

In-class exercise - Inference for rates.

EPIB607 - Inferential Statistics^a

^aFall 2019, McGill University

This version was compiled on November 11, 2019

In this exercise you will practice doing inference for one-sample rates.

Sampling distribution | Standard error | Normal distribution | Quantiles | Percentiles | Z-scores

y	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
6	2.20	13.06	2.61	11.84	3.15	10.53
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
30	20.24	42.83	21.59	40.69	23.23	38.32
32	21.89	45.17	23.3	42.98	25	40.54
33	22.72	46.34	24.15	44.13	25.89	41.65
36	25.21	49.84	26.73	47.54	28.56	44.98

Table 1. Poisson based Confidence Intervals

1. Seismicity before and after hydraulic fracturing in the Horn River Basin, northeast British Columbia

Consult the article *Investigation of regional seismicity before and after hydraulic fracturing in the Horn River Basin, northeast British Columbia* by Farahbod et al. (2014).

- Using just the year 2006 data in the first four columns of Table 2, calculate separate event rates for the periods when hydraulic fracturing (i) was and (ii) was not in operation. Express each rate as the number of events per year, and accompany it with a 95% CI.
- Repeat the calculations for the individual years 2007 to 2011.

- c) Repeat the calculations for the entire period 2006 to 2011.
- d) Store your results in a tidy `data.frame`.
- e) Are you comfortable using Poisson-based CIs for the entire period 2006 to 2011? i.e. does it look like the rate in non-HF days is homogeneous over the 6 years, i.e., do the 6 CIs largely overlap each other?
- f) Does it look like the rate in HF days is homogeneous over the 6 years, i.e., do the 6 CIs largely overlap each other?

2. De-novo Mutation Rates

The [US NCI Dictionary of Genetics Terms](#) defines a *de novo* mutation as

a genetic alteration that is present for the first time in one family member as a result of a variant (or mutation) in a germ cell (egg or sperm) of one of the parents, or a variant that arises in the fertilized egg itself during early embryogenesis.

These point mutations differ from the chromosome-level errors that lead to Down Syndrome, and so the influences of the two parental ages might also be expected to differ.

The article *Rate of de novo mutations and the importance of father's age to disease risk* is a good introduction to the fast-moving research on this. The written news item ([Nature News, 2012](#)), and the audio interview with the senior author ([Stefansson, 2012](#)) provide a good lay introduction. That study involved genetic material from 78 father-mother-offspring trios. The focus was on single nucleotide polymorphism mutations and limited to the autosomal chromosomes 1-22, thereby excluding the sex chromosome 23. Using several rigorous criteria, a total of 4,933 *de novo* mutations, or an average of 63.2 per trio, were called.

- a) In five trios, a child of the offspring was also sequenced, thereby allowing the investigators to determine the parent of origin of each of the 348 *de novo* mutations called. Using Table 1, calculate separate mutation rates for mutations from mothers and from fathers. Express each rate as the number of mutations per parent-year-lived, and accompany it with a 95% CI.

3. Vacancy rates in the US Supreme Court

Refer to the article *Updating a Classic: The Poisson Distribution and the Supreme Court Revisited*.

- a) Update Table 1 and calculate an updated vacancy rate (expressed as vacancies per year) for the period 1933-2019.
- b) Based only on your updated point-estimate of the rate [i.e. without using any information on the health of the current court], what is the probability that the next US president (the one who takes office in 2021) will be able to appoint 0, 1, 2 ... new judges if (s)he stays in office for (i) four (ii) 8 years.