Statistics

Combinatorics

Combinations: order doesn't matter

Permutations: order matters

Permutations

① Case 1: repeats are allowed

Example: how many possible lock "combinations" are there?

($n=10$, $r=3$)

$$\underline{10} \times \underline{10} \times \underline{10}$$

$$= n^r, \quad \begin{array}{l} n = \text{\# of possibilities for each event} \\ r = \text{\# of events} \end{array}$$

② Case 2: repeats not allowed

(2) case 2: repeats not allowed

Example: place pool balls in a certain order

(pool balls have 16 different #s)

$$\underline{16} \times \underline{15} \times \underline{14} \times \ldots$$

to place all items: $n! = 16 \cdot 15 \cdot 14 \cdot \ldots \cdot 3 \cdot 2 \cdot 1$

to place the first r items: $\dfrac{n!}{(n-r)!} = \dfrac{16 \cdot 15 \cdot 14 \cdot \ldots \cdot \cancel{3 \cdot 2 \cdot 1}}{\cancel{13 \cdot 12 \cdot 11 \cdot \ldots \cdot 3 \cdot 2 \cdot 1}}$

$= 16 \cdot 15 \cdot 14$

Combinations

(1) case 1: no repetitions allowed

Example: how many ways to draw r

pool balls from n balls?

$$\left(\frac{n!}{(n-r)!}\right) / r! = \frac{n!}{(n-r)!\,r!}$$

how many unique orderings of 3 balls?

10	8	7
10	7	8
7	10	8

...

3! ways to order the 3 balls.

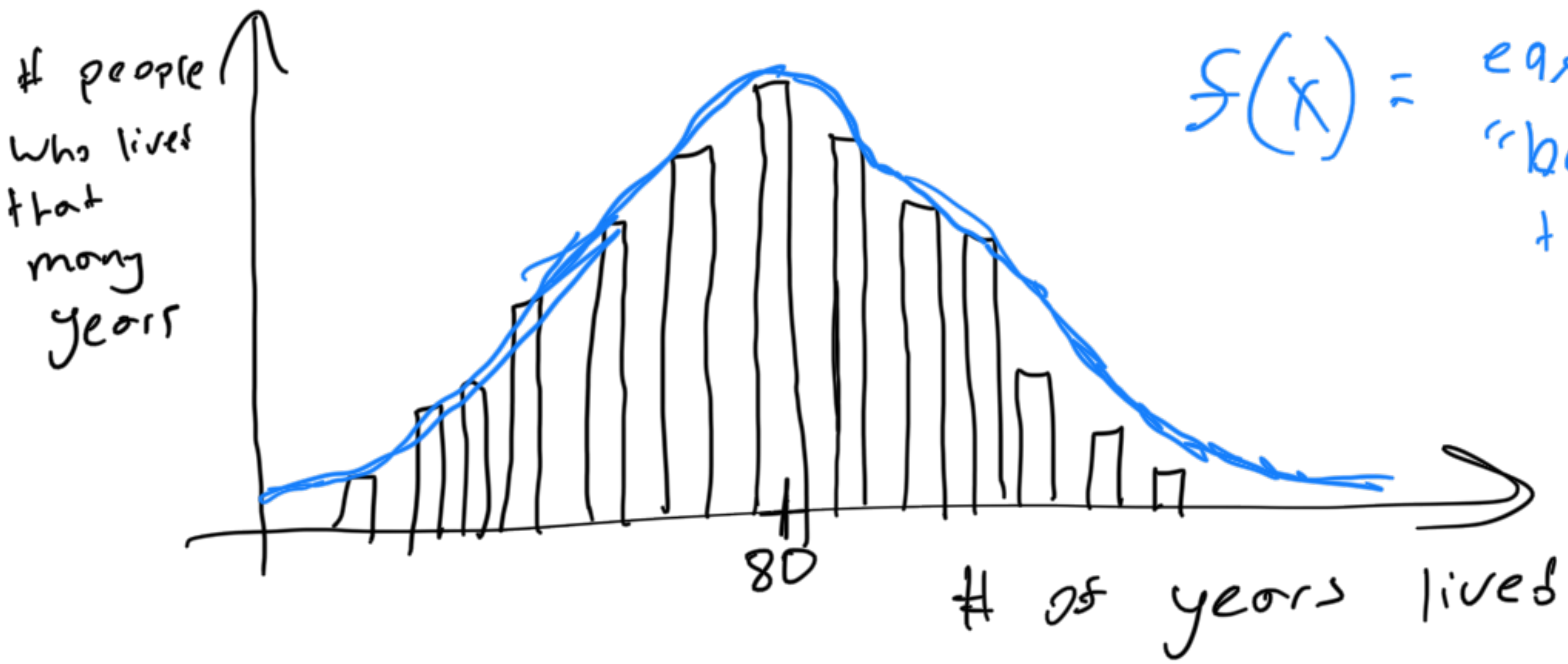
Statistics

Stats

Normal Distribution / "Bell Curve"

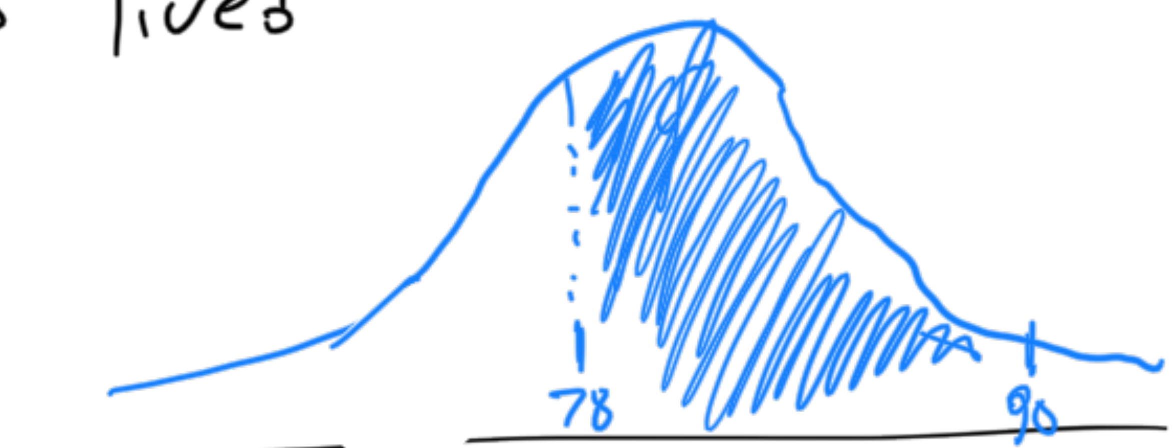
example of a continuous distribution

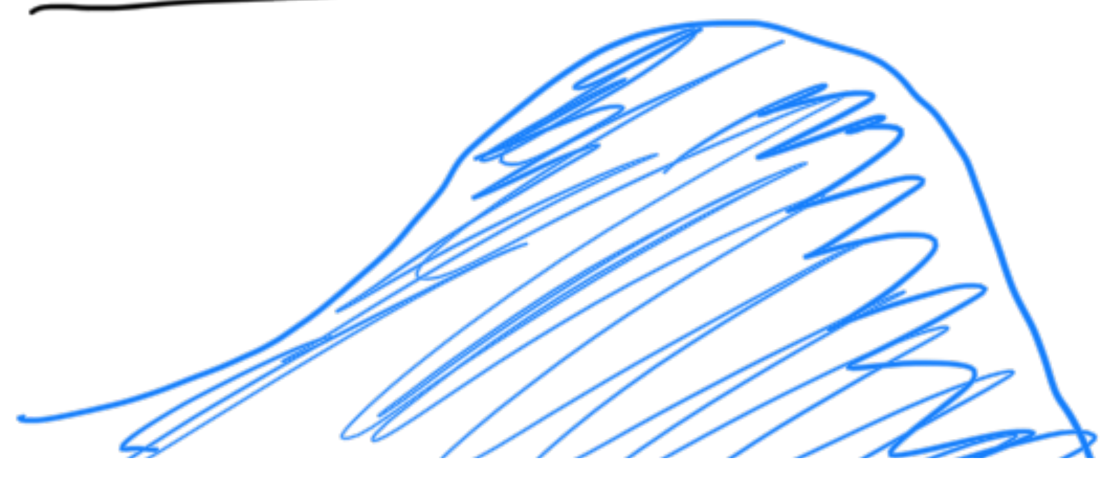
is one example of many probability distributions



$f(x)$ = equation of the continuous "bell curve" that fits the data

$P(78 < X < 90) =$





<90

$$= \int_{78} f(x)\,dx$$

Bernoulli Distribution

example of a <u>discrete</u> distribution

Warriors won over Rockets 70% of the time

In the finals, what is the probability that Warriors will win exactly 4 times?

Bernoulli tells us probability of "flipping a coin" with probability of heads = p exactly n times out of m

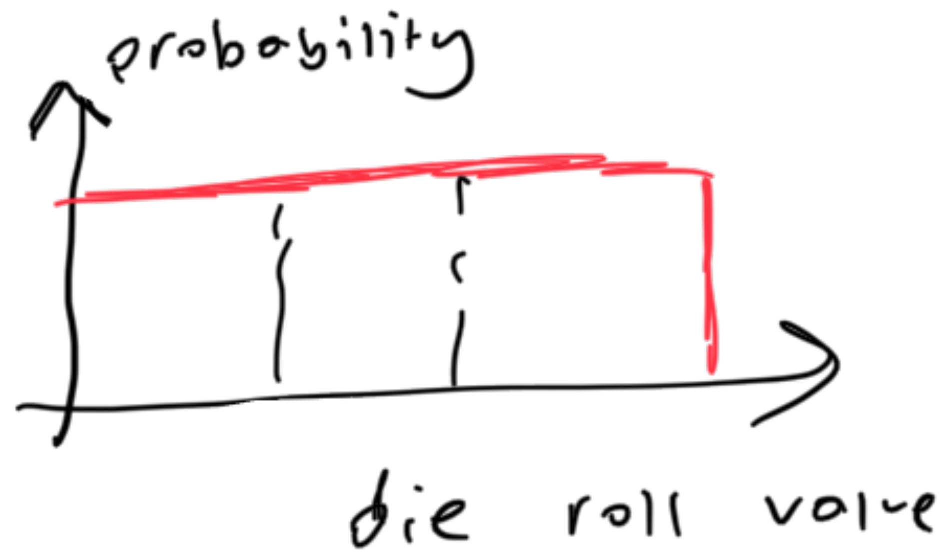
$$0.7 \times 0.7 \times 0.7 \times 0.7 \times 0.3 \times 0.3 \times 0.3$$

$$= 0.7^4 \times 0.3^3$$

$$= p^n (1-p)^{(m-n)}$$

Other popular distributions:

Uniform Distributions:



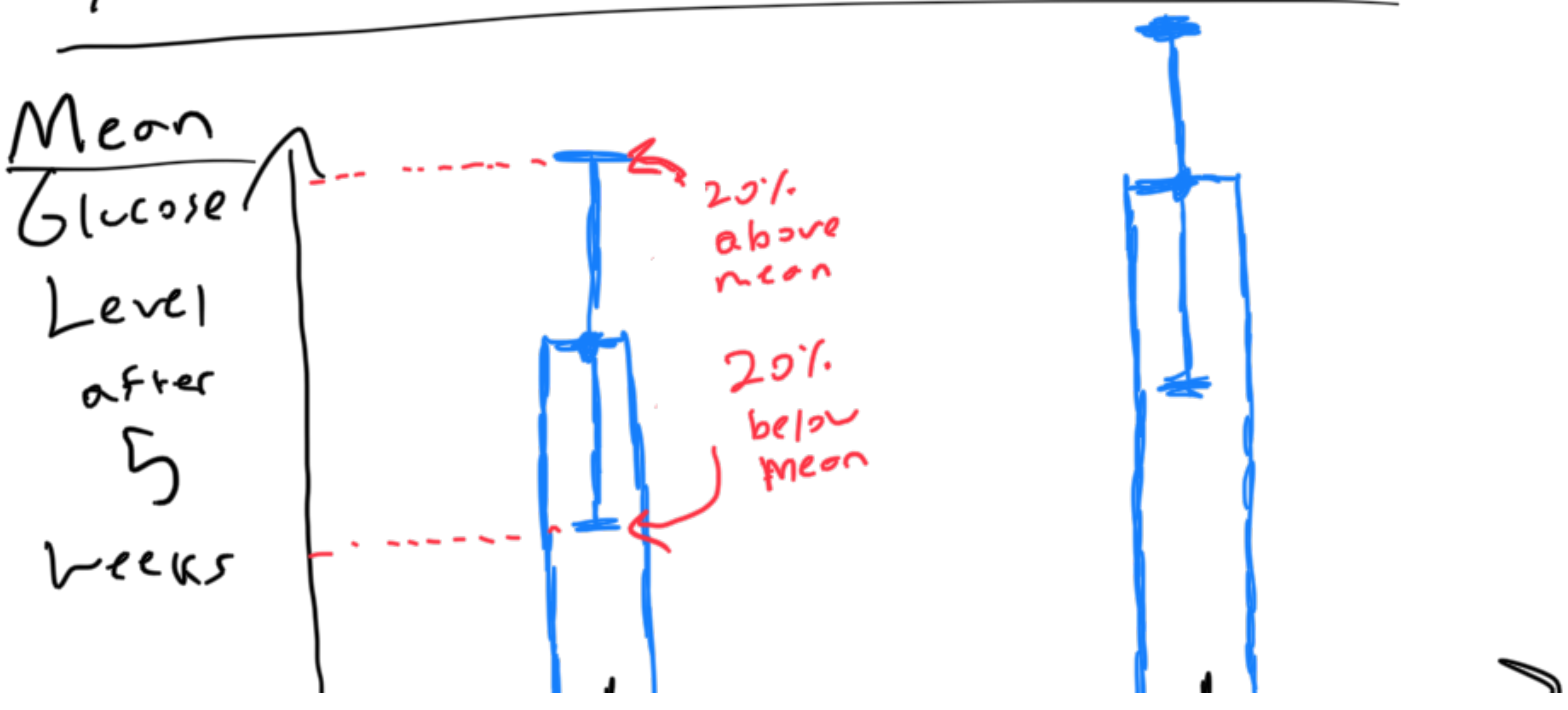
Geometric Distribution:

car successfully turns on 90% of the time. What is P(takes 3 times to get started)?

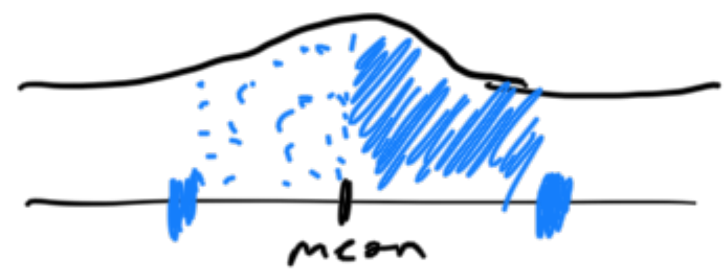
Poisson Distribution: Receive 4.5 telemarketing calls per week. What is P(receiving 6 calls in 1.5 weeks)?

...

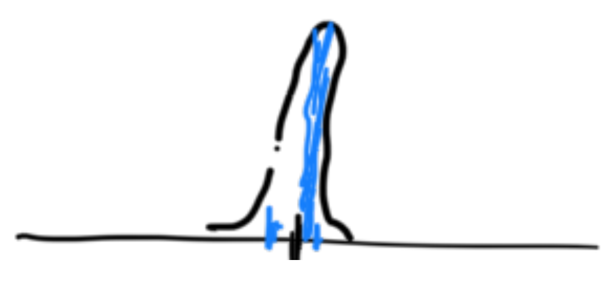
Means + $\boxed{\text{Error}}$

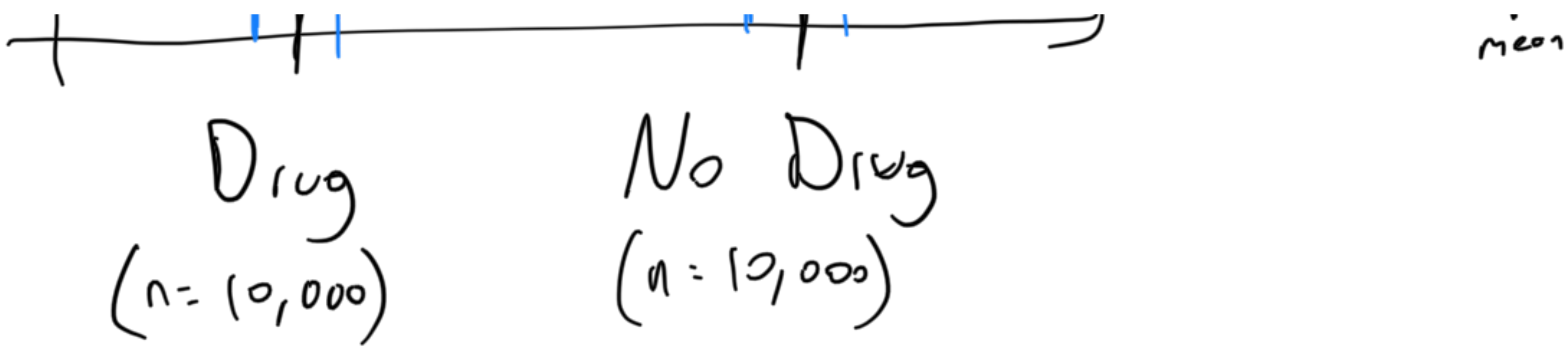


High Error:

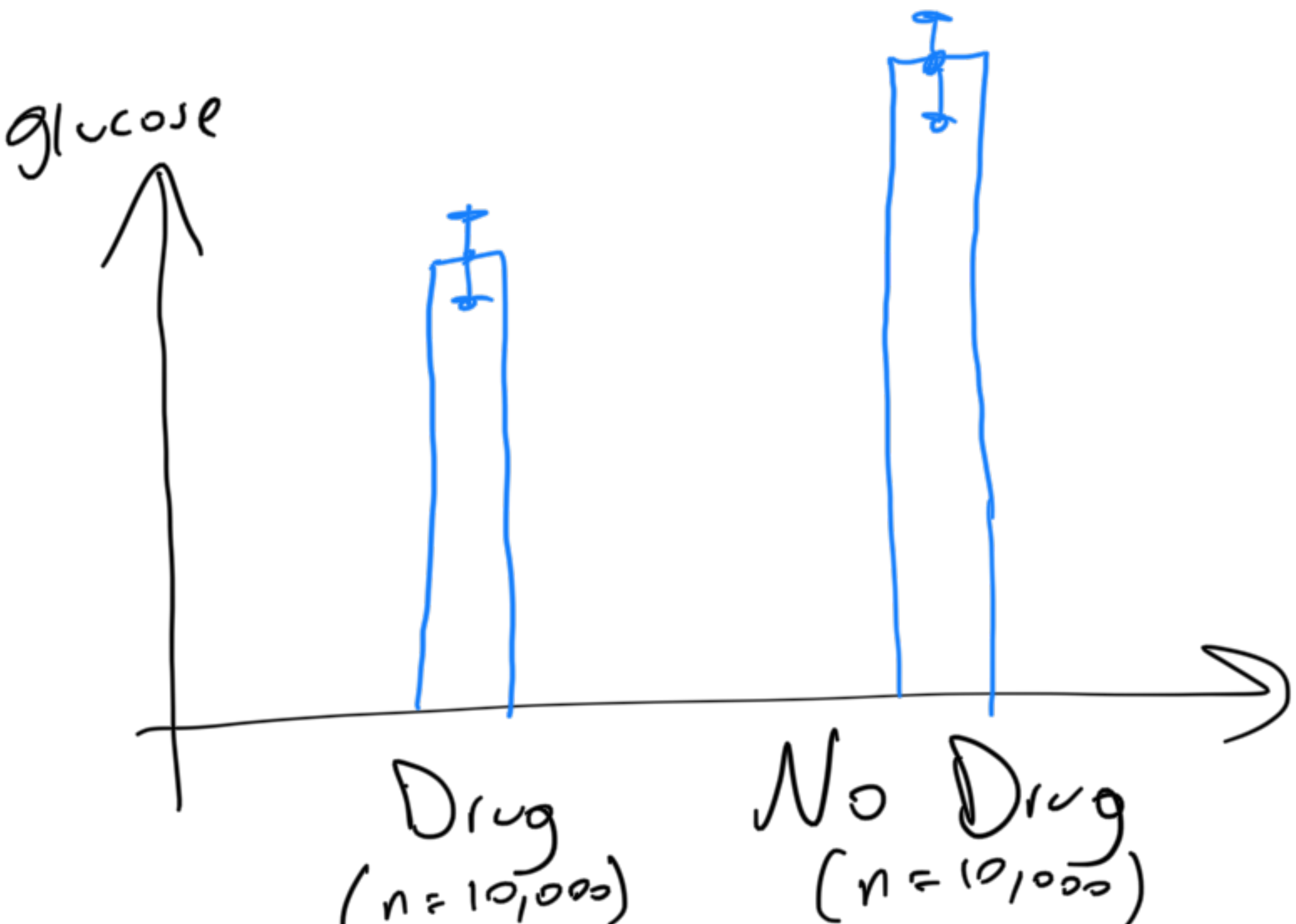


Low Error:





In the above, we are not confident that the drug reduces glucose



In the above example, we are confident that the drug reduces glucose levels

Ways To Measure Error

Variance

$$\text{Variance} = \frac{\sum_{i=1}^{n} (x_i - u)^2}{n}$$

Breaking this down:

$\sum_{i=1}^{n}$ → add up everything in the symbol

$\sum_{i=1}^{n} / n$ → averaging

u → mean

x_i → i^{th} value of X, aka X[i]

$(x_i - u)$ the

If we didn't square $(x_i - u)$, the error value would penalize a difference of 6 from the mean exactly twice as much as a distance of 3

↳ we square it to say "we have more error the more we 'go away' from the mean"

$$\text{Variance} = \underbrace{\frac{1}{n}\sum_{i=n}^{n}}_{\text{average}} \underbrace{(x_i - u)^2}_{\substack{\text{distance from each point} \\ \text{from the mean, squared ...} \\ \text{aka: error}}}$$

$$= \text{Mean} \quad \text{error}$$

Standard Deviation

Square root of the Variance

↳ this gets us error in the original units (e.g, glucose vs glucose^2)