

Håkon K. Gjessing

Professor/Principal Investigator

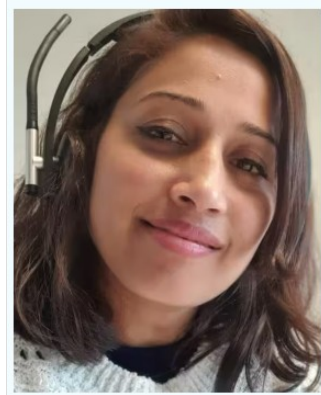
⇒ Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo
– a Norwegian Research Council Centre of Excellence

⇒ Department of Global Public Health and Primary Care, University of Bergen



Kathmandu
Tuesday 6 October 2026





- **Gerhard Sulo**
Associate Professor, IGS UiB
- **Sarala Banjara**
PhD candidate, IGS UiB

Some illustrative analyses by Håkon

Data: Oral health on 1298 patients **Variables:**

- DMFT: Number of teeth Decayed, Missing, or Filled per patient (possible range 0–28)
- Number of retained teeth per patient (possible range 0–28)
- Narrow age range
- Information on income, education etc.
- No information on side, upper/lower jaw, only the total per patient

Tooth decay

dmft = Number of Decayed, Missing, or Filled Teeth

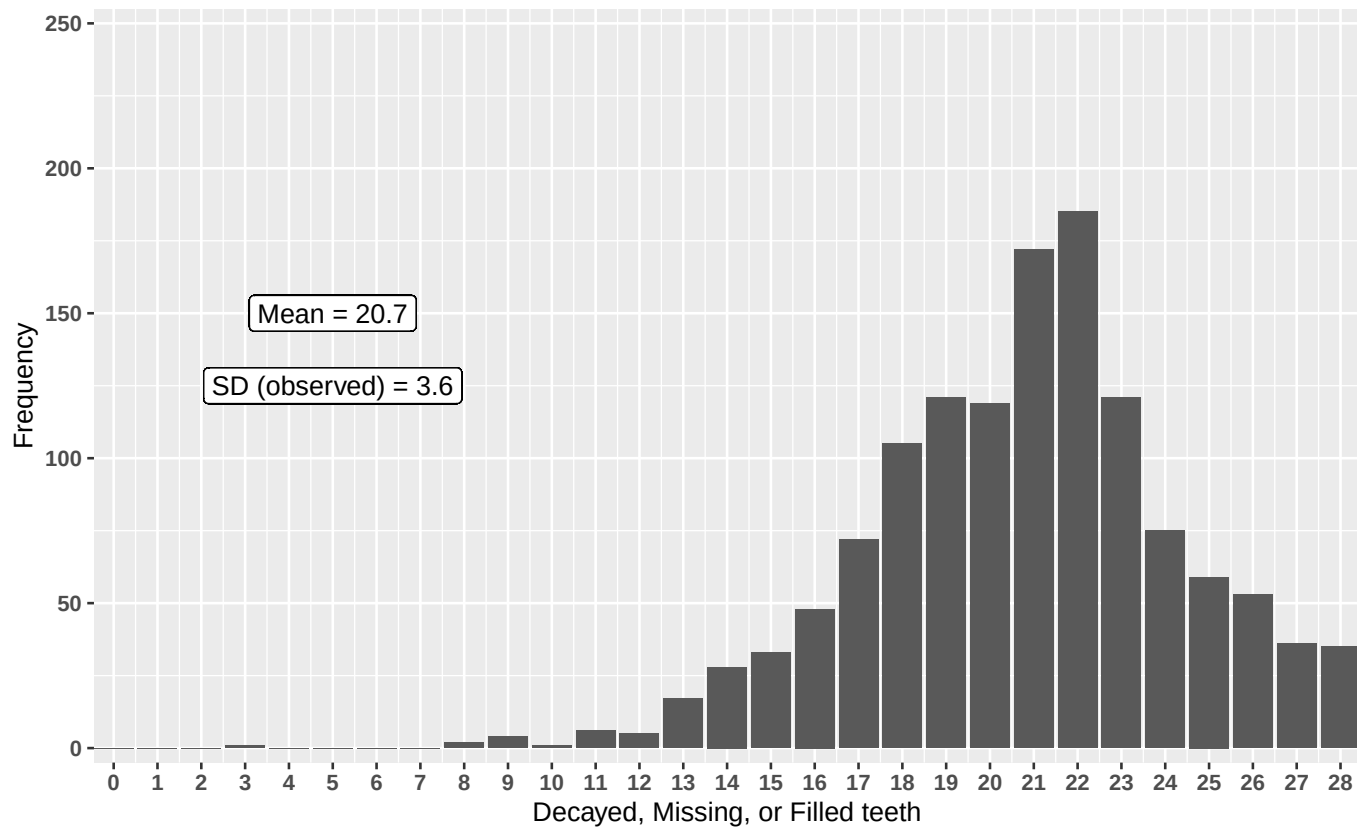
freq = Number of patients with that dmft value

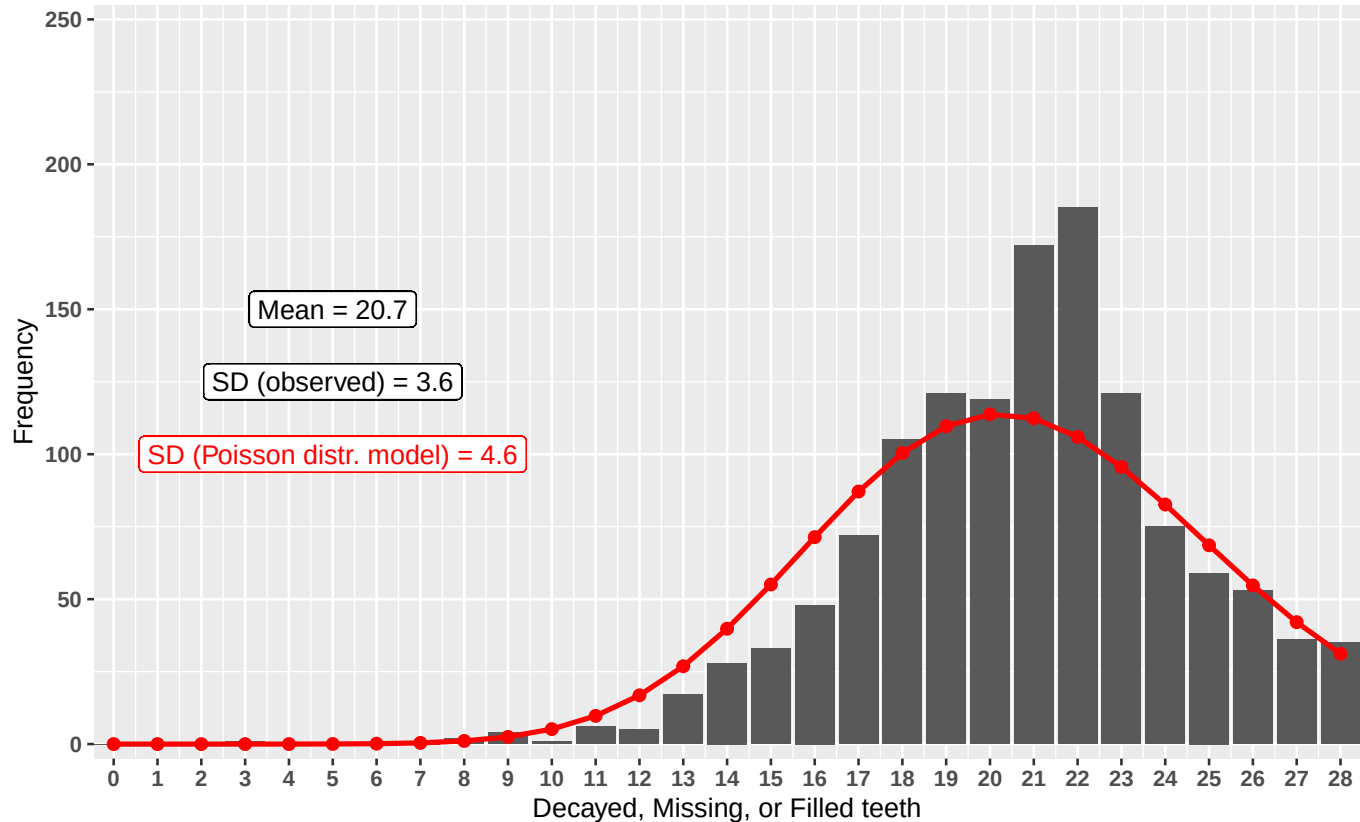
dmft	3	8	9	10	11	12	13	14	15	16
freq	1	2	4	1	6	5	17	28	33	48

dmft	17	18	19	20	21	22	23	24
freq	72	105	121	119	172	185	121	75

dmft	25	26	27	28
freq	59	53	36	35

- A total of 1298 patients





- Fit a Poisson distribution with mean equal to the observed mean

Poisson distribution

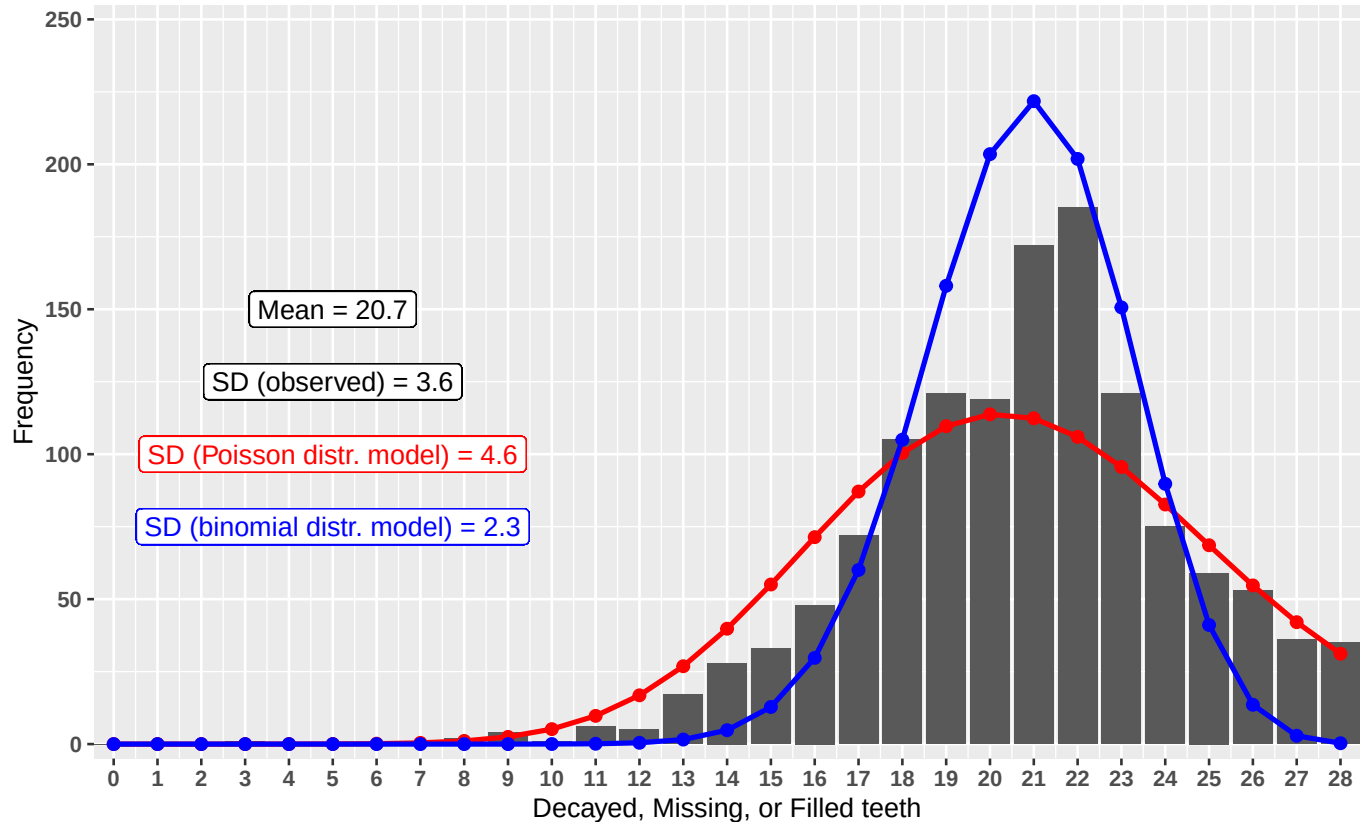
$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, 3, \dots, \quad \lambda \geq 0.$$

$$\text{mean} = E[X] = \lambda$$

$$\text{Variance} = \text{var}[X] = \lambda$$

$$\text{Std. dev.} = \text{SD}[X] = \sqrt{\lambda}$$

- Compared to a Poisson distribution, **overdispersion** means observed variance $> \lambda = E[X]$
- In the present distribution, there is actually **underdispersion**



- Fit a Binomial distribution with mean equal to the observed mean

Binomial distribution

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, 2, 3, \dots, n, \quad 0 \leq p \leq 1.$$

$$\text{mean} = E[X] = np$$

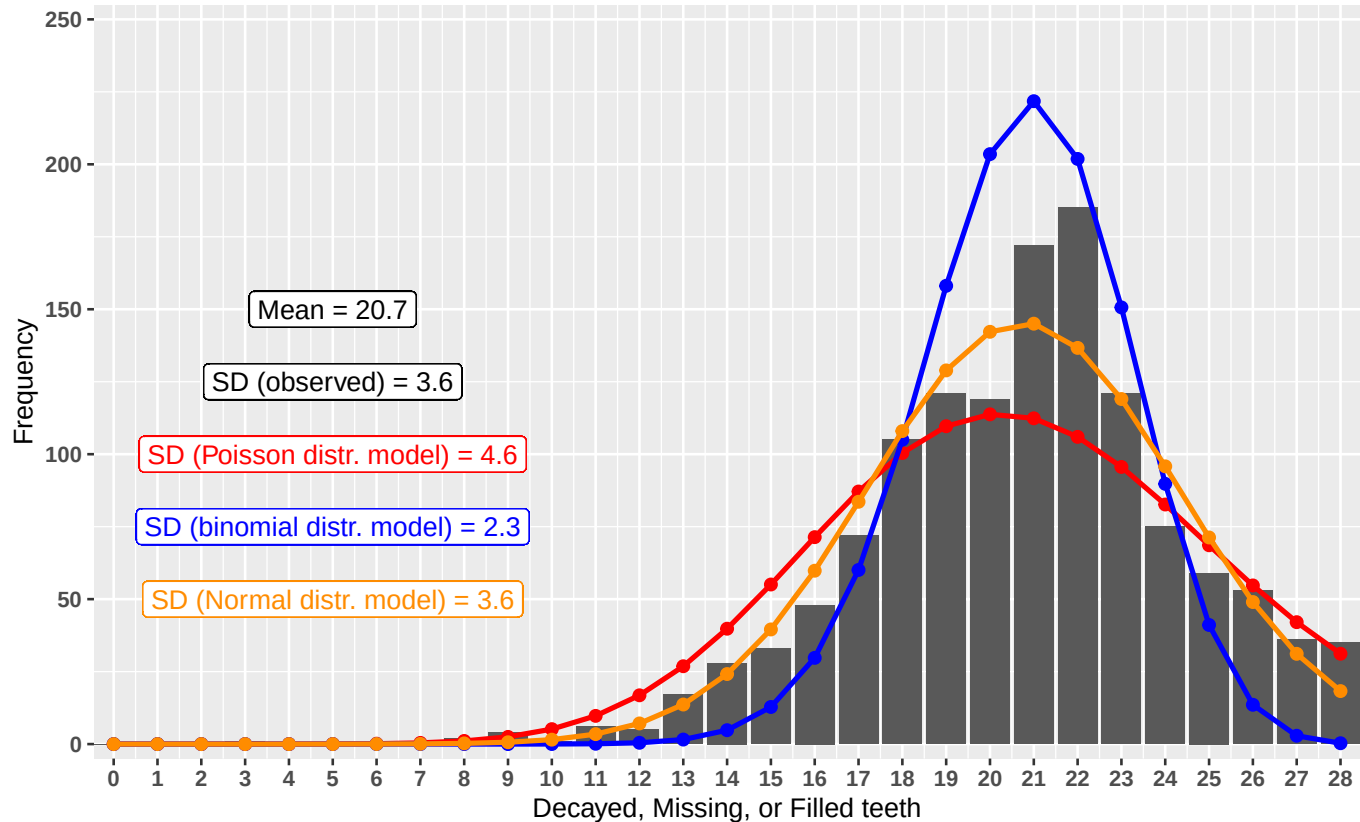
$$\text{Variance} = \text{var}[X] = np(1 - p)$$

$$\text{Std. dev.} = \text{SD}[X] = \sqrt{np(1 - p)}$$

NOTE: $n = 28$ and $p = \frac{20.7}{28} \approx 0.74$ in this case

- Compared to a Binomial distribution, **overdispersion** means observed variance $> np(1 - p)$
- In the present distribution, there is **overdispersion**

NOTE: **Bernoulli variable** is 0/1, and binomial is a sum of independent Bernoulli variables. The independence assumption is the problem here.



- Fit a Normal distribution with mean equal to the observed mean

Normal (Gaussian) distribution

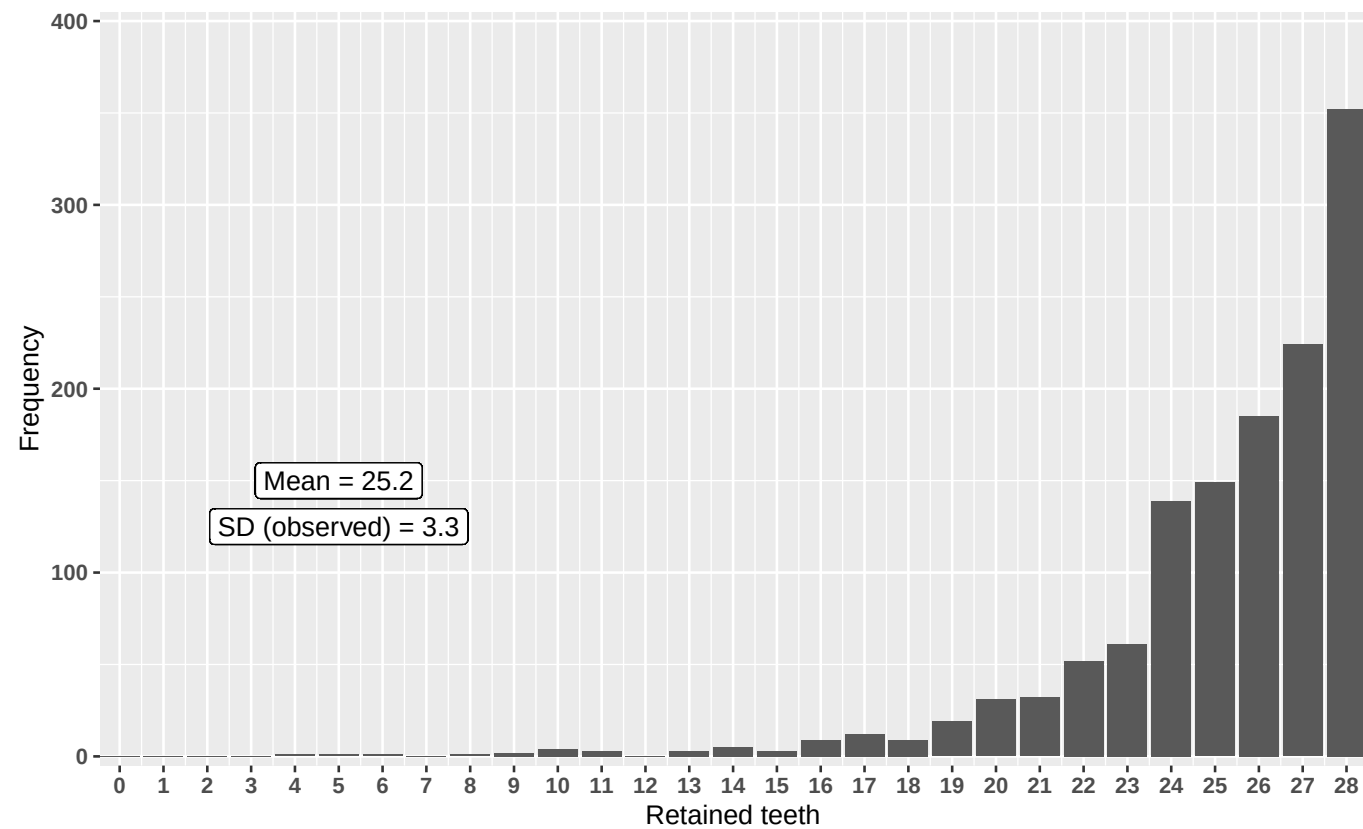
$$P(X \in dx) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) dx$$

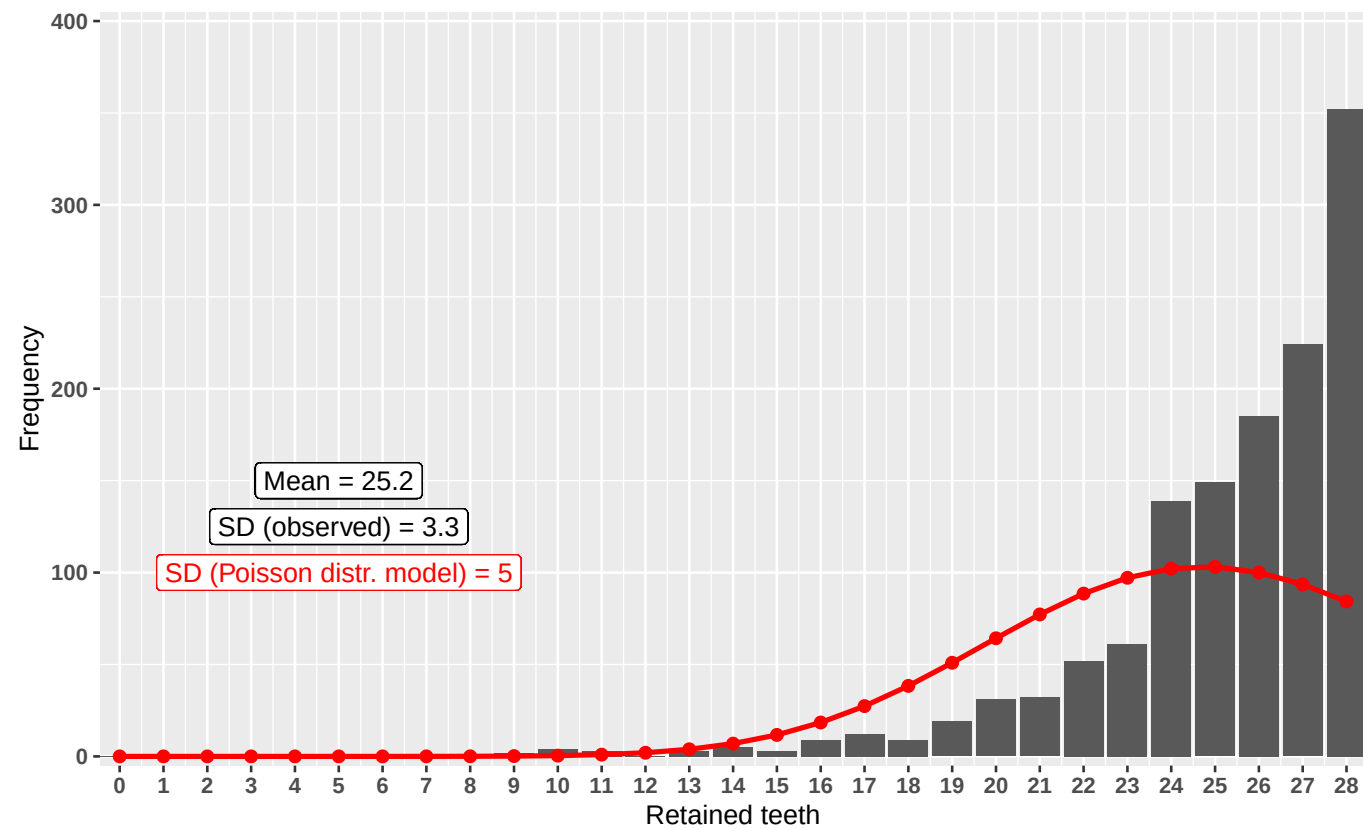
$$\text{mean} = E[X] = \mu$$

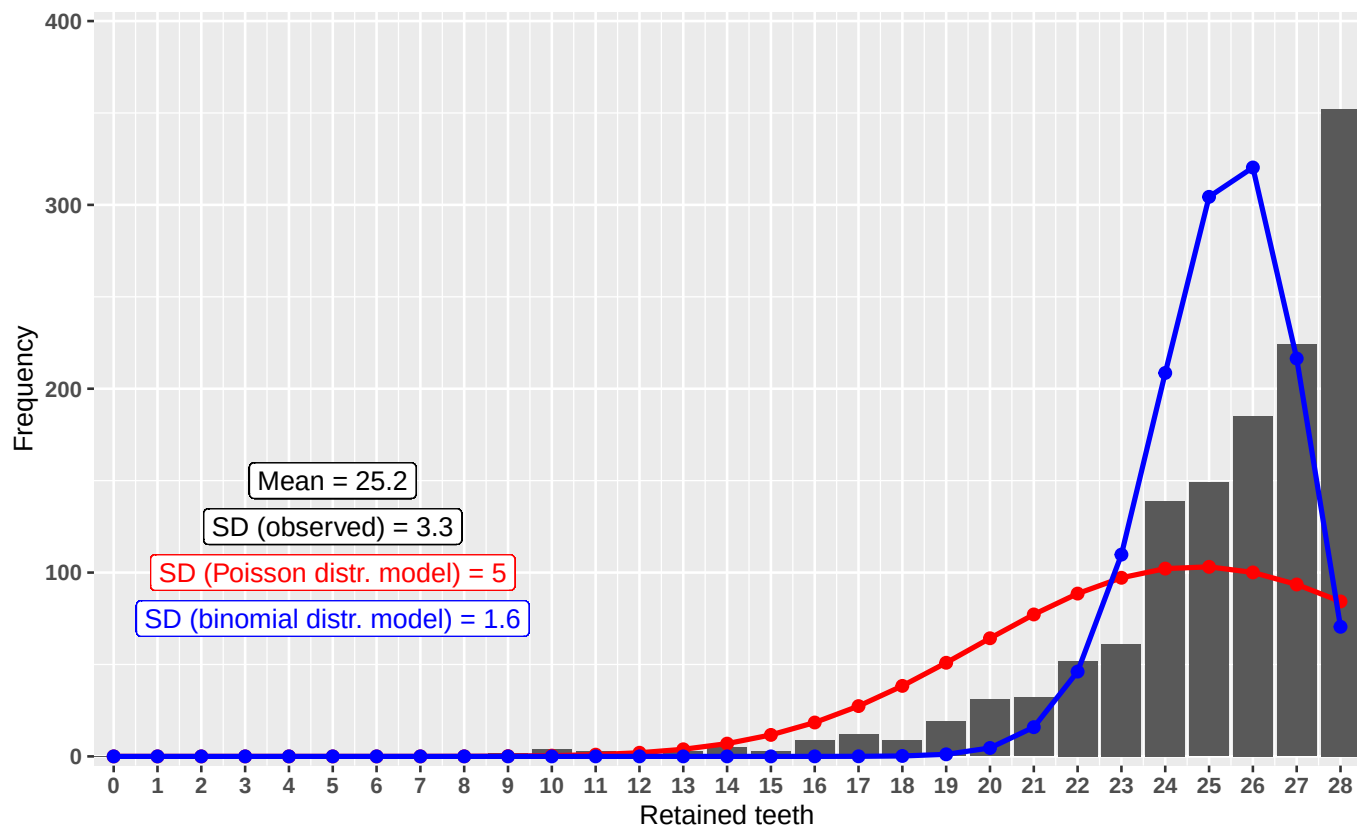
$$\text{Variance} = \text{var}[X] = \sigma^2$$

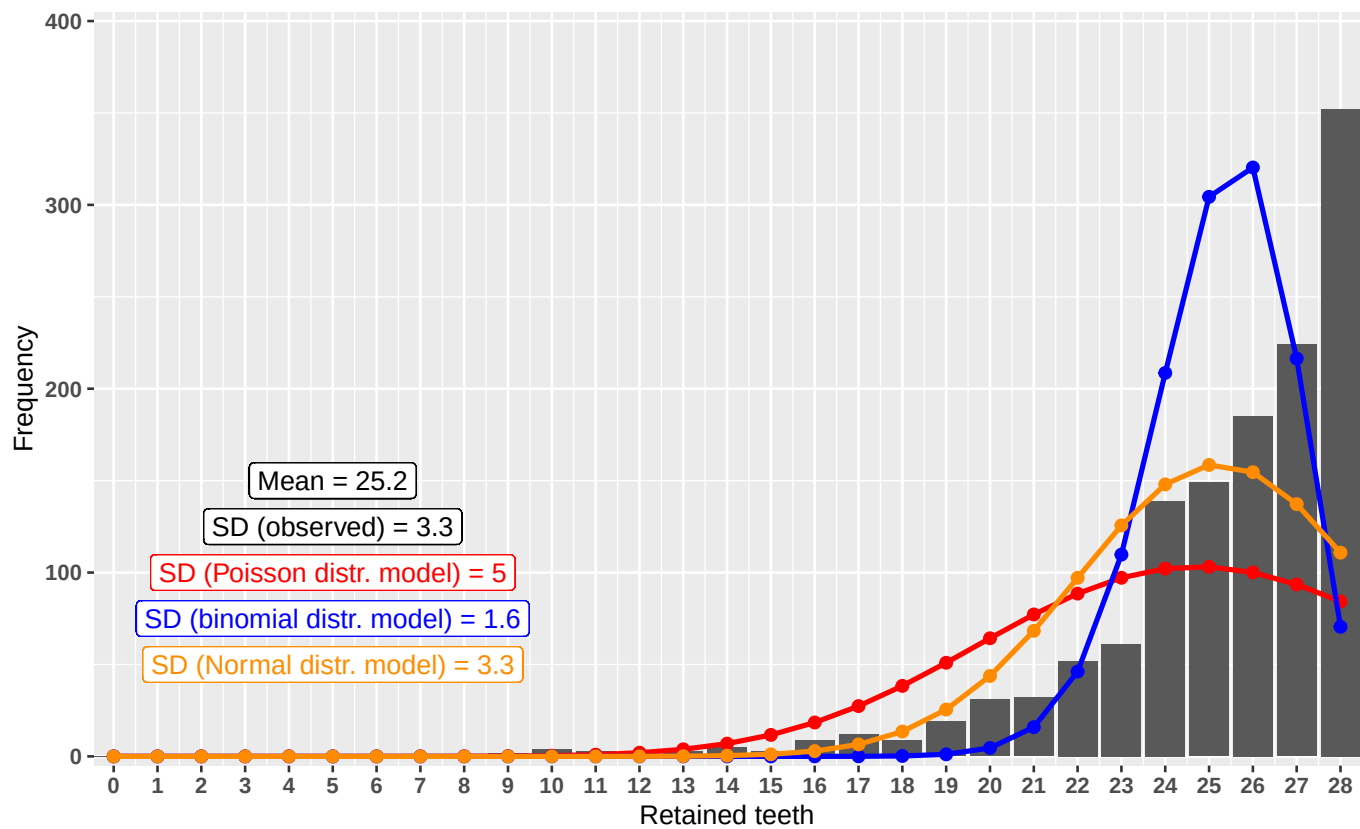
$$\text{Std. dev.} = \text{SD}[X] = \sigma$$

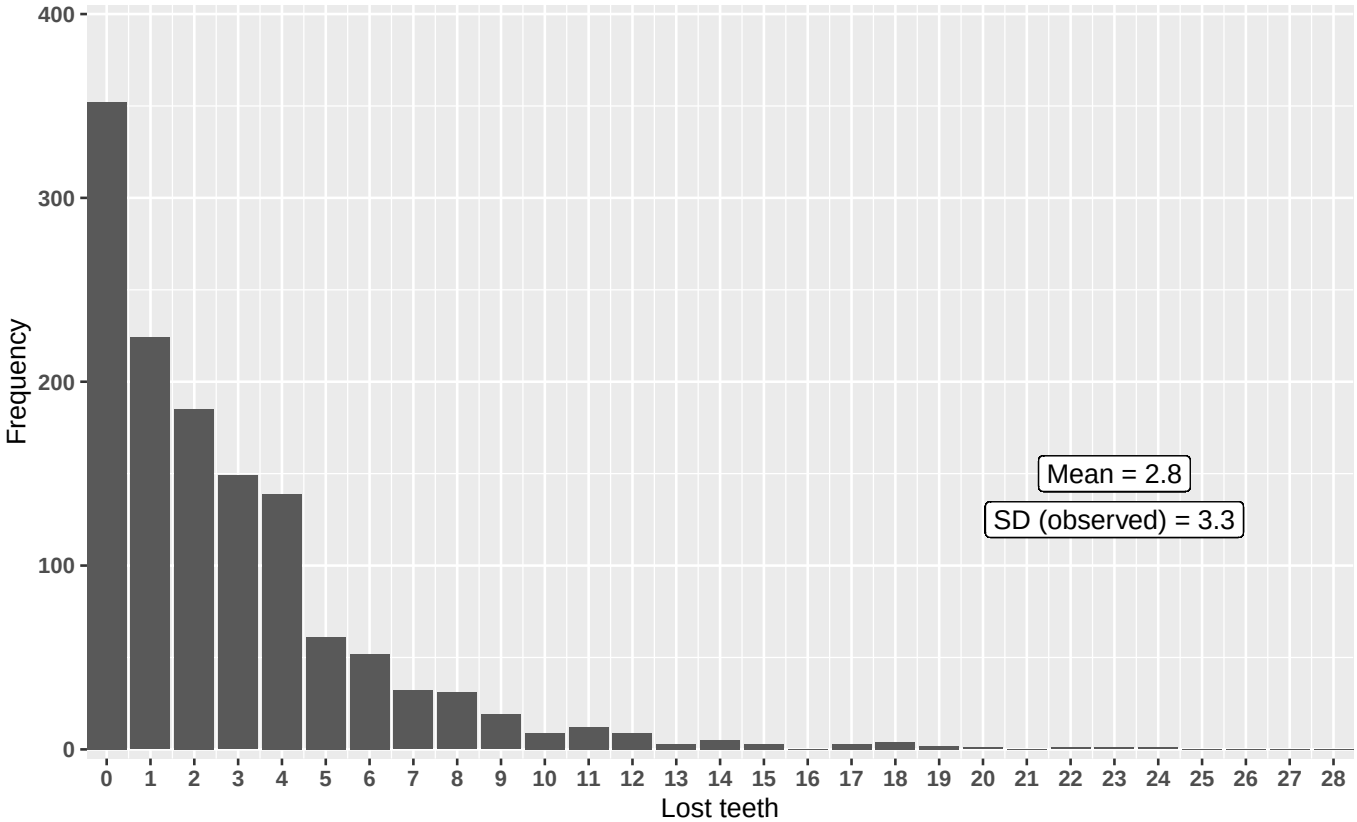
- Compared to a Normal distribution, **overdispersion** would mean observed variance $> \sigma^2$
- but since there is a separate parameter for the standard deviation, it avoids overdispersion

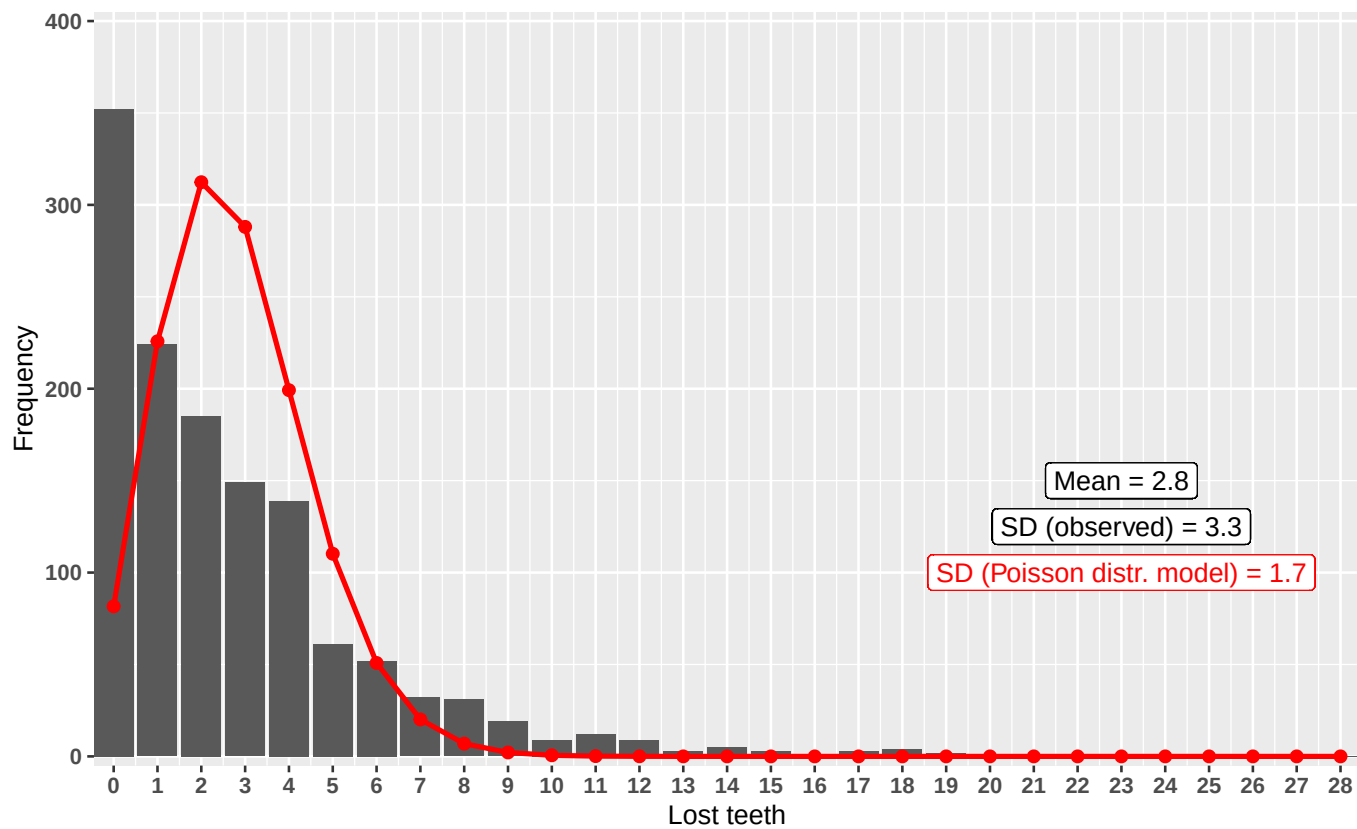


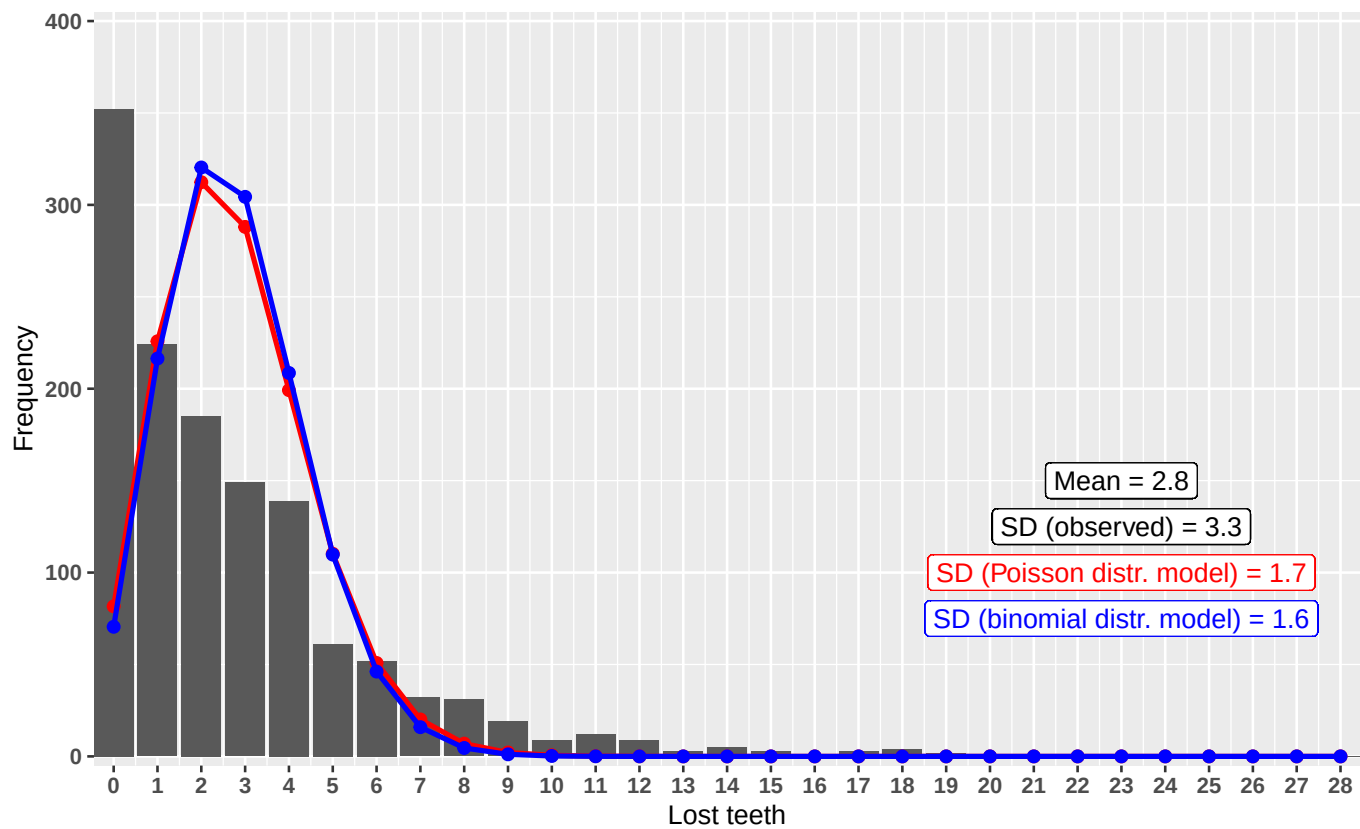


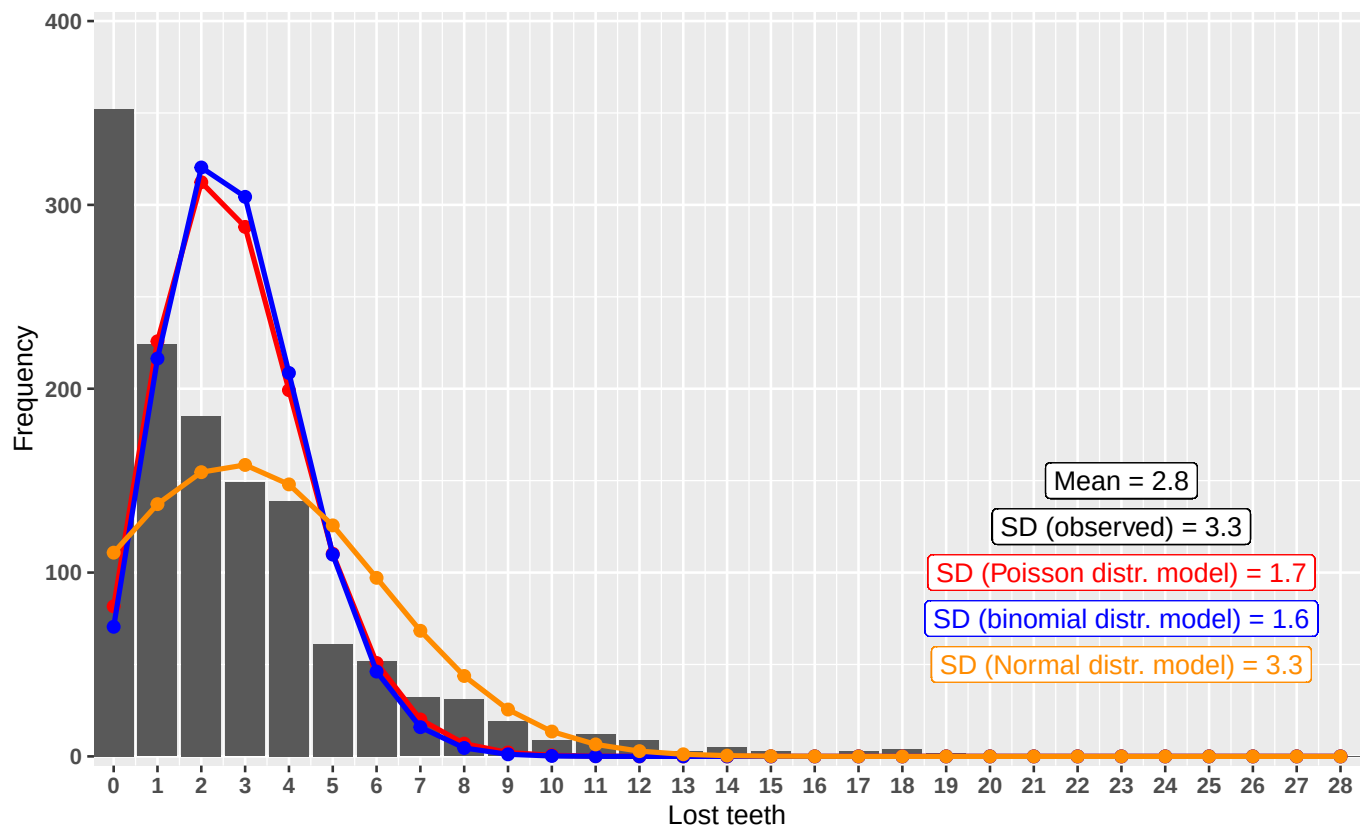












Binomial variable with random p

- Each patient has an underlying, unobserved risk p of DMFT
- We assume the logit risk follows a Normal distribution,

$$\log\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + b_{i0}$$

where b_{i0} is a “random intercept” with a Normal distribution

- Conditional on p_i (i.e. for a given patient), DMFT follows a binomial distribution

$$X \mid p_i \sim \text{Bin}(n, p_i)$$

- The random p provides more variability to the distribution of X
- Each patient has his/her own DMFT risk
- Covers overdispersion
- Known as a **random intercept model**, a special case of mixed effects models

