- Statistics
  - → Discriptive → you have All the Data & you want To Describe it
  - → In Ferential → You have Portion of Data & you wants to Draw conclusion.
- Central tendency of the Dataset.
  - → Arithmetic Mean → sum (All the Number) / # Numbers
  - → Median → middle element In the sorted data
  - → Mode → Most frequenty occuring digit
- Sample and popwlation

Popwlation Mean = $\mu = \frac{\sum_{n=1}^{N} x_n}{N}$ ⋮ Sample Mean = $\bar{x} = \frac{\sum_{n=1}^{n} x_n}{n}$

Conly the Notational Difference.

- Measure of Dispersion

→ variance

$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$ ⋮ For Sample (↳ $\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$) $S_n^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$

For the popwlation

But when we calculate the variance for Sample we divide it By (n-1) To get unbaised estimate

$s_{n-1}^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{(n-1)}$ ← variance for sample.

- Standard deviation

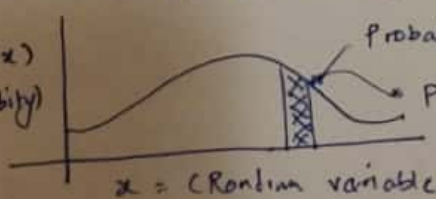
$\sigma = \sqrt{\sigma^2} = \sqrt{\text{variance}}$.

- Random variable
  - → It's Function That maps Random processes To Some value.

$X = \begin{cases} 1, & \text{If its Rain Tommorow} \\ 0, & \text{It it Not Rain} \end{cases}$ ← mapping

X → can be discrete (Probability Distribution) can be continuous. (Probability Density function)

- Probability Density function
  - → In case of continuous Random variable we have mappping function



- Binomial Distribution for Reliable & Non Reliable Points

→ Experiment consists of n Independent Trials. with Two mutually exclusive out comes

→ so probability of success = (P) then Prob. of failure = (1-P)

Now you want to Determine the probability of the success for N trials

x = Discrete random variable. = Number of success In N n trails.

$$f(x) = {}^{n}C_{x} (P)^{x} (1-P)^{(n-x)}$$

The distribution you get is called As a binomial Distribution

- Expected Value [E(x)]

→ It is noting But different representation of the mean.

→ Here you know the frequency of event's occurance In the Discrete case and In continuous case you know the Probability density function.

$$E(x) = \sum_{i=0}^{\infty} x_i P_i$$ → x_i = set of out comes of exp.
P_i = Associated Probabilities.

$$E(x) = \int_{-\infty}^{\infty} x \cdot f(x) \, dx$$ where F(x) = Probability density Function
x = Ran variable at which you want to calculate the P(x)

- Expected value For Binomial Distribution

E(x) = (Probability of success) · (# For which I want to calculate Exp value)

$E(x) = n \cdot p$

- Poisson Process → (continuous Representation of Binomial Dist.)

X = # of car pass in an hour.

E(X) = Let say you scale we measured that λ no. of car passed In hour we can model this Experiment As a Binomial Distribution.

where n = number of minute

p = Poj Probability of car pass In a minute

$E(x) = \lambda = n \cdot p$

So Now I have the Expected value so Let cal Prob. for 1 car per minute

$$P(X=K) = \binom{60}{K} \cdot \left(\frac{\lambda}{60}\right)^{K} \left(1-\frac{\lambda}{60}\right)^{60-K}$$ — Binomial Dist

As The n → ∞ which is 60 above Poisson Distribution
And In Poisson Dist we measure Prob. for Interval and Not for final.

$$\lim_{x \to \infty} \left(1 + \frac{a}{x}\right)^x = e^a$$

We know from Binomial Distribution that

$$P(X=K) = {}^{n}C_{K} \left(\frac{\lambda}{n}\right)^K \left(1 - \frac{\lambda}{n}\right)^{n-K}$$ — where λ = rate of success.

As $n \to \infty$

$$P(x=K) \lim_{n \to \infty} = \frac{n!}{(n-K)!\,K!} \cdot \frac{\lambda^K}{n^K} \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-K}$$

$$= \frac{(n)\cdot(n-1)\cdot(\cdots)(n-K+1)}{K!} \cdot \frac{\lambda^K}{n^K} \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-K}$$

$$P(x=K) \lim_{n \to \infty} = \frac{n \cdot (n-1) \cdots (n-K+1)}{n^K} \cdot \frac{\lambda^K}{K!} \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-K}$$

$\frac{n^K + \cdots}{n^K} = 1$; $\frac{\lambda^K}{K!}$ constant ; $\left(1 - \frac{\lambda}{n}\right)^n \to e^{\lambda}$; $\left(1 - \frac{\lambda}{n}\right)^{-K} \to 1$ → As $n \to \infty$, λ = constant

$$P(x=K) \lim_{n \to a} = \frac{\lambda^K}{K!} e^{-\lambda}$$

The probability of car pass is now trial independent In a way Its now continuous distribution.

So Poisson Distribution Is special Approximation of Binomial Dist. where No. of trails are Infinite.

- Normal Distribution (Gaussian Distribution)

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

→ As opposed to the Binomial dist ~ Poisson dist. the Normal dist is continuous func.

→ Normal distribution Defined by The mean & standard deviation

- Central limit Theorem

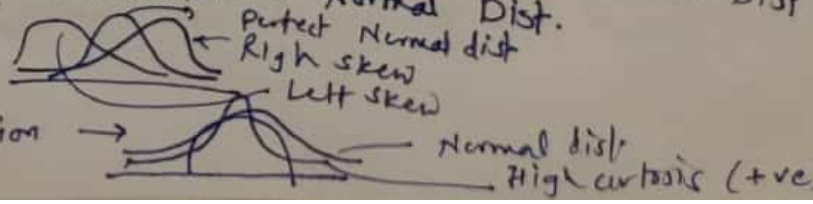
→ So Let's Assume you draw samples from identical same distribution with Re some sample size. the central limit theorem states that As your # of sample Increases (greater than 1/2) the If you find the distribution of sample sum/mean It will Be close to Normal Dist. & As ($n \to \infty$ / $\# n \to \infty$) → Dist Became pure Normal Dist

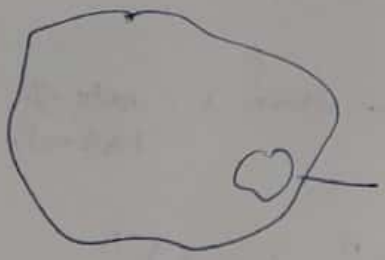
→ As sample size & std. dev. of distribution Dec.

→ Samples → ∞ Sampling distribution → Normal Dist.

→ Skew, In Normal Distribution

→ Kurtosis In Normal Distribution →





Population

Sample → Sample mean
→ sample variance.

↓ Now If you draw Infinite No. of Samples

The mean will Be close to population mean however Standard deviation of Sample mean distribution

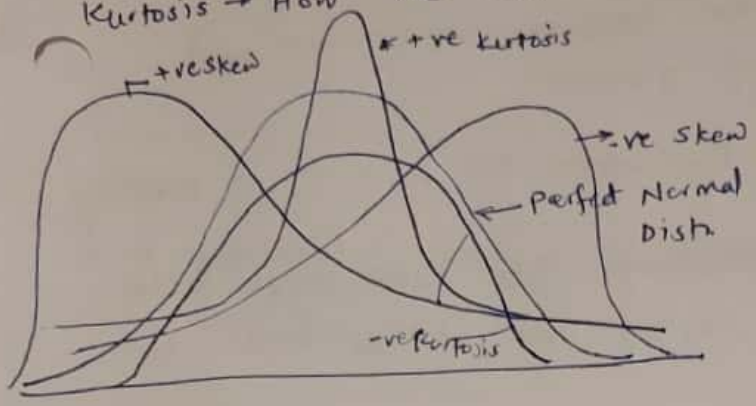
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

← Std for sample
← sample size.

Std for Sample dist

- Sampling Distribution of sample mean.

Skew → How the distribution differ from Normal dis horizontally
Kurtosis → How the distribution differ from Normal Distribution vertically



→ CLT Doesn't work if sample size is equal to 1
→ As the sample size → ∞ The Distribution Looks like Normal Dist.
→ As the #sample → ∞ the Dist goes close To Normal Dist
→ As samples ↑ #sample Inc standard der. decreases.

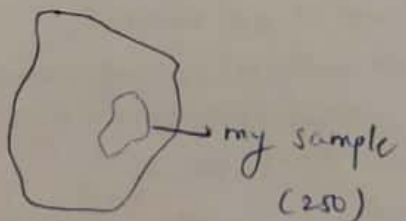
→ So Let's say you have a Non Normal some Random distribution with some mean & variance. ~~the~~ If you ~~calculate~~ Let's say you draw a sample of size n. and plot them on graph
→ The mean of the Sample Distribution ≈ mean of original Distribution
→ The standard deviation of $\sigma_{\bar{x}}^2$ for the sample distribution will be

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$ ← where n is sample size

So ~~$\mu_{\bar{x}} = \mu$~~ $\mu_{\bar{x}} = \mu$ ← of original Dist $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$ → where n sample size.

- Margin of error

→ So Let's say for Pixel (100, 100) I'm getting 2000 values of Distances. How ever I can't use All 2000 values. So I selected ~~100~~ 250 ~~Point~~ values Randomly from 2000 values.



Values for Pixel (100,100)
(2000)

Conclusion:

95% chance that a Random $\bar{x}$ is with in $2\sigma_{\bar{x}}$ of Total population.

- So we are 95% sure that 68% to 91% of Point's are Reliable.
- to shrink the % gap we have to draw More Samples.

- We have total 2000 callbacks of Depth data
- Let's say ground Truth is 1000mm And we consider the point is Reliable If Distance value is betn ± 5mm.
- Now we draw 50 samples Randomly from ~~above~~ for pixel (1,1) Randomly From above 2000 values
- We don't care about The Type of Distribution Thoes 2000 Population ~~Sample~~ follows coz we know If we draw a sample of size > 2 and plot the mean It will Follow Normal Dist enough
- So By central limit Theorem we are Rest Assure that The Distribution of Sample mean will be Normal Dist with some mean & standard dev
- Let say In our Sample of 50 we got 40 → Reliable points & 10 → unreliable points
- Now our Task Is to come up with the 95% Interval such that If we draw x points we are sure that some proportion of Thoes x points are Reliable.
- As per central limit theorem we know

Population mean = mean of Sample distribution ≈ sample mean.

Variance of Sample dist = $\frac{\text{variance of population}}{\text{sample size.}}$ $\therefore \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$

For example

Sample Size = 50

Reliable point = 40

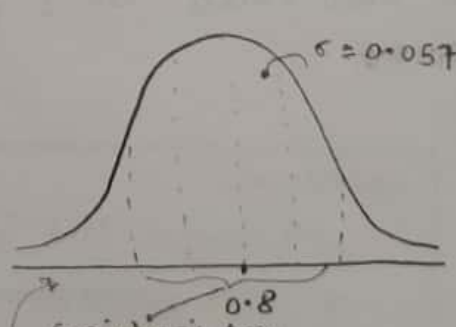
unreliable pts = 10

Sample mean = 0.8

$= \frac{\cancel{50\times1} + 40\times1 + 10\times0}{50}$

sample std der $= \sqrt{\frac{(40\times(1-0.8)^2 + 10(0-0.8)^2}{49}}$

$= 0.404$ → std For sample



std For Sample Dist $= \frac{\text{std For sample}}{\sqrt{\text{sample size}}}$

$= \frac{0.404}{\sqrt{50}}$

$= 0.0571$

Let's say we want 95% confidance either side so it 2σ Level

→ confident 95% of chance that $\bar{x}$ is with in 0.8 Valid points Is

→

Lower limit

$= 0.8 - 2\cdot(0.0571)$

$= 0.6857$

upper limit

$= 0.8 + 2\cdot(0.0571)$

$= 0.9142$

Hypothesis Testing for Reliable & Non Reliable Points

Image

H_0 → Glare has no effect on the depth value for that point

H_1 → Glare has an effect

Let's Assume

→ We have 100 reading for the point under consideration

→ we know the Grand Truth depth is 500mm

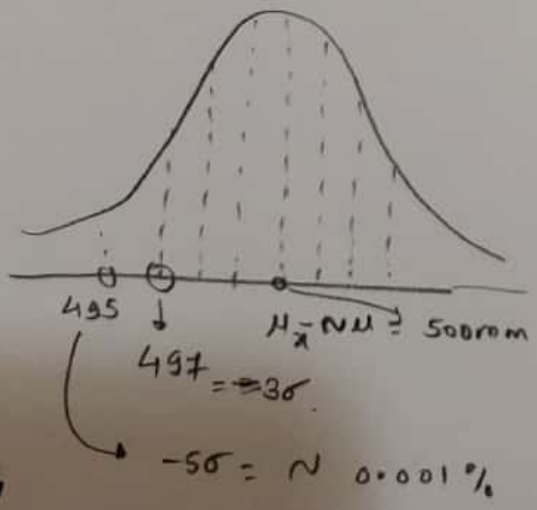
→ for the 100 reading of point on glare Boundary the mean is 495mm & σ = 10mm

NOW the process

- Calculate Standard deviation for Sampling distribution

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{\text{sample size}}} = \frac{10}{\sqrt{100}} = 1\,mm$$

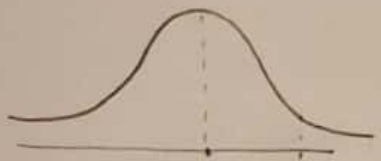
- How many standard deviation away from Ground truth i.e. 500 mm the Sample mean 495mm is. so it's just like calculating the z score.



So the hypothesis that Glare has No effect is Rejected Because it has an Chance of Less than 0.01%. so

Z-statistics vs T-statistics

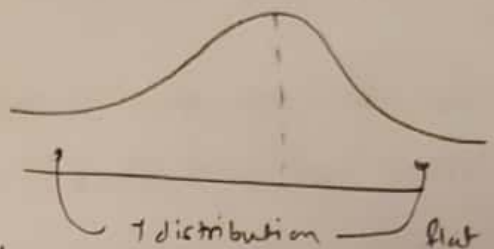
We use the T-statics we calculate the calculation steps are same as Z statistics Most ever in case of T-statistics the distribution will Loook like T distribution



Z statistics $\mu_{\bar{x}}$ $\bar{x}$

Z-score = $\frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$



T distribution — flat tails

In case n is very small we use T-score Instead Z score

⇒ Type-1 error

Rejecting the null hypothesis even if it's true.

→ → → → → → → → → →

So we have

Sample size = 10

mean_of_sample = 17.17

Sample_std = 2.98

Now we want to find the Zone with confidence Level 95%

i.e. Z score = ?

$-2.262 < Z < 2.262$

we know $t = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$

$\sigma_{\bar{x}} = \frac{\sigma_{\bullet}}{\sqrt{n}}$ ← sample std / ← sample size

$\bar{x}$ = sample mean

$Z = \frac{17.17 - \mu}{2.98/\sqrt{10}}$

$-2.262 < Z < +2.262$

$-2.262\,(.942) < 17.17 - \mu < 2.262\,(.942)$

$= -2.13 < 17.17 - \mu < 2.13$

$2.13 > \mu - 17.17 > -2.13$

$19.3 > \mu > 15.04$

95% chance that sample mean is Btn this Range

Rule of thumb

$n > 30$

$np > 5$

$n(1-p) > 5$

Question 1 → Given a Distance value for a pixel with and without glare Prove whether the glare have statistically significant effect on point's depth value

What we know →

(1) we have Sample of 2000 point's for the Depth data of a pixel with & without Glare value.

Our hypothesis

H_0 → Glare has no effect on the depth value for that point

H_1 → Glare has an effect on the depth value for that point

Our methodology

→ First calculate the ground Truth value

Ground Truth value = average (All point's, without introduction of Glare) for that particular Pixel

→ calculate the mean & standard deviation for the Glare Sample

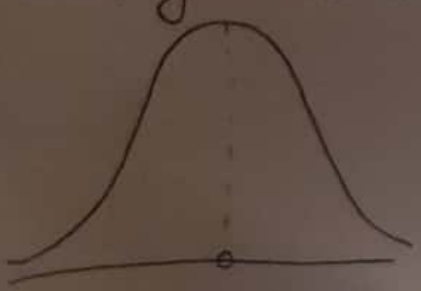
- Collect the Pixel in the Glare boundary Region.
- get the value of Distance for that Pixel. (sample > 200 points)
- Calculate the mean & standard deviation. For the Glare point Sample

→ H_0 ⇒ Glare has no effect

→ H_1 = Glare has effect — Glare Points $\begin{cases} \mu_{distance} = \mu_{true\ value} \text{ (even with glare)} \\ \mu_{distance} \neq \mu_{true\ value} \end{cases}$

→ Assume H_0 is True

Sampling distribution



$\mu_{\bar{x}} = \mu$ = True depth value

mean for points w/o Glare

Sampling dist

Standard dev for sampling dist ⇒ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ → σ → std dev for population; n → sample size

$\sigma_{\bar{x}} \approx \frac{s}{\sqrt{n}}$ → s → std of sample

⇒

So Now we have

(1) Sample distribution mean $\mu_{\bar{x}}$ = True depth value

(2) Sample → std. dev $\sigma_{\bar{x}} = s/\sqrt{n}$

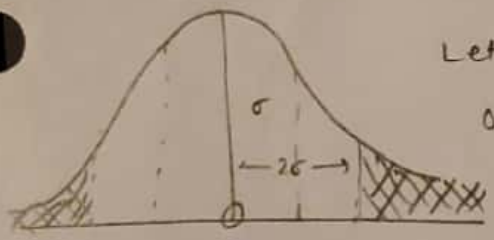
→ How Many std dev away from the sampling distribution Mean is the mean of the glare sample & the what is the prob of getting that Results away from sampling distribution Mean.

→ Need to calculate the Z-Score

$$Z = \frac{(\text{True depth value}) \text{ or } (\text{Sampling dist mean for the pixel where there is No GLARE}) - (\text{Glare depth value}) \text{ mean we are getting for the same pixel where there is a (GLARE)}}{\text{Standard dev of sampling distribution}}$$

$$= \sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{\text{Standard dev of Glare sample}}{\sqrt{\text{sample size}}}$$

→ once we get the Z value we are sure that the Result is Z standard deviation away from the mean. (True depth value)



Lets say our Z=2 then we are focusing on shaded area.

(sampling distribution for the Dataset w/o glare)

→ We know for Z: 2σ The area under Normal curve where ± Std 7 2σ = 5%

→ So Based on the Z score we can say that

→ If the Null Hypothesis (H_0) is true the there is 5% chance of getting this value

→ So we can say that If we are confident 95% time for

Question No- 02

→ Let's say or you have set of 100 points distributed irregularly for the particular pixel

→ you do also know the true depth value for that pixel.

→ you consider the point As Reliable If it's value is in some threshold x

→ so Reliable point's Range is (true value I x)

→ Based on this you divided your sample In Reliable and unreliable Points

→ Let's say P is the probability That point is Reliable

(1-P) ⟶ un Reliable

⟹ question → If we draw a sample of 100 point for particular pixel And Then Base on some threshold 'T' from the mean of sample we decide Whether the current point is reliable or Not Let Say In sample we got x% of Reliable points what is the 95% confidence Region that the Actual percentage of Reliable points In the population is close To sample mean

Methodology :→ [For our Sample $\hat{P}$ = (1-P)= 0.57 (1-P) = 0.43 n=100

① First you have To calculate the mean & std. dev for a sample

mean → $\bar{x} = \frac{57 \cdot 0 + 43 \cdot 1}{100} = 0.43$

std dev =

$$S^2 = \left(57 \cdot (0-0.43)^2 + 43(1-0.43)^2\right) / 99$$

② We know mean of sampling dist of the Sample mean $\mu_{\bar{x}} = \mu =$ of population

③ We know The std. dev of Sample dist of Sample mean = $\sigma_{\bar{x}}$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ $\sigma \approx \sqrt{S^2}$ → coz sample is from same dist

④ In statistic we know (curren scenario)

Sample

mean = $\bar{x} = \frac{0 \cdot q + 1 \cdot P}{n}$

std dev $S = \sqrt{\frac{P(1-\bar{x}) + q(0-\bar{x})}{n-1}}$

size → n

Sample dist sample mean

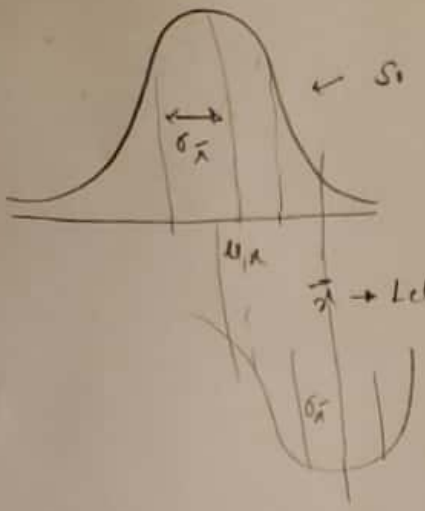
$\mu_{\bar{x}}$ ⟶ equal to

$\sigma_{\bar{x}} = \sigma/\sqrt{n}$

Relationship bet $\bar{x}$ & μ Population μ

$\sigma = S$

so we know value for $\bar{x}, S, \sigma, \sigma_{\bar{x}}$ and Relation that $\mu_{\bar{x}} = \mu$ now Base on this we know To find



← So this is the distribution of our sampling means

$\bar{x}$ → Lets assume $\bar{x}$ for If here So Based on $\bar{x}$ we need To predict possible values for the $\mu_{\bar{x}}$

(5) So Based of It we can calculate the Z score.

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

we know $\bar{x} \pm \sigma$ & Range for Z value Based of normal dist table & we need to cal μ for 95% confidence

$$Z < \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} < Z$$

$Z = 2.58 \cdot \sigma_{\bar{x}}$

$$-2.58\,\sigma_{\bar{x}} < \bar{x} - \mu < 2.58\sigma_{\bar{x}}$$

$$+2.58\,\sigma_{\bar{x}} > \mu - \bar{x} > -2.58\,\sigma_{\bar{x}}$$

$$+2.586_{\bar{x}} + \bar{x} > \mu > -2.586_{\bar{x}} + \bar{x}$$

Results :→

Based of this Results we can find the % of an reliable point

Range of

Based of a sample.

Let's say we draw 100 points for some pixel & 100 points for some other Pixel. Then we took a difference of mean for both samples

Now we want to say whether the point's follow belongs to different surface under consideration.

→ Let's say we have exact same position for the camera & we just change the surface under consideration does the sample data we get is enough to tell us the diff bet surfaces

→ our methodology →

For sample one we know

mean → $\bar{x}_1$

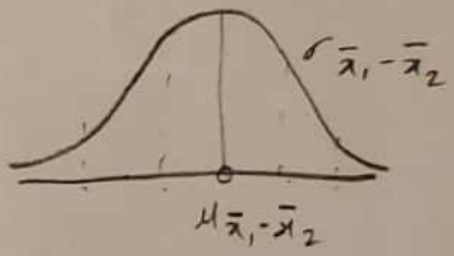
std dev = s_1 → $\sigma_{\bar{x}_1} = \frac{s_1}{\sqrt{n}}$

for sample two we know

mean → $\bar{x}_2$

std dev → s_2 $\sigma_{\bar{x}_2} = \frac{s_2}{\sqrt{n}}$

⇒ Our sampling strategy is diff Now we sample two diff values from two samples & substract them so we get



(2) now we know

$\mu_{\bar{x}-\bar{y}} = \mu_{\bar{x}} - \mu_{\bar{y}}$

$\sigma^2_{\bar{x}-\bar{y}} = \sigma^2_{\bar{x}} + \sigma^2_{\bar{y}}$

$\sigma_{\bar{x}-\bar{y}} = \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}$

t-value.

for this case we know

x_1	x_2	$\mu_{\bar{x}_1 - \bar{x}_2}$	$\mu_{x_1 - x_2}$
s_1	s_2	$\sigma_{\bar{x}_1 - \bar{x}_2}$	σ
		↳	

→ so problem here is given $\sigma_{\bar{x}_1 - \bar{x}_2}$ & $\bar{x}_1 - \bar{x}_2$ → calculate the 95% confidence interval for $\mu_{\bar{x}_1 - \bar{x}_2}$

(3) Now we have to get the 95% confidence Region ~ $1.96\,\sigma_{\bar{x}_1 - \bar{x}_2}$

confidence Region = mean ± $1.96\,\sigma_{\bar{x}_1 - \bar{y}_1}$ of BP

↳ Saying other way = of $(\bar{x}_1 - \bar{x}_2) \pm 1.96\,(\sigma_{\bar{x}_1 - \bar{x}_2})$

↳ terms of Sample mean

→ The confidence Interval will give the Range In terms of the mean value Now Based on the confidence Interval we can write hypothesis

(3) $H_0 : \mu_1 - \mu_2 = 0 \Rightarrow \mu_{\bar{x}_1} - \mu_{\bar{x}_2} \Rightarrow \mu_{\bar{x}_1 - \bar{x}_2} \Rightarrow 0$

Question No-4

given the data of a point determine whether or Not point 2 is in Glore area

---

Bozeman Science (Student's T-test)

once we get T-value → $\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}$

→ We create a null Hypothesis with some confidence Interval Let's say $P = 0.05$ i.e 95% confidence

→ Now we need to find the degrees of freedom

$= n+m-2$

→ we get the σ value for Normal distribution

→ If the σ value is greater than our T value we Reject our Null Hypothesis (H_0)

→ We can do T test in spread sheet Quite easily.

→ Population dist should be Normal.

→ Similar variance

→ If the point ≥ 20-30 data point else go for z test