents: Observed	Expected[$]
0	117	71.5
1	157	164.0
2	158	187.9
3	115	143.6
4	78	82.3
5	44	37.7
6	21	14.4
7	7	4.7
8	6	1.4
9	1	0.3
10	3	0.1
11	1	0.0
	708	708

$\hat{\lambda} = \bar{y} = \frac{1623 \ \textit{accidents}}{708 \ \textit{drivers}} = \frac{2.29}{\textit{driver}}.$

SD of the 708 y's :1.86; $\sqrt{2.29} = 1.51$.

[$] `708 * dpois(0:11,lambda=2.29)`

Colton:

"Comparison of observed and expected frequencies:

- More than the expected number of drivers with no accidents
- More than the expected number of drivers with five or more accidents

Are "accidents" distributed "randomly" over bus drivers?

Observed & "expected" numbers of accidents during a 3-year period among 708 Northern Ireland Transport Authority bus drivers. [Table 2.5]

No. accidents in 3-year period (y)	Number of drivers with y accidents: Observed	Expected[$]
0	117	71.5
1	157	164.0
2	158	187.9
3	115	143.6
4	78	82.3
5	44	37.7
6	21	14.4
7	7	4.7
8	6	1.4
9	1	0.3
10	3	0.1
11	1	0.0
	708	708

$\hat{\lambda} = \bar{y} = \frac{1623\ accidents}{708\ drivers} = \frac{2.29}{driver}$.

SD of the 708 y's :1.86; $\sqrt{2.29} = 1.51$.

[$] `708 * dpois(0:11,lambda=2.29)`

Colton:

"Comparison of observed and expected frequencies:

- More than the expected number of drivers with no accidents
- More than the expected number of drivers with five or more accidents
- These data suggest that the accidents did not occur completely at random;

Are “accidents” distributed “randomly” over bus drivers?

Observed & “expected” numbers of accidents during a 3-year period among 708 Northern Ireland Transport Authority bus drivers. [Table 2.5]

No. accidents in 3-year period (y)	Number of drivers with y accidents: Observed	Expected[$]
0	117	71.5
1	157	164.0
2	158	187.9
3	115	143.6
4	78	82.3
5	44	37.7
6	21	14.4
7	7	4.7
8	6	1.4
9	1	0.3
10	3	0.1
11	1	0.0
	708	708

$\hat{\lambda} = \bar{y} = \frac{1623 \textit{ accidents}}{708 \textit{ drivers}} = \frac{2.29}{\textit{driver}}.$

SD of the 708 y's :1.86; $\sqrt{2.29} = 1.51.$

[$] `708 * dpois(0:11,lambda=2.29)`

Colton:

"Comparison of observed and expected frequencies:

- More than the expected number of drivers with no accidents
- More than the expected number of drivers with five or more accidents
- These data suggest that the accidents did not occur completely at random;

  in fact it appears that there is some indication of accident proneness.

Are "accidents" distributed "randomly" over bus drivers?

Observed & "expected" numbers of accidents during a 3-year period among 708 Northern Ireland Transport Authority bus drivers. [Table 2.5]

No. accidents in 3-year period	Number of drivers with y accidents	
(y)	Observed	Expected$
0	117	71.5
1	157	164.0
2	158	187.9
3	115	143.6
4	78	82.3
5	44	37.7
6	21	14.4
7	7	4.7
8	6	1.4
9	1	0.3
10	3	0.1
11	1	0.0
	708	708

$\hat{\lambda} = \bar{y} = \frac{1623 \ accidents}{708 \ drivers} = \frac{2.29}{driver}.$

SD of the 708 y's :1.86; $\sqrt{2.29} = 1.51.$

$ `708 * dpois(0:11,lambda=2.29)`

Colton:

"Comparison of observed and expected frequencies:

- More than the expected number of drivers with no accidents
- More than the expected number of drivers with five or more accidents
- These data suggest that the accidents did not occur completely at random;

  in fact it appears that there is some indication of accident proneness.
- Ignoring this variation makes for (model-based) Standard Errors (SE's) and CI's that are too narrow, and that can lead to 'false positive' findings

Are "accidents" distributed "randomly" over bus drivers?

Observed & "expected" numbers of accidents during a 3-year period among 708 Northern Ireland Transport Authority bus drivers. [Table 2.5]

No. accidents in 3-year period	Number of drivers with y accidents	
(y)	Observed	Expected$
0	117	71.5
1	157	164.0
2	158	187.9
3	115	143.6
4	78	82.3
5	44	37.7
6	21	14.4
7	7	4.7
8	6	1.4
9	1	0.3
10	3	0.1
11	1	0.0
	708	708

$\hat{\lambda} = \bar{y} = \frac{1623 \ accidents}{708 \ drivers} = \frac{2.29}{driver}.$

SD of the 708 y's :1.86; $\sqrt{2.29} = 1.51$.

$ `708 * dpois(0:11,lambda=2.29)`

Colton:

"Comparison of observed and expected frequencies:

- More than the expected number of drivers with no accidents
- More than the expected number of drivers with five or more accidents
- These data suggest that the accidents did not occur completely at random;

  in fact it appears that there is some indication of accident proneness.
- Ignoring this variation makes for (model-based) Standard Errors (SE's) and CI's that are too narrow, and that can lead to 'false positive' findings
- Encoding this (not identifiable) variation in 'random effects' or 'random-intercept' regression models makes for more realistic SE's and CI's.

http://www.epi.mcgill.ca/hanley/bios601/Intensity-Rate/AccidentsBusDrivers.pdf

DAYLIGHT SAVINGS TIME AND TRAFFIC ACCIDENTS, NEJM 1996.04,04

DAYLIGHT SAVINGS TIME AND TRAFFIC ACCIDENTS

To the Editor: It has become increasingly clear that insufficient sleep and disrupted circadian rhythms are a major public health problem. For instance, in 1988 the cost of sleep-related accidents exceeded $56 billion and included 24,318 deaths and 2,474,430 disabling injuries.[1] Major disasters, including the nuclear accident at Chernobyl, the *Exxon Valdez* oil spill, and the destruction of the space shuttle *Challenger,* have been linked to insufficient sleep, disrupted circadian rhythms, or both on the part of involved supervisors and staff.[2,3] It has been suggested that as a society we are chronically sleep-deprived[4] and that small additional losses of sleep may have consequences for public and individual safety.[2]

https://www.nejm.org/doi/full/10.1056/NEJM199604043341416

We can use noninvasive techniques to examine the effects of minor disruptions of circadian rhythms on normal activities if we take advantage of annual shifts in time keeping. More than 25 countries shift to daylight savings time each spring and return to standard time in the fall. The spring shift results in the loss of one hour of sleep time (the equivalent in terms of jet lag of traveling one time zone to the east), whereas the fall shift permits an additional hour of sleep (the equivalent of traveling one time zone to the west). Although one hour's change may seem like a minor disruption in the cycle of sleep and wakefulness, measurable changes in sleep pattern persist for up to five days after each time shift.[5] This leads to the prediction that the spring shift, involving a loss of an hour's sleep, might lead to an increased number of "microsleeps," or lapses of attention, during daily activities and thus might cause an increase in the probability of accidents, especially in traffic. The additional hour of sleep gained in the fall might then lead conversely to a reduction in accident rates.

We used data from a tabulation of all traffic accidents in Canada as they were reported to the Canadian Ministry of Transport for the years 1991 and 1992 by all 10 provinces. A total of 1,398,784 accidents were coded according to the date of occurrence. Data for analysis were restricted to the Monday preceding the week of the change due to daylight savings time, the Monday immediately after, and the Monday one week after the change, for both spring and fall time shifts. Data from the province of Saskatchewan were excluded because it does not observe daylight savings time. The analysis of the spring shift included 9593 accidents and that of the fall shift 12,010. The resulting data are shown in Figure 1.

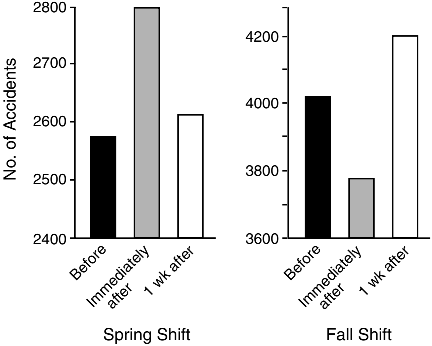

Figure 1. Numbers of Traffic Accidents on the Mondays before and after the Shifts to and from Daylight Savings Time for the Years 1991 and 1992.

There is an increase in accidents after the spring shift (when an hour of sleep is lost) and a decrease in the fall (when an hour of sleep is gained).

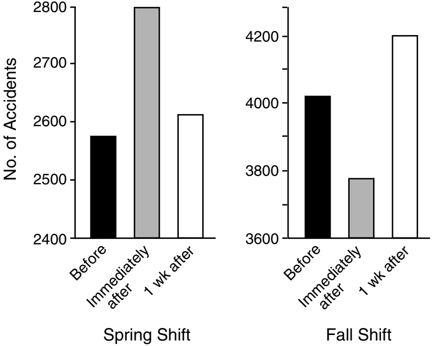

Figure 1. Numbers of Traffic Accidents on the Mondays before and after the Shifts to and from Daylight Savings Time for the Years 1991 and 1992.

There is an increase in accidents after the spring shift (when an hour of sleep is lost) and a decrease in the fall (when an hour of sleep is gained).

to daylight savings time increased the risk of accidents. The Monday immediately after the shift showed a relative risk of 1.086 (95 percent confidence interval, 1.029 to 1.145; $\chi^2 = 9.01$, 1 df; $P<0.01$). As compared with the accident rate a week later, the relative risk for the Monday immediately after the shift was 1.070 (95 percent confidence interval, 1.015 to 1.129; $\chi^2 = 6.19$, 1 df; $P<0.05$). Conversely, there was a reduction in the risk of traffic accidents after the fall shift from daylight savings time when an hour of sleep was gained. In the fall, the relative risk on the Monday of the change was 0.937 (95 percent confidence interval, 0.897 to 0.980; $\chi^2 = 8.07$, 1 df; $P<0.01$) when compared with the preceding Monday and 0.896 (95 percent confidence interval, 0.858 to 0.937; $\chi^2 = 23.69$; $P<0.001$) when compared with the Monday one week later. Thus, the spring shift to daylight savings time, and the concomitant loss of one hour of sleep, resulted in an average increase in traffic accidents of approximately 8 percent, whereas the fall shift resulted in a decrease in accidents of approximately the same magnitude immediately after the time shift.

These data show that small changes in the amount of sleep that people get can have major consequences in everyday activities. The loss of merely one hour of sleep can increase the risk of traffic accidents. It is likely that the effects are due to sleep loss rather than a nonspecific disruption in circadian rhythm, since gaining an additional hour of sleep at the fall time shift seems to decrease the risk of accidents.

Vancouver, BC V6T 1Z4, Canada

STANLEY COREN, PH.D.
University of British Columbia

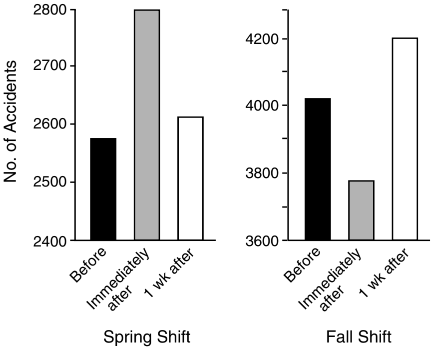

Figure 1. Numbers of Traffic Accidents on the Mondays before and after the Shifts to and from Daylight Savings Time for the Years 1991 and 1992.

There is an increase in accidents after the spring shift (when an hour of sleep is lost) and a decrease in the fall (when an hour of sleep is gained).

to daylight savings time increased the risk of accidents. The Monday immediately after the shift showed a relative risk of 1.086 (95 percent confidence interval, 1.029 to 1.145; $\chi^2 = 9.01$, 1 df; $P<0.01$). As compared with the accident rate a week later, the relative risk for the Monday immediately after the shift was 1.070 (95 percent confidence interval, 1.015 to 1.129; $\chi^2 = 6.19$, 1 df; $P<0.05$). Conversely, there was a reduction in the risk of traffic accidents after the fall shift from daylight savings time when an hour of sleep was gained. In the fall, the relative risk on the Monday of the change was 0.937 (95 percent confidence interval, 0.897 to 0.980; $\chi^2 = 8.07$, 1 df; $P<0.01$) when compared with the preceding Monday and 0.896 (95 percent confidence interval, 0.858 to 0.937; $\chi^2 = 23.69$; $P<0.001$) when compared with the Monday one week later. Thus, the spring shift to daylight savings time, and the concomitant loss of one hour of sleep, resulted in an average increase in traffic accidents of approximately 8 percent, whereas the fall shift resulted in a decrease in accidents of approximately the same magnitude immediately after the time shift.

These data show that small changes in the amount of sleep that people get can have major consequences in everyday activities. The loss of merely one hour of sleep can increase the risk of traffic accidents. It is likely that the effects are due to sleep loss rather than a nonspecific disruption in circadian rhythm, since gaining an additional hour of sleep at the fall time shift seems to decrease the risk of accidents.

Vancouver, BC V6T 1Z4, Canada

STANLEY COREN, PH.D.
University of British Columbia

Vincent: http://www.epi.mcgill.ca/hanley/bios601/Intensity-Rate/Sleep-Vincent.pdf The results of both a graphical analysis and the variability estimation of 10 years of data failed, as had an earlier study, to support Coren's hypothesis. [JH: Coren's χ^2 statistic assumes that the 2 'before' counts are like 2 Geiger counts generated by a uniform source, and ignores all outside influences.]

Other examples of "lumpy" counts [“extra-Poisson” variation]

Other examples of "lumpy" counts [“extra-Poisson” variation]

Examples that don’t fit Poisson distribution (or, not without further aggre-/segre-gation)

Other examples of "lumpy" counts [“extra-Poisson” variation]

Examples that don’t fit Poisson distribution (or, not without further aggre-/segre-gation)

- Yearly variations in numbers of persons killed in plane crashes
  [Yearly variations in numbers of plane crashes may be closer to Poisson –apart from some extra variation over time due to improvements in safety, fluctuations in numbers of flights etc.]

Other examples of "lumpy" counts [“extra-Poisson” variation]

Examples that don’t fit Poisson distribution (or, not without further aggre-/segre-gation)

- Yearly variations in numbers of persons killed in plane crashes
  [Yearly variations in numbers of plane crashes may be closer to Poisson –apart from some extra variation over time due to improvements in safety, fluctuations in numbers of flights etc.]
- Daily/Yearly variations in numbers of deaths [variation over seasons, ’flu’ years]

Other examples of "lumpy" counts [“extra-Poisson” variation]

Examples that don’t fit Poisson distribution (or, not without further aggre-/segre-gation)

- Yearly variations in numbers of persons killed in plane crashes
  [Yearly variations in numbers of plane crashes may be closer to Poisson –apart from some extra variation over time due to improvements in safety, fluctuations in numbers of flights etc.]
- Daily/Yearly variations in numbers of deaths [variation over seasons, ’flu’ years]
- Daily numbers of Sudden Infant Deaths
  `https://www.ncbi.nlm.nih.gov/pubmed/21059188`

Other examples of "lumpy" counts [“extra-Poisson” variation]

Examples that don’t fit Poisson distribution (or, not without further aggre-/segre-gation)

- Yearly variations in numbers of persons killed in plane crashes
  [Yearly variations in numbers of plane crashes may be closer to Poisson –apart from some extra variation over time due to improvements in safety, fluctuations in numbers of flights etc.]
- Daily/Yearly variations in numbers of deaths [variation over seasons, ’flu’ years]
- Daily numbers of Sudden Infant Deaths
  `https://www.ncbi.nlm.nih.gov/pubmed/21059188`
- Yearly numbers/incidence of hospitalized injuries in a region
  `http://www.epi.mcgill.ca/hanley/c609/Material/LidkopingALL.pdf`

Other examples of "lumpy" counts [“extra-Poisson” variation]

Examples that don’t fit Poisson distribution (or, not without further aggre-/segre-gation)

- Yearly variations in numbers of persons killed in plane crashes
  [Yearly variations in numbers of plane crashes may be closer to Poisson –apart from some extra variation over time due to improvements in safety, fluctuations in numbers of flights etc.]
- Daily/Yearly variations in numbers of deaths [variation over seasons, ’flu’ years]
- Daily numbers of Sudden Infant Deaths
  `https://www.ncbi.nlm.nih.gov/pubmed/21059188`
- Yearly numbers/incidence of hospitalized injuries in a region
  `http://www.epi.mcgill.ca/hanley/c609/Material/LidkopingALL.pdf`
- Lethality of Civilian Active Shooter Incidents With and Without Semiautomatic Rifles in the United States [The authors used a negative binomial model that allows extra-Poisson variation.] A bootstrap CI would also be appropriate.
  `https://jamanetwork.com/journals/jama/fullarticle/2702134`

Is Poisson Model appropriate for?

Is Poisson Model appropriate for?

1. Yearly Numbers of Dengue Fever Cases
   https://www.nature.com/articles/d41586-018-05914-3 and here
   https://gatesopenresearch.org/articles/2-36/v1
2. Daily numbers of Sudden Infant Deaths?
   https://www.ncbi.nlm.nih.gov/pubmed/21059188
3. Yearly numbers/incidence of hospitalized injuries in a region?
   http://www.epi.mcgill.ca/hanley/c609/Material/LidkopingALL.pdf
4. Yearly Accidents, Fatalities, and Rates, 1982 - 2000, U.S. Air Carriers
   http://www.epi.mcgill.ca/hanley/c626/airline-data-sas.txt
5. Quarterly & Monthly (prevalence) rates of Spina Bifida and Anencephaly Among Births (in relation to fortification of Foods with Folic Acid)
   http://www.epi.mcgill.ca//hanley/c626/folic_acid.pdf. See more data on webpage
   http://www.epi.mcgill.ca//hanley/c626/.
6. (Yearly) fatal and nonfatal crash rates on a toll highway (following a 5-15 mph (8-24 kph) decrease in speed limits)
   https://www.ncbi.nlm.nih.gov/pubmed/1251837
7. Daily numbers of in-hospital deaths and Daily Maximal Temperatures
   http://www.epi.mcgill.ca/hanley/c626/Heatwave_death_lyon.pdf
8. The (daily) incidence of crimes reported to 3 police stations in different towns (one rural, one urban, one industrial) vis-a-vis the day of the lunar cycle
   http://www.epi.mcgill.ca/hanley/c626/fullmoon.pdf
9. Daily no.s (Postponement of Death Until Symbolically Meaningful Occasions)
   http://www.epi.mcgill.ca/hanley/c626/holidays.pdf
10. Rates of audience fidget. (F Galton)
    http://www.epi.mcgill.ca/hanley/c626/measure_of_fidget_galton.pdf

References [as of 2007]

– Walker A Observation and Inference, p13,107,154
– Armitage P Berry G & Matthews JNS [4th edition, 2002] Statistical Methods in Medical Research sections 3.7 , 5.3, 6.3
– Colton T Statistics in Medicine, pp 16-17 and 77-78
– Kahn HA, Sempos CT Statistical Methods in Epidemiology pp 218-219
– Selvin S Statistical Analysis of Epidemiologic Data Ch 5 (clustering) and Appendix B (Binomial/Poisson)
– Miettinen O. Theoretical Epidemiology p 295
– Breslow N, Day N Statistical Methods in Cancer ResearchVol II: Analysis of Cohort Studies pp68-70 (SMR) pp131-135; sect. 7.2 (power/sample size)
– Statistical Methods in Cancer ResearchVol I: Analysis of Case-Control Studies p134 (test-based CI's)
– Rothman K, Greenland S [1998] Modern Epidemiology pp 234- pp404-5 (overdispersion) 570-571 (Poisson regression)
– Rothman K, Boice J Epidemiologic Analysis on a Programmable Calculator
– Rothman K [1986] Modern Epidemiology
– Rothman K [2002] Introduction to Epidemiology pp 133-4 & 137-139