

## F test or F Statistics

$$y = \beta_0 + \beta_1 x$$

We normally check,  $\beta_1 = 0$  or not. For that we use t test (One variable)

$$H_0: \beta_1 = 0, H_1: \beta_1 \neq 0 \text{ use t test.}$$

But if we have more than one variable, we use F test.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

$$H_0 = \beta_1 = \beta_2 = \beta_3 = 0$$

$$H_1 = \beta_1 \neq \beta_2 \neq \beta_3 \neq 0$$

$$F \text{ test} = \frac{(SSR_R - SSR_U) / P}{SSR_U / (n - P - 1)}$$

this give F statistic value.

F statistic is larger means we are more likely to reject Null hypothesis.

So why we check coefficient = 0?

Answer is if coefficient equal to zero then there will be no regression line means the variable which we took cannot fit regression line.

## Fisher's Exact test and Enrichment analysis

- R. A. Fisher invented this test in 1900's.

- In his experiment design, he explained one example of Lady testing tea.

- In this Lady testing tea, lady is saying that in a tea if milk is added first in a tea or a tea is added first in tea then milk. So she drank 8 cups of tea and starting predicting. 8 cups — [ 4 milk, 4 tea ]

- First problem is, they don't want lady to drink more cup of tea (injurious to health). So sample size is small.

- If the lady guessed correctly then table would be like  
So, when she said Milk is pour first, it poured first predicted  
and whenever she said tea pour first, it poured first.

	Guess.		
	Milk	Tea	
Milk	4	0	4
Tea	0	4	4
	4	4	8

$H_0$ : There is no association between which was pour first and her guess.  
(This association is random).

So total combination will be  $8C_4 = 70$ . 8 is the total cup she will take. 4 is the max she can get right for milk or tea because 8 cups — [ 4 milk, 4 tea ]

Success count	Permutation count	No of permutations
0	0000	$1 \times 1 = 1$
1	000X, 00X0, 0X00, X000	$4 \times 1 = 16$
2	00XX, 0X0X, X00X, XX00, 0XX0, X0X0	$6 \times 6 = 36$
3	0XXX, X0XX, XX0X, XXX0	$4 \times 1 = 16$
4	XXXX (Tea, Tea, Tea, Tea)	$1 \times 1 = 1$
Total		70

So there is only one way to get all 4 guesses correctly (Success count = 4). probability =  $\frac{1}{70}$ .

$P = \frac{1}{70} \approx 0.14 < 0.5$ . So we reject null hypothesis means there is association.

But the sample is too small 1 out of 70 does that.

If Success count = 3,  $P = \frac{16}{70} \approx 0.22 > 0.5$ . So pvalue > confidence interval. It accepts the null hypothesis and say there is no association.