

# Homework 1

---

In an experiment a marker was transfected into a population of  $n = 3 \times 10^5$  cells. The marker functionally integrates into the genome at a rate of 1 in  $10^5$ . What is the probability of obtaining at least one marked cell after this procedure?

- $n = 3 \times 10^5$  cells, integration rate of  $r = 10^{-5}$
- $P(X \geq 1) = ?$
  
- $P(X \geq 1) = P(X = 1) + P(X = 2) + P(X = 3) + \dots$
- $P(X \geq 1) = 1 - P(X = 0)$

# Homework 1

## Binomial distribution

- Series of 'trials', for each cell
- $n = 3 \times 10^5$ ,  $p = r = 10^{-5}$

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

$$\begin{aligned} P(X = 0) &= \binom{3 \times 10^5}{0} (10^{-5})^0 \\ &\quad \times (1 - 10^{-5})^{3 \times 10^5} \\ &= 0.99999^{3 \times 10^5} \approx 0.05 \end{aligned}$$

$$P(X \geq 1) = 1 - 0.05 = 0.95$$

## Poisson distribution

- Rare, independent transfection events
- $\mu = nr = 3$

$$P(X = k) = \frac{\mu^k e^{-\mu}}{k!}$$

$$P(X = 0) = \frac{3^0 e^{-3}}{0!} \approx \frac{1 \times 0.05}{1} = 0.05$$

$$P(X \geq 1) = 1 - 0.05 = 0.95$$

# Homework 2

---

Consider a population of mice with normally distributed body weight of  $\mu = 20$  g and  $\sigma = 5$  g. The variance of the population is  $\sigma^2 = 25$  g<sup>2</sup>.

Using a computer program (R, Python, whatever!) make a simple simulation.

- Draw 100,000 samples of size  $n = 5$  from this population
- For each sample find two estimators of the variance,  $SD_n^2$  and  $SD_{n-1}^2$
- Plot the distributions of  $SD_n^2$  and  $SD_{n-1}^2$
- Find the *mean* of  $SD_n^2$  and  $SD_{n-1}^2$  across all samples
  
- Which variance estimator represents the true variance,  $\sigma^2$ , better?
- Repeat these calculations for the two standard deviations estimators

$$SD_n^2 = \frac{1}{n} \sum_i (x_i - M)^2$$

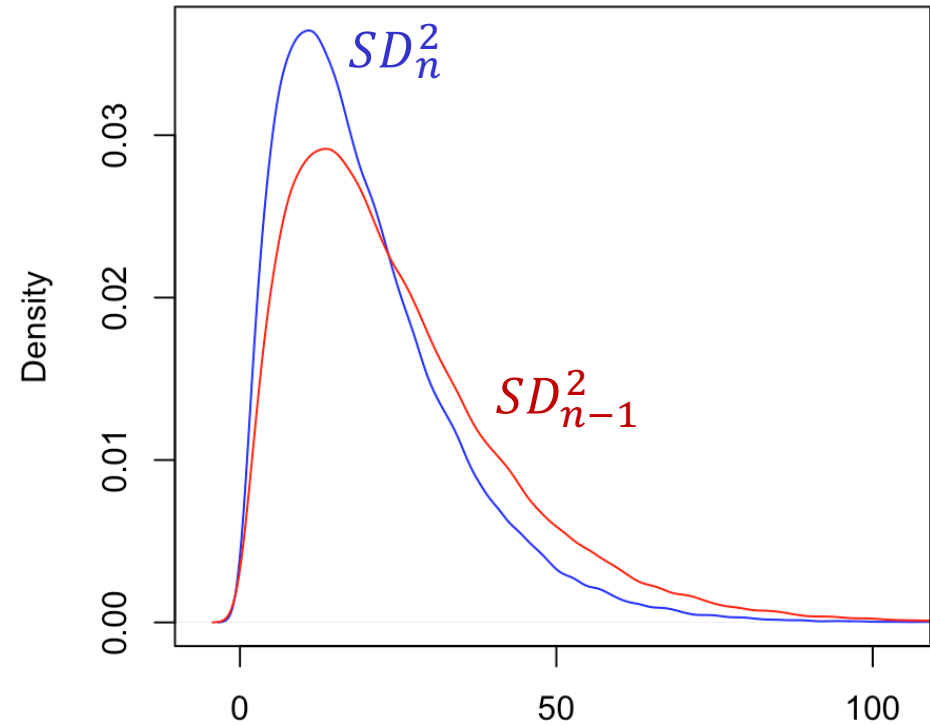
$$SD_{n-1}^2 = \frac{1}{n-1} \sum_i (x_i - M)^2$$

# Homework 2

```
nsim = 100000
n = 5
mu = 20
sigma = 5

Varn = c()
Varn1 = c()
for (i in 1:nsim) {
  x = rnorm(n, mean=mu, sd=sigma)
  M = mean(x)
  S2 = sum((x - M)^2)
  Varn[i] = S2 / n
  Varn1[i] = S2 / (n - 1)
}

mean(Varn)
## [1] 19.98505
mean(Varn1)
## [1] 24.98131
plot(density(Varn), col="blue")
lines(density(Varn1), col="red")
```



# Homework 2

## Variance

- We draw samples from a normal population with  $\sigma^2 = 25 \text{ g}^2$
- The mean values (over 100,000 simulated samples) of variance estimators are

$$\langle S_n^2 \rangle = 20.0 \text{ g}^2$$

$$\langle S_{n-1}^2 \rangle = 25.0 \text{ g}^2$$

- The “n” estimator is *biased*
- The “n – 1” estimator is *unbiased* and represents the true variance well

## Standard deviation

- True value is  $\sigma = 5 \text{ g}$
- This time we use standard deviations:

$$\langle S_n \rangle = 4.21 \text{ g}$$

$$\langle S_{n-1} \rangle = 4.71 \text{ g}$$

- Even the “n – 1” estimator of the standard deviation is biased!

# Homework 3

---

- In a test of a new drug we found the following half-maximal inhibitory concentrations ( $IC_{50}$ ):

No.	1	2	3	4	5	6	7	8	9	10	11	12
$IC_{50}$ (nM)	46	64	30	158	42	28	182	39	292	148	173	61

- Find the mean and the median with their corresponding 95% confidence intervals
- What can you say about how these data are distributed?

# Homework 3

No.	1	2	3	4	5	6	7	8	9	10	11	12
$IC_{50}$ (nM)	28	30	39	42	46	61	64	148	158	173	182	292

## Mean

$$M = 105.25 \text{ nM}$$

$$SD = 83.75 \text{ nM}$$

$$SE = 24.18 \text{ nM}$$

$$d.o.f. = 11$$

$$t^* = 2.201$$

$$CI = t^*SE = 53.21 \text{ nM}$$

$$M = 100 \pm 50 \text{ nM}$$

## Median

$$\tilde{M} = 62.5 \text{ nM}$$

$$L = [6] - [\sqrt{3}] = 4$$

$$U = 8$$

$$\widehat{SE} = \frac{x_{(8)} - x_{(5)}}{2} = \frac{148 - 46}{2} = 51 \text{ nM}$$

$$d.o.f. = 8 - 4 - 1 = 3$$

$$t^* = 3.182$$

$$\widehat{CI} = t^*\widehat{SE} = 162.3 \text{ nM}$$

$$\tilde{M} = 60 \pm 160 \text{ nM}$$

$$\tilde{M} = 60_{-20}^{+110} \text{ nM ("exact" estimator)}$$

# Homework 3

No.	1	2	3	4	5	6	7	8	9	10	11	12
$pIC_{50}$	6.53	6.74	6.76	6.80	6.83	7.19	7.21	7.34	7.38	7.41	7.52	7.55

## Mean

$$M = 7.106$$

$$SD = 0.352$$

$$SE = 0.101$$

$$d. o. f. = 11$$

$$t^* = 2.201$$

$$CI = t^* SE = 0.223$$

$$M = 7.1 \pm 0.2$$

$$pIC_{50} = -\log \frac{IC_{50}}{1 M}$$

## Median

$$\tilde{M} = 7.204$$

$$L = [6] - [\sqrt{3}] = 4$$

$$U = 8$$

$$\tilde{SE} = \frac{x_{(8)} - x_{(5)}}{2} = \frac{7.34 - 6.83}{2} = 0.254$$

$$d. o. f. = 8 - 4 - 1 = 3$$

$$t^* = 3.182$$

$$\tilde{CI} = t^* \tilde{SE} = 0.808$$

$$\tilde{M} = 7.2 \pm 0.8$$

$$\tilde{M} = 7.2_{-0.4}^{+0.2} \text{ ("exact" estimator)}$$



# Homework 4

---

In a study of a new antibiotic a sample of bacterial cells was treated with the drug, while the control sample remained untreated. Both treatment and control were done in three replicates. An aliquot of each replicate was placed in a counting chamber and living cells counted.

Replicate	1	2	3
Control	111	104	123
Treatment	16	18	14

1. Pool the counts together and find the proportion of cells surviving the treatment and its 95% confidence interval.
2. Do the same using replicated data. Compare the results.

# Homework 4

---

Replicate	1	2	3	Pooled
Control	111	104	123	338
Treatment	16	18	14	48

## Ratio

$$r = \frac{48}{338} \approx 0.14$$

$$S' = 49.9 \quad n' = 341.8$$

$$p' = \frac{49.9}{341.8} = 0.146$$

$$W = 0.0374$$

$$CI = [0.11, 0.18]$$

$$r = 0.15^{+0.03}_{-0.04}$$

**WRONG!**

## Proportion

$$\hat{p} = \frac{48}{338 + 48} \approx 0.12$$

$$S' = 49.9 \quad n' = 389.8$$

$$p' = \frac{49.9}{389.8} = 0.128$$

$$W = 0.0332$$

$$CI = [0.09, 0.16]$$

Error propagation,  $r = \hat{p}/(1 - \hat{p})$ :

$$r = 0.14^{+0.05}_{-0.04}$$

# Homework 4

---

Replicate	1	2	3
Control	111	104	123
Treatment	16	18	14

## Mean and CI

$$M_c = 113 \quad M_t = 16.0$$

$$SE_c = 5.55 \quad SE_t = 1.15$$

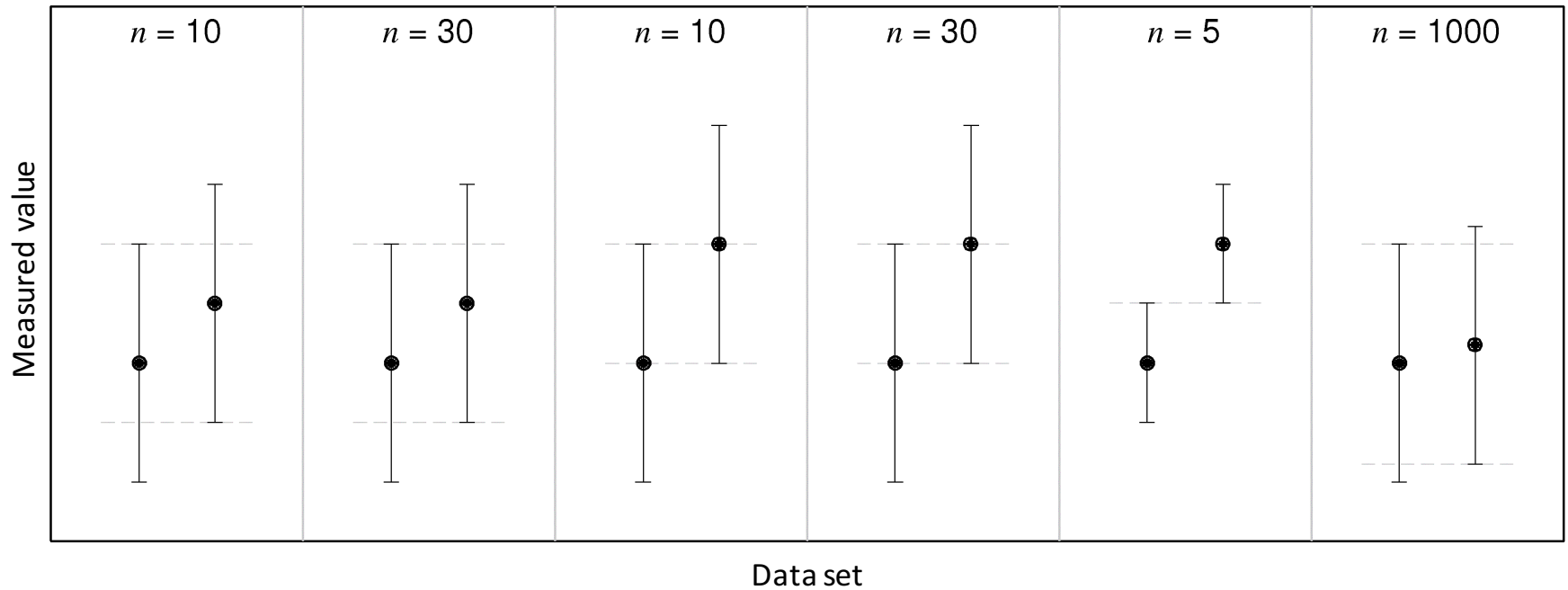
$$t^* = 4.303$$

$$\Delta M_c = 24 \quad \Delta M_t = 5.0$$

Error propagation,  $r = M_t/M_c$ :

$$r = 0.15 \pm 0.05$$

# Homework 5



- Each panel shows results from a pair of samples of the same size
- **Mean and standard deviation** are shown
- Consider a t-test between each pair of samples
- Guess p-values for all of them!
- Which of these pairs of means are significantly different?