

Johns Hopkins Engineering

625.464 Computational Statistics

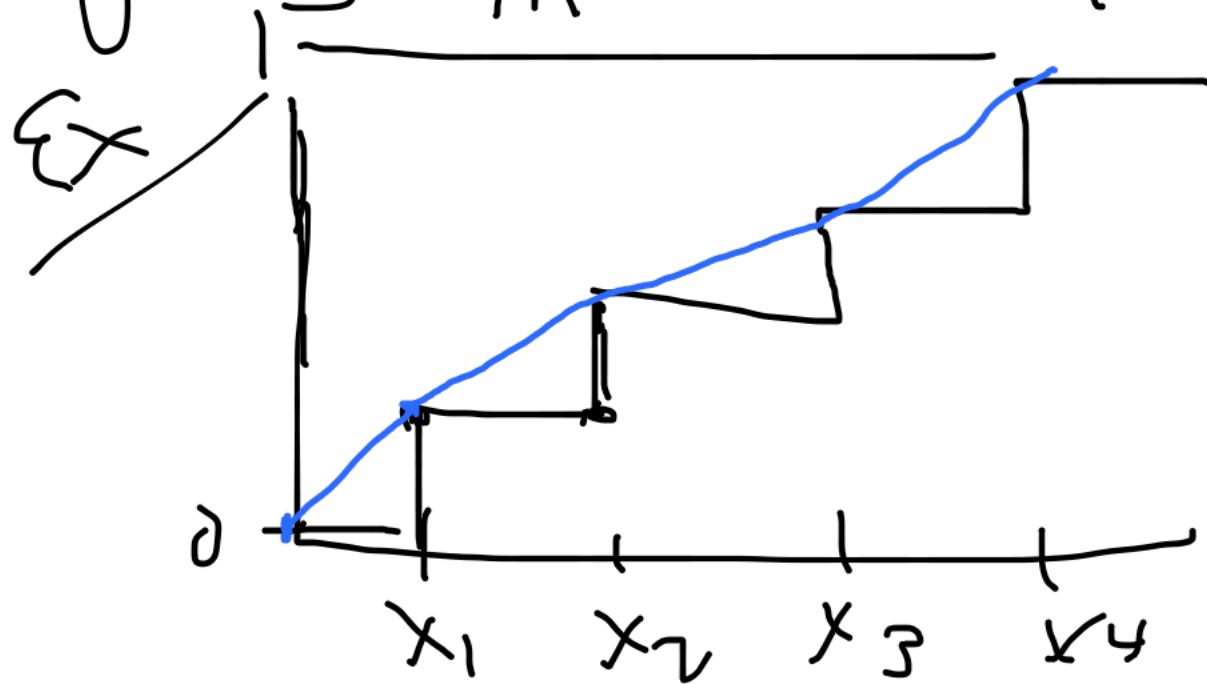
Graphical Methods The ECDF and Q-Q Plots

Module 13 Lecture 13B



The Empirical Cumulative Distribution Function (ECDF)

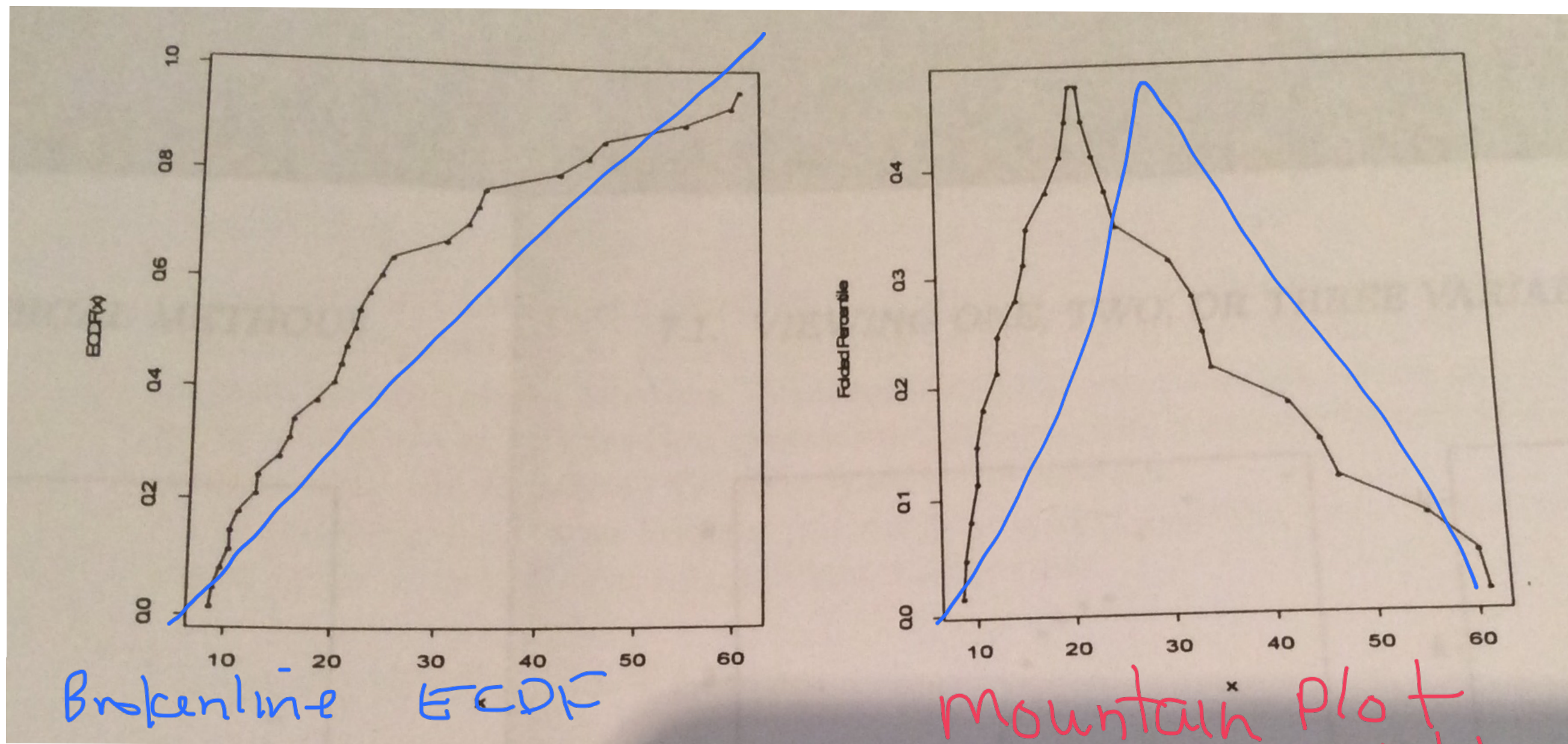
— It is a step function w/ an increase of size $1/n$ at each point in a sample of size n .



Broken line ECDF

connect the points $(x_i, i/n)$
 x_i are sorted

Broken line ECDF and the Mountain Plot



Data from Γ dist.

if $X \sim U$ the B-L ECDF
is a straight line

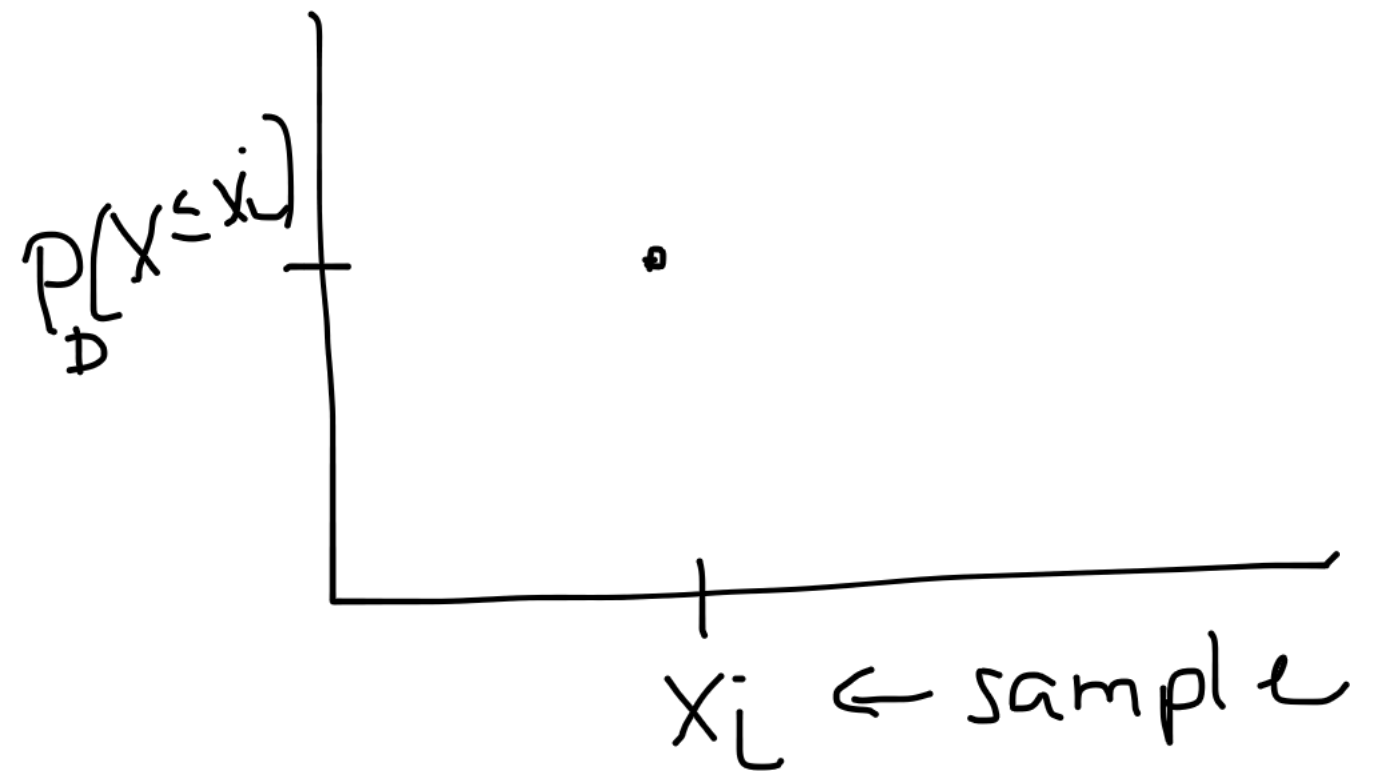
Mountain Plot
- folds ECDF at the
median

ECDF of unimodal is
concave, multimodal is a mix

Probability Plots

The vertical axis is transformed so that it corres. to the cum. dist. function of a given dist. D . ie we plot

$(x_i, \text{eval of CDF of } D \text{ at } x_i)$



If the B-L \in CDF
Plotted on this vert
axis yields a straight
line, then the sample is
probably from D .

Quantile-Quantile (Q-Q) Plot

Basically: the $\frac{j}{n}$ th quantile is plotted against the j th order stat. in the sample and so on to calculate empirical quantiles, the k th smallest value should correspond to a value of p approx equal to $\frac{k}{n}$ or $\frac{(k-1)}{n}$ (Also can split)

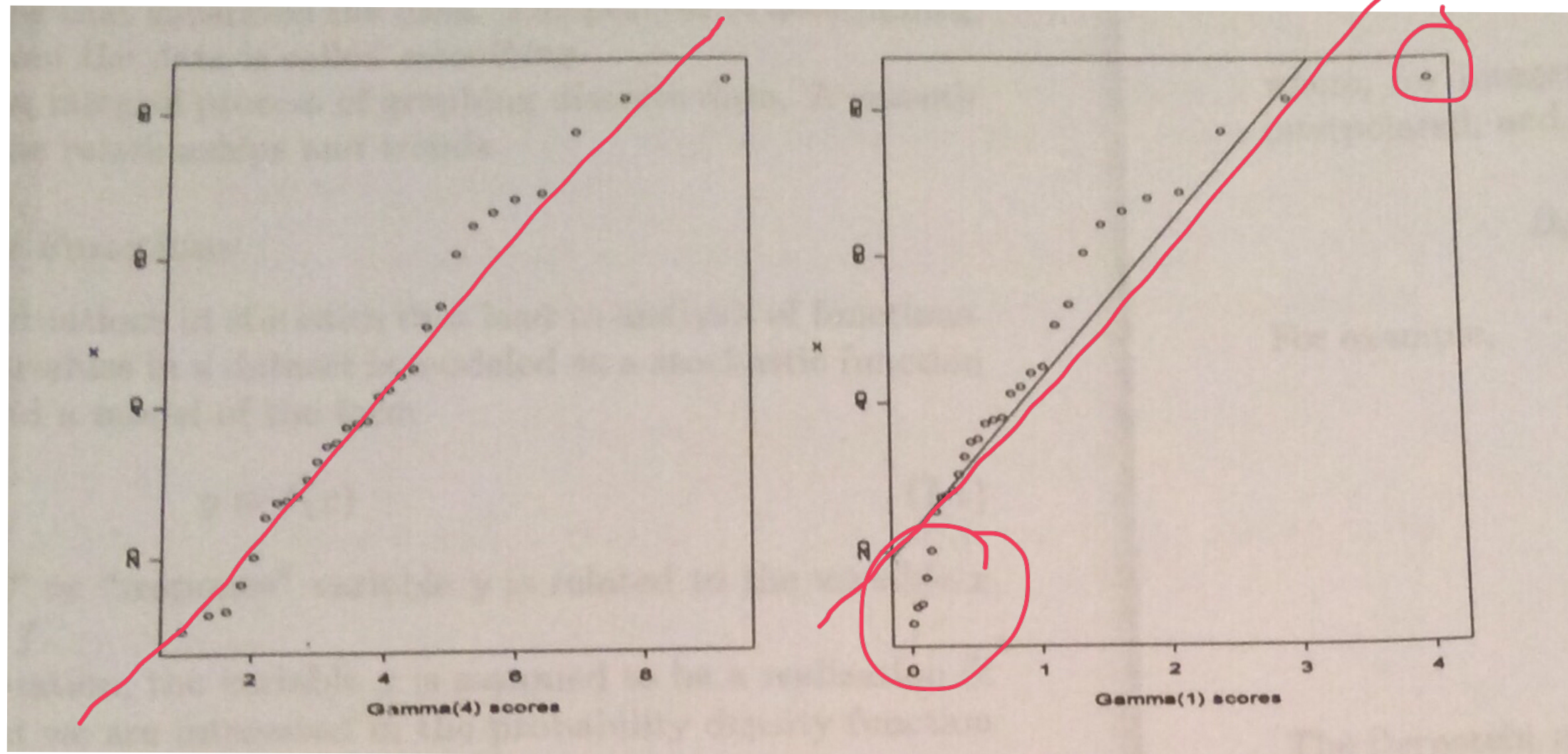
$$P_k[(k-1/2)/n]$$

Then the k th smallest value is the p_k th sample quantile

Q-Q Example

Data is Gamma(4)

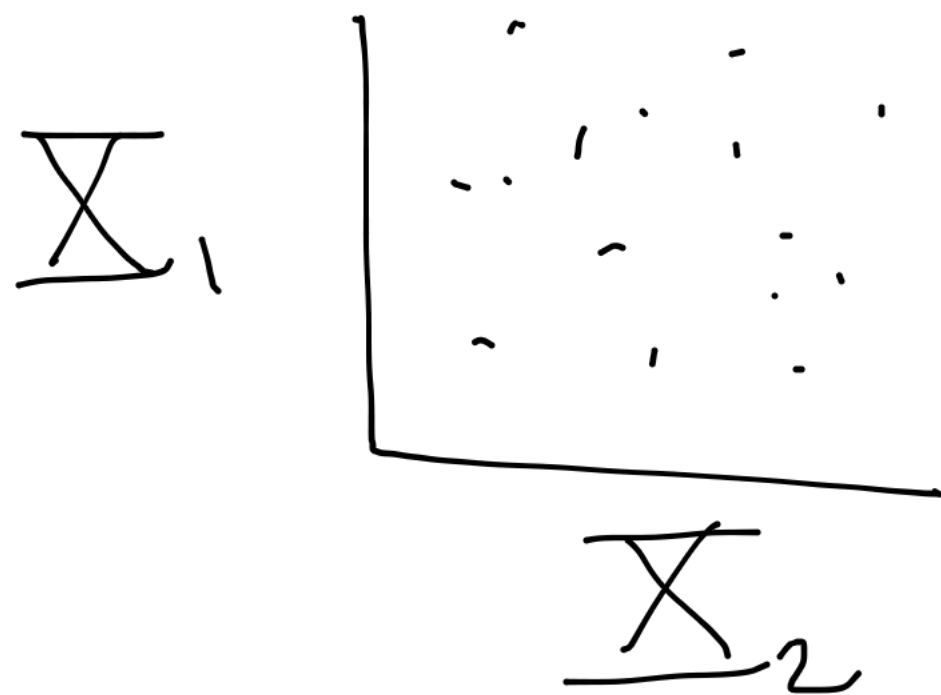
Don't use
Q-Q Plot func.
built in to R.



Points are $(\text{quantile } i/n, x(i))$
 small values below line \Rightarrow heavier left tail
 large values below line \Rightarrow lighter right tail

Scatter Plots

plots points on a cartesian axis
rep the variables



same dim as
the data