8th July 2023

Advanced distributions - 1

Normal

height

→ histogram.

→ math fn

= Normal db^n

60 65

1 - norm.cdf(z_{60})

P[X < 60] Area under curve.

Binomial

- Tossing a coin 10 times.

H → Success → p

T → failure → 1-p

X: # of Success.

0, 1, 2, 3, - - - , 10

R.V P(X=0) P(X=10)

P(X=1)

:

Geometric ✓

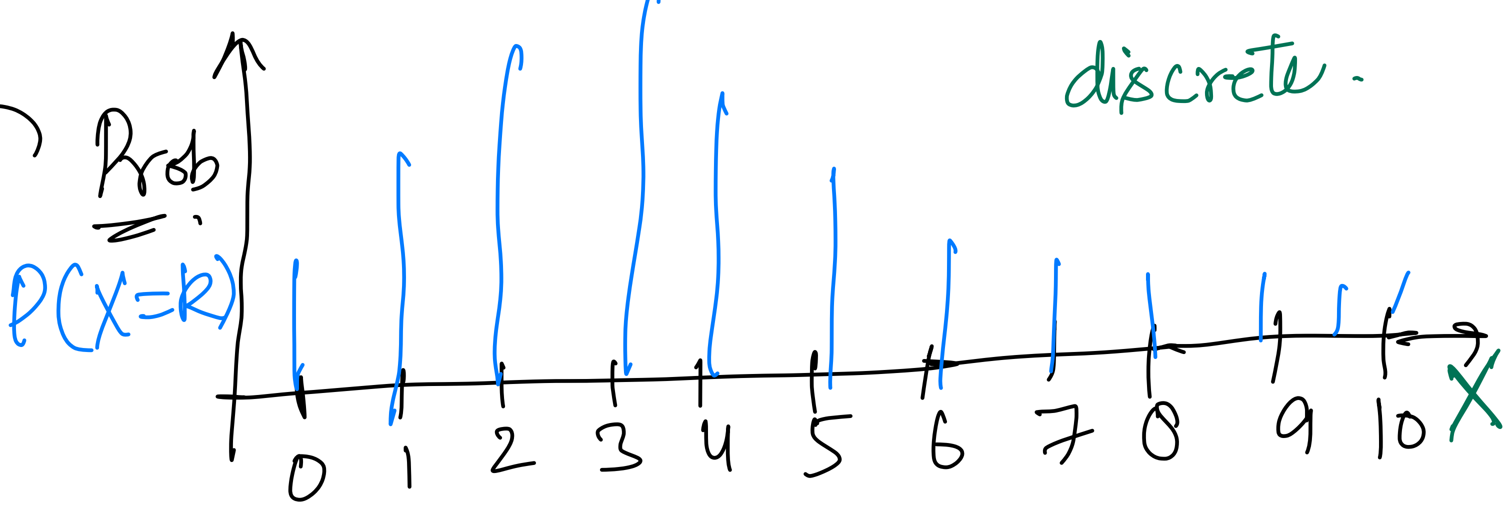
What is the prob. of success in 'n' trials.

⇒ 5th trial of coin Toss

S → H FFFFS

$(1-p)^4 \cdot p$

discrete.



Binomial db^n

$$P(X = k)$$

$0, 1, 2, 3, 4, \ldots, 10$

$$^{n}C_{k} \cdot (p)^{k} \cdot (1-p)^{n-k}$$ ✓

$$P(X = 5)$$

$n = 10$

$k = 5$

$p = 0.5$

What is the prob. I obs. exactly 5 heads.

$$= {}^{10}C_{5} (0.5)^{5} \cdot (1-0.5)^{5}$$

Poi svron

s

Poisson → Scientist

Eg: # of Accidents every night.

5, 3, 0, 2, 10, 7, 1, 1, 2, 0, 1, 0, 2, 3 . . .

↓

avg. value

Rate = # of Events / time.

Expected # of acc / day = $\boxed{3 \text{ acc} / \text{day}}$

- What is the prob. of observing 5 acc next night.

X : # of accidents

P(X = 5) = ? P(X = 2) = ?

P(X = 10) = ? P(X = 0) = ?

X : Random variable.

→ follows some Prob. distribution.

→ 0, 1, 2, 3, 4, - - - - - - -, 100

Process : observing # of Accidents follows

Poisson Process (discrete distⁿ)

X ~ Poisson distribution **

If Process follows below Assumption ~ (Poisson Distⁿ)

Accident

(i) Counting : (Am I able to count # of occurences or not)

(ii) Independence: (occurance of one event shouldn't impact occ. of another event)

(iii) Constant Rate Assumption :

Rate (3 acc/day)

Avg.

SHO

How many accident to Expect next day → 3

— next next day → 3

— — → 3

→ 3

$P(X = 10)$

(iv) No Simultaneous Events:

(2 events can't occur together)

A_1 : 9:05 PM

A_2 : 9:05 PM

ns, ms, μs ✓

Poisson says:

- Rate
- Avg # of Events / time.
- Expected # of Events / time

→ 3 acc / day.

'λ' → theory.

$\boxed{\lambda = 3}$

'mu' → python

$\boxed{k = 0, 1, 2, 3, \ldots}$

$P(X = k)$

0, 1, 2, 3, 4, 5, 6, 7, ...

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

exp → e ≃ 2.7

- $P(X=0)$: Prob. of obs. 0 acc in next night.

$k=0, \lambda=3$

$$P(X=0) = \frac{(3)^0 \cdot e^{-3}}{(0!)} = e^{-3} = 0.049$$

$= 4.9$

$\approx 5\%$

poisson.pmf(mu=3, k=0)

- $P(X=1)$: Prob. of obs 1 acc.

$k=1, \lambda=3$ $P(X=1) = ((3)^1 \cdot e^{-3})/(1!)$

$$P(X=1) = 3 \cdot e^{-3} = 0.149$$

$$= 14.9\%$$

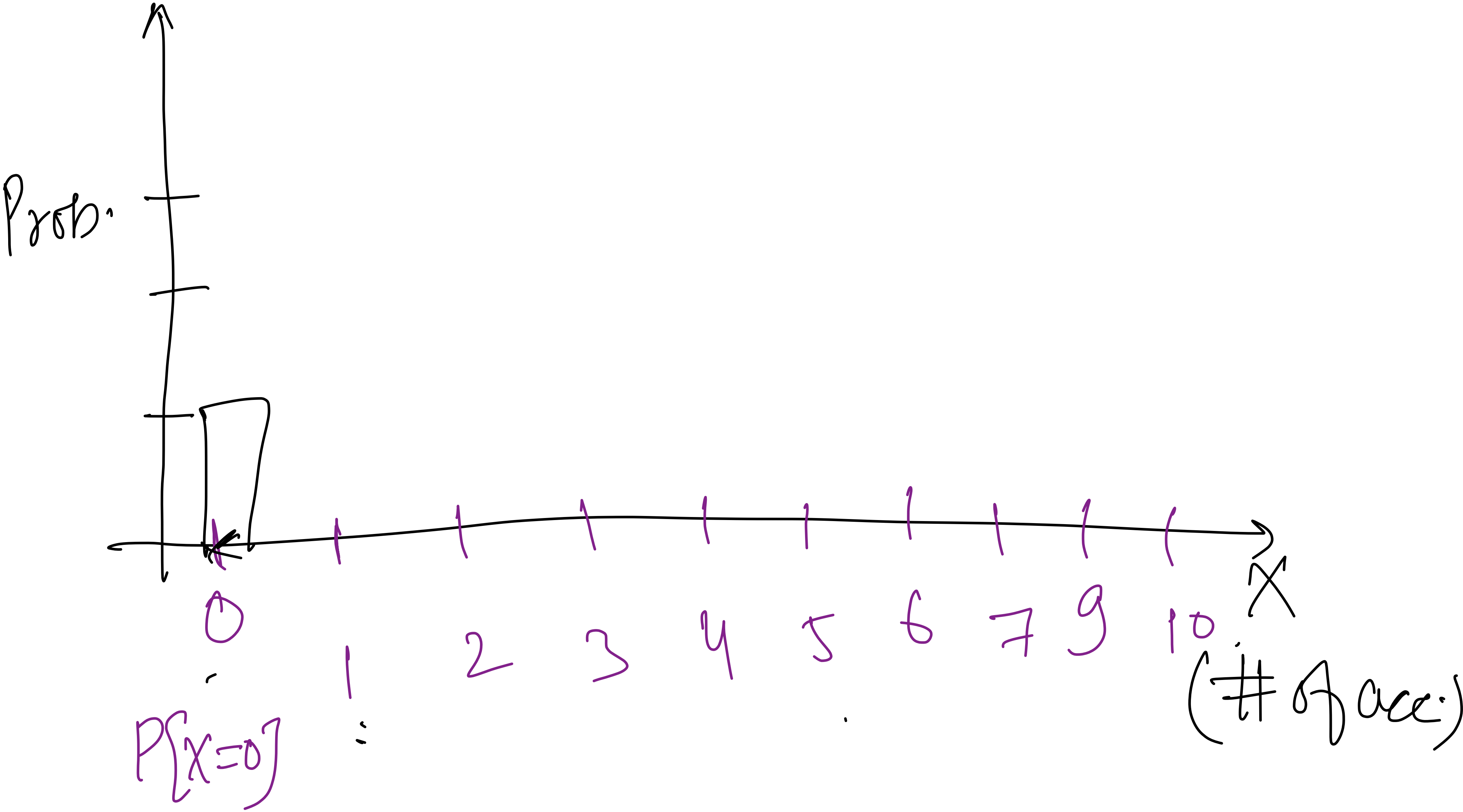
$$\simeq 15\%$$

- $P(X=100)$: Prob. of obs 100 acc next night

$k=100$

$\lambda=3$

$$P(X=100) = \frac{(3)^{100} \cdot e^{-3}}{(100)!}$$



Prob.
X
0
1
2
3
4
5
6
7
9
10
(# of acc.)
P[X=0]

* **Incoming Messages:** (Poisson Process).

whatsapp/telegram/Snapchat/Instagram/Text.

. On avg. you see 240 msg/hr. Assume

~ Poisson Process.

- → Count ✓
- → Independent (somewhat okay)
- → ~240 const. rate. (Expected # of msgs.)
- → Shouldn't simultaneous occ.

You receive 240 messages per hour on average - assume Poisson distributed. What is the probability of one message arriving over a 30 second time interval?

* 240 msg $\rightarrow$ 1 hr $\rightarrow$ 3600 s.

Rate $\rightarrow$ λ_{1hr} = 240 msgs.

3600 s $\rightarrow$ 240

1 s $\rightarrow$ $\left(\frac{240}{3600}\right)$ msg. $\rightarrow$ $\lambda_{1sec.}$

[given time interval 30 s.]

30 s $\rightarrow$ $\left(\frac{240}{3600}\right) \times 30$ $\rightarrow$ $\lambda_{30sec.}$

$\lambda_{30\,sec} = 2$ msgs.

Rate for 30sec $\longrightarrow$ 2 msgs/30 sec.

$\boxed{\lambda = 2}$ ✓

$\underbrace{\text{30sec}}_{\text{time interval}}$

$\boxed{k = 1}$ ✓

$$P[X=k] = \frac{\lambda^k \cdot e^{-\lambda}}{k!}$$

$P[X=1]$ in 30 sec.

λ, k

$$= \frac{(2)^1 \cdot e^{-2}}{(1!)} = 2 \cdot e^{-2}$$

$= 0.2706$

$\simeq 27\%$.

What is the probability that there are no messages in 15 seconds?

$3600\,s \longrightarrow 240$ msgs.

$1\,s \longrightarrow \dfrac{240}{3600}$

$\lambda_{15\,sec}$

$15\,s \longrightarrow \dfrac{240}{3600} \times 15 = 1$

$\boxed{\lambda = 1}$ for 15 sec. $=$ poisson.pmf$(mu=1, k=0)$

$= 0.367$

Poisson $D b^n$ → $P[X=k]$

↳ optimizing some business decisions.

# of acc → Police staff optimization (Resources)

On avg. Ecommerce website :

↳ 1000 cust/hr. → logins/hr

$P[X \geq 2000]$

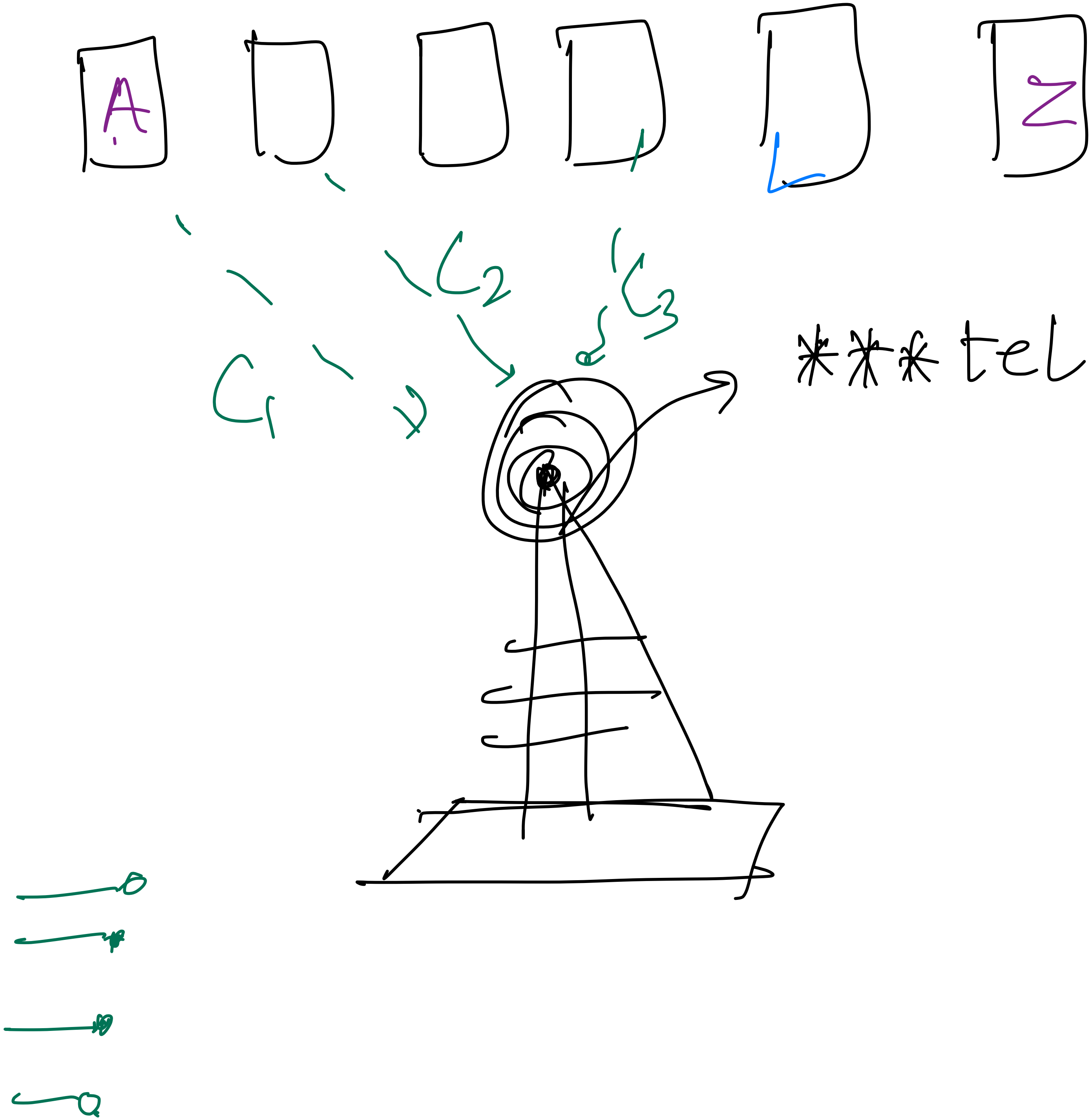
e.g. hospitalization

On avg. in hospital 10 patients arrive every hr.

$P[X=0]$

$P[X=3]$

$P[X \leq 5]$



\# of connections made / hr.

Poisson Process

20 years Data

of acc.

1. 31st dec '01 → 58

2. ⋮ → 49

⋮

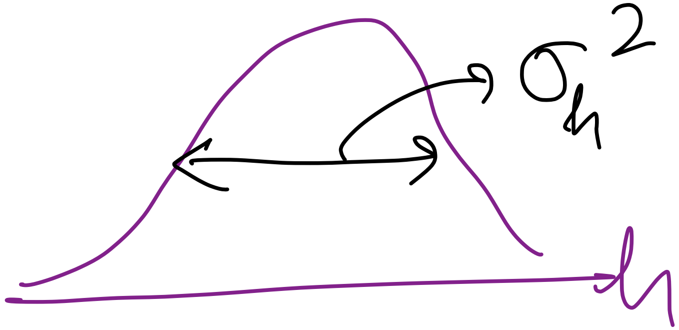
20. 31st dec 2020 →

$\lambda_{31^{st}\,dec.}$

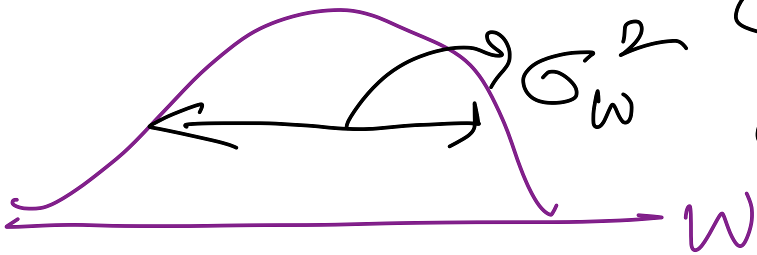
height / weight

1 Variance (1D)

- spread



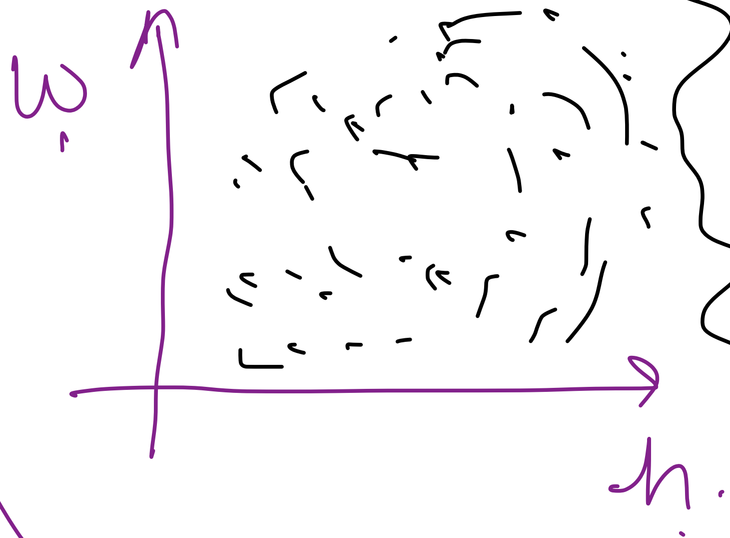
σ_h^2



σ_w^2

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{N}$$

Covariance (2D)



spread of data in 2D.

$$\sigma_{xy}^2 = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{N}$$

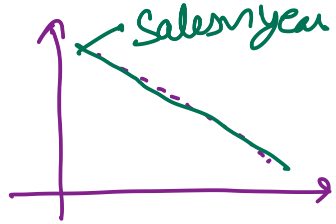
Correlation (2D)

Strength of linear relationship

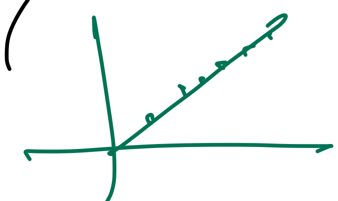
[-1, 1]

Sales vs year

-0.8



corr = -1



* $$\boxed{\operatorname{Cov}(x,y) = \frac{1}{N}\sum(x_i - \bar{x})(y_i - \bar{y})}$$

* $$\boxed{\operatorname{Corr}(x,y) = \frac{1}{N}\sum \frac{(x_i - \bar{x})}{\sigma_x} \cdot \frac{(y_i - \bar{y})}{\sigma_y}}$$