✱ Power law Distribution

pareto Distribution

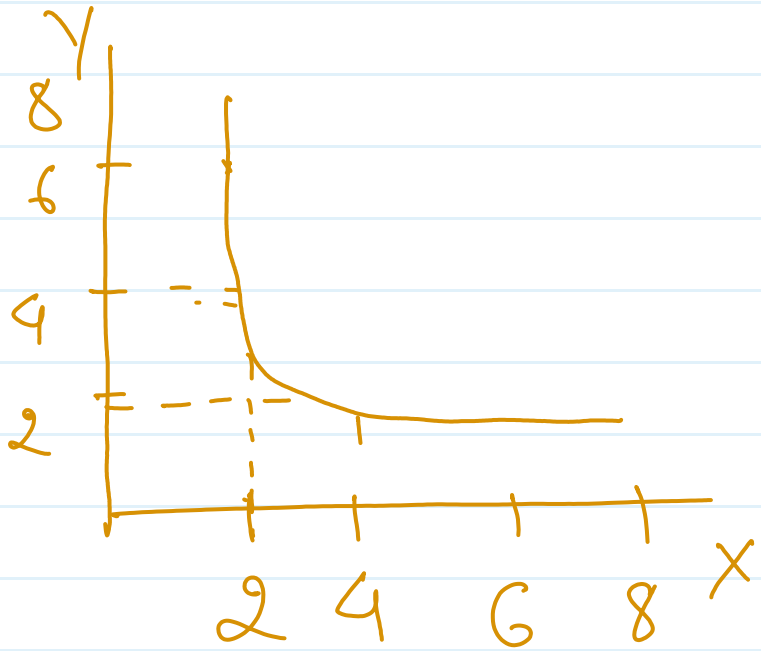
80 - 20

80% player - 20% run.

80% Run - 20% player

80% work - 20% employee

80% employee - 20%.



Pareto Dist. to normal Dist.

Box-Cox transformation

Pareto Dist. $X = \{X_1, X_2, X_3, \ldots, X_n\}$

Gaussian $y = \{Y_1, Y_2, Y_3, \ldots, Y_n\}$

① Box-Cox $(x) = \alpha$

α is height of the dist.

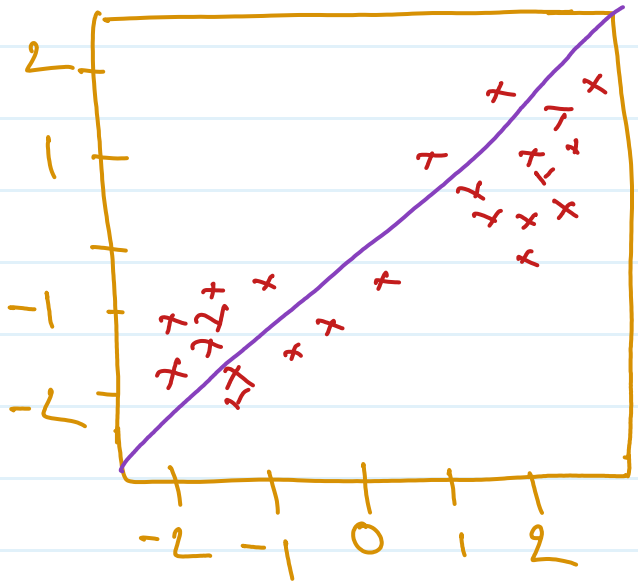
$$y_i = \frac{x_i^{\alpha} - 1}{\alpha}$$

if $\alpha \neq 0$, for Box-Cox transform

if $\alpha = 0$, then, $\log_e(x)$ = lognormal transformation

Q-Q. plot

(Quartile - Quantile plot)



This is non-Gaussian Dist

This is Gaussian Dist.

✱ chi square test

Def. – It is a non parametric test that is performed on categorical data

E.g. – In the year 2000 USA censun, the age of individual in a small town were found to be the following.

< 18	18 - 35	> 35
20%	30%	50%

In 2010, age of n = 500 individual were sample below are result

< 18	18 - 35	> 35
121	288	91

using $\alpha = 0.05$ can you conclud dist. of ages has been changed in

10 years.

Soln ⇒ (i) Null hypothesis

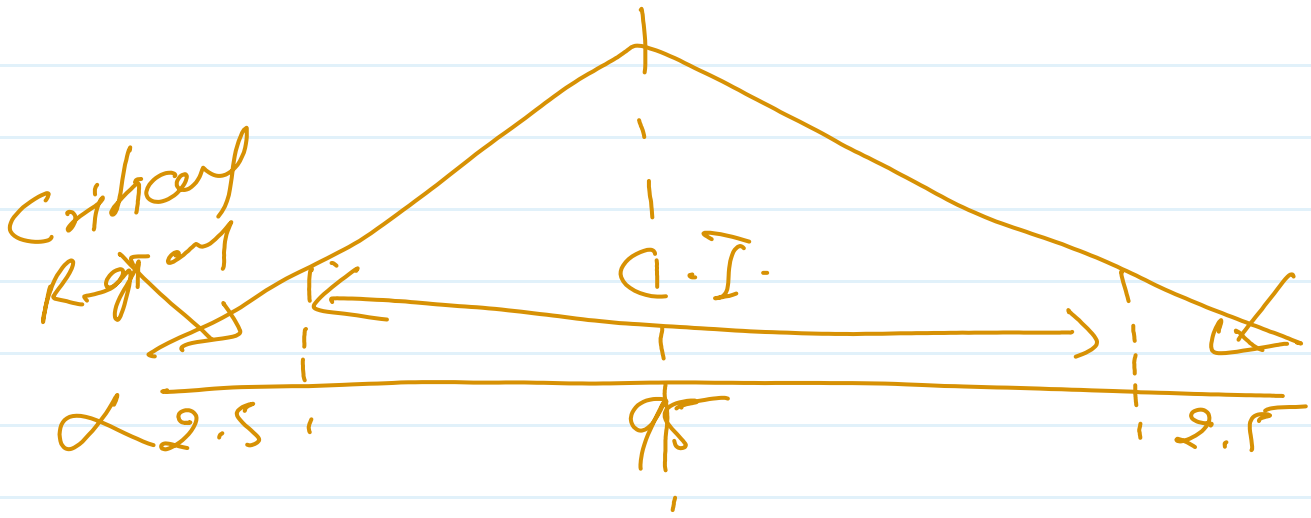
H_0 = the data dist. remain same

H_1 = the data dist. of ages has changed

$\alpha = 0.05$

means C.I. = $1 - 0.05$

⇒ 0.95



d.f. = $3 - 1$

$= 2$

for chi-square test ⇒

if x^2 is greater than 5.99, reject H_0

calculate

$$x^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

f_e = expected value

f_o = observed value

	<18	18-35	>35
f_o	121	288	91
f_e	$\frac{500 \times 20}{1000}$	$\frac{500 \times 30}{1000}$	$\frac{500 \times 50}{1000}$
	100	150	250

$$X^2 = \frac{(121-100)^2}{100} + \frac{(288-150)^2}{150} + \frac{(91-250)^2}{250}$$

$$\boxed{X^2 = 232.94}$$

value fall greater than 5.99,
so we can say distr. has changed
$\Rightarrow$ We reject null-hypothesis
$\Rightarrow$ Accept Alternet hypothesis

If value = less than 5.99
$\Rightarrow$ We fail to reject null-hypothesis.

E.g. (Poisson Dist)

A student receive mean. 7 text msg. in 2 hours.

(a) what is the probability that the student will recieve exactly 9 text msg. in a 2 hours.

$\Rightarrow$ $\mu = 7$

$x = 9$

$$P(x=9) = \frac{\mu^x e^{-\mu}}{x!}$$

$$= \frac{7^9 e^{-7}}{9!}$$

$$= \frac{40353607 \times -4.2817}{9\times8\times7\times6\times5\times4\times3\times2\times1}$$

$\Rightarrow$ 0.104 $\Rightarrow$ 10%.

Imp Type-I error , Type-II error

(I) Cancer → defected] H_0 = ✓
money] H_1 ✓

(II) Cancer → no defect] H_0 = ✓
money / wealth] H_1 ✗