## Chi-Square Distribution $(x^2)$

↳ A continuous prob. distribution.

↳ we do sum of squares of independent normal distribution

$$x^2 = z_1^2 + z_2^2 + z_3^2 \Rightarrow df = 3$$

$$x^2 = \sum_{i=1}^{K} z_i^2 \qquad df = K$$

↳ If we keep on adding then finally we get a normal dist.

↳ It has one parameter called dof

### Properties -

* only for non-negative value
* Positively skewed, degree of skewness decreases as dof↑
* mean = dof
  variance = 2 * dif

### Chi square test -

↳ It is used to determine -

→ If there is sig. association b/w categorical colmn

→ If an observed dist. of categorical data differs from expected menitand dist.

They are of 2 types -

1) Chi-Square goodness of fit test
   ↳ used only for 1 caty column.

↳ If observed dist of single caty-variable matches an expected theoritical dist.

↳ It helps to evaluate weather the data follows specific prob. dist.

## Steps –

→ $H_0$ = observed data follows expected theoritical dist.
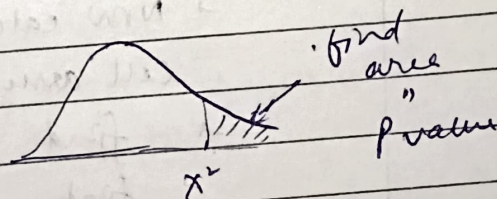
$H_1$ = Not follow $H_0$

→ Assuming it follows uniform dist.
→ Calculate expected freq

| Die roll | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 60 times | 5 | 15 | 5 | 15 | 5 | 15 |
| | 10 | 10 | 10 | 10 | 10 | 10 |

→ $X^2 = \dfrac{(5-10)^2}{10} + \dfrac{(15-10)^2}{10}$ & ...

test statistic



·find area
" p value

$X^2$

$df = \# \text{categories} - 1$

→ Assumptions –

↳ observation in sample must be indep.
↳ Categorized data
↳ Expected freq. must be atleast 5
↳ we must assume a theoritical dist
   that it follows –

· Parametric test ⇒ we assume dist follows a particular dist. & uses its parameters

**· Non-parametric -**
↳ we don't use any of the parameters.

**i)** Chi-square is non-parametric.
 goodness of fit

**ii)** Chi-Square test for independence (Asso..
 ↳ This test is used to determine
 wheather there is significant associetn
 b/w two categorized variables

**Steps -**
→ H0 ! There is no associet" b/w Caty. Glum,
 H1 !

 → Create contengency table with observed fre
 ~~assuming null hypo is true~~

 → Now calculate expected freq. for each
 cell assuming null hypo. is true
 → find chi-Square statis
 → find dof = ($\#$ rows $-1$) + (no. of col $-1$)
 → find p-value & compare with $\alpha$

**Assumptn -**
 · Independence of observers
 · must be categorized column
 · Adequate sample size -
   ↳ each cell value $> 5$
 · fixed marginal totals
   ↳ total of each row & column