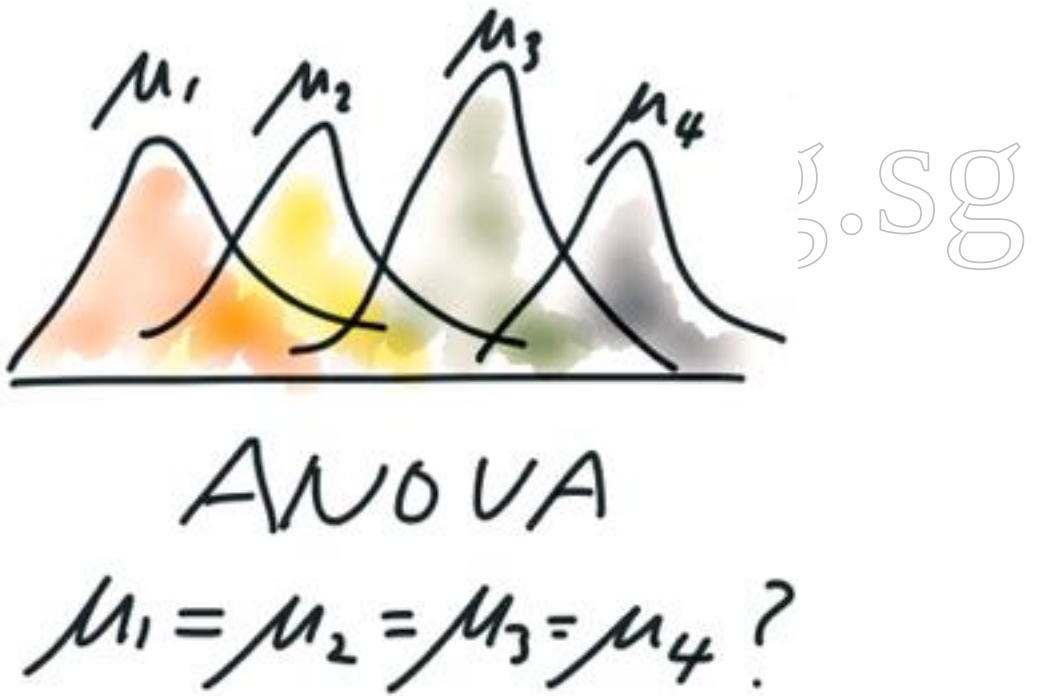
ANALYSIS OF VARIANCE (ANOVA)

DR. ALVIN ANG



CONTENTS

PART I

WHAT IS ANOVA?

1. ANOVA = Analysis of Variance
2. Main purpose of ANOVA is to check → is there a difference between Population Means?
3. It deals only with the F Distribution.
4. Recall Hypothesis Testing, the publication by me in Ang (2019)
    a. The greatest number of samples we can test is 2 samples (be it the Z or t test).
    b. What if we have more than 2 samples.. 3, 4, 5 etc.?
    c. This is where the usefulness of ANOVA comes in – to test several population means simultaneously (where Z and t test cannot do).
5. Also recall Hypothesis Testing, the publication by me in Ang (2019), under section "C. Using F-Test to Determine Equal or Unequal Variances"
    a. This is another function of the F Distribution.

ANOVA Table				
Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	***F***
Treatments	SST	$k-1$	SST/ $(k-1)$ = MST	MST / MSE
Error	SSE	$n-k$	SSE/ $(n-k)$ = MSE	
Total	SS Total	$n-1$		

$$\text{SST} = \text{SS total} - \text{SSE}$$

$$SS\,total = \sum\left(X - \bar{X}_G\right)^2$$

where

X is each sample observation.

n is the total number of observations.

$\bar{X}_G$ is the grand mean, calculated by summing up all the data values and dividing by n.

$$SSE = \sum(X - \bar{X}_c)^2$$

where

X is each sample observation.

$\bar{X}_c$ is the sample mean for each treatment c .

Figure 1: ANOVA Table, SSE, SST, SS Total (SUSS, 2014)

ANOVA EXAMPLE

Given:

Master Card	Visa	Store
$61	$85	$61
28	56	25
42	44	42
33	72	31
51	98	29
56	56	
	72	

- 3 types of credit card: Master Card, Visa, Store Card
- Are their means the same?

Answer:

STEP 1: STATE THE NULL AND ALTERNATE HYPOTHESIS

$$H_0: \mu_1 = \mu_2 = \mu_3$$

H_1: the means are not all equal

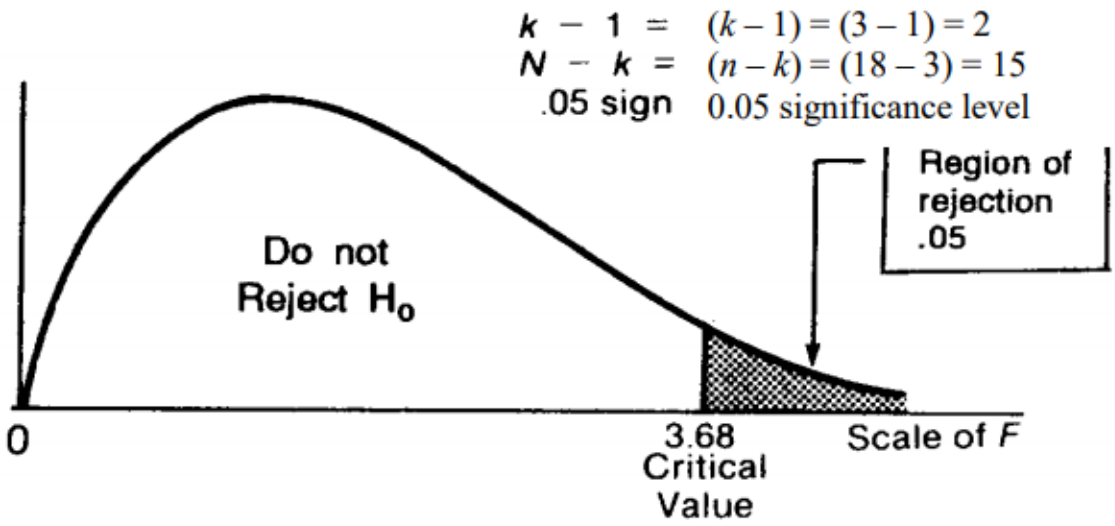
STEP 2: LEEL OF SIGNIFICANCE, ALPHA = 0.05

STEP 3: WE CHOOSE THE F STATISTIC

STEP 4: FORMULATE THE DECISION RULE

- Refer to F Distirbution Table (5%)
- Numerator = k-1 = 3-1 = 2 (where k is master card + visa + store)
- Denominator = n-k = 18-3 = 15 (where n is the total number of samples)

- Thus, F critical = 3.68



STEP 5: MAKE THE DECISION (BY HAND)

- To find SS Total:

	A	B	C	D	E	F	G	H	I	J
	MC			Visa			Store			
	X	$(X-\bar{X}_G)$	$(X-\bar{X}_G)^2$	X	$(X-\bar{X}_G)$	$(X-\bar{X}_G)^2$	X	$(X-\bar{X}_G)$	$(X-\bar{X}_G)^2$	**Totals**
	$61	8.67	75.17	$85	32.67	1067.33	$61	8.67	75.17	
	28	−24.33	591.95	56	3.67	13.47	25	−27.33	746.93	
	42	−10.33	106.71	44	−8.33	69.39	42	−10.33	106.71	
	33	−19.33	373.65	72	19.67	386.91	31	−21.33	454.97	
	51	−1.33	1.77	98	45.67	2085.75	29	−23.33	544.29	
	56	3.67	13.47	56	3.67	13.47				
				72	19.67	386.91			SS Total	
∑	271			483			188			942
∑			1162.72		plus	4023.23		plus	1928.07	7114.02
N	6			7			5			18

- To find SSE:

	A	B	C	D	E	F	G	H	I
	MC	$\bar{X} = 45.17$		Visa	$\bar{X} = 69.00$		Store	$\bar{X} = 37.60$	
	X	$(X - \bar{X})$	$(X - \bar{X})^2$	X	$(X - \bar{X})$	$(X - \bar{X})^2$	X	$(X - \bar{X})$	$(X - \bar{X})^2$
	$61	15.83	250.59	$85	16.00	256.00	$61	23.40	547.56
	28	–17.17	294.81	56	–13.00	169.00	25	–12.60	158.76
	42	–3.17	10.05	44	–25.00	625.00	42	4.40	19.36
	33	–12.17	148.11	72	3.00	9.00	31	–6.60	43.56
	51	5.83	33.99	98	29.00	841.00	29	–8.60	73.96
	56	10.83	117.29	56	–13.00	169.00			
				72	3.00	9.00			
Σ	271			483			188		
Σ			854.84		plus	2078.00		plus	843.20
n	6			7			5		SSE = 3776.04

- To find F statistic:

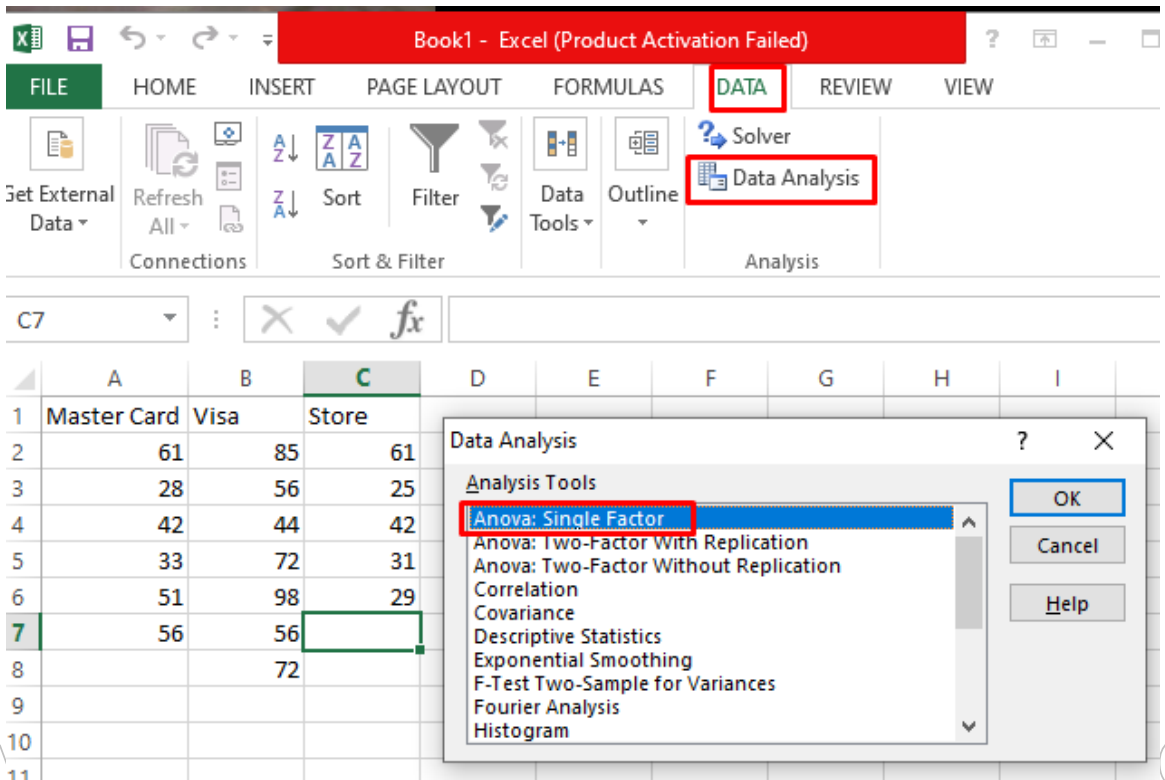
Source Variation	Sum of Squares	Degrees of Freedom	Mean Squares
Treatment	SST = 3,337.98	$(k-1) = (3-1) = 2$	$\text{MST} = \frac{SST}{(k-1)} = \frac{3,337.98}{2} = 1,668.99$
Error	SSE = 3,776.04	$(n-k) = (18-3) = 15$	$\text{MSE} = \frac{SSE}{(n-k)} = \frac{3,776.04}{15} = 251.736$
Total	SS Total = 7,114.02	17	

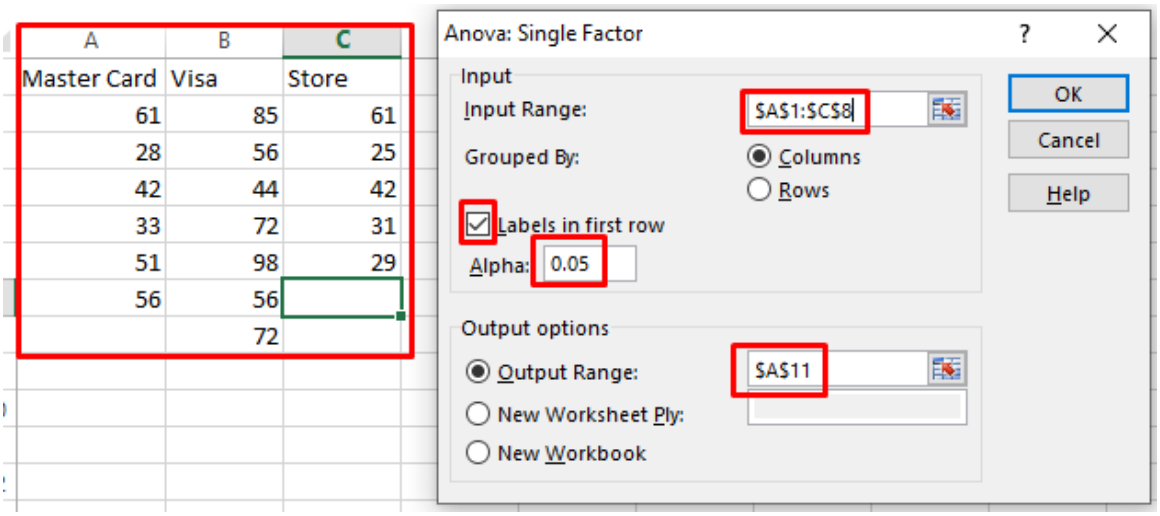
$$F = \frac{\text{MST}}{\text{MSE}} = \frac{1,668.98}{251.73} = 6.63$$

- Since F statistic (=6.63) > F critical (=3.68), Accept H1

- Conclusion: The means are not equal

By Excel





Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
MasterCard	6	271	45.16667	170.9667
Visa	7	483	69	346.3333
Store	5	188	37.6	210.8

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	3337.967	2	1668.983	6.629907	0.008648	3.68232
Within Groups	3776.033	15	251.7356			
Total	7114	17				

- Since p-value (=0.0086) < Alpha (=0.05) → Accept H1
- Conclusion: The means are not equal.

Confidence Interval for the Difference in Treatment Means	$(\bar{X}_1 - \bar{X}_2) \pm t\sqrt{MSE\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$

where

$\bar{X}_1$ is the mean of the first treatment or sample.

$\bar{X}_2$ is the mean of the second treatment or sample.

t is obtained from the t table in Appendix B.2. The degree of freedom is equal to $(n - k)$.

MSE is the mean square error term, which is obtained from the ANOVA table. It equals SSE/$(n - k)$ and is an estimate of the common population variance.

n_1 is the number of observations in the first sample.

n_2 is the number of observations in the second sample.

Figure 2: CI for ANOVA (SUSS, 2014)

- Now we know that $\mu 1 \neq \mu 2 \neq \mu 3$
- But we do not know: Which ones have the BIGGEST difference?
- Figure 2 shows how we can use the CI to detect whether or not is there a difference between 2 means.
- RULE:
    - If CI Tolerance limit include 0 → NO difference between means
    - If CI limit doesn't include 0 (i.e. both answers are on 1 side of 0) → there's a DIFFERENCE between the means.

Given:

Master Card	Visa	Store
$61	$85	$61
28	56	25
42	44	42
33	72	31
51	98	29
56	56	
	72	

- Mean: 45.17 69 37.6
- Biggest Difference!

- Since the biggest difference is VISA vs STORE, Thus we conduct ANOVA CI Test on them to check if there's a difference or not.

Answer:

$$(\bar{X}_1 - \bar{X}_2) \pm t\sqrt{MSE\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$(69.00 - 37.60) \pm 2.131\sqrt{252\left(\frac{1}{7} + \frac{1}{5}\right)}$$

$$31.40 \pm 19.81$$

$$\$11.59 \text{ to } \$51.21$$

where

$\bar{X}_1$ equals 69.00.

$\bar{X}_2$ equals 37.60.

N_1 equals 7.

n_2 equals 5.

t is 2.131 from Appendix B.2 with 15 degrees of freedom and the 95 percent level of confidence.

MSE is 252, which is in the ANOVA table constructed to calculate *F*.

- Since \$11.59 to \$51.21 does NOT include 0 → There's a difference in mean between VISA vs STORE.

REFERENCES

Ang, A. (2019). *Hypothesis Testing*. Singapore.

SUSS. (2014). *BUS105e Study Guide - Business Statistics*. Singapore: Singapore University of Social Sciences (SUSS).

ABOUT DR. ALVIN ANG

Dr. Alvin Ang earned his Ph.D., Masters and Bachelor degrees from NTU, Singapore. He is a scientist, entrepreneur, as well as a personal/business advisor. More about him at www.AlvinAng.sg.