boys

1 → raw data جدول الانوفا فاضي معطى — build Anova table　　Q2

2 → fill in the missing الجدول ناقص
boys|Girls — Components of Anova table

3

ch9
- 1pop
- Mean
- prop
- ↓
- Z test
- t test

ch10
- 2pop
- Mean
- prop
- ↓
- Z test
- t test

ch11
- 1pop
- 2pop
- Variance
- χ^2 test
- F test

ch12
- more than 2pop
- Mean
- F test
- تحليل ال variance

→ ① ② ③ ④

$\mu_1 = \mu_2 = \mu_3 = \mu_4$

$\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$

بسم الله الرحمن الرحيم

ANOVA table

(one way analysis of variance)

It's a common situation when someone needs to determine whether three or more populations have equal means.

ANOVA → An alysis Of Va riance

Completely randomized design

• An experiment that consists of the independent random selection of observations representing each level of one factor

• An analysis of variance design in which independent samples are obtained from two or more levels of a single factor for the purpose of testing whether the levels have equal means.

- **An ANOVA (Analysis of Variance), sometimes called an F test, is closely related to the t test.**
- **The major difference:**

t - test (ٍEst)	**ANOVA tests**
measures the difference between the means of two groups	**measures the difference between the means of more than two groups**

factor

متوسط درجات ميدترم للثلاث

B HR MIS

3 pop

عدد الـ pop

Level

Example

① Single factor

Expected mile age → Levels: brand 1, brand 2, brand 3, brand 4

متوسط الاميال المتوقعة

4 pop

factor

average grade of English → class A, class B, class C

متوسط درجات الـ E

3 pop

Factor

A quantity under examination in' an experiment as a possible cause of variation in the response variable

Levels

The categories, measurements or strata of a factor of interest in the current experiment

Balanced design

An experiment has a balanced design if the factor levels have equal sample sizes.

$n_1 = n_2 = n_3 = \checkmark$

$n_1 \neq n_2 \neq n_3 \longrightarrow$ unbalanced design

what are the assumption of one way ANOVA?

① population should be normally distributed
② Samples are independent
③ variances of data should be equal
④ Data are interval or ratio level
⑤ The residuals are normally distributed

الأخطاء

1 2 3

ANOVA Does	ANOVA Does Not
compare several Means with each other	compare several Means with the overall Mean
say whether or not there is a difference among Means	say which Means differ
require Continuous data	handle Discrete data
require roughly Normal Distributions	handle very Non-Normal Distributions
require somewhat equal Sample Variances	handle very unequal Sample Sizes and Variances

يحدد اذا كان هناك فروق في ال Mean ولا

1 2 3

$\mu_1 = \mu_2$

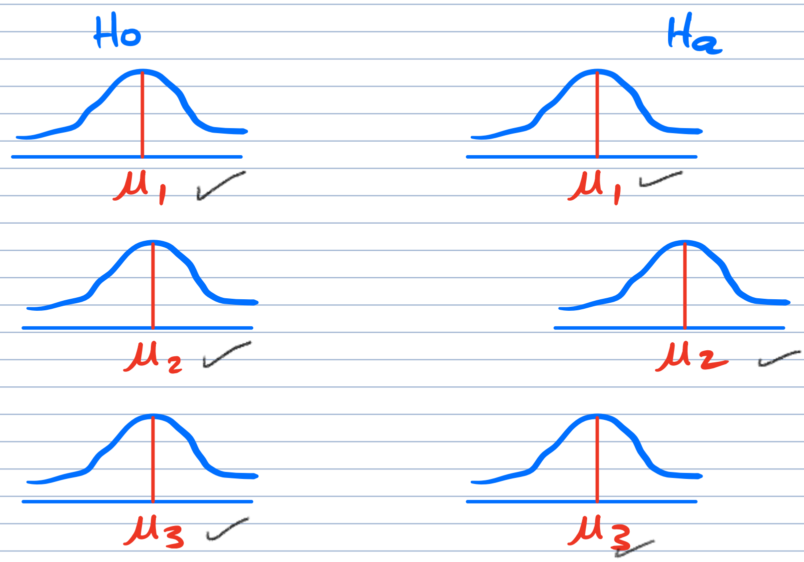
$\boxed{\mu_1 \neq \mu_3}$

Hypotheses of one way Anova [F test is conducted]

① $H_0 : \mu_1 = \mu_2 = \mu_3 = - - - - - - \mu_k$ ثابتة

Ha: at least two population means are different

Not all population have equal mean



$\mu_1 = \mu_2 = \mu_3$

at least 2 pop means are different

ch11,2

~~Ftest = $\frac{S_1^2}{S_2^2}$~~

② Complete Anova table to find F ratio = Ftest

Source of Variation	SS	df	MS	F-Ratio
Between Samples	SSB +	$k-1$	MSB	$\frac{MSB}{MSW}$
Within Samples	SSW =	$n_T - k$	MSW	
Total	SST	$n_T - 1$		

factor, Level

k - Number of populations
n_T - Sum of the sample sizes from all populations
df - Degrees of freedom
MSB - Mean square between
MSW - Mean square within

Example. The Chicago Connection Sandwich Company is a privately held company that operates in four locations in Columbus, Ohio. The VP of sales for the company is interested in knowing whether the dollar values for orders made by individual customers differ, on average, between the four locations.

To answer this question, staff at the VP's office have selected a random sample of eight customers at each of the four locations and recorded the order amounts. These are shown in Table 12.1.

TABLE 12.1 Chicago Connection Sandwich Company Order Data

	Store Locations				
	X_1	X_2	X_3	X_4	
	$n=8$	$n=8$	$n=8$	$n=8$	
Customer	1	2	3	4	
1	\$4.10	\$6.90	\$4.60	\$12.50	
2	5.90	9.10	11.40	7.50	
3	10.45	13.00	6.15	6.25	
4	11.55	7.90	7.85	8.75	
5	5.25	9.10	4.30	11.15	
6	7.75	13.40	8.70	10.25	
7	4.78	7.60	10.20	6.40	
8	6.22	5.00	10.80	9.20	
					Grand Mean
Mean	$\bar{x}_1 = \$7.00$	$\bar{x}_2 = \$9.00$	$\bar{x}_3 = \$8.00$	$\bar{x}_4 = \$9.00$	$\bar{\bar{x}} = \$8.25$
Variance	$s_1^2 = 7.341$	$s_2^2 = 8.423$	$s_3^2 = 7.632$	$s_4^2 = 5.016$	

Note: Data are the dollar values of the orders by customers at the four locations.

➢ Partitioning the Sum of Squares

To understand the logic of ANOVA, you should note several things about the data in Table 12.1.

1. The dollar values of the orders are different throughout the data table. Some values are higher; others are lower. Thus, variation exists across all customer orders. This variation is called the **total variation** in the data.

2. Within any particular location (i.e., factor level), not all customers had the same dollar order. For instance, within Level 1, order amounts ranged from \$4.10 to \$11.55.

Similar differences occur within the other levels. The variation within the factor levels is called the **within-sample variation**.

3. The sample means for the four restaurant locations are not all equal. Thus, variation exists between the four averages. This variation between the factor level means is referred to as the **between-sample variation**.

Partitioned Sum of Squares

$$SST = SSB + SSW \tag{12.1}$$

where:

SST = Total sum of squares
SSB = Sum of squares between
SSW = Sum of squares within

Note

$SST = SSB + SSW$
$SSB = SST - SSW$
$SSW = SST - SSB$

Total Sum of Squares

$$\checkmark\ SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{\bar{x}})^2 \qquad (12.2)$$

grand Mean

where:

SST = Total sum of squares
k = Number of populations (treatments)
n_i = Sample size from population i
x_{ij} = jth measurement from population i
$\bar{\bar{x}}$ = Grand mean (mean of all the data values)

SST → total Sum of squares the aggregate dispersion of the individual data values across the various factor levels

Equation 12.2 can also be restated as

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{\bar{x}})^2 = (n_T - 1)s^2$$

Variance

sample size of all data

$(32-1)\ s^2 = \checkmark$

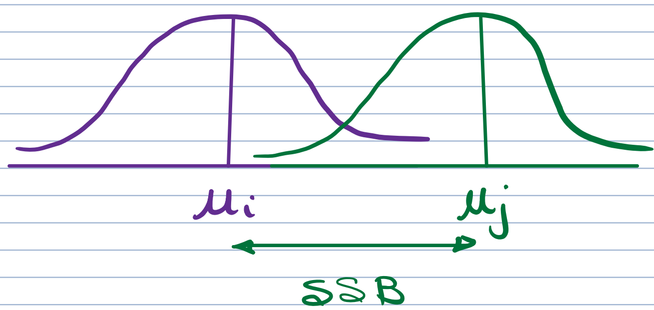
where

s^2 is the sample variance for all data combined, and

n_T is the sum of the combined sample sizes.

SSB $\longrightarrow$ sum of squares between samples dispersion among the factor sample means

التشتت بين عوامل العينة



variation due to differences among groups

Sum of Squares Between ضرب

$$SSB = \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{\bar{x}})^2 \qquad (12.3)$$

mean of each Level — Grand Mean

where:

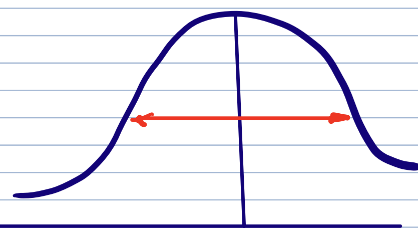
SSB = Sum of squares between samples
k = Number of populations
n_i = Sample size from population i
$\bar{x}_i$ = Sample mean from population i
$\bar{\bar{x}}$ = Grand mean

SSW = SST − SSB

SST = SSB + SSW

SSW → Sum of squares within samples the dispersion that exists among the data values within a particular factor level

التشتت الموجود بين قيم البيانات ضمن مستوى عامل معين



$D_1 + D_2 = \cancel{K} - 1 + n_T - \cancel{K}$

$= n_T - 1$

TABLE 12.2 One-Way ANOVA Table: The Basic Format

Source of Variation	SS	df	MS	F-ratio
Between samples	SSB	$D_1 = \boxed{k-1}$	MSB	$\frac{MSB}{MSW}$
Within samples	SSW	$D_2 = \boxed{n_T - k}$	MSW	
Total	SST	$n_T - 1$		

where:

k = Number of populations

n_T = Sum of the sample sizes from all populations

df = Degrees of freedom

MSB = Mean square between $= \frac{SSB}{\boxed{k-1}}$ D_1 → $d.f. = k - 1$

MSW = Mean square within $= \frac{SSW}{\boxed{n_T - k}}$ D_2 → $d.f. = n_T - k$

Decision Rule: We reject H_0 if $F > F_\alpha$.

خطوات الاختبار

1) Complete ANOVA table ✓ → $F_{test} = ✓$

2) formulate Ho, Ha ✓

Ho: $\mu_1 = \mu_2 = \mu_3 = ----------- = \mu_k$

Ha: at least 2 population means are different

4) critical value is F ✓ → ch 11.2

هناك ٣ جداول لل F على حسب قيمة α

$F \rightarrow \alpha = ✓$, $D_1 = K-1$, $D_2 = n_T - K$ } جدول الانوفا

5) Decision rule ✓

ch 11.2 شابتر جميع المسائل

Reject Ho if F >Fα otherwise don't reject

6) F test = Fratio $= \dfrac{MSB}{MSW}$

7) Compare Fratio with Fα reach the decision and draw the conclusion
✓ ✓

Exercise 7. Examine the three samples obtained independently from three populations:

Item	Group 1	Group 2	Group 3
1	14	17	17
2	13	16	14
3	12	16	15
4	15	18	16
5	16		14
6			16

5

(a) Conduct a one-way analysis of variance on the data. Use $\alpha = 0.05$. Given $s^2 = r$

Group ①

$$\bar{X}_1 = \frac{\sum x}{n_1}$$

$$= \frac{14+13+12+15+16}{5}$$

$$\boxed{\bar{X}_1 = 14}$$

Group ②

$$\bar{X}_2 = \frac{\sum x}{n_2}$$

$$= \frac{17+16+16+18}{4}$$

$$\boxed{\bar{X}_2 = 16.75}$$

Group ③

$$\bar{X}_3 = \frac{17+14+\cdots+16}{6}$$

$$\boxed{\bar{X}_3 = 15.33}$$

Grand mean $\bar{\bar{X}} = \dfrac{14+13+12+ \cdots +16}{15} = 15.27$

$$\boxed{\bar{\bar{X}} = 15.27}$$

$$SST = \sum\sum (x_{ij} - \bar{\bar{x}})^2$$

$$= (14-15.27)^2 + (13-15.27)^2 + (12-15.27)^2 + (15-15.27)^2$$

$$+ \text{------} + (16-15.27)^2$$

$$\boxed{SST \simeq 36.93}$$

$SSB = \sum n_i (\bar{x}_i - \bar{\bar{x}})^2$

$= 5(14 - 15.27)^2 + 4(16.75 - 15.27)^2 + 6(15.33 - 15.27)^2$

$SSB = 16.85$

$SSW = SST - SSB$

$= 36.93 - 16.85 = 20.083$

$df_1 = K - 1 = 3 - 1 = 2$

$df_2 = n_T - K = 15 - 3 = 12$

Test

[1] $H_0: \mu_1 = \mu_2 = \mu_3$

H_a: at least two population means are different

[2] Critical value is F $\alpha = 0.05$, $D_1 = 2$, $D_2 = 12$

$F(0.05, 2.12) = 3.885$

[3] Decision rule: Reject Ho if Ftest > 3.885 otherwise don't reject

[4] Test $F_{test} = \frac{MSB}{MSW} = \frac{8.425}{1.67} = 5.045$

[5] Since $F_{test} = 5.045 > 3.885$ reject Ho

[6] There is sufficient evidence to conclude that at least two pop means are different

	ANOVA					
source of variation	SS	df	MS	Fratio	p-value	critical value
Between groups	16.85	2	$\frac{16.85}{2} = 8.425$	$\frac{8.425}{1.67} = 5.045$		3.885
within groups	20.083	12	$\frac{20.083}{12} = 1.67$			
Total	36.93	14				

	Stat				
	201	202	203	204	205
1	80.50	94	88.40	69.50	85.50
2	75.90	89.01	77	100	67
3	98	78.50	90	96	76.30
4	100	62	63.50	67	60
5	88.75	85	73.10	79.30	50.10
6	54	76.70	69	84	96

Calculate SSW, SSB, SST Given

$\overline{x}_1 = 82.86$ $\overline{x}_2 = 80.87$ $\overline{x}_3 = 76.83$ $\overline{x}_4 = 82.63$ $\overline{x}_5 = 72.48$

$\overline{\overline{x}} = 79.14$ $S^2 = 187.78$

$SST = (n_T - 1) S^2 \longrightarrow n_T = 30$

$= (30-1)(187.78)$

$= 5445.62$

$SSB = \sum n_i (\bar{x}_i - \bar{\bar{x}})^2$

$= 6(82.86 - 79.14)^2 + 6(80.87 - 79.14)^2 + 6(76.83 - 79.14)^2$

$+ 6(82.63 - 79.14)^2 + 6(72.48 - 79.14)^2$

$= 471.77$

$SSW = SST - SSB$

$= 5445.62 - 471.77 = 4973.85$

	Stat				
	201	202	203	204	205
1	80.50	94	88.40	69.50	85.50
2	75.90	89.01	77	100	67
3	98	78.50	90	96	76.30
4	100	62	63.50	67	60
5	88.75	85	73.10	79.30	50.10
6	54	76.70	69	84	96

$\sum\sum x = 79.135$

$\underbrace{\sum n\left(\bar{x}_i - \bar{\bar{x}}\right)^2}_{SSB} = 471.89$

$\underbrace{\sum\sum\left(x_{ij} - \bar{\bar{x}}\right)^2}_{SST} = 5445.62$

$\sum\sum\left(x_{ij} - \bar{\bar{x}}\right) = 2540$

$\sum\left(\bar{x} - \bar{\bar{x}}\right)^2 = 130$

Calculate SSW, SSB, SST

$\boxed{1}$ $SST = \sum\sum(x_{ij} - \bar{\bar{x}})^2 = 5445.62$

$\boxed{2}$ $SSB = \sum n(\bar{x}_i - \bar{\bar{x}})^2 = 471.89$

$\boxed{3}$ $SSW = SST - SSB = 5445.62 - 471.89 = 4973.85$