(4) Univariable Analysis of Numerical Data

Dr. Wan Nor Arifin

Biostatistics and Research Methodology Unit
Universiti Sains Malaysia
wnarifin@usm.my / wnarifin.github.io

Last update: Jul 16, 2023

Outlines

- Introduction
- Independent *t*-test
- Paired *t*-test
- ANOVA

Learning outcomes

- Understand the concept of parametric test
- Familiarize with selected parametric tests for a numerical outcome
- Understand and able to interpret the results of the selected parametric tests

Introduction

Parametric Test

- Statistical test that requires:
  - Sample data come from population data that can be modeled by specific statistical distribution.
  - e.g. SBP of sample ← Normally distributed SBP of population.
  - Fixed set of parameters for chosen distribution.
  - e.g. normal distribution ← mean, SD.

Parametric Test

- Statistical test that requires (cont.):
  - Specific parameters to be tested.
  - e.g. MEAN is different or not.
  - Several assumptions to be tested before performing analysis.
  - Less flexible, BUT powerful and commonly used.

Parametric Test

- Parametric tests for comparison of means:
  - Two independent samples: Independent *t*-test
  - Two related samples: Paired *t*-test
  - More than two independent samples: ANOVA

Two independent samples: Independent *t*-test

Independent *t*-test

- Purpose: Compare MEANS of TWO independent samples/groups.
- Assumptions:
    1. Numerical outcome.
    2. Normal data distribution for each group.
    3. Equal variance between groups.

Independent *t*-test

Research objective:

To compare mean cholesterol level between male and female.

Research question:

Is there any difference in mean cholesterol level between male and female populations?

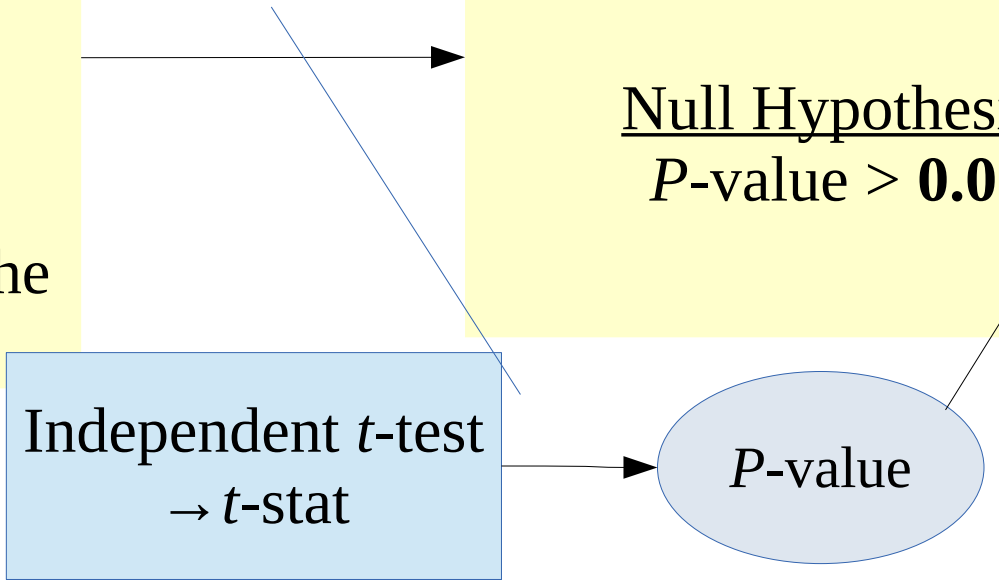
Independent *t*-test

RQ: Is there any difference in mean cholesterol level between male and female populations?

Alternative Hypothesis:
Mean cholesterol level of male population is different from female population

Null Hypothesis:
No difference in mean cholesterol level between the populations

Statistical Test



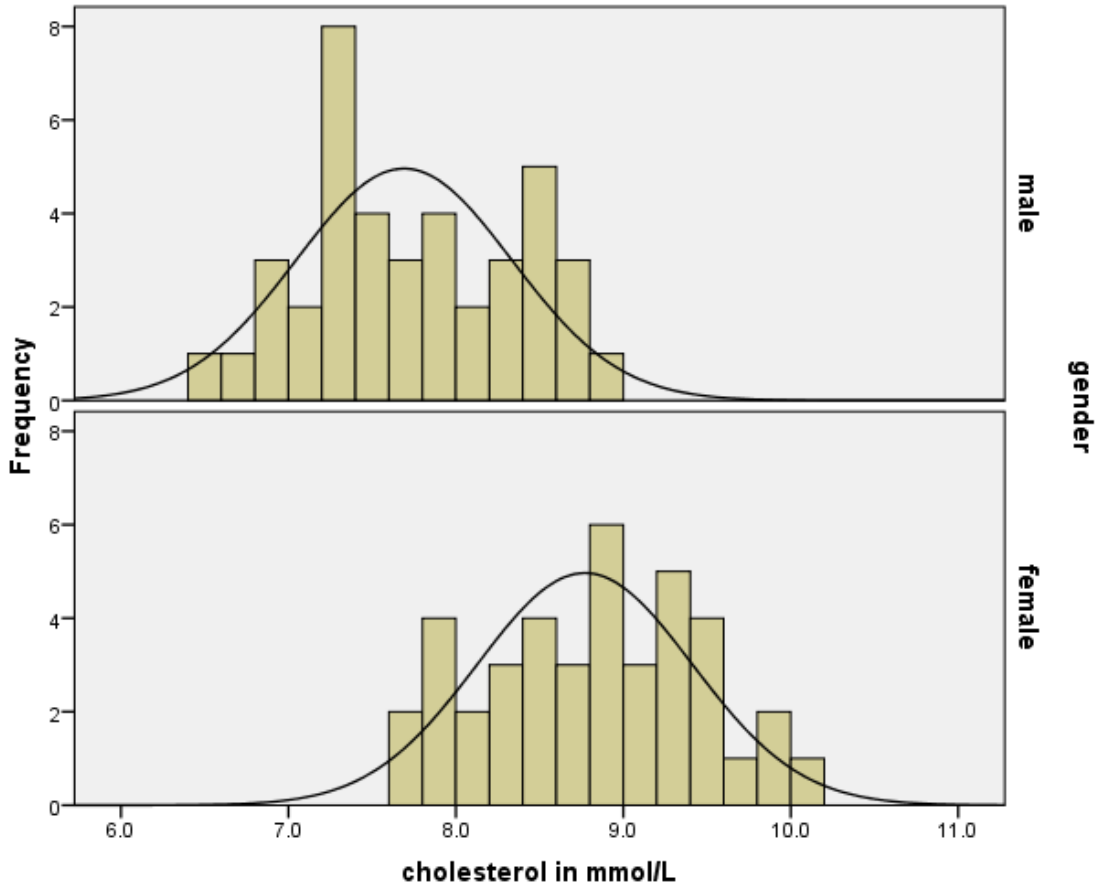
Alternative Hypothesis:
P-value ≤ **0.05**

Null Hypothesis:
P-value > **0.05**

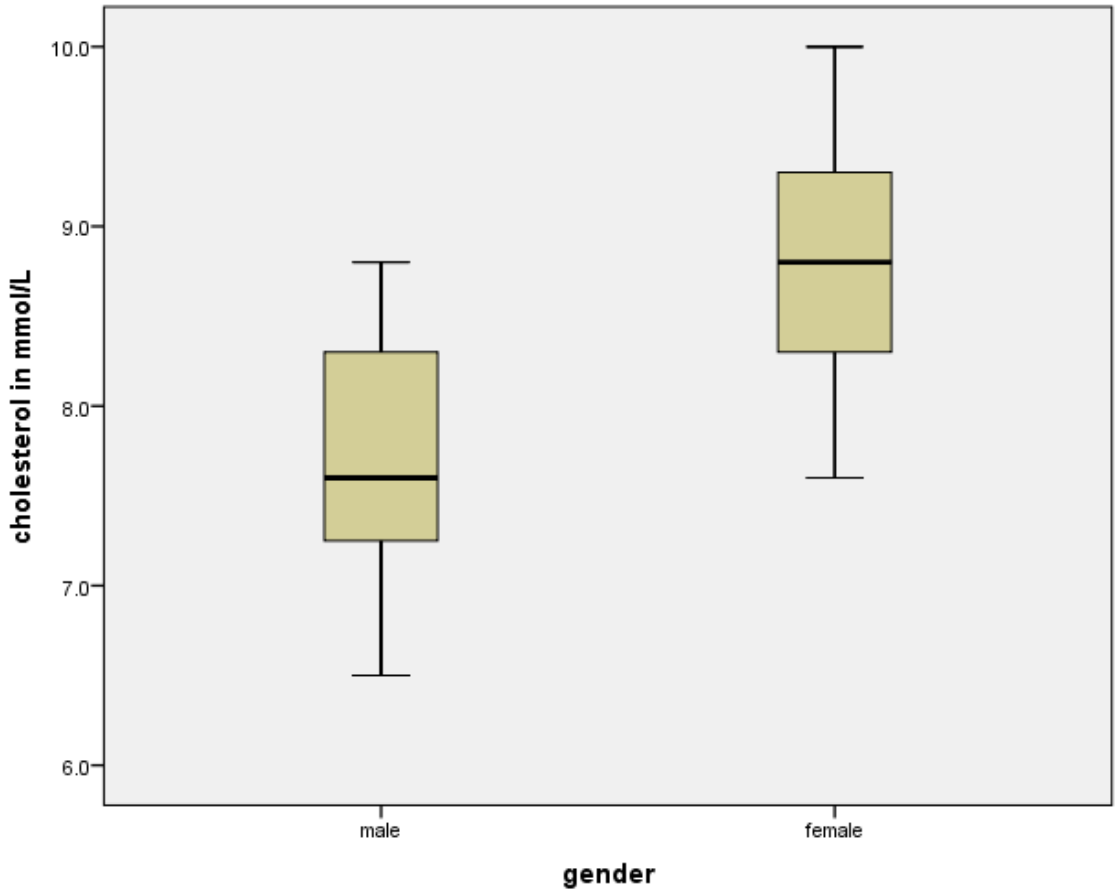
Independent *t*-test: Example

- Sample size: 40/group
- Group: 2 (male and female)
- Outcome: cholesterol level in mmol/L

Normality: Histogram



Normality: Boxplot



Independent t-test: Results

Group Statistics

	gender	N	Mean	Std. Deviation	Std. Error Mean
cholesterol in mmol/L	male	40	7.693	.6439	.1018
	female	40	8.768	.6462	.1022

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
cholesterol in mmol/L	Equal variances assumed	.076	.783	-7.453	78	.000	-1.0750	.1442	-1.3622	-.7878
	Equal variances not assumed			-7.453	77.999	.000	-1.0750	.1442	-1.3622	-.7878

Equal: $p \geq 0.05$
Unequal: $p < 0.05$

Use Welch *t*-test when variance not equal

Independent t-test: Results

Table 1: Comparison of cholesterol level between male and female.

Variable	Mean (SD)		Mean difference (95% CI)	*t*-statistic (df)	*P*-value[a]
Cholesterol (mmol/L)	Male $n = 40$	Female $n = 40$	-1.08 (-1.36, -0.79)	-7.45 (78)	< 0.001
	7.69 (0.644)	8.77 (0.646)			

[a]Independent *t*-test.

Two related samples: Paired *t*-test

Paired *t*-test

- Purpose: Compare MEAN DIFFERENCE between TWO related samples, i.e. equal to ZERO if there is no difference.
- Assumptions:
    1. Numerical outcome.
    2. Normal distribution of the DIFFERENCES between TWO paired observations (e.g. SBP after treatment – SBP before treatment).

Paired *t*-test

Research objective:

To compare mean cholesterol level of hypertensive patients before and after treatment.

Research question:

Is there any difference in mean cholesterol level of hypertensive patients before and after treatment?

Paired *t*-test

RQ: Is there any difference in mean mean cholesterol level of hypertensive patients before and after treatment?

Alternative Hypothesis:
Mean cholesterol level of HPT patients is different before and after treatment

Null Hypothesis:
No difference in mean cholesterol level of HPT patients before and after treatment

Statistical Test

Paired *t*-test → *t*-stat

P-value

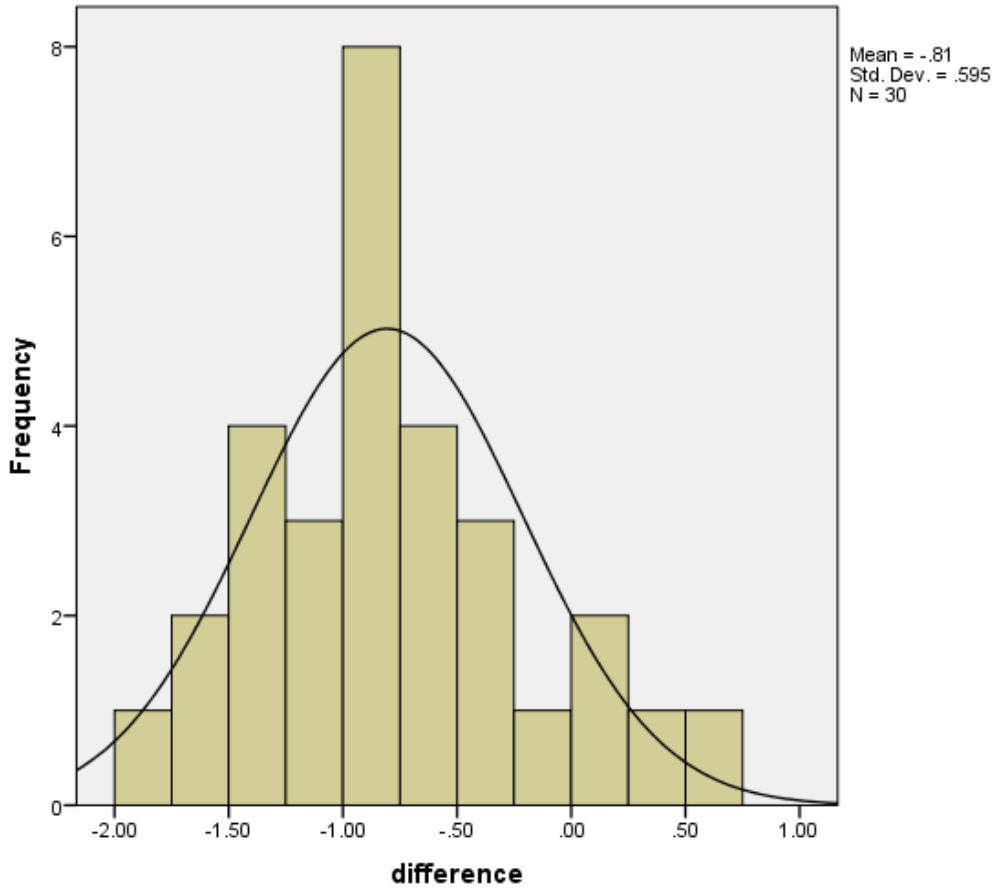
Alternative Hypothesis:
P-value ≤ **0.05**

Null Hypothesis:
P-value > **0.05**

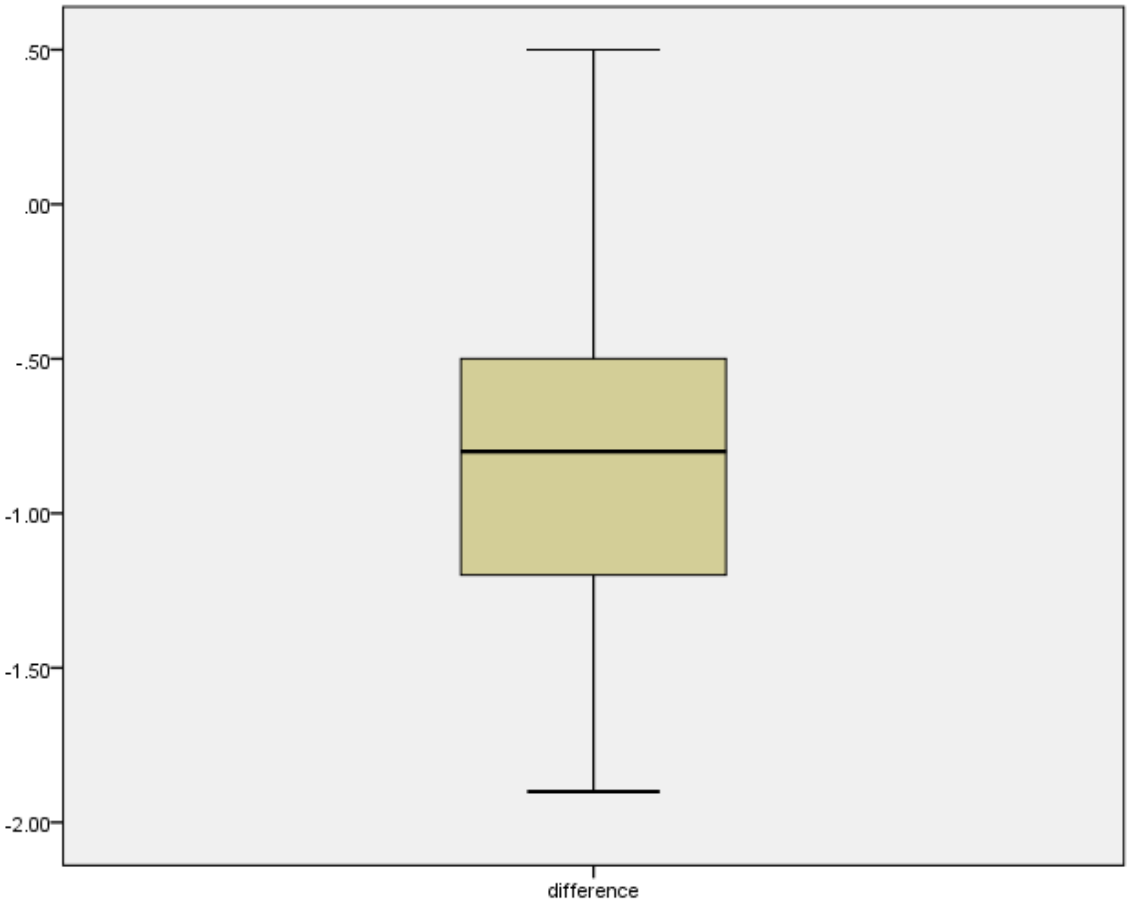
Paired *t*-test: Example

- Sample size: 30 paired observations
- Repetition: 2 (before and after treatment)
- Outcome: cholesterol level in mmol/L

Normality: Histogram



Normality: Boxplot



Paired *t*-test: Results

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	cholesterol in mmol/L before treatment	8.247	30	.3277	.0598
	cholesterol in mmol/L post treatment	7.440	30	.6806	.1243

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	cholesterol in mmol/L before treatment & cholesterol in mmol/L post treatment	30	.485	.007

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	cholesterol in mmol/L before treatment - cholesterol in mmol/L post treatment	.8067	.5953	.1087	.5844	1.0290	7.421	29	.000

mean difference = 8.247 – 7.440

Paired *t*-test: Results

Table 2: Comparison of cholesterol level before and after treatment.

Variable	Mean (SD) $n = 30$		Mean difference (95% CI)	*t*-statistic (df)	*P*-value[a]
Cholesterol (mmol/L)	Before	After	0.81 (0.58, 1.03)	7.421 (29)	< 0.001
	8.25 (0.328)	7.44 (0.681)			

[a]Paired *t*-test.

More than two independent samples: ANOVA

ANOVA

- ANalysis Of VAriance.
- Purpose: Compare MEANS of THREE/MORE independent samples/groups.
- Assumptions:
    1. Numerical outcome.
    2. Normal data distribution for each group.
    3. Equal variance between groups.

ANOVA

Research objective:

To compare mean cholesterol level between Group A, B and C treatment groups.

Research question:

Is there any difference in mean cholesterol level between Group A, B and C treatment groups?

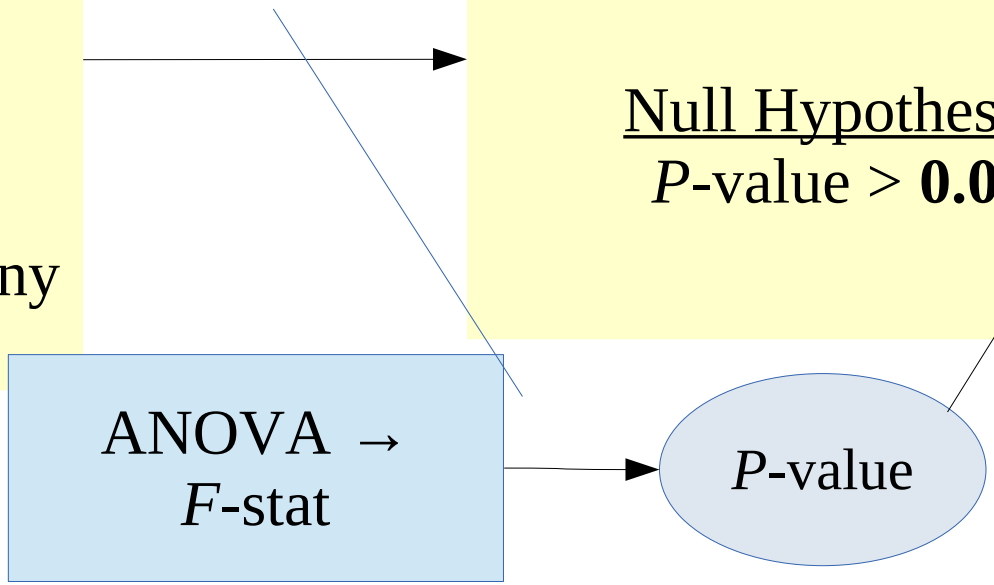
ANOVA

RQ: Is there any difference in mean cholesterol level between Group A, B and C treatment groups?

Alternative Hypothesis:
Mean cholesterol level between any of the populations are different.

Null Hypothesis:
No difference in mean cholesterol level between any of the populations

Statistical Test



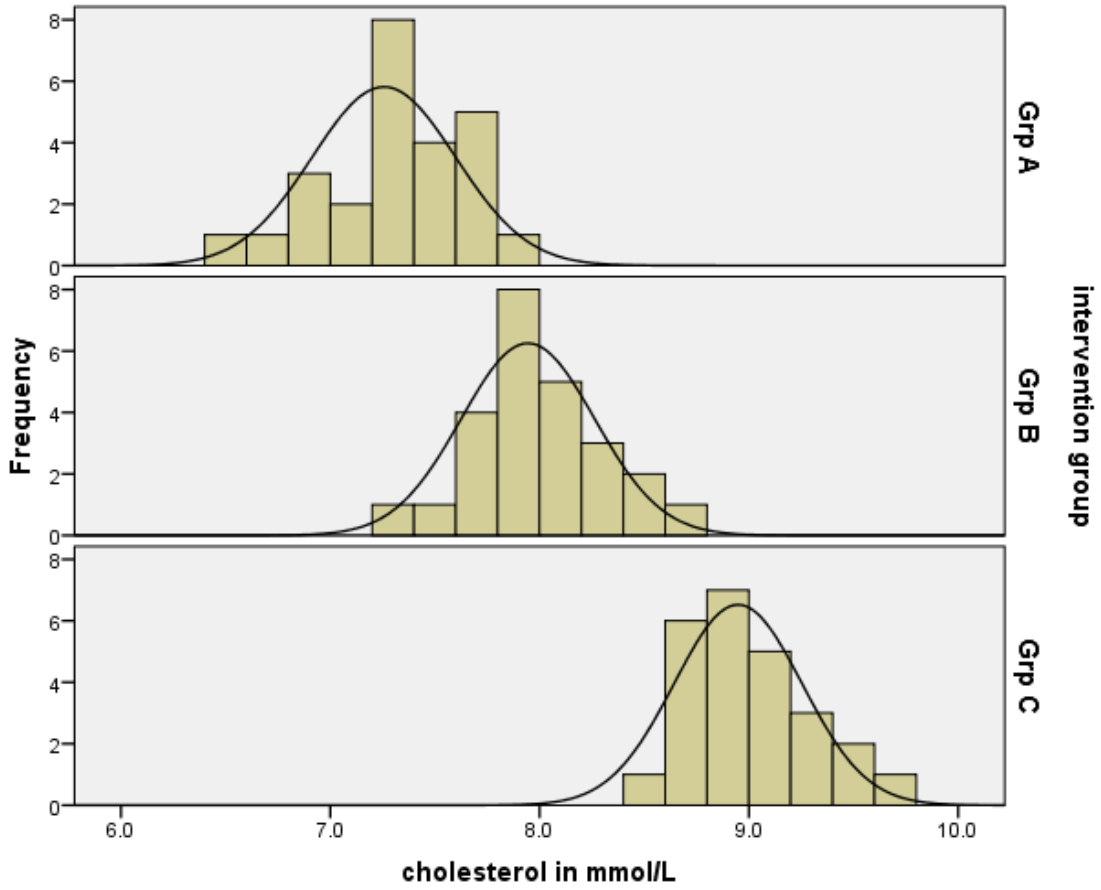
Alternative Hypothesis:
P-value ≤ **0.05**

Null Hypothesis:
P-value > **0.05**

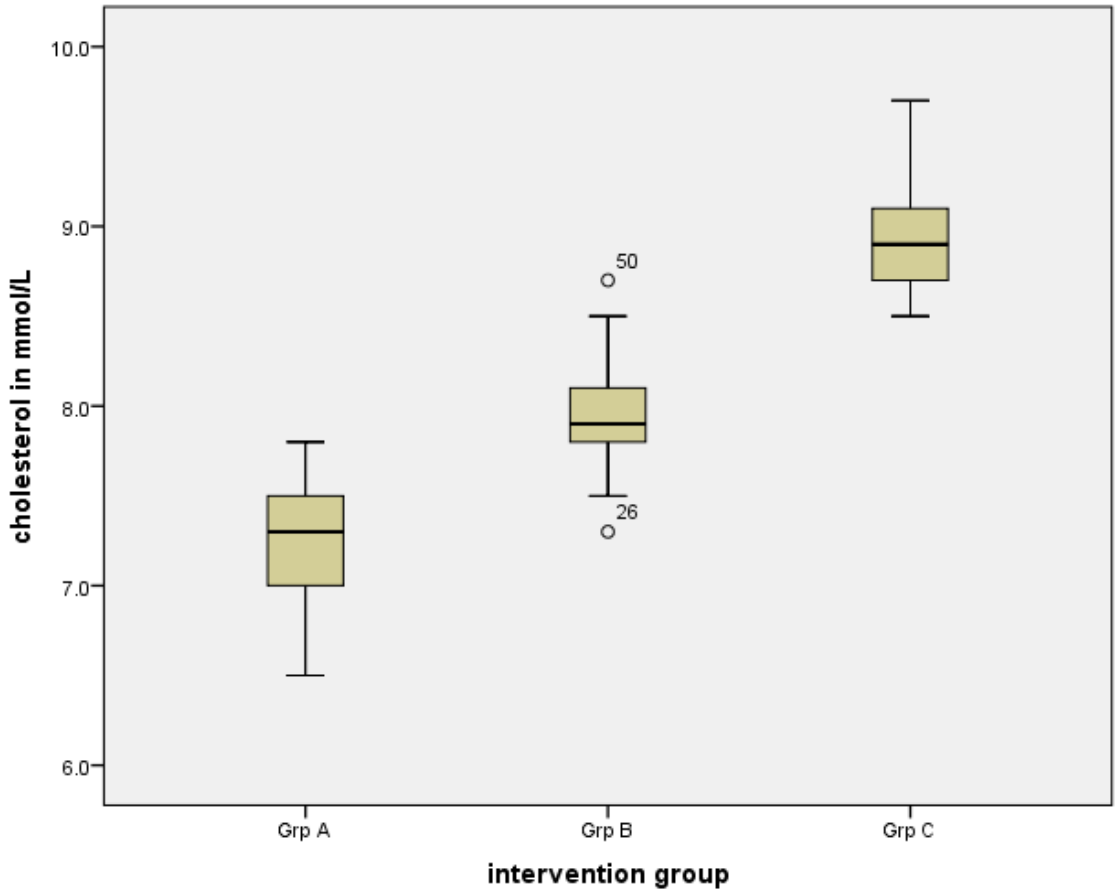
ANOVA: Example

- Sample size: 25/group
- Group: 3 (Grp A, B and C)
- Outcome: cholesterol level in mmol/L

Normality: Histogram



Normality: Boxplot



ANOVA: Results

Descriptives

cholesterol in mmol/L

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Grp A	25	7.256	.3429	.0686	7.114	7.398	6.5	7.8
Grp B	25	7.944	.3190	.0638	7.812	8.076	7.3	8.7
Grp C	25	8.948	.3057	.0611	8.822	9.074	8.5	9.7
Total	75	8.049	.7685	.0887	7.873	8.226	6.5	9.7

ANOVA: Results

Test of Homogeneity of Variances

cholesterol in mmol/L

Levene Statistic	df1	df2	Sig.
.105	2	72	.900

Equal: $p \geq 0.05$
Unequal: $p < 0.05$

ANOVA

cholesterol in mmol/L

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	36.202	2	18.101	173.639	.000
Within Groups	7.506	72	.104		
Total	43.707	74			

df1 = 2
df2 = 72

Robust Tests of Equality of Means

cholesterol in mmol/L

	Statistic[a]	df1	df2	Sig.
Welch	172.475	2	47.896	.000

a. Asymptotically F distributed.

Use Welch ANOVA when variance not equal

ANOVA: Results

Multiple Comparisons

Dependent Variable: cholesterol in mmol/L

	(I) intervention group	(J) intervention group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Sidak	Grp A	Grp B	-.6880*	.0913	.000	-.911	-.465
		Grp C	-1.6920*	.0913	.000	-1.915	-1.469
	Grp B	Grp A	.6880*	.0913	.000	.465	.911
		Grp C	-1.0040*	.0913	.000	-1.227	-.781
	Grp C	Grp A	1.6920*	.0913	.000	1.469	1.915
		Grp B	1.0040*	.0913	.000	.781	1.227
Games-Howell	Grp A	Grp B	-.6880*	.0937	.000	-.915	-.461
		Grp C	-1.6920*	.0919	.000	-1.914	-1.470
	Grp B	Grp A	.6880*	.0937	.000	.461	.915
		Grp C	-1.0040*	.0884	.000	-1.218	-.790
	Grp C	Grp A	1.6920*	.0919	.000	1.470	1.914
		Grp B	1.0040*	.0884	.000	.790	1.218

*. The mean difference is significant at the 0.05 level.

Equal variance

Unequal variance

ANOVA: Results

Table 3: Comparison of cholesterol level between the three intervention groups.

Groups	*n*	Cholesterol (mmol/L) Mean (SD)	*F*-statistic (df1, df2)[a]	*P*-value[a]
Grp A	25	7.26 (0.343)	173.64 (2, 72)	< 0.001[b]
Grp B	25	7.94 (0.319)		
Grp C	25	8.95 (0.306)		

[a]One-way ANOVA, [b] Post-hoc multiple comparison with Sidak correction shows significant difference between all intervention groups ($P < 0.001$).

Quiz

- Briefly describe about parametric test
- Describe the purpose of testing by independent *t*-test
- Describe the purpose of testing by paired *t*-test
- Describe the purpose of testing by ANOVA

Quiz

Table 2. Comparison of FBS pre- and post-intervention for exercise and control groups

Variables	Mean (SD)	Mean (SD)	Mean difference (95% CI)	*t*-statistic (df)	*P*-value
FBS level (mg/dl)	Pre-trial	Post-trial			
Exercise group (n = 25)	175.3 (17.2)	124.0 (7.7)	51.3 (44.9, 57.6)	16.67 (24) [a]	0.001[c]
Control group (n = 25)	165.0 (18.0)	148.6 (15.2)	16.4 (11.2, 21,6)	6.48 (24) [a]	0.001[c]
Mean difference (95% CI)	10.3 (0.3, 20.1)	-24.6 (-31.4, -17.1)	**34.9 (26.8, 42.8)**	**8.75 (48)** [b]	**0.001**[c]

FBS = fasting blood sugar, [a] Paired *t*-test, [b] Independent *t*-test, [c] *P*-value is significant at $P < 0.05$

Ezema, C. I., Omeh, E., Onyeso, O. K. K., Anyachukwu, C. C., Nwankwo, M. J., Amaeze, A., ... & Ugwuanyi, I. (2019). The effect of an aerobic exercise programme on blood glucose level, cardiovascular parameters, peripheral oxygen saturation, and body mass index among Southern Nigerians with type 2 diabetes mellitus, undergoing concurrent sulfonylurea and metformin treatment. The Malaysian journal of medical sciences: MJMS, 26(5), 88.

Quiz

Table 2. Comparison of means of motives of PA participation between types of co-curricular activity

Motives of participating in PA	Co-curricular	Mean (SD)	*F*-stat (df)	*P*-value
Enjoyment	Arts	3.91 (0.59)	2.790 (2,579)	0.062
	Uniform	3.78 (0.63)		
	Sports	3.98 (0.57)		
Mastery	Arts	3.55 (0.53)	0.771 (2,578)	0.463
	Uniform	3.47 (0.48)		
	Sports	3.57 (0.51)		
Competition	Arts	3.72 (0.65)	2.336 (2,581)	0.098
	Uniform	3.57 (0.70)		
	Sports	3.78 (0.65)		
Affiliation	Arts	4.17 (0.56)	8.223 (2,584)	< 0.001
	Uniform	3.99 (0.62)		
	Sports	4.30 (0.49)		
Appearance	Arts	3.79 (0.50)	4.833 (2,582)	0.008
	Uniform	3.62 (0.48)		
	Sports	3.85 (0.53)		
Physical	Arts	3.32 (0.48)	4.634 (2,575)	0.010
	Uniform	3.31 (0.32)		
	Sports	3.44 (0.43)		
Psychological	Arts	3.74 (0.62)	1.736 (2,579)	0.177
	Uniform	3.79 (0.53)		
	Sports	3.83 (0.58)		
Other's expectation	Arts	3.02 (0.65)	1.673 (2,576)	0.189
	Uniform	2.97 (0.44)		
	Sports	3.10 (0.62)		

Kuan, G., Abdullah, N., Kueh, Y. C., Ismail, M., Shafei, M. N., & Morris, T. (2019). Co-curricular activities and motives for participating in physical activity among health sciences students at Universiti Sains Malaysia, Malaysia. The Malaysian journal of medical sciences: MJMS, 26(1), 138.

Thank You