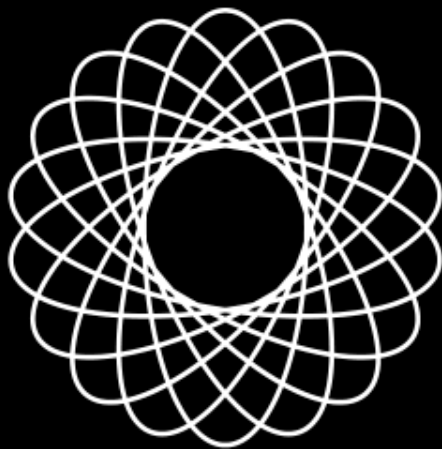# DATA
SCIENCE

Agenda

Anova
- One Way
- Two Way
- Post Hoc Tests

Chi Square
- **Association Tests**
- Goodness-of-Fit Tests

Chi Square Parametric
- Tests of Variance

Chi-Square Tests

A more complex example:

You look at preferences for beverages by age to understand if there is an association between age and brand preference, in order to decide if you need differentiated marketing strategies by age

Chi-Square Tests

A more complex example:

You look at preferences for beverages by age to understand if there is an association between age and brand preference, in order to decide if you need differentiated marketing strategies by age

You do a survey on a random sample, and get the following results:

Observed	Preference			
Brand	M 15-25	M 26-40	M 41-55	Total
Coke	49	50	69	168
Pepsi	24	36	38	98
Sprite	19	22	28	69
Total	92	108	135	335

Chi-Square Tests

A more complex example:

We want to check if the preference for a brand changes as age changes i.e., is there an association between brand preference and age, or are they independent?

Observed	Preference			
Brand	M 15-25	M 26-40	M 41-55	Total
Coke	49	50	69	168
Pepsi	24	36	38	98
Sprite	19	22	28	69
Total	92	108	135	335

Chi-Square Tests

A more complex example:

We want to check if the preference for a brand changes as age changes i.e., is there an association between brand preference and age, or are they independent?

Expected value – calculate the expected values under the assumption that the null hypothesis is true

Observed	Preference			
Brand	M 15-25	M 26-40	M 41-55	Total
Coke	49	50	69	168
Pepsi	24	36	38	98
Sprite	19	22	28	69
Total	92	108	135	335

Chi-Square Tests

A more complex example:

Mathematical calculation :

Observed	Preference			
Brand	M 15-25	M 26-40	M 41-55	Total
Coke	49	50	69	168
Pepsi	24	36	38	98
Sprite	19	22	28	69
Total	92	108	135	335

Expected	Preference			
Brand	M 15-25	M 26-40	M 41-55	Total
Coke	46.14	54.16	67.70	168
Pepsi	26.91	31.59	39.49	98
Sprite	18.95	22.24	27.81	69
Total	92	108	135	335

Chi-Square Tests

<u>A more complex example:</u>

Mathematical calculation :

Expected Values = (Row Total * Column Total) / n,

Observed	Preference			
Brand	M 15-25	M 26-40	M 41-55	Total
Coke	49	50	69	168
Pepsi	24	36	38	98
Sprite	19	22	28	69
Total	92	108	135	335

Expected	Preference			
Brand	M 15-25	M 26-40	M 41-55	Total
Coke	46.14	54.16	67.70	168
Pepsi	26.91	31.59	39.49	98
Sprite	18.95	22.24	27.81	69
Total	92	108	135	335

Chi-Square Tests

A more complex example:

Mathematical calculation :

**Expected Values = (Row Total * Column Total) / n,
where n is total number of observations in sample**

Observed	Preference			
Brand	M 15-25	M 26-40	M 41-55	Total
Coke	49	50	69	168
Pepsi	24	36	38	98
Sprite	19	22	28	69
Total	92	108	135	335

Expected	Preference			
Brand	M 15-25	M 26-40	M 41-55	Total
Coke	46.14	54.16	67.70	168
Pepsi	26.91	31.59	39.49	98
Sprite	18.95	22.24	27.81	69
Total	92	108	135	335

Chi-Square Tests

A more complex example:

Now that we have observed and expected values, we can either manually calculate a Chi Square test statistic,

Or use a tool – like Excel

Observed	Preference				Expected	Preference			
Brand	M 15-25	M 26-40	M 41-55	Total	Brand	M 15-25	M 26-40	M 41-55	Total
Coke	49	50	69	168	Coke	46.13731	54.16119	67.70149	168
Pepsi	24	36	38	98	Pepsi	26.91343	31.59403	39.49254	98
Sprite	19	22	28	69	Sprite	18.94925	22.24478	27.80597	69
Total	92	108	135	335	Total	92	108	135	335

=chitest(B3:D5,H3:J5

CHITEST(actual_range, **expected_range**)

Agenda

Anova
- One Way
- Two Way
- Post Hoc Tests

Chi Square
- Association Tests
- **Goodness-of-Fit Tests**

Chi Square Parametric
- Tests of Variance

Chi-Square Tests

Very popular use of Chi Square, Goodness-of-fit tests if the data follows a particular distribution or not

Chi-Square Tests

Very popular use of Chi Square, Goodness-of-fit tests if the data follows a particular distribution or not

Example:

A gambler is playing a new game in a casino, which
involves rolling three dice at a time. Winnings are
directly proportional to the number of 6's rolled.

Chi-Square Tests

Very popular use of Chi Square, Goodness-of-fit tests if the data follows a particular distribution or not

Example:

A gambler is playing a new game in a casino, which involves rolling three dice at a time. Winnings are directly proportional to the number of 6's rolled.

This is what is observed in 100 rolls of the dice

Number of 6's	Rolls
0	48
1	35
2	15
3	2

Chi-Square Tests

Very popular use of Chi Square, Goodness-of-fit tests if the data follows a particular distribution or not

Example:

A gambler is playing a new game in a casino, which involves rolling three dice at a time. Winnings are directly proportional to the number of 6's rolled.

This is what is observed in 100 rolls of the dice

Would you have cause to believe that the gambler is maybe "too" lucky, and is playing with loaded dice?

Number of 6's	Rolls
0	48
1	35
2	15
3	2

Chi-Square Tests

What distribution would you expect the outcome of seeing a 6 on rolled dice to follow?

- **Binomial**

Chi-Square Tests

What distribution would you expect the outcome of seeing a 6 on rolled dice to follow?

– **Binomial**

What should the expected probabilities be of number of 6's in three rolled dice?

Chi-Square Tests

What distribution would you expect the outcome of seeing a 6 on rolled dice to follow?

– **Binomial**

What should the expected probabilities be of number of 6's in three rolled dice?

– **Calculate using the Binomial Distribution formula**

Chi-Square Tests

What distribution would you expect the outcome of seeing a 6 on rolled dice to follow?

– **Binomial**

What should the expected probabilities be of number of 6's in three rolled dice?

– **Calculate using the Binomial Distribution formula**

=BINOM.DIST(O26,3,1/6,FALSE)

M	N	O	P	Q
	Expected	Number of 6's	Expected Prob	Expected 6's in 100 throws
		0	0.5787	57.8704
		1	0.34722	34.7222
		2	0.06944	6.94444
		3	0.00463	0.46296

Chi-Square Tests

Testing for a normal distribution or any type of distribution

- – Calculate expected distribution using the probability distribution formula

Chi-Square Tests

Testing for a normal distribution or any type of distribution

- Calculate expected distribution using the probability distribution formula

- If checking for normality, for example, follow these steps:

   1. Calculate mean and std deviation of your data

Chi-Square Tests

Testing for a normal distribution or any type of distribution

- Calculate expected distribution using the probability distribution formula

- If checking for normality, for example, follow these steps:

    1. Calculate mean and std deviation of your data
    2. Bin the data into sub-intervals

Chi-Square Tests

Testing for a normal distribution or any type of distribution

- Calculate expected distribution using the probability distribution formula

- If checking for normality, for example, follow these steps:

    1. Calculate mean and std deviation of your data
    2. Bin the data into sub-intervals
    3. Calculate expected probability of those sub-intervals (using normal probability function)

Chi-Square Tests

Testing for a normal distribution or any type of distribution

- Calculate expected distribution using the probability distribution formula

- If checking for normality, for example, follow these steps:

  1. Calculate mean and std deviation of your data
  2. Bin the data into sub-intervals
  3. Calculate expected probability of those sub-intervals (using normal probability function)
  4. Compare that to frequency observed in data

Chi-Square Tests

Testing for a normal distribution or any type of distribution

- Calculate expected distribution using the probability distribution formula

- If checking for normality, for example, follow these steps:

    1. Calculate mean and std deviation of your data
    2. Bin the data into sub-intervals
    3. Calculate expected probability of those sub-intervals (using normal probability function)
    4. Compare that to frequency observed in data
    5. Construct Chi Square and test

Coming Up

Chi Square Parametric:

- Tests of Variance

THANK YOU