Chapter 3 part 2

Lamiae HANA

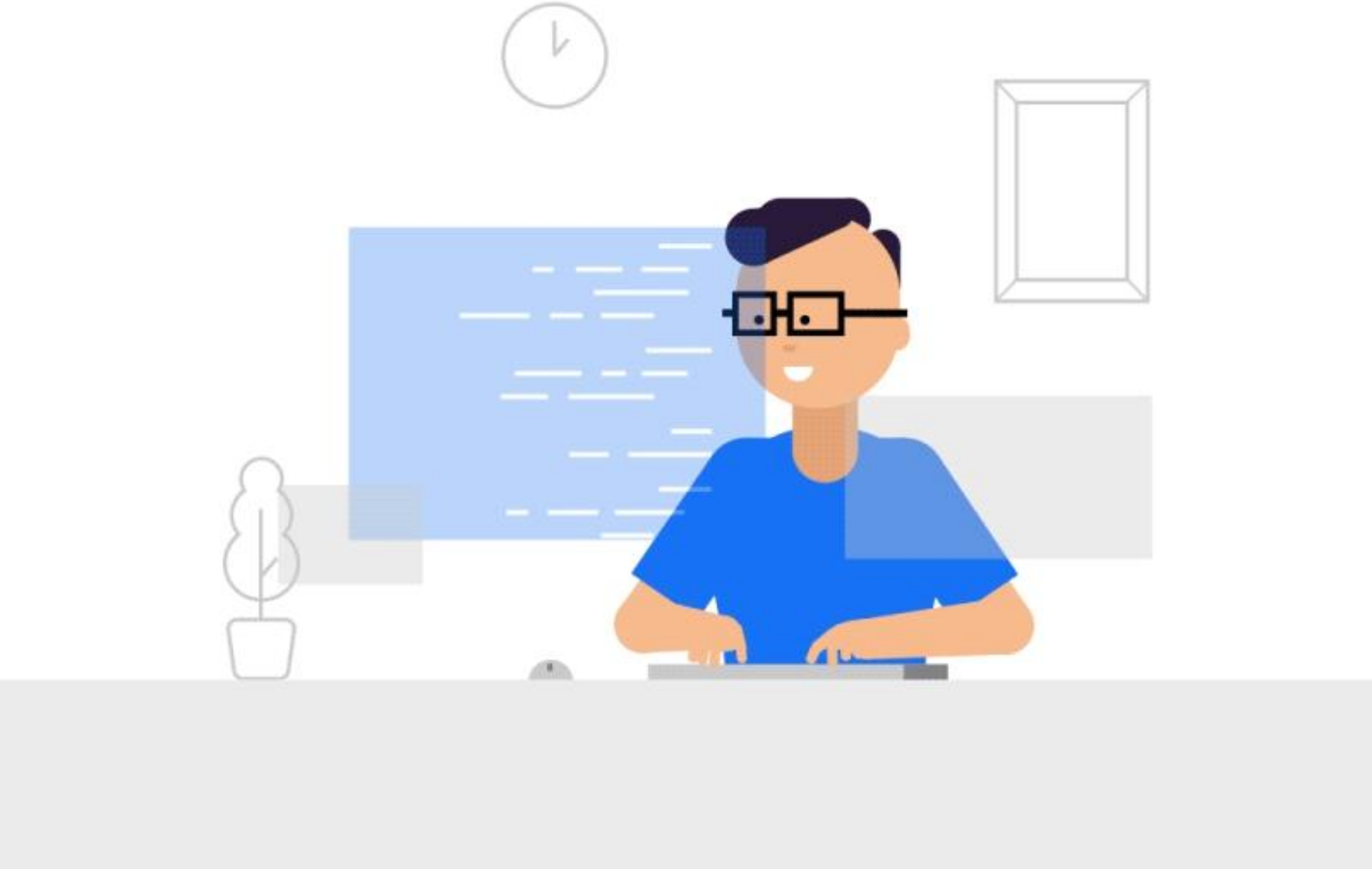
Second Part

- Degrees of Freedom
- ANOVA
- Chi-Square Test
- Multi-Arm Bandit Algorithm
- Power and Sample Size

Degrees of Freedom

Degrees of freedom are often broadly defined as the number of "observations" (pieces of information) in the data that are free to vary when estimating statistical parameters.

> *Degrees of Freedom = n – 1*



Test the population mean with a sample of 10 values, using a 1-sample t test

Constraint

The estimation of the mean

What is that constraint, exactly?

The sum of all values in the data must equal *n* x mean

$$\boxed{X} + \boxed{X} + X + X + X + X + X + X + X + X = 10 * 3{,}5 = 35$$

mean = 3.5

In fact, the first 9 values could be anything, including these two examples:

34, -8.3, -37, -92, -1, 0, 1, -22, 99
0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9

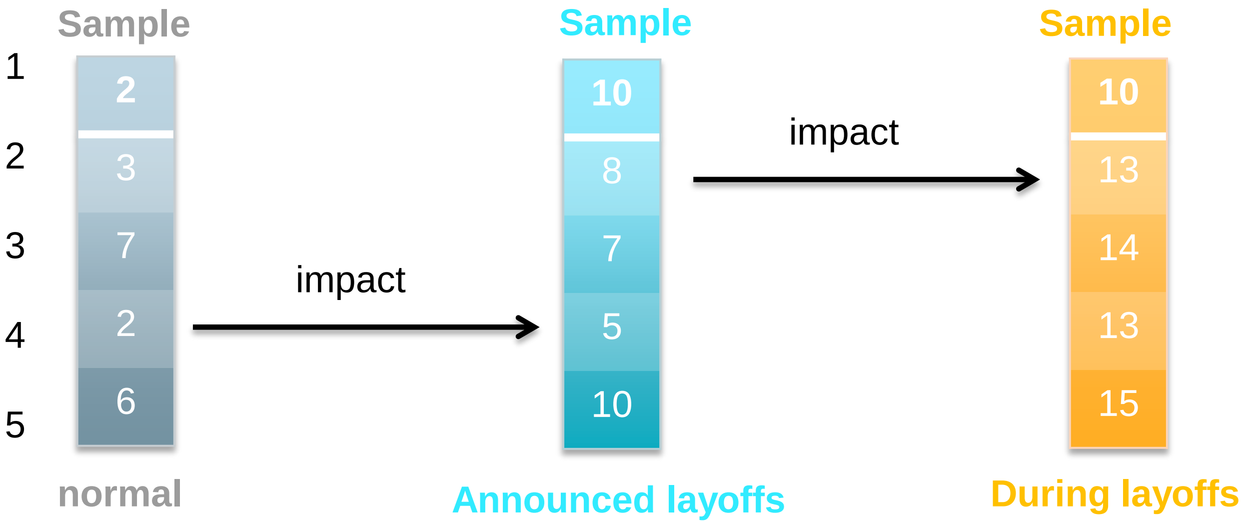
But to have all 10 values sum to 35, and have a mean of 3.5, the 10^{th} value *cannot* vary. It must be a specific number:

34, -8.3, -37, -92, -1, 0, 1, -22, 99 → 10^{TH} value *must* be 61.3
0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 → 10^{TH} value *must* be 30.5

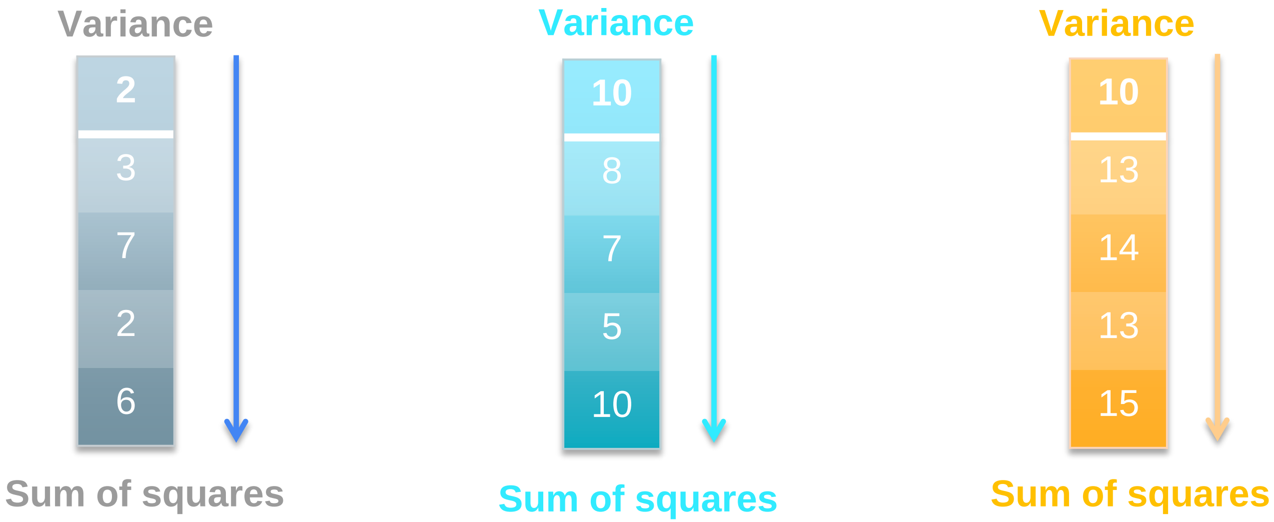
The number of degrees of freedom equals:

The number of "observations" — **The number of required relations among the observations**

ANOVA : **Analysis of Variance**

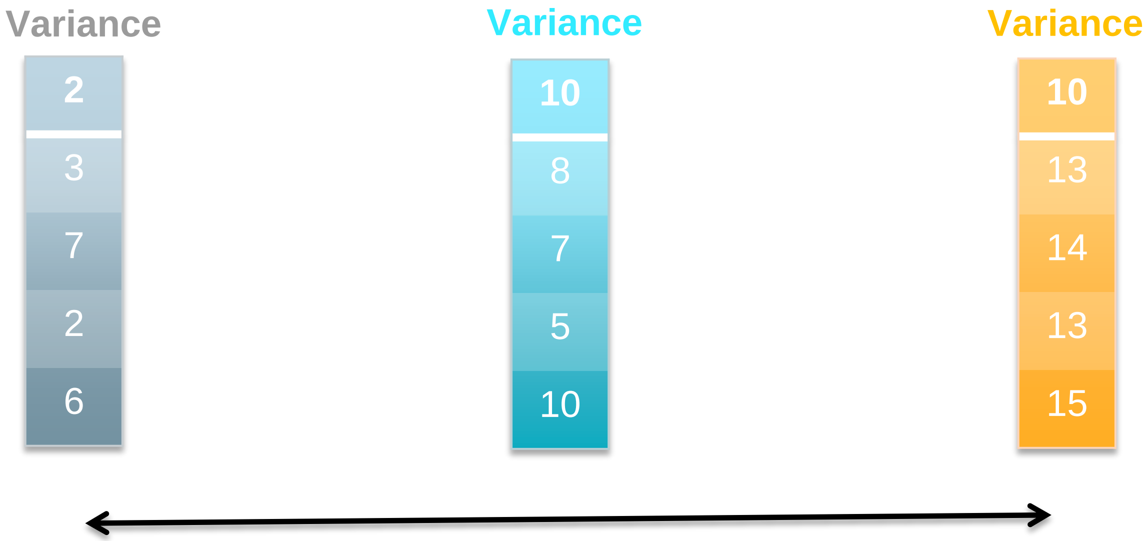


ANOVA : **Analysis of Variance**



Sum of squares Within groups

ANOVA : **Analysis of Variance**



Variance between groups

Sum of squares

Sum of squares Between Groups

ANOVA : **Analysis of Variance**

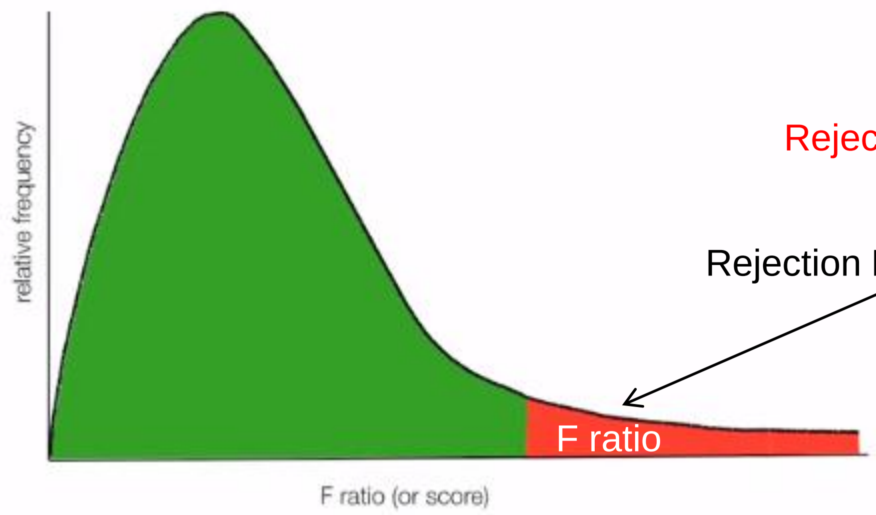
2	**10**	**10**
3	8	13
7	7	14
2	5	13
6	10	15

Variance from the mean

Total Sum of Squares

F ratio = F statistic

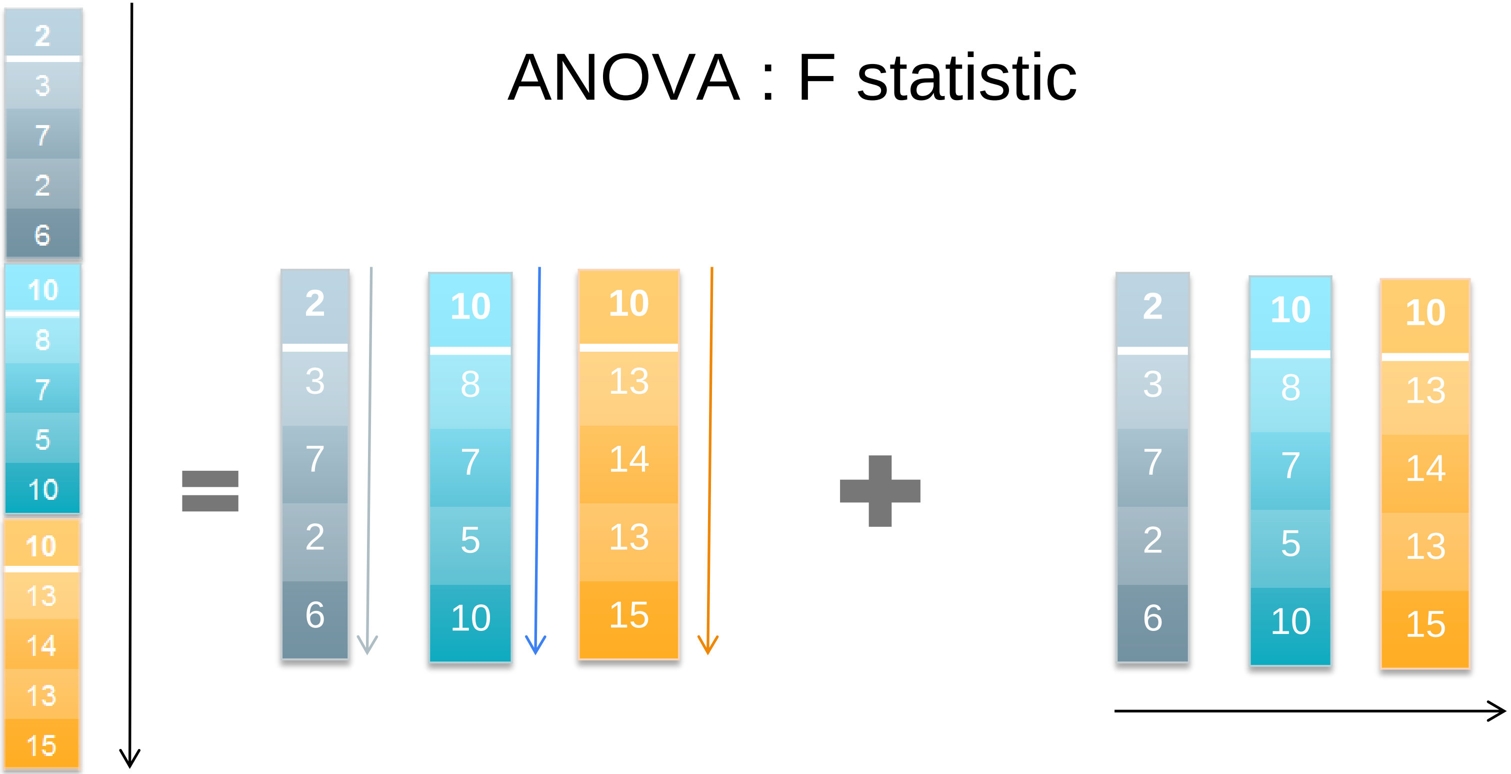
Fall to reject null hypothesis



Reject nul hypothesis

Rejection Region

ANOVA : F statistic



Total Sum of Squares =	Sum of Squares Within Groups	+	Sum of Squares Between Groups
257.3 =	54	+	203.3

ANOVA : F statistic

Total Sum of Squares = Sum of Squares Within Groups + Sum of Squares Between Groups

ANOVA : **Analysis of Variance**

	Sample	Sample	Sample
	2	**10**	**10**
	3	8	13
	7	7	14
	2	5	13
	6	10	15
Mean	**4**	**8**	**13**

ANOVA : **Analysis of Variance**

Sample

2	-	$= -2^2$	4
3	- 4	$= -1^2$	1
7	- 4	$= 3^2$	9
2	- 4	$= -2^2$	4
6	- 4	$= 2^2$	4
4	**sum**		22

$\sum(\text{Observation} - \text{mean})^2$

Sample

10	-	$= 2^2$	4
8	- 8	$= 0^2$	0
7	- 8	$= -1^2$	1
5	- 8	$= -3^2$	9
10	- 8	$= 2^2$	4
8			18

Sample

10	-	$= -3^2$	9
13	- 13	$= 0^2$	0
14	- 13	$= 1^2$	1
13	- 13	$= 0^2$	0
15	- 13	$= 2^2$	4
13			14

Sum of squares within Groups (SSW)= + + = 54

ANOVA : F statistic

Total Sum of Squares = Sum of Squares Within Groups + Sum of Squares Between Groups

54

observation		mean	observation - mean	(observation - mean) ^{2}
2	-	8.3	= -6.3	40.1
3	-	8.3	= -5.3	28.4
7	-	8.3	= -1.3	1.8
2	-	8.3	= -6.3	40.1
6	-	8.3	= -2.3	5.4
10	-	8.3	= 1.7	2.7
8	-	8.3	= -0.3	0.1
7	-	8.3	= -1.3	1.8
5	-	8.3	= -3.3	11.1
10	-	8.3	= 1.7	2.8
10	-	8.3	= 1.7	2.8
13	-	8.3	= 4.7	21.8
14	-	8.3	= 5.7	32.1
13	-	8.3	= 4.7	21.8
15	-	8.3	= 6.7	44.4

Total **S**um of **S**quares

SST = 257.3

ANOVA : F statistic

Total Sum of Squares = Sum of Squares Within Groups + Sum of Squares Between Groups

257.5

54

Analysis of Variance

Sum of Squares Between Groups

2
3
7
2
6
10
8
7
5
10
10
13
14
13
15
mean

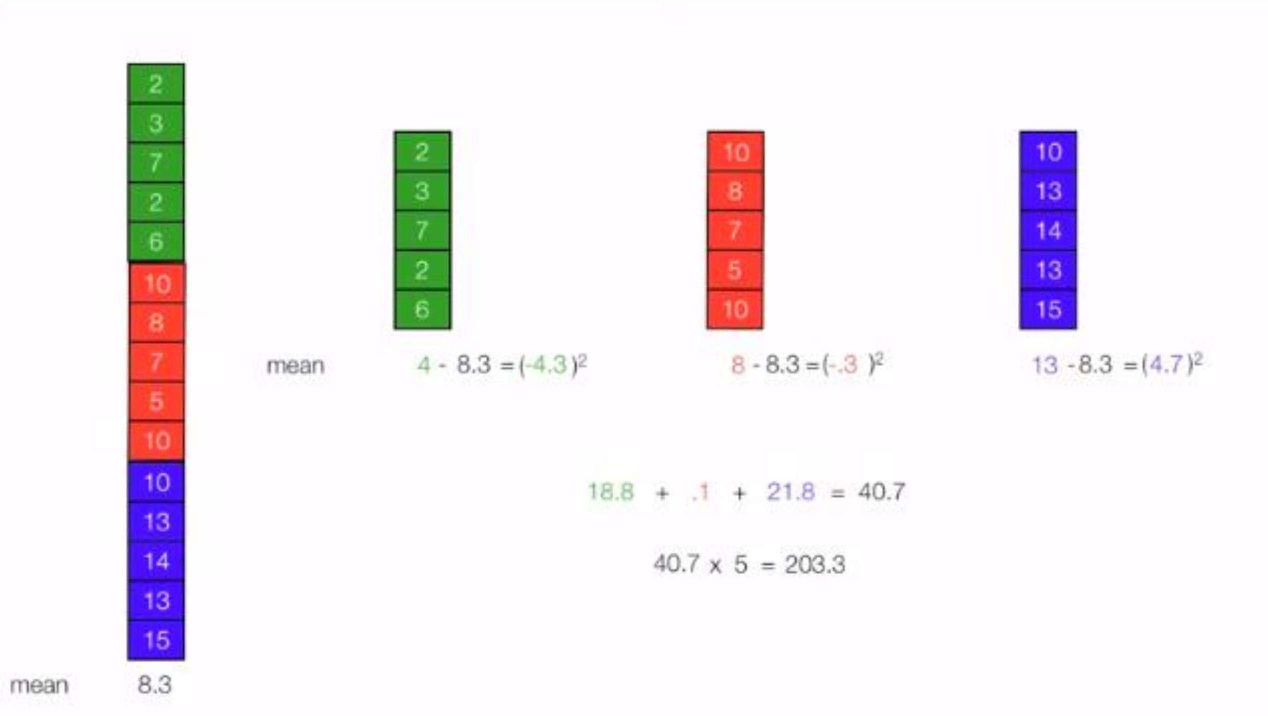
2	10	10
3	8	13
7	7	14
2	5	13
6	10	15
mean	mean	mean

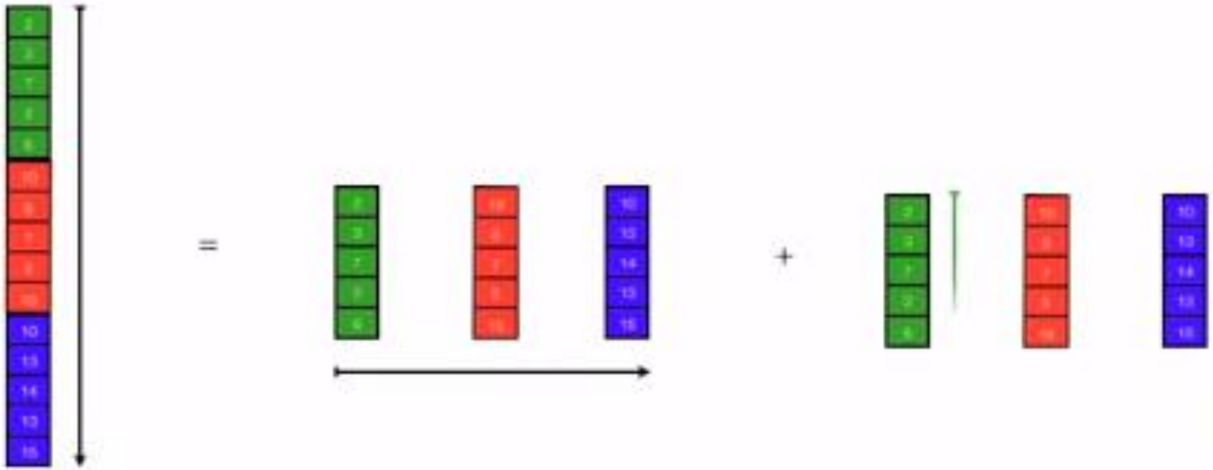
1. mean - mean   mean - mean   mean - mean

2. (mean - mean)^2 (mean - mean)^2 (mean - mean)^2

3. (mean - mean)^2 + (mean - mean)^2 + (mean - mean)^2

4. (mean - mean)^2 + (mean - mean)^2 + (mean - mean)^2 x 5


2
3
7
2
6
10
8
7
5
10
10
13
14
13
15
mean 8.3
2
3
7
2
6
10
8
7
5
10
10
13
14
13
15
mean
4 - 8.3 = (-4.3)²
8 - 8.3 = (-.3)²
13 - 8.3 = (4.7)²
18.8 + .1 + 21.8 = 40.7
40.7 x 5 = 203.3



Total Sum of Squares = Sum of Squares Between Groups + Sum of Squares Within Groups

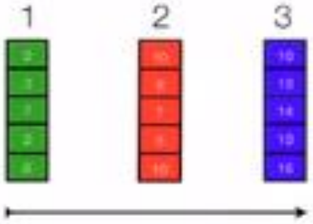
257.3 = 203.3 + 54

Final Calculations

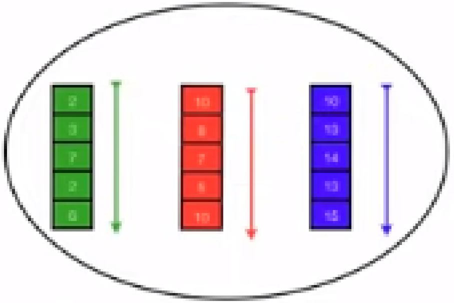
$$\frac{\text{Sum of Squares Between Groups}}{\text{degrees of freedom}} = \frac{203.3}{2} = 101.667$$

Groups - 1

3 - 1 =



Final Calculations



15 - 3 = 12

Observations - Groups

$$\frac{\text{Sum of Squares Within Groups}}{\text{degrees of freedom}} = \frac{54}{12}$$

Final Calculations

$$\frac{\text{Sum of Squares Between Groups}}{\text{degrees of freedom}} = \frac{203.3}{2} = 101.667$$

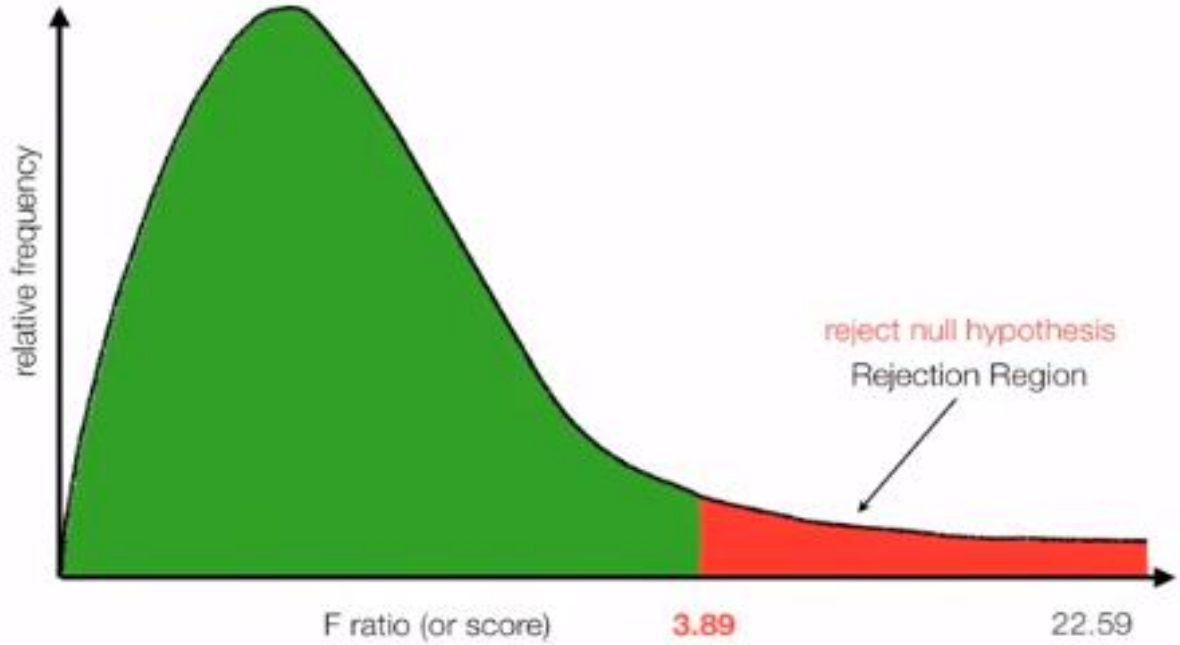
$$\boxed{F = \frac{101.667}{4.5} = 22.59}$$

$$\frac{\text{Sum of Squares Within Groups}}{\text{degrees of freedom}} = \frac{54}{12} = 4.5$$

F Distribution $F(2, 12) = 22.59$, $p<.05$

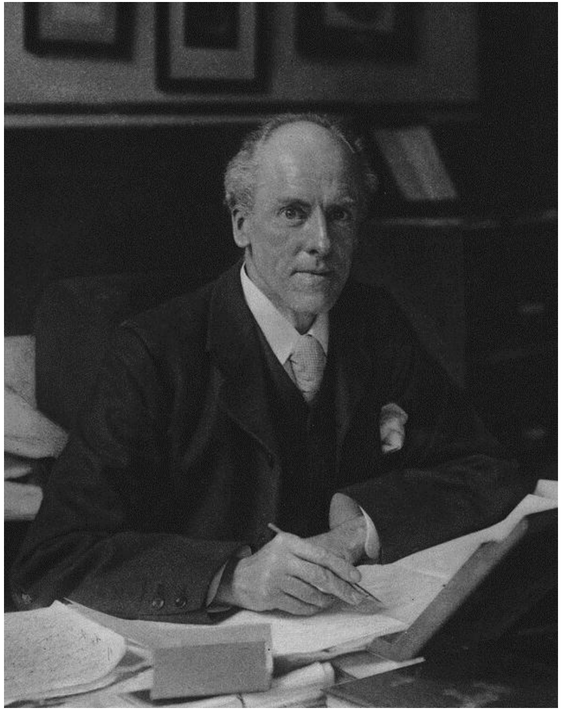
degrees of freedom numerator

degrees of freedom denominator	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	161.5	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	246.0	248.0	249.1	250.1
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25


relative frequency
reject null hypothesis
Rejection Region
F ratio (or score)
3.89
22.59

Chi-square test

X



What is a Chi-square test?

What is a Chi-square test?

A Chi-square test is a hypothesis testing method. Two common Chi-square tests involve checking if observed frequencies in one or more categories match expected frequencies.

What are my choices?

Chi-square test

Chi-square goodness of fit test

Chi-square test of independence

You can use the test when you have counts of values for (1 or 2) categorical variable.

	Chi-Square Goodness of Fit Test	Chi-Square Test of Independence
Number of variables	One	Two
Purpose of test	Decide if one variable is likely to come from a given distribution or not	Decide if two variables might be related or not
Example	Decide if bags of candy have the same number of pieces of each flavor or not	Decide if movie goers' decision to buy snacks is related to the type of movie they plan to watch
Hypotheses in example	H_o: proportion of flavors of candy are the same H_a: proportions of flavors are not the same	H_o: proportion of people who buy snacks is independent of the movie type H_a: proportion of people who buy snacks is different for different types of movies
Theoretical distribution used in test	Chi-Square	Chi-Square
Degrees of freedom	Number of categories minus 1 •In our example, number of flavors of candy minus 1	Number of categories for first variable minus 1, multiplied by number of categories for second variable minus 1 •In our example, number of movie categories minus 1, multiplied by 1 (because snack purchase is a Yes/No variable and 2-1 = 1)

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

χ^2 = chi squared

O_i = observed value

E_i = expected value

Chi-square test answer the question: is that a due to chance or something wrong with my method

100 times: 65 head and 35 tail

Null hypothesis: there is no significant difference between the observed and expected frequencies

Observed value

50 head and 50 tail

Expected value

Degree of freedom = 2 – 1 = 1

Critical value we will use is 0.05

Degree of Freedom	Probability of Exceeding the Critical Value								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38

Accept

Reject

Multi-Arm Bandit Algorithm


Ethiopia
Colombia
Kenya
Guatemala
Wine
Berry
Fruit
Spice
Floral
New
Guinea
5000 Ft
Very high
Altitude
Java
Sumatra
Citrus
Vanilla
Chocolate
Nut
4000 Ft
High
Altitude
Nicaragua
Mexico
Altura
Brazil
Bourbon
Santos
Hawaiian
Kona
Medium
Altitude
3000 Ft
Smooth
Sweet
Low Altitude
2500 Ft
Subtle
Soft
Mild
Simple
Bland
Coffee
SIMONELLI

Multi-Arm Bandit Algorithm

- The dilemma in our coffee tasting experiment arises from incomplete information. In other words, we need to gather enough information to formulate the best overall strategy and then explore new actions. This will eventually lead to minimizing the overall bad experiences.

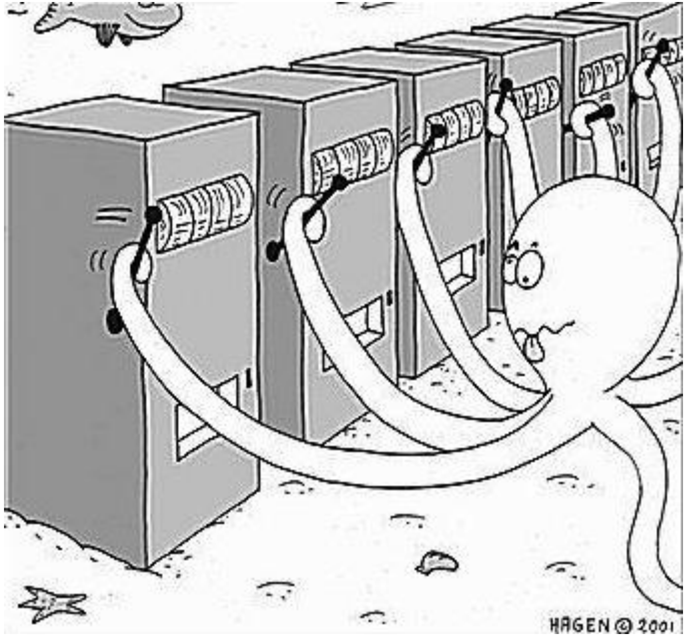
- **A multi-armed bandit is a simplified form of this analogy.** It is used to represent similar kinds of problems and finding a good strategy to solve them is already helping a lot of industries.


WELCOME
TO Fabulous
LAS VEGAS
NEVADA
SUPER SLOTS
JACKPOT
JACKPOT
PLAY TO WIN

Multi-Arm Bandit Algorithm

- A multi-armed bandit is a complicated slot machine wherein instead of 1, there are several levers which a gambler can pull, with each lever giving a different return. The probability distribution for the reward corresponding to each lever is different and is unknown to the gambler.

- The task is to identify which lever to pull in order to get maximum reward after a given set of trials.



Exploration Exploitation in the context of Bernoulli MABP

- The below table shows the sample results for a 5-armed Bernoulli bandit with arms labelled as 1, 2, 3, 4 and 5:

Arm	Reward
1	0
2	0
3	1
4	1
5	0
3	1
3	1
2	0
1	1
4	0
2	0

This is called Bernoulli, as the reward returned is either 1 or 0. In this example, it looks like the arm number 3 gives the maximum return and hence one idea is to keep playing this arm in order to obtain the maximum reward (pure exploitation).

Just based on the knowledge from the given sample, 5 might look like a bad arm to play, but we need to keep in mind that we have played this arm only once and maybe we should play it a few more times (exploration) to be more confident. Only then should we decide which arm to play (exploitation).

Use Cases: Clinical Trials

- The well being of patients during clinical trials is as important as the actual results of the study. Here, exploration is equivalent to identifying the best treatment, and exploitation is treating patients as effectively as possible during the trial.

Online Advertising

- The goal of an advertising campaign is to maximise revenue from displaying ads. The advertiser makes revenue every time an offer is clicked by a web user. Similar to MABP, there is a trade-off between exploration, where the goal is to collect information on an ad’s performance using click-through rates, and exploitation, where we stick with the ad that has performed the best so far.

Power and Sample Size

Power and Sample Size

- If you run a web test, how do you decide how long it should run (i.e., how many impressions per treatment are needed)? Despite what you may read in many guides to web testing, there is no good general guidance—it depends, mainly, on the frequency with which the desired goal is attained.

Key Ideas

• Finding out how big a sample size you need requires thinking ahead to the statistical test you plan to conduct.
• You must specify the minimum size of the effect that you want to detect.
• You must also specify the required probability of detecting that effect size (power).
• Finally, you must specify the significance level (alpha) at which the test will be conducted.

References

- https://blog.minitab.com/en/statistics-and-quality-data-analysis/what-are-degrees-of-freedom-in-statistics
- https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/anova/
- https://www.coursera.org/lecture/inferential-statistics-intro/anova-KoTvZ
- https://www.analyticsvidhya.com/blog/2018/01/anova-analysis-of-variance/#:~:text=Introduction%20to%20ANOVA,-A%20common%20approach&text=Analysis%20of%20variance%20(ANOVA)%20is,the%20means%20of%20different%20samples.
- https://www.youtube.com/watch?v=lgs7d5saFFc
- Support de cours: https://www.youtube.com/watch?v=WXPBoFDqNVk
- https://www.analyticsvidhya.com/blog/2019/11/what-is-chi-square-test-how-it-works
- https://www.youtube.com/watch?v=s1xEB9XknBI&list=PLD2EE7A6284364CA2