

An Example,

Zinc concentrations were measured on 12 random samples taken from two industrial effluents, A and B, and the results in 10^{-8} M zinc are given below. Given that the data are not normally distributed, test for a significant difference in zinc concentrations between the two sites:

- a) Site A: 12.5 12.1 13.4 13.0 12.1 12.15 12.25 12.7 11.4 11.7 12.5 13.7
- b) Site B: 9.8 11.6 10.4 10.3 10.4 11.1 9.8 10.15 10.2 10.8 10.9 11.2

Model Answer:

H_0 : Zinc concentration of A effluent = Zinc concentration of B effluent

H_A : Zinc concentration of A effluent \neq Zinc concentration of B effluent

As the data are not normally distributed, a non-parametric Mann-Whitney test should be used. All data are ranked as below:

A	A	A	A	A	A	A	A	A	A	A	B	A
13.7	13.4	13.0	12.7	12.5	12.5	12.25	12.15	12.1	12.1	11.7	11.6	11.4
B	B	B	B	B	B	B	B	B	B	B		
11.2	11.1	10.9	10.8	10.4	10.4	10.3	10.2	10.15	9.8	9.8		

$$\sum R_1 (\text{Site A}) = 1 + 2 + 3 + 4 + 5.5 + 5.5 + 7 + 8 + 9.5 + 9.5 + 11 + 13 = 79$$

$$U = n_1 n_2 + (1/2)n_1(n_1+1) - \sum R_1 = 143$$

$$U_{0.05 (2), 12, 12} = 107, p < 0.001$$

\Rightarrow Calculated $U > 107$, therefore H_0 is rejected.

In conclusion, the zinc concentration in the effluent of Site A is significantly higher than that of Site B ($U = 143, p < 0.001$).



Answer ALL the questions

(Tips: Remember to check homogeneity of variances and run an appropriate post hoc test if there are significant differences among different means or medians when comparing > 2 groups)

- 1) **Use of Calculator:** An investigation of the effects of two different atmospheric carbon dioxide concentrations on the growth of a weed *Senecio jacobaea* was undertaken under controlled conditions. Seventeen plants from the same seed stock were grown from seed to maturity in 12 weeks under CO_2 concentrations of 370 and 450 ppm. Each plant was then harvested, dried and weighed in grams. Given that the data are normally distributed, test whether the 'high CO_2 ' plants (450 ppm) produced more mean biomass than the control (370 ppm) plants.

- a) Control (in g): 22.9 17.0 20.2 21.8 22.8 23.6 24.1 26.3 24.7 23.1 23.9 25.2
- b) High CO_2 (in g): 25.8 25.4 22.3 26.1 26.0 19.9 18.7 26.3 27.7 26.4 17.1 27.7

- 2) **Use of Calculator:** Concentrations of lead (Pb in ppb) were measured in the plasma of a random sample of 8-year-old boys from a city and another sample of 8-year-old boys from the surrounding countryside. It was hypothesised that the city boys would have elevated levels of lead compared with the country boys. **Given that the data of country boys are not normally distributed (but the data of city boys follow normal distribution),** test this null hypothesis.

City boys: 5.36 6.48 7.19 5.55 4.89 4.12 10.00 6.65 6.80 5.00 3.50 5.72
Country boys: 4.40 5.25 4.50 4.80 5.49 2.95 3.25 7.93 4.95 5.65 4.15 3.41

A further study was made of 14 8-year-old girls from the same city. The girls had mean plasma Pb level of 5.51 ppb with sample standard deviation 0.9318 ppb. How did they compare with the city boys?

- 3) **Use of Calculator:** A comparison was made between the total bacteria counts obtained from swabs taken from two designs of telephone mouthpiece, A and B. Both mouthpieces were subjected to two months' use by randomly selected volunteers so that all headpieces received the same amount of use. Given that the data are not normally distributed, determine whether the level of bacterial contamination ascertained by the swabs differed between the two designs. Total bacterial counts ($\times 10^5$):
- a) Design A: 6.8 3.2 10.9 7.8 14.1 12.1 15.7 1.1 3.7 8.4 9.6 10.6
b) Design B: 8.7 14.4 15.1 12.9 6.8 14.8 11.3 16.3 18.0 11.1 13.1 14.3
- 4) **Use of SPSS or Excel:** Houses were surveyed for raising moisture using electrical resistance (relative units) in the north and south side of an east-west running street. Given that both datasets are not normally distributed, test if there is a significant difference in rising moisture between the two sides.
- a) North side: 1842 1041 495 842 1541 1980 222 395 1615 2042 970 890 572 1315 741 1251 1092 1695 1225 1138 1161 132 675
b) South side: 1265 563 868 449 1205 1090 1155 1072 672 1131 1008 1070 1032 1031 1088 1444 721 610 756 1180 1389
- 5) **Use of Calculator:** The following temperatures ($^{\circ}\text{C}$) were taken 10 m above and 10 m below a sewage outfall on a stream. Temperatures were taken at a fixed time on 8 occasions. Use a suitable parametric test to determine if there is a significant increase in temperature below the outfall.

Sample no.	Above outfall	Below outfall
1	6.8	6.9
2	5.1	5.4
3	7.2	6.8
4	8.8	8.9
5	9.2	9.6
6	9.5	9.8
7	10.6	9.8
8	12.2	11.7

- 6) **Use of SPSS or Excel:** As a quality control measure, 10 ammonia-sensing electrodes were each dipped into two solutions containing the same concentration of ammonia but different concentrations of salt (sodium chloride). The data below show the level of ammonia (ppm) detected by the electrodes (in ppm) without salt (A) and with 1 g l⁻¹ salt (B). Perform a suitable parametric test to determine if the presence of salt affects the readings.

Electrode no.	A	B
1	4.363	4.960
2	4.428	4.792
3	4.380	4.578
4	4.409	4.800
5	4.306	4.708
6	4.319	4.212
7	4.336	4.256
8	4.408	4.310
9	4.388	4.860
10	4.367	4.461

- 7) **Use of Calculator:** Ten nurses were randomly selected to test the efficacy of a germicidal handcream. Total viable bacterial counts were determined by pressing the fingertips of the nurses onto nutrient agar plates before an 8-hr shift. The cream was applied and the procedure was repeated at the end of the shift. Use an appropriate non-parametric test to determine if the cream significantly reduced bacterial counts.

Total viable count x 10000		
Nurse	Before	After
1	3.61	2.04
2	2.82	1.41
3	1.98	2.71
4	4.21	1.31
5	0.82	1.20
6	0.41	0.62
7	2.25	1.33
8	6.41	0.68
9	1.23	0.91
10	2.07	0.22

- 8) **Use of Calculator:** Four measurements of chloride (ppm) were taken randomly at six sampling stations in a water distribution system. Given that all data are normally distributed, determine if the chloride concentration changes significantly along the system.

Station/ Chlorine (ppm)					
1	2	3	4	5	6
1.72	1.54	1.48	1.46	1.32	1.17
1.66	1.49	1.53	1.51	1.50	1.18
1.50	1.42	1.39	1.63	1.30	1.01
1.68	1.41	1.45	1.43	1.38	1.19

- 9) **Use of SPSS or Excel:** Five samples of soil were collected randomly from each of five fields, A-E, as part of a metal contamination study. Each soil was dried, digested and analysed for lead (Pb). The results are shown below. Given that all data are normally distributed, test if Pb concentration (mg/kg) differs among the five fields.

A	B	C	D	E
6.30	5.00	4.20	6.10	8.30
6.80	5.90	4.10	7.20	8.20
6.50	5.80	4.30	7.50	8.20
5.90	6.70	3.90	6.10	8.20
6.10	6.70	4.50	6.90	8.10

- 10) **Use of Calculator:** A bacteriological survey of seawater taken five sandy bays, A-E, was undertaken. Four random samples were taken from each bay and viable counts of *Escherichia coli* were determined. Use a suitable test to determine if there is a significant difference in bacteria numbers. Note that bacterial counts are rarely normally distributed so the use of ANOVA is questionable in this case.

Numbers of <i>E. coli</i> in 100 ml seawater				
A	B	C	D	E
1500	11300	12200	13750	1353
1560	9800	17400	17200	1002
7900	11800	51200	8000	2300
1550	11860	15000	14000	4200

- 11) **Use of SPSS:** In a study of acid rain effects on HK soils, the pH of three soils was determined on 10 random samples from each of three sites, A-C. Use a suitable test to determine if there is a significant difference in the soil pH at the **three** sites. Note that pH is a logarithmic quantity which would not be expected to follow a normal distribution.

A	B	C
6.42	7.01	6.95
5.95	6.98	7.14
5.80	6.86	7.27
6.06	7.23	7.23
5.95	6.98	7.14
5.80	6.86	7.27
6.06	7.65	6.23
6.15	6.41	7.11
6.15	6.41	7.11
5.92	6.81	7.00

- 12) **Use SPSS or Excel:** In a study of nitrogen dioxide pollution inside vehicles, three models of cars were driven for 1 hr in a town area (A) and an adjacent country area (B). Each car interior carried a triethanolamine disc, which adsorbs nitrogen dioxide, permitting the mean nitrogen dioxide concentration to be calculated over the 1 hr period. Perform a **two-way ANOVA** to determine whether the pollutant concentration differs significantly between vehicle model and location of the drive.

Vehicle model	Nitrogen dioxide (ppb)	
	Area A	Area B
1	17.4	12.3
1	21.8	18.2
1	20.5	19.0
2	42.3	21.8
2	38.1	41.0
2	44.2	38.6
3	21.6	16.4
3	36.0	19.8
3	23.0	15.7

- 13) **Use of Calculator:** The data below are for Secchi disc transparency (m) and phytoplankton production ($\text{mg C m}^{-3} \text{ h}^{-1}$) in 12 locations of Victoria Harbour. Obtain the Pearson's product-moment correlation coefficient and determine if the two parameters have a significant correlation.

Secchi disc transparency (m)	Phytoplankton production ($\text{mg C m}^{-3} \text{ h}^{-1}$)
7.95	390
7.25	235
6.65	590
8.05	605
4.35	710
3.55	335
5.95	360
6.45	205
5.75	770
4.45	650
3.50	520
1.85	560

- 14) **Use of Calculator:** Determine whether there is a significant correlation between phosphoric acid annual production (PAAP; k tonnes) and phosphate annual disposal to sea (PADS; tonnes) in nine European countries using Spearman Rank Correlation Analysis.

PAAP	243	65	199	240	430	470	180	45	1000	491
PADS	20	490	520	670	3440	50	45	150	520	3390

- 15) **Use SPSS or Excel:** In general, solid waste generation rate is positively correlated with the population size. The following table presents the data of population and the production rate of municipal solid waste (tonnes per day) during 1987-2007 in Hong Kong. It is assumed that (1), the population data are measured very accurately by the Government and (2), the data of solid waste generation rate are bivariately normally distributed. Recently, the Government has announced that the population size will reach to 9,000,000 by year 2020.
- a) Plot a scatter diagram to illustrate the association between the two variables.
 - b) Test whether there is a significant linear relationship between these two variables using regression analysis.
 - c) What is the percentage of variance in the waste production rate that can be explained by the population size?
 - d) Perform a regression analysis to predict the production rate of municipal solid waste by 2020 and calculate its 95% confidence interval.

Year	Population (millions)	Municipal solid waste (tonnes per day)
1987	5.7	7100
1989	5.8	7400
1991	5.8	7950
1993	5.9	8450
1995	6.0	8400
1997	6.2	7800
1999	6.5	8200
2001	6.6	8700
2003	6.7	8750
2005	6.7	9300
2007	6.8	9900

End of the Assignment