Remarks about α-TM:- $\alpha \in [0, \frac{1}{2})$.

i) Already mentioned, $\alpha \to 0$, α-TM = Sample mean → NOT robust against the outliers.

$\alpha \to 1/2$, α-TM → Sample median → Robust.

It gives an impression that, the tuning parameter "α" controls the robustness against the outliers.

ii) The choice of α is an issue of research.

iii) α-TM is also equivariant under arbitrary affine transformation.

Remark:- Sample mean, Sample median & Sample α-TM all of them are asymptotically normal after normalization by $\sqrt{n}$.

Measure of spreadness:-

i) variance :- $\mathcal{X} = \{x_1, \ldots, x_n\}$

Sample variance $= \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2$, where $\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i$.

Sample standard deviation $= \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2}$

A few other measures:-

Mean ~~absolute~~ absolute deviation about mean

$$= \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}_n|.$$

Remark:- All these ~~a~~ measures are NOT robust.

Robust version of measure of spreadness:-

Median absolute deviation about median

$X = \{x_1, \ldots, x_n\}$, It is defined as

$$\text{Med}\left\{ \left| x_i - \text{Median}(x_i) \right| \right\}.$$

<u>Remark:-</u> It is robust against the outliers.

<u>Remark:-</u> All these measure are invariant under location transformation $(x \to x + c)$ but equivariant under scale transformation $(x \to ax)$.

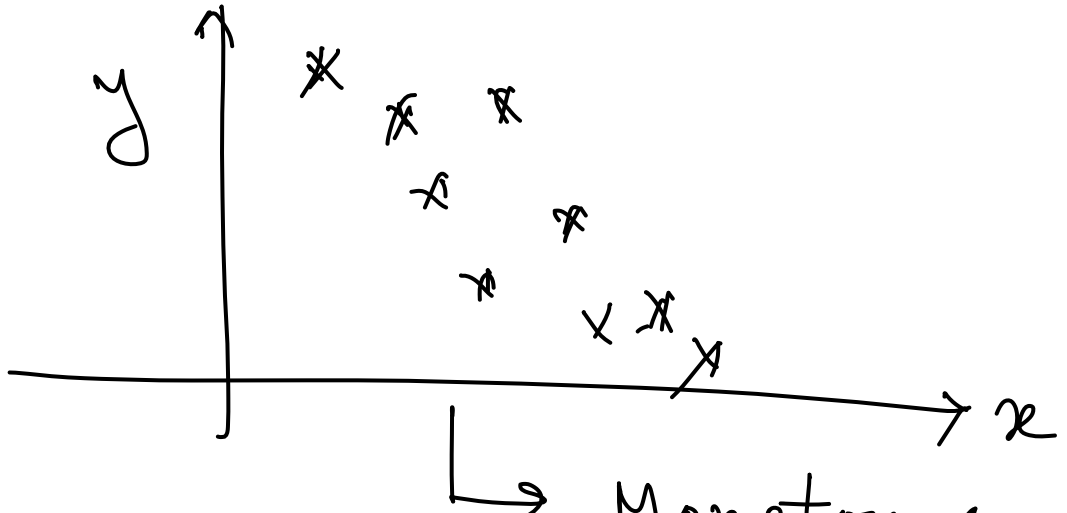
A few more order statistic based measures:-

Range $= x_{(n)} - x_{(1)}$.

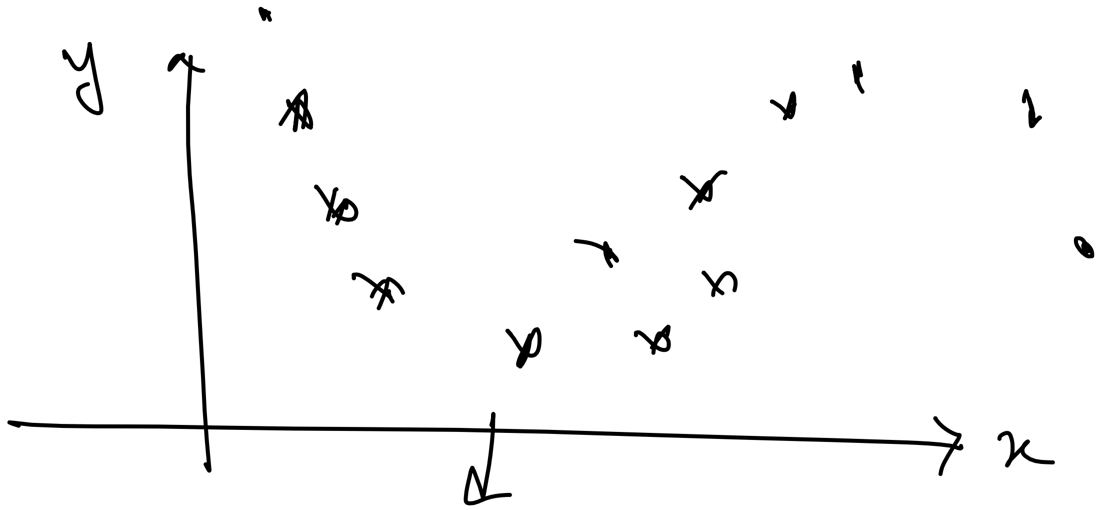
Inter quartile range $= x_{([0.75n])} - x_{([0.25n])}$, where n is the sample size.

Analysis of data based on graphical device:-

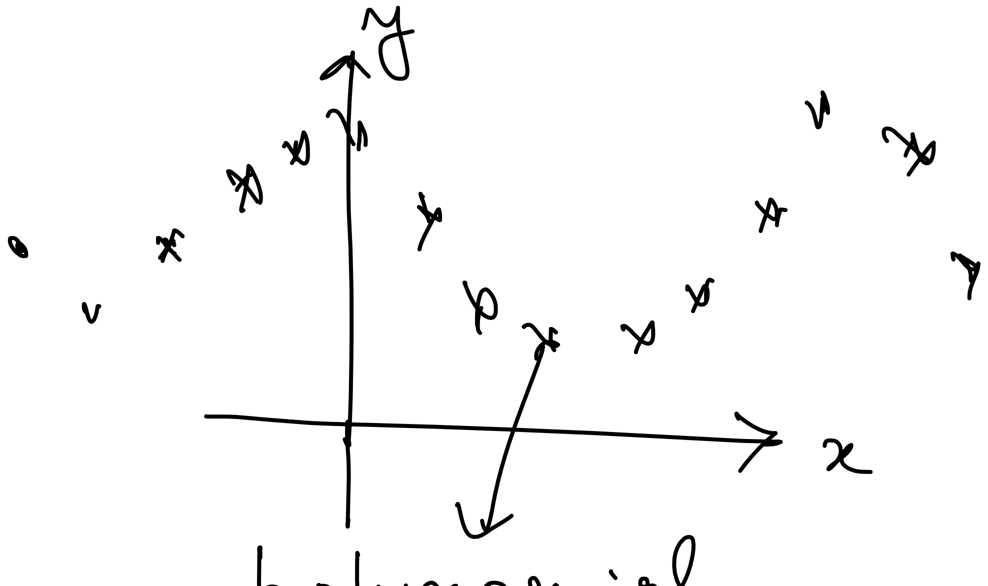
i) Scatter plot (Plot on 2-dimensional plane).



Monotone association

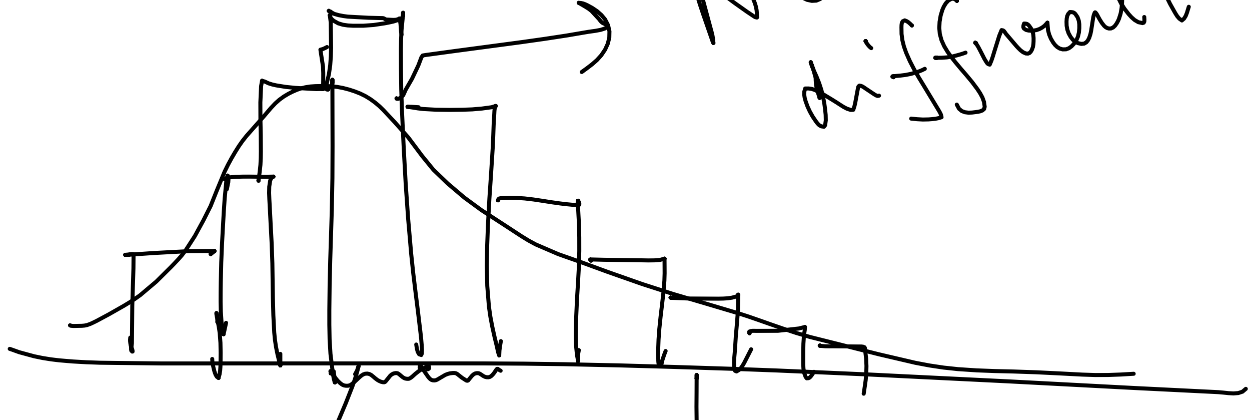


convex association



polynomial association.

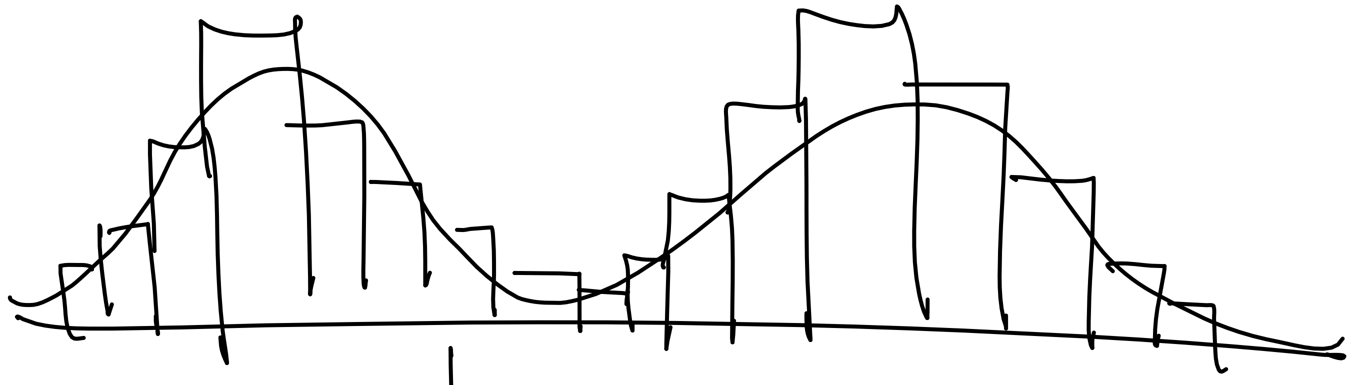
ii) Histogram :-



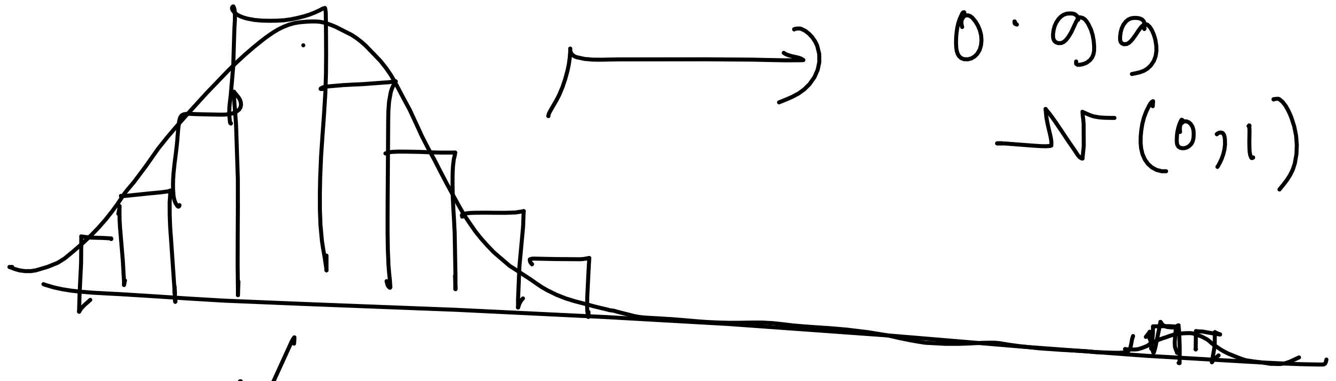
NOT differentiable.

Histogram

Original density is a skewed density



Original density is a bimodal density.



$0.99\, N(0,1) + 0.01\, N(100,1)$.

Original density is a mixture of two densities. (One is probably from outlier density)

Remark :- Although under some conditions, histogram is asymptotically unbiased estimator of the p.d.f. but it cannot capture the smoothness of the p.d.f.

Quantile - Quantile plot:-

2- sample problem:-

$\mathcal{X} = \{x_1, \dots, x_n\} \sim F$

$\mathcal{Y} = \{y_1, \dots, y_m\} \sim G.$

independent data.

Want to test:- $H_0: F = G$ ag. $H_1: F \neq G$.

Formulation:-

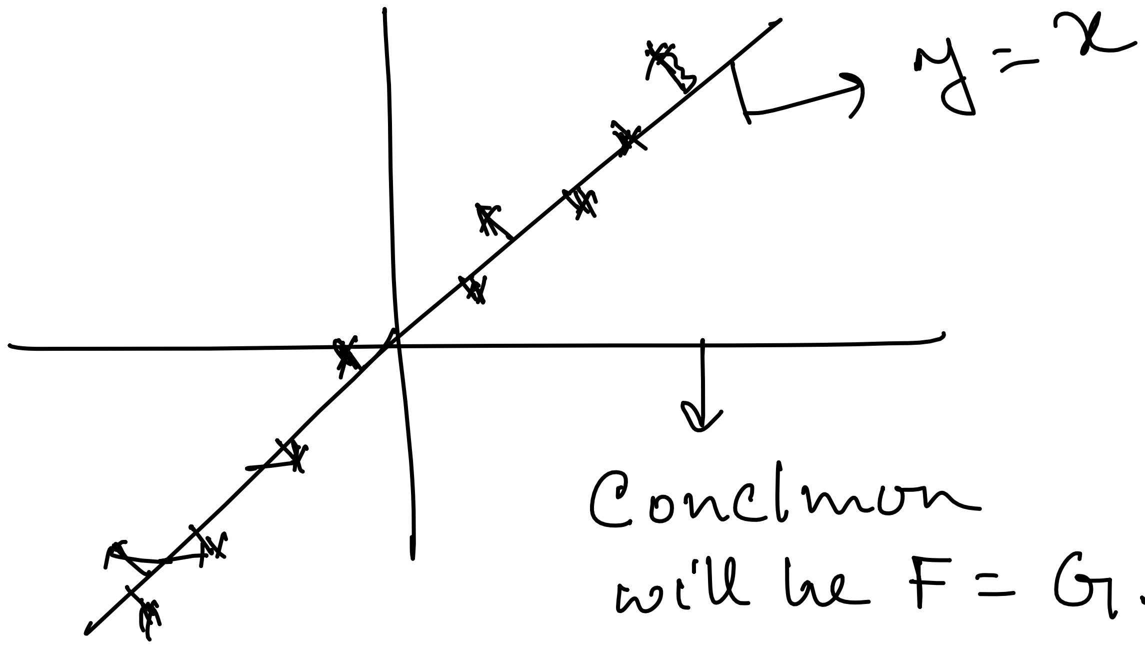
Step 1:- Compute $\frac{i}{n}$-th quantile and $\frac{j}{m}$-th quantile from both data set, where $i = 1, \dots, n$ & $j = 1, \dots, m$.

Hence, altogether, we will have $(n+m)$ many quantiles.

Step 2:- Plot those $(n+m)$ many quantiles.

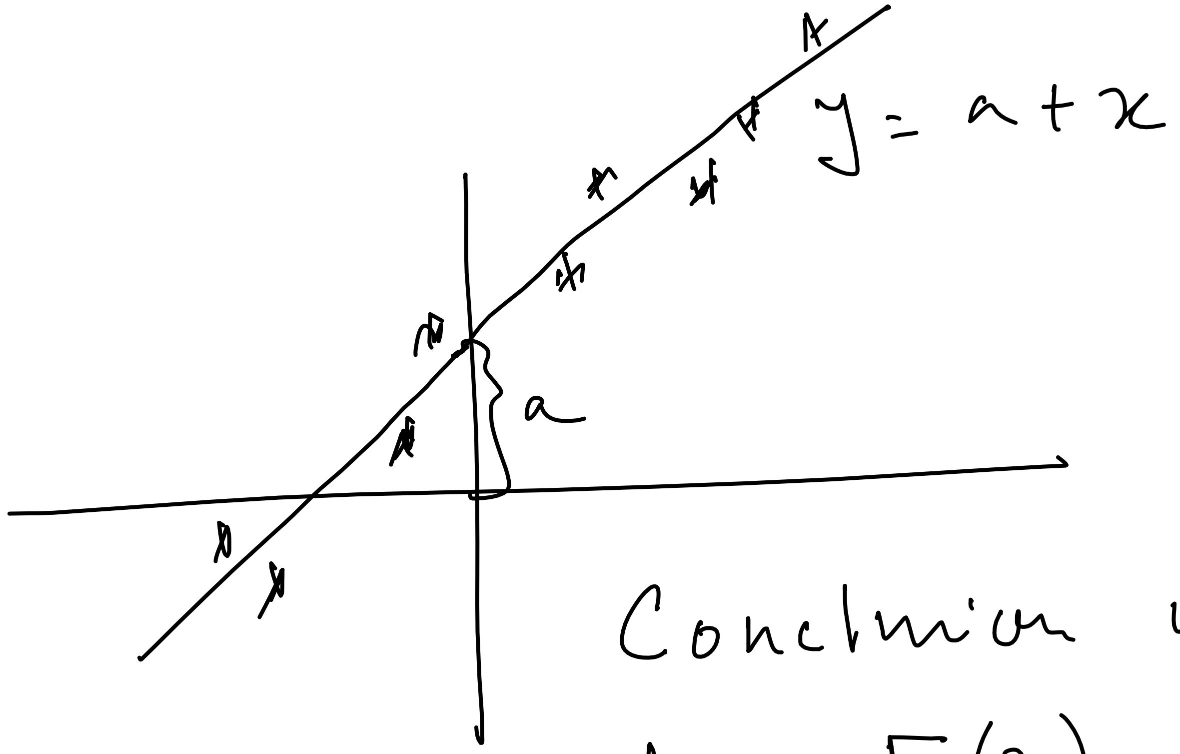
Implication:- (In the Q-Q plot).

(i)



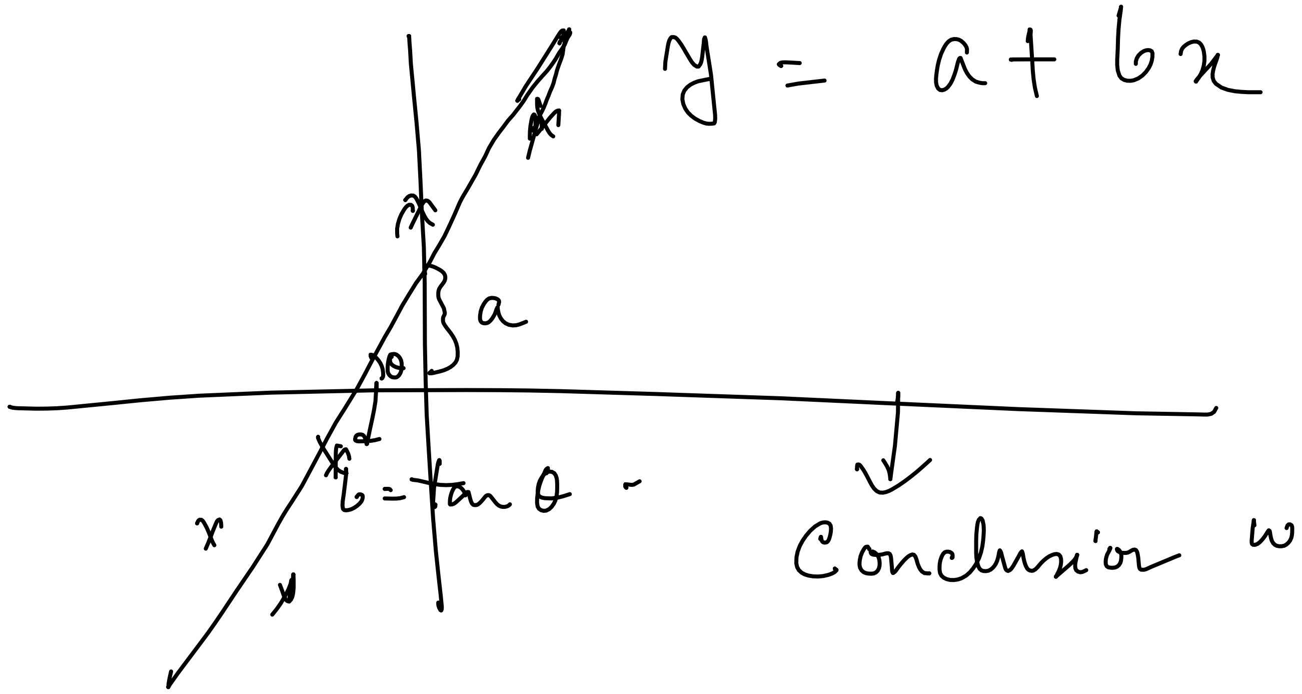
Conclusion will be $F = G$.

ii)



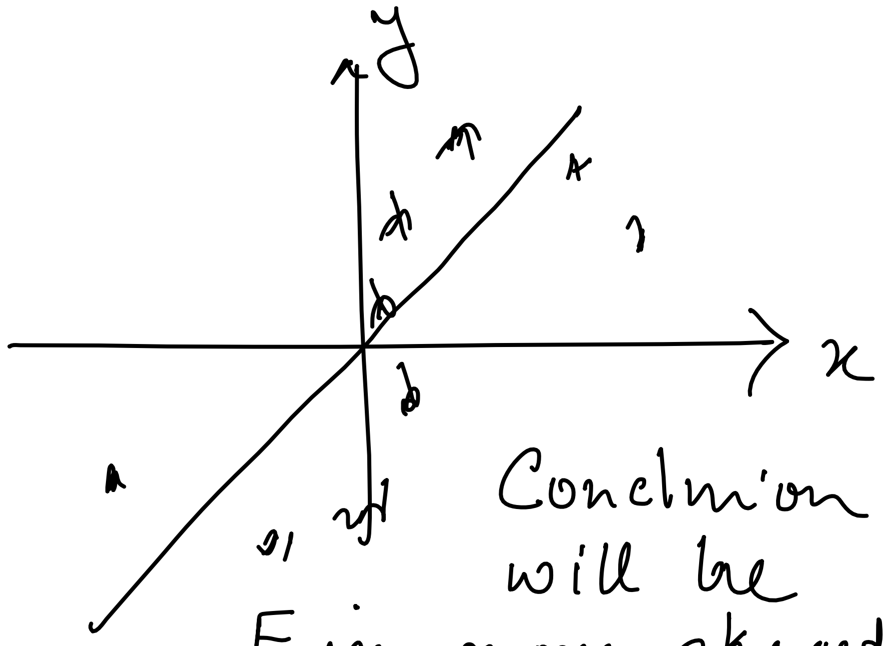
Conclusion will be $F(x) = G(x-a)$

iii)



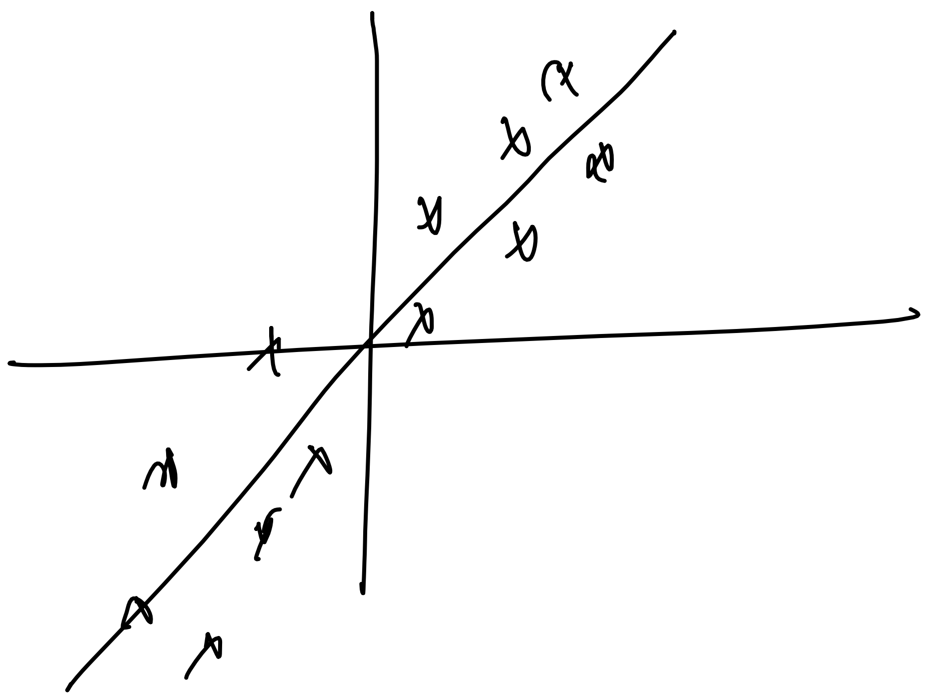
Conclusion will be $F(x) = G\left(\frac{x-a}{b}\right)$.

iv)



Conclusion will be F is more skewed than G.

v)



$$G(x) = (1-\beta)\,F(x) + \beta\, H(x),$$

where β is a small positive number.