# Inference about a Population Rate $(\lambda)$ IH notes on rates

#### Sahir Bhatnagar and James Hanley

EPIB 607 Department of Epidemiology, Biostatistics, and Occupational Health McGill University

sahir.bhatnagar@mcgill.ca
https://sahirbhatnagar.com/EPIB607/

November 1, 2018



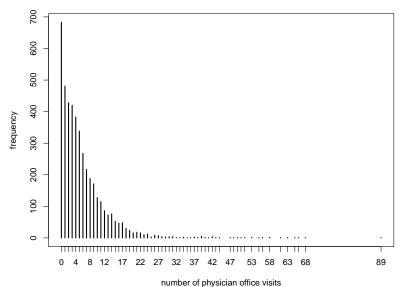
# Poisson Model for Sampling Variability of a Count in a Given Amount of "Experience"

- 2

#### Motivating example: Demand for medical care

- Data from the US National Medical Expenditure Survey (NMES) for 1987/88
- 4406 individuals, aged 66 and over, who are covered by Medicare, a public insurance program
- The objective of the study was to model the demand for medical care - as captured by the number of physician/non-physician office and hospital outpatient visits - by the covariates available for the patients.

# Motivating example: Demand for medical care



number of physician office visits

■ Discrete outcome  $\rightarrow$  1, 2, 3, ... visits

- Discrete outcome  $\rightarrow$  1, 2, 3, ... visits
- There are rare events, e.g. 1 individual with 89 visits

- Discrete outcome  $\rightarrow$  1, 2, 3, ... visits
- There are rare events, e.g. 1 individual with 89 visits
- The data are far from normally distributed

- Discrete outcome  $\rightarrow$  1, 2, 3, ... visits
- There are rare events, e.g. 1 individual with 89 visits
- The data are far from normally distributed
- Can theoretically go on forever

#### The Poisson Distribution

The binomial distribution was derived by starting with an experiment consisting of trials or draws and applying the laws of probability to various outcomes of the experiment.

#### The Poisson Distribution

- The binomial distribution was derived by starting with an experiment consisting of trials or draws and applying the laws of probability to various outcomes of the experiment.
- There is no simple experiment on which the Poisson distribution is based, although we will shortly describe how it can be obtained by certain limiting operations.

■ The (infinite number of) probabilities  $P_0, P_1, ..., P_y, ...$ , of observing Y = 0, 1, 2, ..., y, ... events in a given amount of "experience."

-

- The (infinite number of) probabilities  $P_0, P_1, ..., P_y, ...$ , of observing Y = 0, 1, 2, ..., y, ... events in a given amount of "experience."
- These probabilities,  $P(Y = k) \rightarrow \texttt{dpois()}$ , are governed by a single parameter, the mean  $E[Y] = \mu$  which represents the expected **number** of events in the amount of experience actually studied.

- The (infinite number of) probabilities  $P_0, P_1, ..., P_y, ...$ , of observing Y = 0, 1, 2, ..., y, ... events in a given amount of "experience."
- These probabilities,  $P(Y = k) \rightarrow \texttt{dpois()}$ , are governed by a single parameter, the mean  $E[Y] = \mu$  which represents the expected **number** of events in the amount of experience actually studied.
- We say that a random variable  $Y \sim Poisson(\mu)$  distribution if

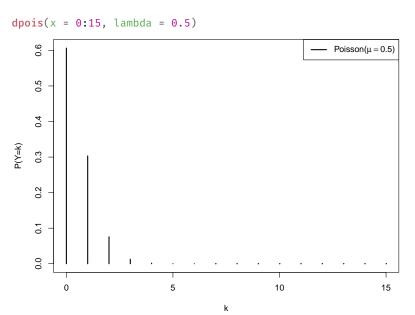
$$P(Y = k) = \frac{\mu^k}{k!} e^{-\mu}, \quad k = 0, 1, 2, \dots$$

- The (infinite number of) probabilities  $P_0, P_1, ..., P_y, ...$ , of observing Y = 0, 1, 2, ..., y, ... events in a given amount of "experience."
- These probabilities,  $P(Y = k) \rightarrow \texttt{dpois()}$ , are governed by a single parameter, the mean  $E[Y] = \mu$  which represents the expected **number** of events in the amount of experience actually studied.
- We say that a random variable  $Y \sim Poisson(\mu)$  distribution if

$$P(Y = k) = \frac{\mu^k}{k!} e^{-\mu}, \quad k = 0, 1, 2, \dots$$

- Note: in **dpois()**  $\mu$  is referred to as **lambda**
- Note the distinction between  $\mu$  and  $\lambda$ 
  - $\blacktriangleright \mu$ : expected **number** of events
  - $\triangleright$   $\lambda$ : **rate** parameter

# The probability mass function for $\mu=0.5$

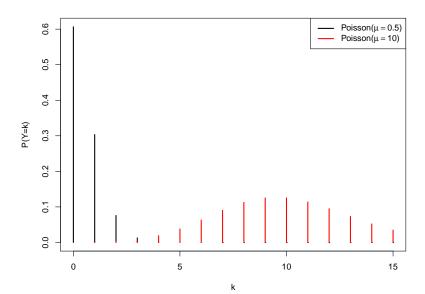


# The probability mass function for $\mu = 10$

dpois(x = 0:15, lambda = 10)Poisson( $\mu = 10$ ) 0.12 0.10 0.08 90.0 0.04 0.02 0.00 0 5 10 15

k

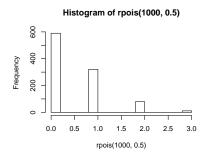
#### The probability mass function

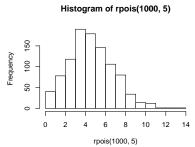


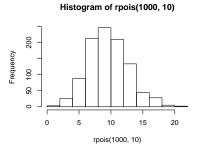
■ Approximated by  $\mathcal{N}(\mu, \sqrt{\mu})$  when  $\mu >> 10$ 

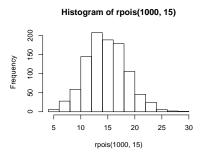
- Approximated by  $\mathcal{N}(\mu, \sqrt{\mu})$  when  $\mu >> 10$
- Open-ended (unlike Binomial), but in practice, has finite range.
- Poisson data sometimes called "numerator only": (unlike Binomial) may not "see" or count "non-events"

# Normal approximation to Poisson is the CLT in action









■ Count of events or items that <u>occur randomly</u>, with <u>low homogeneous intensity</u>, in time, space, or 'item'-time (e.g. person-time).

- Count of events or items that <u>occur randomly</u>, with low homogeneous intensity, in time, space, or 'item'-time (e.g. person-time).
- Binomial $(n, \pi)$  when  $n \to \infty$  and  $\pi \to 0$ , but  $n \times \pi = \mu$  is finite.

- Count of events or items that <u>occur randomly</u>, with low homogeneous intensity, in time, space, or 'item'-time (e.g. person-time).
- Binomial $(n, \pi)$  when  $n \to \infty$  and  $\pi \to 0$ , but  $n \times \pi = \mu$  is finite.
- Y ~ Poisson( $\mu_Y$ ) if time (T) between events follows an  $T \sim \text{Exponential}(\mu_T = 1/\mu_Y)$ . http://www.epi.mcgill.ca/hanley/bios601/Intensity-Rate/Randomness\_poisson.pdf

- Count of events or items that <u>occur randomly</u>, with low homogeneous intensity, in time, space, or 'item'-time (e.g. person-time).
- Binomial $(n, \pi)$  when  $n \to \infty$  and  $\pi \to 0$ , but  $n \times \pi = \mu$  is finite.
- Y ~ Poisson( $\mu_Y$ ) if time (T) between events follows an  $T \sim \text{Exponential}(\mu_T = 1/\mu_Y)$ . http://www.epi.mcgill.ca/hanley/bios601/Intensity-Rate/Randomness\_poisson.pdf
- As sum of  $\geq 2$  independent Poisson random variables, with same **or different**  $\mu$ 's:  $Y_1 \sim \operatorname{Poisson}(\mu_1) \ Y_2 \sim \operatorname{Poisson}(\mu_2) \Rightarrow Y = Y_1 + Y_2 \sim \operatorname{Poisson}(\mu_1 + \mu_2).$

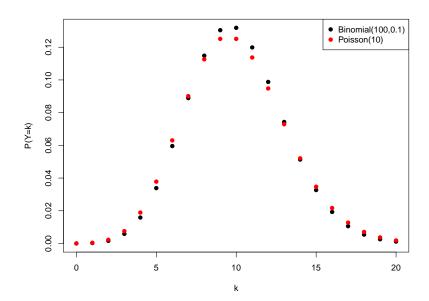
#### Poisson distribution as a limit

The rationale for using the Poisson distribution in many situations is provided by the following proposition.

#### Proposition 1 (Limit of a binomial is Poisson)

Suppose that  $Y \sim Binomial(n,\pi)$ . If we let  $\pi = \mu/n$ , then as  $n \to \infty$ ,  $Binomial(n,\pi) \to Poisson(\mu)$ . Another way of saying this: for large n and small  $\pi$ , we can approximate the  $Binomial(n,\pi)$  probability by the  $Poisson(\mu = n\pi)$ .

# Poisson approximation to the Binomial



#### Examples

- numbers of asbestos fibres
- deaths from horse kicks\*
- needle-stick or other percutaneous injuries
- bus-driver accidents\*
- twin-pairs\*
- radioactive disintegrations\*
- flying-bomb hits\*
- white blood cells
- typographical errors
- cell occupants in a given volume, area, line-length, population-time, time, etc. <sup>1</sup>

<sup>1\*</sup> included in http://www.epi.mcgill.ca/hanley/bios601/Intensity-Rate/

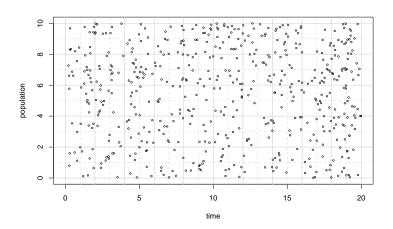


Fig.: Events in Population-Time randomly generated from intensities that are constant within (2 squares high by 2 squares wide) 'panels', but vary between such panels. In Epidemiology, each square might represent a number of units of population-time, and each dot an event.

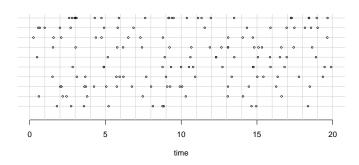


Fig.: Events in Time: 10 examples, randomly generated from constant over time intensities. Simulated with 1000 Bernoulli( ${
m small}\ \pi$ )'s per time unit.

#### Does the Poisson Distribution apply to..?

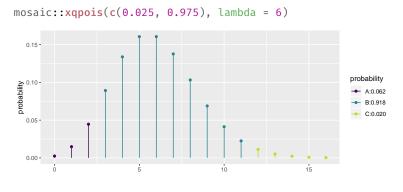
- 1. Yearly variations in numbers of persons killed in plane crashes
- 2. Daily variations in numbers of births
- 3. Weekly variations in numbers of births
- 4. Daily variations in numbers of deaths
- 5. Daily variations in numbers of traffic accidents
- 6. Variations across cookies/pizzas in numbers of chocolate chips/olives

Inference regarding  $\mu$ , based on observed count y

# Confidence interval for $\mu$

■ If the CLT hasn't kicked in, then the usual CI might not be appropriate:

point-estimate  $\pm z^* \times \text{standard error}$ 



## [1] 2 11

# Confidence interval for $\mu$

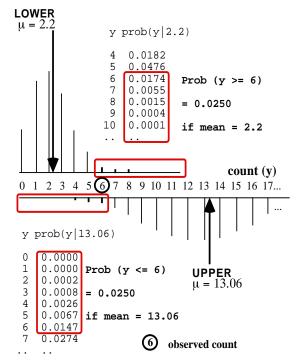
```
manipulate::manipulate(
mosaic::xqpois(c(0.025, 0.975), lambda = LAMBDA),
LAMBDA = manipulate::slider(1, 200, step = 1))
```

# Confidence interval for $\mu$

■ Similar to the binomial (Clopper-Pearson CI), we consider a first-principles  $100(1-\alpha)\%$  CI  $[\mu_{LOWER},\ \mu_{UPPER}]$  such that

$$P(Y \ge y \mid \mu_{LOWER}) = \alpha/2$$
 and  $P(Y \le y \mid \mu_{UPPER}) = \alpha/2$ .

■ For example, the 95% CI for  $\mu$ , based on y = 6, is  $[\underline{2.20}, \underline{13.06}]$ .



# Poisson 95% CI for $\mu$ when y = 6

```
# upper limit --> lower tail needs 2.5%
manipulate::manipulate(
mosaic::xppois(6, lambda = LAMBDA),
LAMBDA = manipulate::slider(0.01, 20, step = 0.01))

# lower limit --> upper tail needs 2.5%
# when lower.tail=FALSE, ppois doesnt include k, i.e., P(Y > k)
manipulate::manipulate(
mosaic::xppois(5, lambda = LAMBDA, lower.tail = FALSE),
LAMBDA = manipulate::slider(0.01, 20, step = 0.01))
```

# Confidence interval for $\mu$

- For a given confidence level, there is one CI for each value of *y*.
- Each one can be worked out by trial and error, or as has been done for the last 80 years directly from the (exact) link between the tail areas of the Poisson and Gamma distributions.
- These Cl's for y up to at least 30 were found in special books of statistical tables or in textbooks.
- As you can check, z-based intervals are more than adequate beyond this y. Today, if you have access to R (or Stata or SAS) you can obtain the first principles CIs directly for any value of y.

# 80%, 90% and 95% CI for mean count $\mu$ if we observe 0 to 30 events in a certain amount of experience

у	95%		90%		80%	
0	0.00	3.69	0.00	3.00	0.00	2.30
1	0.03	5.57	0.05	4.74	0.11	3.89
2	0.24	7.22	0.36	6.30	0.53	5.32
3	0.62	8.77	0.82	7.75	1.10	6.68
4	1.09	10.24	1.37	9.15	1.74	7.99
5	1.62	11.67	1.97	10.51	2.43	9.27
	2.20	13.06	2.61	11.84	3.15	10.53
<u>6</u> 7	2.81	14.42	3.29	13.15	3.89	11.77
8	3.45	15.76	3.98	14.43	4.66	12.99
9	4.12	17.08	4.70	15.71	5.43	14.21
10	4.80	18.39	5.43	16.96	6.22	15.41
11	5.49	19.68	6.17	18.21	7.02	16.60
12	6.20	20.96	6.92	19.44	7.83	17.78
13	6.92	22.23	7.69	20.67	8.65	18.96
14	7.65	23.49	8.46	21.89	9.47	20.13
15	8.40	24.74	9.25	23.10	10.30	21.29
16	9.15	25.98	10.04	24.30	11.14	22.45
17	9.90	27.22	10.83	25.50	11.98	23.61
18	10.67	28.45	11.63	26.69	12.82	24.76
19	11.44	29.67	12.44	27.88	13.67	25.90
20	12.22	30.89	13.25	29.06	14.53	27.05
21	13.00	32.10	14.07	30.24	15.38	28.18
22	13.79	33.31	14.89	31.41	16.24	29.32
23	14.58	34.51	15.72	32.59	17.11	30.45
24	15.38	35.71	16.55	33.75	17.97	31.58

# 95% CI for mean count $\mu$ with **q** function

- To obtain these in **R** we use the natural link between the Poisson and the gamma distributions.<sup>2</sup>
- In R, e.g., the 95% limits for  $\mu$  based on y=6 are obtained as

```
qgamma(p = c(0.025, 0.975), shape = c(6, 7))
## [1] 2.201894 13.059474
```

■ More generically, for any y, as

```
qgamma(p = c(0.025, 0.975), shape = c(y, y+1))
```

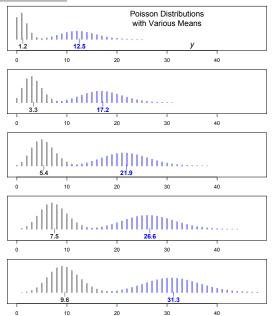
details found here

# 95% CI for mean count $\mu$ with canned function

■ These limits can <u>also</u> be found using the canned function in R

```
stats::poisson.test(6)
##
## ^^IExact Poisson test
##
## data: 6 time base: 1
## number of events = 6, time base = 1, p-value = 0.0005942
## alternative hypothesis: true event rate is not equal to 1
## 95 percent confidence interval:
## 2.201894 13.059474
## sample estimates:
## event rate
## 6
```

once  $\mu$  is in the upper teens, the Poisson ightarrow the Normal



■ Thus, a plus/minus CI based on SE =  $\hat{\sigma} = \sqrt{\hat{\mu}} = \sqrt{y}$ , is simply

$$[\mu_{\mathsf{L}},\ \mu_{\mathsf{U}}] = \mathsf{y}\ \pm\ \mathsf{z}^{\star} \times \sqrt{\mathsf{y}}.$$

Equivalently we can use the **q** function:

$$qnorm(p = c(0.025, 0.975), mean = y, sd = \sqrt{y})$$

■ Thus, a plus/minus CI based on SE =  $\hat{\sigma} = \sqrt{\hat{\mu}} = \sqrt{y}$ , is simply

$$[\mu_{\mathsf{L}}, \ \mu_{\mathsf{U}}] = \mathsf{y} \ \pm \ \mathsf{z}^{\star} \times \sqrt{\mathsf{y}}.$$

Equivalently we can use the **q** function:

$$qnorm(p = c(0.025, 0.975), mean = y, sd = \sqrt{y})$$

■ From a single realization y of a  $N(\mu, \sigma_Y)$  random variable, we can't estimate **both**  $\mu$  and  $\sigma_Y$ : for a SE, we would have to use *outside* information on  $\sigma_Y$ .

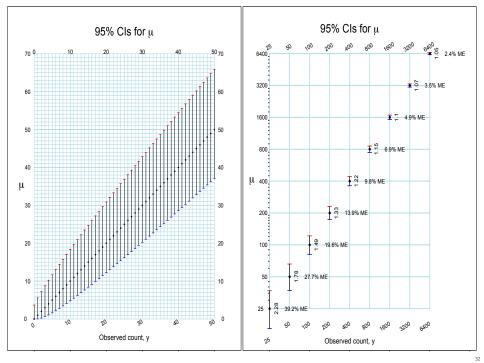
■ Thus, a plus/minus CI based on SE =  $\hat{\sigma} = \sqrt{\hat{\mu}} = \sqrt{y}$ , is simply

$$[\mu_L, \ \mu_U] = y \pm z^* \times \sqrt{y}.$$

Equivalently we can use the **q** function:

$$qnorm(p = c(0.025, 0.975), mean = y, sd = \sqrt{y})$$

- From a single realization y of a  $N(\mu, \sigma_Y)$  random variable, we can't estimate **both**  $\mu$  and  $\sigma_Y$ : for a SE, we would have to use *outside* information on  $\sigma_Y$ .
- In the Poisson( $\mu$ ) distribution,  $\sigma_Y = \sqrt{\mu}$ , so we calculate a "model-based" SE.



Inference regarding an event rate parameter  $\lambda$ , based on observed number of events y in a known amount of population-time (PT)

year	deaths (y)
1971	33
2002	211

Table: Deaths from lung cancer in the age-group 55-60 in Quebec in 1971 and 2002

year		deaths (y)
	1971	33
	2002	211

Table: Deaths from lung cancer in the age-group 55-60 in Quebec in 1971 and 2002

A researcher asks: Is the situation getting worse over time for lung cancer in this age group?

year	deaths (y)	
1971	33	
2002	211	

Table: Deaths from lung cancer in the age-group 55-60 in Quebec in 1971 and 2002

A researcher asks: Is the situation getting worse over time for lung cancer in this age group?

Your reply: What's the denominator??

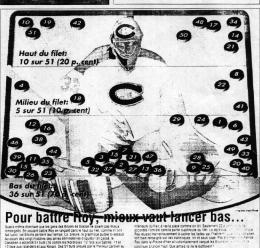
A: 91 - 1-4 (537 A 5



monde. 35, soit 70 p. cent d'entre eux, ont vu la rondelle pénétrer dans la partie.

#### Sutter a trop parlé; personne ne va toucher à Roy, foi de Carbo

Pages 2 à 5



- So far, we have focused on inference regarding  $\mu$ , the expected **number** of events in the amount of experience actually studied.
- However, for <u>comparison</u> purposes, the frequency is more often expressed as a rate, intensity or incidence density (ID).

- So far, we have focused on inference regarding  $\mu$ , the expected **number** of events in the amount of experience actually studied.
- However, for <u>comparison</u> purposes, the frequency is more often expressed as a <u>rate</u>, <u>intensity</u> or <u>incidence density</u> (ID).

year	deaths (y)	person-time (PT)	rate $(\hat{\lambda})$
1971	33	131,200 years	25 per 100,000 women-years
2002	211	232,978 years	91 per 100,000 women-years

Table: Deaths from lung cancer in the age-group 55-60 in Quebec in 1971 and 2002

■ The *statistic*, the empirical rate or empirical incidence density, is

$$rate = \hat{ID} = \hat{\lambda} = y/PT.$$

- where y is the observed number of events and PT is the amount of Population-Time in which these events were observed.
- We think of  $\hat{ID}$  or  $\hat{\lambda}$  as a point estimate of the (theoretical) Incidence Density parameter, ID or  $\lambda$ .

To calculate a CI for the ID parameter, we **treat the PT** denominator as a constant, and the <u>numerator</u>, y, as a Poisson random variable, with expectation  $E[y] = \mu = \lambda \times PT$ , so that

$$\lambda = \mu \div PT$$
$$\hat{\lambda} = \hat{\mu} \div PT$$
$$= y \div PT$$

CI for 
$$\lambda = \{ \text{CI for } \mu \} \div \text{PT.}$$
 (1)

■ y=211 deaths from lung cancer in 2002 leads to a 95% CI for  $\mu$ :

```
qgamma(p = c(0.025, 0.975), shape = c(211, 212))
## [1] 183.4885 241.4725
```

• y = 211 deaths from lung cancer in 2002 leads to a 95% CI for  $\mu$ :

```
qgamma(p = c(0.025, 0.975), shape = c(211, 212))
## [1] 183.4885 241.4725
```

From this we can calculate the 95% CI **per 100,000 WY** for  $\lambda$  using a PT=232978 years:

```
qgamma(p = c(0.025, 0.975), shape = c(211, 212)) / 232978 * 1e5
## [1] 78.75788 103.64607
```

• y = 211 deaths from lung cancer in 2002 leads to a 95% CI for  $\mu$ :

```
qgamma(p = c(0.025, 0.975), shape = c(211, 212))
## [1] 183.4885 241.4725
```

From this we can calculate the 95% CI **per 100,000 WY** for  $\lambda$  using a PT=232978 years:

```
qgamma(p = c(0.025, 0.975), shape = c(211, 212)) / 232978 * 1e5
## [1] 78.75788 103.64607
```

• y=33 deaths from lung cancer in 131200 women-years in 1971 leads to a 95% CI per 100,000 WY for  $\lambda$  of

```
qgamma(c(0.025,0.975), c(33,34)) / 131200 * 1e5
## [1] 17.31378 35.32338
```

# CI for the rate parameter $\lambda$ using canned function

```
stats::poisson.test(x = 33, T = 131200)

##
## ^^IExact Poisson test
##
## data: 33 time base: 131200
## number of events = 33, time base = 131200, p-value < 2.2e-16
## alternative hypothesis: true event rate is not equal to 1
## 95 percent confidence interval:
## 0.0001731378 0.0003532338
## sample estimates:
## event rate
## 0.0002515244</pre>
```

Test of  $H_0: \mu = \mu_0 \quad \Leftrightarrow \quad \lambda = \lambda_0$ 

# Statistical evidence and the p-value

#### Recall:

- P-Value = Prob[y or more extreme |  $H_0$ ]
- With 'more extreme' determined by whether  $H_{alt}$  is 1-sided or 2-sided.
- lacksquare For a **formal test**, at level lpha, compare this P-value with lpha.

■ Cancers in area surrounding the Douglas Point nuclear station

- Cancers in area surrounding the Douglas Point nuclear station
- Denote by {CY1, CY2,...} the numbers of Douglas Point child-years of experience in the various age categories that were pooled over.
- Denote by  $\{\lambda_1^{Ont}, \lambda_2^{Ont}, \dots\}$  the age-specific leukemia incidence rates during the period studied.

- Cancers in area surrounding the Douglas Point nuclear station
- Denote by {CY<sub>1</sub>, CY<sub>2</sub>,...} the numbers of Douglas Point child-years of experience in the various age categories that were pooled over.
- Denote by  $\{\lambda_1^{Ont}, \lambda_2^{Ont}, \dots\}$  the age-specific leukemia incidence rates during the period studied.
- If the underlying incidence rates in Douglas Point were the same as those in the rest of Ontario, the Expected total number of cases of leukemia for Douglas Point would be

$$E = \mu_0 = \sum_{aqes} CY_i \times \lambda_i^{Ont} = 0.57.$$

- Cancers in area surrounding the Douglas Point nuclear station
- Denote by {CY<sub>1</sub>, CY<sub>2</sub>,...} the numbers of Douglas Point child-years of experience in the various age categories that were pooled over.
- Denote by  $\{\lambda_1^{Ont}, \lambda_2^{Ont}, \dots\}$  the age-specific leukemia incidence rates during the period studied.
- If the underlying incidence rates in Douglas Point were the same as those in the rest of Ontario, the Expected total number of cases of leukemia for Douglas Point would be

$$E = \mu_0 = \sum_{\text{gaps}} CY_i \times \lambda_i^{\text{Ont}} = 0.57.$$

The actual total number of cases of leukemia **O**bserved in Douglas Point was

$$O = y = \sum_{ages} O_i = 2.$$

- Cancers in area surrounding the Douglas Point nuclear station
- Denote by {CY<sub>1</sub>, CY<sub>2</sub>,...} the numbers of Douglas Point child-years of experience in the various age categories that were pooled over.
- Denote by  $\{\lambda_1^{Ont}, \lambda_2^{Ont}, \dots\}$  the age-specific leukemia incidence rates during the period studied.
- If the underlying incidence rates in Douglas Point were the same as those in the rest of Ontario, the Expected total number of cases of leukemia for Douglas Point would be

$$E = \mu_0 = \sum_{ones} CY_i \times \lambda_i^{Ont} = 0.57.$$

The actual total number of cases of leukemia **O**bserved in Douglas Point was

$$O = y = \sum_{opes} O_i = 2.$$

Age Standardized Incidence Ratio (SIR) = O/E = 2/0.57 = 3.5.

# Q: Is the O=2 significantly higher than E=0.57

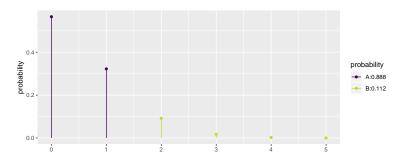
#### Question:

- Is the *y* = 2 cases of leukemia observed in the Douglas Point experience statistically significantly <u>higher</u> than the *E* = 0.57 cases "expected" for this many child-years of observation if in fact the rates in Douglas Point and the rest of Ontario were the same?
- Or, is the y=2 observed in this community compatible with  $H_0: y \sim \text{Poisson}(\mu=0.57)$ ?

# A: Is the O=2 significantly higher than E=0.57

Answer: Under  $H_0$ , the age-specific numbers of leukemias  $\{y_1=O_1,\,y_2=O_2,\,\dots\}$  in Douglas Point can be regarded as independent Poisson random variables, so their sum y can be regarded as a single Poisson random variable with  $\mu=0.57$ .

```
mosaic::xppois(1, lambda = 0.57, lower.tail = FALSE)
```



## [1] 0.1121251

# 95% CI for the SIR by hand

■ To get the <u>CI for the SIR</u>, divide the CI for Douglas Point  $\mu_{DP}$  by the null  $\mu_0 = 0.57$  (Ontario scaled down to the same size and age structure as Douglas Point.) We treat it as a constant because the Ontario rates used in the scaling are measured with much less sampling variability that the Douglas Point ones.

# 95% CI for the SIR by hand

- To get the <u>CI for the SIR</u>, divide the CI for Douglas Point  $\mu_{DP}$  by the null  $\mu_0=0.57$  (Ontario scaled down to the same size and age structure as Douglas Point.) We treat it as a constant because the Ontario rates used in the scaling are measured with much less sampling variability that the Douglas Point ones.
- The y = 2 cases translates to
  - ▶ 95% CI for  $\mu_{DP}$  → [0.24, 7.22]
  - ▶ 95% CI for the SIR  $\rightarrow$  [0.24/0.57, 7.22/0.57]=[0.4, 12.7].

# 95% CI for the SIR using canned function

■ We can *trick* **stats::poisson.test** to get the same CI by putting time as 0.57:

```
stats::poisson.test(x=2,T=0.57)

##
## ^^IExact Poisson test
##
## data: 2 time base: 0.57
## number of events = 2, time base = 0.57, p-value = 0.1121
## alternative hypothesis: true event rate is not equal to 1
## 95 percent confidence interval:
## 0.4249286 12.6748906
## sample estimates:
## event rate
## 3.508772
```

(2) と (2) は (3) は (4) は (4

One definition of epidemiology:

• 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?

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?

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?

から中 4百×4百×4回×4□×

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, other none or more of them will be killed in this way. If R. A. Esher, 1925, using just 10 of 14 corps used by Borkhewiz in its 1938 work, Das Gesetz der kleinen Zahlen.]

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, other none or more of them will be killed in this way. If R. A. Esher, 1925, using just 10 of 14 corps used by Borkhewiz in its 1938 work, Das Gesetz der kleinen Zahlen.]

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

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, other none or more of them will be killed in this way. If R. A. Esher, 1925, using just 10 of 14 corps used by Borkhewiz in its 1938 work, Das Gesetz der kleinen Zahlen.]

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

	3
Deaths	3000
<u>.</u>	

Frequencies)

		y × Obs. Freq.	0	92	44	6	4	•	No. Deaths: 122
	s)	H	7	m	2	1	9		200 NC
(00:0	with y death	Expected <sup>5</sup>	108.7	.99	20.2	4.	0	0.1	20
(n)	(No. "corps-years" with y deaths)	Observed	109	92	22	က	-		200
	in corps-year	×	0	-	7	က	4	2	Sum:

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, other none or more of them will be killed in this way. If R. A. Esher, 1925, using just 10 of 14 corps used by Borkhewiz in its 1938 work, Das Gesetz der kleinen Zahlen.]

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

No. Deaths

Frequencies)

	y × Obs. Freq.	0	9	44	6	4	•	No. Deaths: 122
s" with y deaths)	Expected <sup>\$</sup>	108.7	66.3	20.2	4.1	9.0	0.1	200
(No. "corps-years" with y deaths)	Observed	109	65	22	က	-		200
in corps-year	×	0	-	7	က	4	2	Sum:

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

The chance of a man being killed by horsekkid on any one day is exceedingly small but if an army corps of men are exposed to this risk for a year, offen one on one of them will be killed in this way. If A. Fisher, 1925, using just 10 of 14 corps used by Bornkewsz in his 1939 work, Las Geaerz der kleinen Zahlen.]

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

	1
Deaths	000
ġ	

Frequencies)

	y × Obs. Freq.	0	92	44	6	4	•	No. Deaths: 122
(No. "corps-years" with y deaths)	Expected <sup>\$</sup>	108.7	66.3	20.2	4.1	9.0	0.1	200
(No. "corps-ye	Observed	109	9	22	က	-	•	200
in corps-year	×	0	-	7	က	4	2	Sum:

dpois (0:6, lambda=0.61) \* \$200  $\frac{122 \text{ deaths}}{200CY} = 0.61/CY$ ; Ш 1> ) | |

The charce 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, offen on no more of them will be killed in this way. If A. Fisher, 1925, using just 10 of 14 corps used by Bornkews in its 1998 work, Das Geetz der kilenen Zahlen].

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

	_
	4
'n	i
Ĕ	ò
a	000
Φ	
2	0
÷	(
ġ	•

Frequencies)

	y × Obs. Freq.	0	92	44	6	4	•	No. Deaths: 122
(No. "corps-years" with y deaths)	Expected <sup>5</sup>	108.7	66.3	20.2	4.1	9.0	0.1	200
(No. "corps-ye	Observed	109	65	22	က	-	•	200
in corps-year	У	0	-	7	က	4	2	Sum:

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

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

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.]

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

SL	
Deaths	
_ 2	

Frequencies)

	y × Obs. Freq.	0	92	44	6	4		No. Deaths: 122	
(No. "corps-years" with y deaths)	Expected <sup>5</sup>	108.7	66.3	20.2	4.1	9.0	0.1	200	
(No. "corps-year	Observed	109	65	22	က	-		200	
in corps-year	y	0	-	7	က	4	2	Sum:	

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

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

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.]

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

	1
Deaths	
Des	-
ġ	,

Frequencies)

	y × Obs. Freq.	0	65	44	6	4	•	No. Deaths: 122
(No. "corps-years" with y deaths)	Expected <sup>5</sup>	108.7	66.3	20.2	4.1	9.0	0.1	200
(No. "corps	Observed	109	92	22	ო	-	•	200
in corps-year	×	0	-	7	က	4	2	Sum:

\* dpois(0:6,lambda=0.61) \$200 = 0.61/CY; 122 deaths 200CY  $\parallel$ 1> ) | |

- 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 =  $\sqrt{mean}$ "  $\Rightarrow$  Poisson Distribution, but there is close agreement ('fit') between the observed and expected/theoretical distributions.

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 Geserz der kleinen Zahlen.]

orps-year" (CY).

ō
per
_
observations,
200
.e.
years,
20
ģ
army-corps
0 Prussian
_

Frequencies)

No. Deaths

	y × Obs. Freq.	0	92	44	6	4	•	No. Deaths: 122
in corps-year (No. "corps-years" with y deaths)	Expected\$	108.7	66.3	20.2	4.1	9.0	0.1	200
(No. "corps-ye	Observed	109	9	22	က	-	,	200
in corps-year	У	0	-	7	ო	4	2	Sum:

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

- 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 =  $\sqrt{mean}$ "  $\Rightarrow$  Poisson Distribution, but there is close agreement ('fit') between the observed and expected/theoretical distributions.

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

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

Decade		Saffir-Si	Saffir-Simpson Category	egory		Ψ	Major
	1	2	3	4	2	1,2,3,4,5	3,4,5
1851-1860	8	9	2	-	0	19	9
1861-1870	8	9	1	0	0	15	1
1871-1880	7	9	7	0	0	20	7
1881-1890	8	6	4	1	0	22	9
1891-1900	8	9	2	3	0	21	8
1901-1910	10	4	4	0	0	18	4
1911-1920	10	4	4	3	0	21	7
1921-1930	9	3	3	2	0	13	9
1931-1940	4	- 4	9	1	1	19	8
1941-1950	8	9	6	-	0	24	10
1951-1960	8	1	5	3	0	17	8
1961-1970	3	9	4	-	-	14	9
1971-1980	9	2	4	0	0	12	4
1981-1990	6		4	-	0	15	2
1991-2000	3	9	4	0	-	14	9
2001-2004	4	2	2	1	0	6	3
1851-2004	109	72	7.1	18	3	273	82
Average Per Decade	7.1	4.7	4.6	1.2	0.2	17.7	6.0

Number of hunicanes by Saffir-Smpson Category to strike the maintand U.S. each decade

Major	3,4,5	9	-	7	9	8	4	7	9	8	10	8	9	4	9	9	3	92	6.0	
All	1,2,3,4,5	19	15	20	22	21	18	21	13	19	24	17	14	12	15	14	6	273	17.7	
	2	0	0	0	0	0	0	0	0	-	0	0	-	0	0	-	0	3	0.2	
egory1	4	-	0	0	-	3	0	3	2	-	-	3	-	0	-	0	-	18	12	
Saffir-Simpson Category <sup>1</sup>	က	2	-	7	4	2	4	4	3	9	6	2	4	4	4	4	2	7.1	4.6	
Saffir-Si	2	2	9	9	6	2	4	4	3		9	-	2	2	-	9	2	72	4.7	
	-	8	8	7	8	8	10	10	2	4	80	8	3	9	6	3	4	109	7.1	
- Cooperation		1851-1860	1861-1870	1871-1880	1881-1890	1891-1900	1901-1910	1911-1920	1921-1930	1931-1940	1941-1950	1951-1960	1961-1970	1971-1980	1981-1990	1991-2000	2001-2004	1851-2004	Average Per Decade	

5.9(2.3) 2.4 :: 17.6(3.2) 4.2 **15 FULL DECADES** 4.6(1.8) 1.1(1.1) 0.2(0.4) 2.1 4.7(2.2) 2.2 7.0(2.3) 5.6 mean(sd) √mean: Ob(A 重 4重 4 重 4 4 4 4 ロ

nland U.S. each decade nber of h

open-od		Saffir-Si	Saffir-Simpson Category <sup>1</sup>	egory1		ΑII	Major
2000	-	2	က	4	2	1,2,3,4,5	3,4,5
1851-1860	8	22	2	-	0	19	9
1861-1870	8	9	-	0	0	15	-
1871-1880		9		0	0	20	7
1881-1890	8	6	4	-	0	22	9
1891-1900	8	2	9	3	0	21	8
1901-1910	10	4	4	0	0	18	4
1911-1920	10	4	4	3	0	21	7
1921-1930	9	3	3	2	0	13	9
1931-1940	4	7	9	-	-	19	8
1941-1950	8	9	6	-	0	24	10
1951-1960	8	-	40	3	0	17	8
1961-1970	3	2	4	-	-	14	9
1971-1980	9	2	4	0	0	12	4
1981-1990	6	-	4	-	0	15	2
1991-2000	3	9	4	0	-	14	9
2001-2004	4	2	2	-	0	6	3
1851-2004	109	72	7.1	18	3	273	92
Average Per Decade	7.1	4.7	4.6	1.2	0.2	17.7	6.0

5.9(2.3) 17.6(3.2) 4.2 **15 FULL DECADES** 3) 1.1(1.1) 0.2(0.4) 0.4 4.6(1.8) 2.1 4.7(2.2) 7.0(2.3) 5.6 mean(sd) √mean:

2.4 !!

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

Ö

mber of hunicanes by Saffr-Simpson Category to strike the mainland U.S. each decade

1	Sa	Sa	ffir-Si	Saffir-Simpson Category	egory <sup>1</sup>		All	Major
1 0   19   19   19   19   19   19   19		-	7	က	4	co	1,2,3,4,5	0,4,0
1	1851-1860	8	2	9	-	0	19	9
7   6   7   0   0   0   20     0   0   4   1   0   0   22     10   4   4   0   0   0   21     10   4   4   0   0   0   21     10   4   4   0   0   0   18     4   7   0   1   1   10     5   3   5   1   0   24     6   1   0   2   1   10     7   1   4   1   10     7   4   4   1   0   0     7   4   4   1   1   0     7   4   4   1   1   0     7   4   4   1   1   0     7   4   4   1   1   0     7   4   4   1   1   0     7   4   4   1   1   0     7   4   4   1   1   0     7   4   4   1   1   0     7   4   4   4   1   0     7   4   4   6   1   1     7   4   7   4   6   1     7   7   7   7   7     7   7   7   7	1861-1870	8	9	1	0	0	15	1
0	1871-1880	7	9	7	0	0	20	7
10	1881-1890	8	6	4	-	0	22	2
10	1891-1900	80	2	2	3	0	21	80
10   4   4   5   0   21     4   7   6   1   1   1     4   7   6   1   1   1     8   1   6   3   3   1   1     9   1   6   3   1   1     10   2   4   1   1     11   4   2   2   1     10   72   71   18   3   273     10   10   12   14     10   12   13   14     10   12   13   14     10   12   13     10   12   13     11   14   14     12   14   14     13   15   15     14   15   15     15   17   18   17     17   18   12   12     17   18   17   17     18   18   18     18   18   18     19   19   19     10   10   10     10   10   10     11   11	1901-1910	10	4	4	0	0	18	4
5   3   3   2   0   13     8   7   6   1   1   1   1     9   1   5   3   0   17     9   1   4   1   0   15     9   1   4   1   0   15     100   72   71   18   3   273     11   12   13   14   14     100   72   71   18   3   273     11   12   12   17     12   13   14   15   15     13   14   15   15   15     14   15   15   15   17     15   15   15   17     16   17   17   18   17     17   18   18   17     18   18   18   18     19   19   19   19     10   10   10   10     11   12   12   13   17     12   13   13   13     13   14   15   15     14   15   15   15     15   15   15   17     16   17   18   18     17   18   18   18     18   18   18   18     19   19   19     10   10   10     10   10   10     10   10	1911-1920	10	4	4	3	0	21	7
4   7   6   1   1   19     8   1   6   3   0   24     9   1   6   3   0   17     9   2   4   0   0   12     9   1   4   1   0   15     10   4   2   2   1   0   9     10   72   71   18   3   273     11   45   12   02   177     12   45   46   12   02   177     13   47   46   12   02   177     14   7   7   7   7   7   7   7     15   7   7   7   7   7   7     16   7   7   7   7   7   7     17   7   7   7   7   7   7     18   7   7   7   7   7     19   7   7   7   7   7     10   7   7   7   7   7     10   7   7   7   7   7     11   12   7   7   7     12   7   7   7   7     13   7   7   7   7     14   7   7   7   7     15   7   7   7   7     16   7   7   7     17   7   7   7     17   7   7   7     18   7   7   7     19   7   7   7     10   7   7   7     10   7   7   7     11   7   7   7     12   7   7   7     13   7   7     14   7   7   7     15   7   7     16   7   7     17   7   7     17   7   7   7     18   7   7     19   7   7     10   7   7     10   7   7     11   7   7     12   7   7     13   7   7     14   7   7     15   7   7     16   7   7     17   7   7     18   7   7     18   7   7     19   7   7     10   7   7     10   7   7     11   7   7     12   7   7     13   7   7     14   7   7     15   7   7     16   7   7     17   7   7     18   7   7     18   7   7     18   7   7     18   7   7     18   7   7     18   7   7     18   7   7     18   7   7     18   7   7     18   7   7     18   7   7     18   7   7     18   7   7     18   7   7     18   7     18   7   7     18   7	1921-1930	2	3	3	2	0	13	2
0   0   0   1   0   24	1931-1940	4	7	9	-	-	19	80
8   1   5   3   0   17     9   2   4   1   1   1     9   1   4   1   0   15     9   1   4   1   0   15     10   2   2   3   0   15     10   2   2   1   0   0     10   72   71   18   3   273     11   12   48   12   02   173	1941-1950	8	9	6	-	0	24	10
3   5   4   1   1   14   14   19   19   19	1951-1960	8	1	2	3	0	17	8
6   2   4   0   0   12     9   1   4   1   0   15     3   6   4   0   1   14     4   2   2   1   0   9     100   72   71   18   3   273     71   47   46   12   02   177	1961-1970	3	2	4	-	-	14	9
9   1   4   1   0   15     1   2   2   2   2   3     1   3   2   2   3     1   4   2   2   3     1   6   72   71   18   3     2   7   47   48   12   02   177	1971-1980	9	2	4	0	0	12	4
3   6   4   0   1   14   14   14   14   14	1981-1990	6	-	4	-	0	15	2
4   2   2   1   0   9	1991-2000	3	9	4	0	1	14	9
(100         72         71         18         3         273           71         47         46         12         02         177	2001-2004	4	2	2	1	0	6	3
109 72 71 18 3 273 7.1 4.7 4.6 12 0.2 17.7								
7.1 4.7 4.6 1.2 0.2 17.7	1851-2004	109	72	7.1	18	3	273	92
	Average Per Decade	7.1	4.7	4.6	1.2	0.2	17.7	6.0

Contraction of the Contraction o

5.9(2.3) 2.4 ⊞ 17.6(3.2) 4.2 0.2(0.4) 15 FULL DECADES 3) 1.1(1.1) 0.21 4.6(1.8) 4.7(2.2) 7.0(2.3) mean(sd) √mean:

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

jilji 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

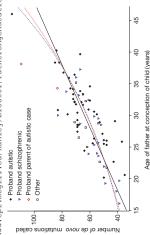


Figure 2 | Father's age and number of de novo mutations. The number of denovo mutations called is plotted against father's age at conception of child for the 78 trios. The solid black line denotes the linear fit. The dashed red curve is assumed to have a constant rate of 14.2 and paternal mutations are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

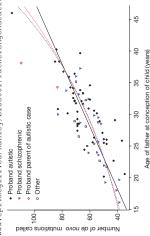


Figure 2 | Father's age and number of de novo mutations. The number of denovo mutations called is plotted against father's age at conception of child for the 78 trios. The solid black line denotes the linear fit. The dashed red curve is assumed to have a constant rate of 14.2 and paternal mutations are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

#### What events are these?

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

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 as aggested that law by York was expediencing a backby boom. The columns below show the numbers of bables bon every day during a 25-day period, Augi Augi5, centreed 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.

On November 9: 1990, the power went out in two York OLY, and it stayed out for ally—The <b>CHEST BEXACUI.</b> Unite mounts 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. Augi-Augi5, centreed nine months and len days after the Great Blackout. These numbers average out to 436. This turns out not to be unusually high for New York.	Births	451	468	429	448	466	377	344	448	438	455	468		405								351				426
#∃∃ C	Aug.	-	7	က	4	S	9	7	80	6	9	Ξ	12	5	4	15	16	17	8	6	8	2	83	g	24	S

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 as aggested that law by York was expediencing a backby boom. The columns below show the numbers of bables bon every day during a 25-day period, Augi Augi5, centreed 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.

Day Mon Tue Wed	Sat	Sun Mon Tue	Thu Fri	Sur Sun Mon	Tue Wed	Sar	Mon	Tue Wed Thu
Births 451 468 429	448 466 377	344 4 48 4 38	4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4	405 377 451	497 458	2 4 4 2 4 0 4 0 5 0 7 0	351 467	508 4 432 426
Aug.	4100	<b>~</b> 8 6	2 T 2	t 4 t	5 t a	5 th 8	2 23	8 2 8

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 as aggested that law by York was expediencing a backby boom. The columns below show the numbers of bables bon every day during a 25-day period, Augi Augi5, centreed 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	<b>Day</b> Mon	Тие	Wed	Thu	E	Sat	Sun	Mon	Tue	Wed	과	Œ	Sat	Sun	Mon	Tue	Wed	과	Œ	Sat	Sun	Mon	Тие	Wed	Thu
	451 P																								
	Aug.	7	က	4	ιΩ	9	7	ω	6	9	Ξ	12	5	4	5	16	17	9	19	8	5	83	ន	54	83

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 as aggested that law by York was expediencing a backby boom. The columns below show the numbers of bables bon every day during a 25-day period, Augi Augi5, centreed 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.

Otationics (And Ed) by D Headman, It in sail of all	<ul> <li>"Apparently, the New York Times, NYT sent a reporter around to a few</li> </ul>	hospitals on Monday, August 8, and Tuesday, August 9, nine months	after the blackout.																						
Day	Mon	Tue	Wed	모	Έ	Sat	Sun	Mon	Tue	Wed	Τh	Œ	Sat	Sun	Mon	Tue	Wed	Thu	Œ	Sat	Sun	Mon	Tue	Wed	무
Births	451	468	429	448	466	377	344	448	438	455	468	462	405	377	451	497	458	429	434	410	351	467	208	432	426

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 after the Great Blackout. These numbers average out to 436. This turns out not to be unusually high for New York the numbers of babies born every day during a 25-day period, Aug1-Aug25, centered nine months and ten days

### 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.

Wed

Έ

Line Line 롣

468 429 84 466 377 <del>4</del> 84 438 455 468 462 377 45 497 458 429 434 5 351 467 508 432

Aug.

Non Mo

Sun Mon Tue Wed Non Sat Wed

Tue 귤

르

Œ

9 7 8 6 0 - 2 5 4 5 9 7 8

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.

	į	

Wed

e Le

World Sat

Ē

<u>o</u> 828888

On November 9. 1965, the power went out in New York City, and is tayed out for a day—the **Great Blackout**. Nine months later, the newspapers out as uggested that level York was uggested that be york was expediencing a busby boom. The outburns bedown show the numbers of basbes born every day duming a 25-day period, Aug 1-Aug5, centered mine 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.

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.

hospitals on Monday, August 8, and Tuesday, August 9, nine months

after the blackout.

Wed

Έ

Line Line 롣

468 429 84 466 377 <del>8</del> 84 438 455 468 462 405 377 451 497 458 429 434 5 351 467 508 432

Aug.

Non Mo

Headline, p.1, Wed, Aug10, '66: "Births Up 9 Months After the

Sun Mon Tue Wed Sun Wed

Tue 귤

2

Œ

297860-2545978

Blackout http://www.epi.mcgill.ca//hanley/bios601 Intensity-Rate/BirthsUp9MonthsAfterBlackout.pdf

Wed

e Le

World Sat

Ē

<u>o</u> 828888

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 <b>Great Blackout</b> . Nine months start, the tewespapers suggested that New York was experienting a bady brown. The outbrusts below show from the number of bades beto newer day during a 25-day period. Augh-Augh, centered nine morths and lein days after the Great Blackout. These numbers average out to 436. This turns out not to be unusually high for New York.	Statistics (znd Ecl) by D Freedman, R Pisani et all  • "Apparently, the New York Times WYT sent a reporter around to a few hoopilates on Monday. August 8, and Tugsday, August 9, rine months after the backout.  • The hospilate sopred that their Obstetrics wards were <u>Duser than use</u> probably because of the general pattern that weekends are slow, probably because of the general pattern that weekends are slow.  • Headine, p.1, Wed, Augit G. Ser. "Eithe Up 9 Months After the Buckout, http://www.epi.mogill.com/honlay/bicos601/ Internativy Rate/BilthalpysportthafteesBackout, pdf  • This joverloaded succe of variation) seems to be the origin of the baby-boom myth.
of BIF	tin New that New ring a 25- average	Statistik
about Daily No.s	, 1965, the power went or newspapers suggested pabies born every day duri lackout. These numbers	Day We do
Whate	November 9, ths later, the numbers of b the Great B	Births 4 68 8 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4
	On f mon the r after	Aug. 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2

Windt BOOLI DAIII/ NO.S of BIFFI HS 9 WION ITS AFTER BLACK/DUI! ?  November 9, 1965, the powerwant out in New York (by, and it staged out for a day – the <b>creat Backout.</b> Nine monthis late, the newspapers augrested that why York was experienting 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 len days.	atter the Great Blackout. These numbers <u>average</u> out to 436. This turns out not to be unusually high for New York. Consider to the Consider on the Institute of the Consider of the Consideration of the Consi	atistas (zna Eu) dy d' Freedman, n' Fisain et al	<ul> <li>"Apparently, the New York Times, NYT sent a reporter around to a few</li> </ul>	hospitals on Monday, August 8, and Tuesday, August 9, nine months	after the blackout.	<ul> <li>The hospitals reported that their obstetrics wards were busier than usual</li> </ul>	probably because of the general pattern that weekends are slow,	Mondays and Tuesdays are busy.	<ul> <li>Headline, p.1, Wed, Aug10, '66: "Births Up 9 Months After the</li> </ul>	Blackout http://www.epi.mcgill.ca//hanley/bios601/	Intensity-Rate/BirthsUp9MonthsAfterBlackout.pdf	This foverlooked source of variation seems to be the origin of the	haby-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/ 言, 言 つくへ
DOUT DAIIY NO.S OF 1965, the power went out in newspapers suggested the abies born every day during	ackout. These numbers aw	Day	Mon	Tue	Wed	2 3	Sat	Sun	Mon	Tue	Wed	Thu	Ē	Sat		-			_			Sun			_	nd F	ч
What a November 9, 1 oths later, the 1 numbers of be	r the Great Bla	Births	451	468	429	448	377	344	448	438	455	468	462	405	377	451	497	458	429	434	410	351	467	208	432	426	
P E E	afte	Aug.	-	7	က	4 п	9	7	80	6	9	Ξ	12	5	4	5	16	17	8	19	8	5	83	ន	24	S	

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

eded" numbers of accidents eriod among 708 Northern	rivers. [Table 2.5]	Number of drivers	with y accidents	Expected <sup>\$</sup>	71.5	164.0	187.9	143.6	82.3	37.7	14.4	4.7	1.4	0.3	0.1	0.0	708	
ed" numbers od among 70	uthority bus di	Number	with y a	Observed	117	157	158	115	78	44	21	7	9	-	ო	-	208	
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	3	0	-	2	က	4	2	9	7	80	6	10	=		

## domly" over bus drivers?

Obsedurir Irelau

Are "accidents" distributed "randor cted" numbers of accidents eriod among 708 Northern Authority bus drivers. [Table 2.5]	Number of drivers	with $y$ accidents	Expected <sup>\$</sup>	71.5	164.0	187.9	143.6	82.3	37.7	14.4	4.7	1.4	0.3	0.1	0.0	708	
e "accident ed" numbers od among 70 ithority bus d	Number	with y	Observed	117	157	158	115	78	44	21	7	9	-	က	-	208	$\frac{s}{driver}$ .
Are "accidents" distribute served & "expected" numbers of accidents ing a 3-year period among 708 Northern and Transport Authority bus drivers. [Table.]	o. accidents in	3-year period	Ś	0	-	2	က	4	2	9	7	80	6	10	1		$= \bar{y} = \frac{1623 \ accidents}{708 \ drivers}$

jilji

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

Number of drivers

No. accidents in

nts	Expected <sup>\$</sup>	71.5	164.0	187.9	143.6	82.3	37.7	14.4	4.7	1.4	0.3	0.1	0.0	708		
ocide	Expe															
with y accidents	<u>D</u> bserved	117	157	158	115	78	44	2	7	9	_	က	-	208	2.29 driver	0
3-year period	ğ (3)	0	-	2	က	4	2	9	7	80	6	10	Ξ		$\hat{\lambda} = \bar{y} = \frac{1623 \ accidents}{708 \ drivers} =$	100 to 11 100 to 11 100

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

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

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

frequencies:

164.0 143.6 71.5 82.3 1.4 Expected\$ 187.9 4 14 Number of drivers with y accidents 157 158 115 Observed 78 4 No. accidents in 3-year period 3 9

 $\hat{\lambda} = \bar{y} = \frac{1623}{708} \frac{accidentis}{arivers} = \frac{2.29}{ariver}$ . SD of the 708 y's :1.86;  $\sqrt{2.29} = 1.51$ .

708

0.3

<sup>\$708 \*</sup> dpois(0:11,lambda=2.29)

Observed & "expected" numbers of accidents during a 3-year period among 708 Northern

"Comparison of observed and expected frequencies: Ireland Transport Authority bus drivers. [Table 2.5] Number of drivers with v accidents No. accidents in 3-year period

<ul> <li>More than the expected number of</li> </ul>	drivers with no accidents														
with y accidents	Expected\$	71.5	164.0	187.9	143.6	82.3	37.7	14.4	4.7	1.4	0.3	0.1	0.0	708	
with y a	Observed	117	157	158	115	78	44	21	7	9	-	က	_	208	ts _ 2.29
3-year period	Ś	0	-	2	က	4	5	9	7	80	6	10	Ξ		. T. 1623 accidents

	5.
	<del>-</del> -
	Ш
	√2.29
	lc4
	ĺΩ
	>
	'n
	æ
	<del>-</del> -
i	
	့်လ
i	y's :1.86;
	708
ı	0
	^
	<u>e</u>
	the
	ot
	Ö

<sup>\$708 \*</sup> dpois(0:11,lambda=2.29)

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

Nore than the expected number of Nore than the expected number of rivers with five or more accidents

rivers with no accidents

"Comparison of observed and expected

Number of drivers	with y accidents • More th	Observed Expected\$ drivers	117 71.5 • More th	157 164.0 drivers	158 187.9	115 143.6	78 82.3	44 37.7	21 14.4	7 4.7	6 1.4	1 0.3	3 0.1	1 0.0	708 708	2.29	driver ·
	3-year period w	iesqO (A)	0	-	Ø	ო	4	5	9	7	80	6	10	11		જા 	708 drivers

l	
l	
2.29 driver	$\sqrt{2.29} = 1.51$ .
1623 accidents = 708 drivers =	SD of the 708 y's :1.86;
1623	200
$=\bar{y}=$	of the
( <u>~</u>	SD

dpois (0:11, lambda=2.29) \$708 ×

(0) ð

### 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]

Number of drivers

No. accidents in

	یا	l.,	0	0	.0	~	7	511	7	511	<u>~</u>	_	0	l_	l
with y accidents	Expected\$	71.5	164.	187.	143.	82.	37.	14.	4.	1.	0	0.1	0.0	708	
with y	Observed	117	157	158	115	78	44	21	7	9	-	ო	-	708	$\frac{16}{2.29} = \frac{2.29}{2.29}$
3-year period	3	0	-	2	က	4	5	9	7	80	6	10	Ξ		$= \overline{V} = \frac{1623 \ accidents}{7}$

708		
708	2.29 driver	$\sqrt{2.29} = 1.51$ .
	$\hat{\lambda} = \bar{y} = \frac{1623 \ accidents}{708 \ drivers} = $	SD of the 708 y's :1.86;

<sup>\$708 \*</sup> 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 of dri not occur completely at random;

#### Are "accidents" distributed "randomly" over bus drivers? Colton: Observed & "expected" numbers of accidents

during a 3-year period among 708 Northern Ireland Transport Authority bus drivers. [Table 2.5]

cidents  Expected (	ected*		71.5	164.0	187.9	143.6	82.3	37.7	14.4	4.7	1.4	0.3	0.1	0.0	708	
with y accidents		Observed Exp	117	157	158	115	78	44	21	7	9	-	က	-	208	its = 2.29
COLLOC ACCOST	o-year period	3	0	-	2	က	4	2	9	7	80	6	10	Ξ		$= \bar{v} = \frac{1623 \ accidents}{1}$

 $\hat{\lambda} = \bar{y} = \frac{1623 \ accidents}{708 \ drivers} = \frac{2.29}{6 \ driver}$ SD of the 708 y's:1.86;  $\sqrt{2.29} = 1.51$ .

"Comparison of observed and expected frequencies:

- More than the expected number of drivers with no accidents
- orivers with no accidents

   More than the expected number of drivers with five or more accidents

   These data suggest that the accidents

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

did not occur completely at random;

<sup>\$708 \*</sup> dpois(0:11,lambda=2.29)

#### Are "accidents" distributed "randomly" over bus drivers? Colton: Observed & "expected" numbers of accidents

during a 3-year period among 708 Northern Ireland Transport Authority bus drivers. [Table 2.5]

		its = 2.29	$\lambda = \overline{\nu} = 1623 \ accidents$
	708	208	
can le	0.0	-	=
and C	0.1	က	10
pom)	0.3	-	6
Ignor	1.4	9	80
Indica	4.7	7	7
In rac	14.4	21	9
1	37.7	44	2
ala no	82.3	78	4
seul	143.6	115	က
5	187.9	158	2
a kind	164.0	157	-
More	71.5	117	0
drive	Expected <sup>5</sup>	Observed	3
More	with y accidents	with y a	3-year period
edadiones industrial	Number of drivers	Number	No. accidents in

 $\hat{\lambda} = \bar{y} = \frac{1629}{708} \frac{\text{accidents}}{\text{drivers}} = \frac{2.29}{\text{driver}}.$ SD of the 708 y's :1.86;  $\sqrt{2.29} = 1.51.$ 

"Comparison of observed and expected frequencies:

- More than the expected number of drivers with no accidents
- 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;

at it appears that there is some

indication of accident proneness.

Ignoring this variation makes for (model-based) Standard Frors (SE's) and CI's that are too narrow, and that can lead to 'false positive' findings

<sup>\$708 \*</sup> dpois(0:11,lambda=2.29)

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

Number of drivers	with y accidents	Expected <sup>5</sup>	71.5	164.0	187.9	
Number	with y	Observed	117	157	158	1
No. accidents in	3-year period	3	0	-	2	

with y accidents	Expected\$	71.5	164.0	187.9	143.6	82.3	37.7	14.4	4.7	1.4	0.3	0.1	0.0
with y	Observed	117	157	158	115	78	44	21	7	9	-	ო	-
3-year period	3	0	-	2	က	4	2	9	7	80	6	10	Ξ

i	
÷	
Ш	
2.29	
N	
lمi	
~	
_	
'n	
æ	
Τ.	
'n	
٠	
$\overline{}$	
ω	
0	
708 y's :1.86	
Φ	
_	
-	
of the	
~	,

1623 accidents = 708 drivers =

<u>آ</u> ا

"Comparison of observed and expected frequencies: Colton:

More than the expected number of drivers with no accidents

More than the expected number of

These data suggest that the accidents did not occur completely at random; drivers with five or more accidents

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 •

makes for more realistic SE's and CI's. 'random-intercept' regression models http://www.epi.mcgill.ca/hanley/bios601/ Intensity-Rate/AccidentsBusDrivers.pdf Encoding this (not identifiable) variation in 'random effects' or •

708

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

# DAYLIGHT SAVINGS TIME AND TRAFFIC ACCIDENTS, NEJM 1996.04,04

#### 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 sleeprelated accidents exceeded \$56 billion and included 24,318 deaths and 2,474,430 disabling injuries.<sup>1</sup> 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 sleepdeprived and that small additional losses of sleep may have consequences for public and individual safety.2

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

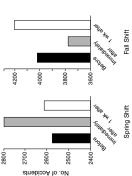
NEJM199604043341416

《冊》《冊》《**門**》《**日**》

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.

Canada as they were reported to the Canadian Ministry of day preceding the week of the change due to daylight savings cause it does not observe daylight savings time. The analysis We used data from a tabulation of all traffic accidents in 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 Monweek after the change, for both spring and fall time shifts. Data from the province of Saskatchewan were excluded beof the spring shift included 9593 accidents and that of the fall time, the Monday immediately after, and the Monday one shift 12,010. The resulting data are shown in Figure 1.





Spring Shrift
Figure 1. Numbers of Traffic Accidents on the Mondays before
and after the Shrifts to and from Daylight Savings Time for the
Years 1999! 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).

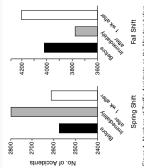


Figure 1. Numbers of Traffic Accidents on the Mondays before and after the Shifts to and from Daylaght Savings Time for the Years 1991 and 1992.

There is an increase in accidents after the spring shift (when an

time shift seems to decrease the risk of accidents. 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)

Monday immediately after the shift showed a relative risk of .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 1 sk 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 con-P<0.001) when Thus, the spring hour of sleep, resulted in an average increase in traffic acci dents of approximately 8 percent, whereas the fall shift result to daylight savings time increased the risk of accidents. shift to daylight savings time, and the concomitant fidence interval, 0.858 to 0.937;  $\chi^2 = 23.69$ ; compared with the Monday one week later.

ed in a decrease in accidence of approximately the same magniande immediately after the time shift.

These data slow that small changes in the amount of sleep that people get can have migner consequences in everytha sactivities. The loss of merely one hour of sleep can intercase the rivines. The loss of merely one hour of sleep can intercase the rivines. The state that an a nonspecific disruption in circadian sleep loss rather than a nonspecific disruption in circadian sleep loss rather than a nonspecific disruption in circadian tylythm, since gaining an additional hour of sleep at the fall Vancouver, BC V6T 1Z4, STANLEY COREN, PH.D. Canada University of British Columbia

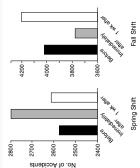


Figure 1. Numbers of Traffic Accidents on the Mondays before and after the Shifts to and from Daylght Savings Time for the Years 1991 and 1992.

There is an increase in accident safter the spirit When an hour of sleep is lost) and edecrease in the fall (when an hour of sleep is lost) and a decrease in the fall (when an hour

of sleep is gained)

11986 (95) percut confidence interval, (1999 to  $1137 \times 7=301$ , coll. Percut confidence interval, 1999 to  $1137 \times 7=301$ , the reducer Table when Monday immediately after the shift was 1100 (95) percent confidence interval, 1015 to  $1120 \times 7=61$  g. 1100 (95) percent confidence interval, 1016 to  $1120 \times 7=61$  g. 1017 (95) percent confidence interval, 1015 to  $1120 \times 7=61$  g. 1017 (95) percent confidence interval, 1035 to  $1120 \times 7=61$  g. 1017 (95) percent confidence interval, 1039 to 0390  $\times 7=92$ , 1 off. Pe-Only when compared with the preceding Monday and to 808 (95) percent confidence interval, 1038 to 0390,  $\times 7=25$  g. Pe-Only when compared with the Monday one week four. Thus, the spring fine, and the concominant loss of one high exclusion time, and the concominant loss of one high exclusion time, and the concominant loss of one high exclusion time, and the concominant loss of one high exclusion time, and the concominant loss of one high exclusion time, and the concominant loss of one had easily exclusive whereas in reliable and

Monday immediately after the shift showed a relative risk of

to daylight savings time increased the risk of accidents.

ed in a decrease in accelerate of approximately the same magnitude inmediately after the time shift.

There data show the small changes in the mount of short that people each and the mount of short that proper in curreda, are taking. The loss of merely one floating that the decrease the risk of traffic accelerate. It is likely that the effects are the to size post rather than a nonspecific disruption in circulant thythm, since gaining an additional hour of sleep at the fall time shift seems to decrease the risk of accidents. Vancouver, BC V6T 1Z4, STANLEY COREN, PH.D. Canada University of British Columbia

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  $\chi^2$  statistic assumes that the 2 before' counts are like 2 Geiger counts Vincent http://www.epi.mcgill.ca/hanlev/bios601/Intensity-Rate/Sleep-Vincent.pdf The generated by a uniform source, and ignores all outside influences.]

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

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

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

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

- [Yearly variations in numbers of plane crashes may be closer to Poisson -apart from some extra variation Yearly variations in numbers of persons killed in plane crashes 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]

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

- [Yearly variations in numbers of plane crashes may be closer to Poisson -apart from some extra variation Yearly variations in numbers of persons killed in plane crashes 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

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

- [Yearly variations in numbers of plane crashes may be closer to Poisson -apart from some extra variation Yearly variations in numbers of persons killed in plane crashes 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
- http://www.epi.mcgill.ca/hanley/c609/Material/LidkopingALL.pdf Yearly numbers/incidence of hospitalized injuries in a region https://www.ncbi.nlm.nih.gov/pubmed/21059188

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

- [Yearly variations in numbers of plane crashes may be closer to Poisson -apart from some extra variation Yearly variations in numbers of persons killed in plane crashes 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
- Yearly numbers/incidence of hospitalized injuries in a region https://www.ncbi.nlm.nih.gov/pubmed/21059188
- 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?

- https://www.nature.com/articles/d41586-018-05914-3 and here https://gatesopenresearch.org/articles/2-36/v1 Yearly Numbers of Dengue Fever Cases ÷
  - Daily numbers of Sudden Infant Deaths? αi

https://www.ncbi.nlm.nih.gov/pubmed/21059188

- http://www.epi.mcgill.ca/hanley/c609/Material/LidkopingALL.pdf Yearly numbers/incidence of hospitalized injuries in a region?
- Yearly Accidents, Fatalities, and Rates, 1982 2000, U.S. Air Carriers http://www.epi.mcgill.ca/hanley/c626/airline-data-sas.txt 4
- Quarterly & Monthly (prevalence) rates of Spina Bifida and Anencephaly Among http://www.epi.mcgill.ca//hanley/c626/folic\_acid.pdf. See more data on webpage Births (in relation to fortification of Foods with Folic Acid) 5
  - (Yearly) fatal and nonfatal crash rates on a toll highway (following a 5-15 mph http://www.epi.mcgill.ca//hanley/c626/. (8-24 kph) decrease in speed limits) 9
- Daily numbers of in-hospital deaths and Daily Maximal Temperatures https://www.ncbi.nlm.nih.gov/pubmed/1251837
- 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/Heatwave\_death\_lyon.pdf œ
- Daily no.s (Postponement of Death Until Symbolically Meaningful Occasions) http://www.epi.mcgill.ca/hanley/c626/fullmoon.pdf http://www.epi.mcgill.ca/hanley/c626/holidays.pdf 6
  - ${\tt http://www.epi.mcqill.ca/hanlev/c626/measuze\_of\_fidget\_galton.pdf} \stackrel{\leftarrow}{\equiv} {\tt http://www.epi.mcd.ca/hanlev/c626/measuze\_of\_fidget\_galton.pdf} \stackrel{\leftarrow}{\equiv} {\tt http://www.epi.mcd.ca/hanlev/c626/$ Rates of audience fidget. (F Galton) 9

#### 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 Cl'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