UNIVERSITI KUALA LUMPUR
INSTITUTE OF MEDICAL SCIENCE TECHNOLOGY

FINAL EXAMINATION
OCTOBER 2025 SEMESTER

COURSE CODE	:	HGB 30103
COURSE NAME	:	BIOSTATISTICS
PROGRAMME NAME	:	BACHELOR OF ENVIRONMENTAL HEALTH (Hons) BACHELOR OF BIOMEDICAL SCIENCE (HONOURS) BACHELOR OF OCCUPATIONAL SAFETY & HEALTH (HONOURS)
DATE	:	26 JANUARY 2026
TIME	:	9:00 AM – 12:00 PM
DURATION	:	3 HOURS

INSTRUCTIONS TO CANDIDATES

1. Please CAREFULLY read the instructions given in the question paper.
2. This question paper has information printed on both sides of the paper.
3. This question paper consists of TWO sections.
4. Answer ALL questions for Section A.
5. Section B consist of four questions. Answer THREE (3) questions only.
6. Please write your answers on the answer booklet provided.
7. Answer all questions in English language ONLY.
8. Please answer MCQ/EMQ questions using OMR sheet.
9. Refer to the attached Formula/ Appendices. [✓] *Tick if applicable*

THERE ARE 15 PAGES OF QUESTIONS, EXCLUDING THIS PAGE.

SECTION A (Total: 40 marks)

INSTRUCTION: Answer ALL questions.
Please use the answer booklet provided.

Question 1

Answer all the following questions.

a. According to a report from the National Health Surveillance Unit dated 10th July 2024, 72% of dengue fever patients showed full recovery within the first week of diagnosis. If 25 dengue patients are randomly selected across Malaysia, what is the probability that at least 80% of them have recovered within the first week after receive treatment?

(5 marks)

b. A large logistics company is evaluating workplace strain issues among its employees. Recent internal assessments showed that 42.5% of workers in the loading department reported experiencing work-related muscle strain, compared to 28.3% of employees in the dispatch department. To further assess the situation, an occupational safety officer randomly selected 300 workers from the loading department and 250 workers from the dispatch department for a detailed health survey.

What is the probability that the difference between the sample proportions of workers experiencing muscle strain in the two groups ($\hat{p}_1 - \hat{p}_2$) will be less than 5%?

Refer below – Table 1: Prevalence of asthma among workers from morning and evening shift

(5 marks)

Table 1: Prevalence of workplace strain among workers from loading and dispatch department

Department	*n*	%
Loading	300	42.5
Dispatch	250	28.3

Question 2

Many neurological symptoms, including cognitive impairment, memory loss, and personality changes have been increasingly reported among automotive spray painters. All affected workers stated that they had been exposed to solvent-based paints and thinners over long periods (chronic exposure). The prevalence of neurological symptoms among the spray painters is shown in the table below.

Refer below – Table 2: The prevalence of neurological symptoms among automotive spray painters

Table 2: The prevalence of neurological symptoms among automotive spray painters

Chronic exposure to solvent-based	**Neurological symptoms**		**Total**
	Yes	**No**	
Yes	35	85	120
No	17	155	172
Total	52	240	292

a. Compute the relative risk of neurological symptoms associated with the chronic exposure to solvent-based paint. Explain your result.

(3 marks)

b. Calculate the attributable risk. Explain your result.

(4 marks)

c. Compute the attributable risk percentage in the exposed group. Explain your result.

(3 marks)

Question 3

Environmental health officers measured lead (Pb) concentrations (ng/m³) around a battery recycling center. Measurements were conducted in February and July after engineering improvements were implemented to minimize Pb emissions. The collected data were analysed using SPSS and are illustrated in the figure below.

Refer below – Figure 1: A.

Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	February	257.73	30	49.040	8.954
	July	174.87	30	15.558	2.840

Correlations

		N	Correlation	Sig.
Pair 1	February & July	30	-.371	.043

Test

					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	February - July	82.867	56.687	10.349	61.700	104.034	8.007	29	.000

Figure 1: A

a) Assuming the data is normally distributed, determine the test used to analyse the above data and state the justification of using the selected test.

(2 mark)

b) State the null hypothesis for this analysis.

(1 mark)

c) Construct a reporting table to summarize the findings and interpret the results.

(5 mark)

d) Determine the decision-making process and state the conclusion based on the findings.

(2 marks)

Question 4

A researcher conducted a cross-sectional study to examine the association between total cholesterol levels and age among adult patients at Putrajaya Hospital who are following a carbohydrate-controlled diet. Total cholesterol levels were obtained through fasting blood tests, while age information was collected through medical records. The collected data were analysed using SPSS and are illustrated in the figures below.

Refer below – Figure 2: B.

Descriptive Statistics

	Mean	Std. Deviation	N
Total cholesterol (mmol/L)	4.151	.2748	35
Age (year)	26.63	6.353	35

Correlations

		Total cholesterol (mmol/L)	Age (year)
Pearson Correlation	Total cholesterol (mmol/L)	1.000	-.990
	Age (year)	-.990	1.000
Sig. (1-tailed)	Total cholesterol (mmol/L)	.	.000
	Age (year)	.000	.
N	Total cholesterol (mmol/L)	35	35
	Age (year)	35	35

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.990[a]	.979	.978	.0403

a. Predictors:

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.514	1	2.514	1547.198	.000[b]
	Residual	.054	33	.002		
	Total	2.567	34			

a. Dependent Variable:

b. Predictors: (Constant),

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	5.291	.030		177.758	.000
	Age (year)	-.043	.001	-.990	-39.334	.000

a. Dependent Variable:

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Age (year)	35	26.63	6.353	18	38
Total cholesterol (mmol/L)	35	4.151	.2748	3.7	4.6

Ranks

		N	Mean Rank	Sum of Ranks
Total cholesterol (mmol/L) - Age (year)	Negative Ranks	35[a]	18.00	630.00
	Positive Ranks	0[b]	.00	.00
	Ties	0[c]		
	Total	35		

a. Total cholesterol (mmol/L) < Age (year)

b. Total cholesterol (mmol/L) > Age (year)

c. Total cholesterol (mmol/L) = Age (year)

Test Statistics[a]

	Total cholesterol (mmol/L) - Age (year)
Z	-5.167[b]
Asymp. Sig. (2-tailed)	.000

a.

b. Based on positive ranks.

Descriptives

			Statistic	Std. Error
Age (year)	Mean		26.63	1.074
	95% Confidence Interval for Mean	Lower Bound	24.45	
		Upper Bound	28.81	
	5% Trimmed Mean		26.48	
	Median		25.00	
	Variance		40.358	
	Std. Deviation		6.353	
	Minimum		18	
	Maximum		38	
	Range		20	
	Interquartile Range		12	
	Skewness		.342	.398
	Kurtosis		-.928	.778
Total cholesterol (mmol/L)	Mean		4.151	.0464
	95% Confidence Interval for Mean	Lower Bound	4.057	
		Upper Bound	4.246	
	5% Trimmed Mean		4.152	
	Median		4.200	
	Variance		.076	
	Std. Deviation		.2748	
	Minimum		3.7	
	Maximum		4.6	
	Range		.9	
	Interquartile Range		.5	
	Skewness		-.016	.398
	Kurtosis		-.863	.778

Figure 2: B

a) Assuming the data is not normally distributed. From the SPSS output above, select the correct analysis test used to analyse the above situation. Provide justification of the test selection and state the null hypothesis for the analysis.

(3 marks)

b) Identify the dependent and independent variables used in this analysis.

(1 mark)

c) Using the SPSS Output, identify the relevant findings and construct a reporting table to summarize the findings. Provide clear interpretation of the results.

(4 marks)

d) Determine if there is a statistically significant association between total cholesterol levels and age among adult patients at Putrajaya Hospital who are following a carbohydrate-controlled diet. Justify your determination.

(2 marks)

SECTION B (Total: 60 marks)

Answer THREE (3) questions only.

Please use the answer booklet provided.

Question 1

A medical researcher conducted a study to determine the average fasting blood glucose level (mg/dL) among patients with prediabetes attending two districts government health clinics. The data collected from all participating patients are summarized in the table below.

Refer below – Table 3: Summary of fasting blood glucose level (mg/dL) among prediabetic patients

Table 3: Summary of fasting blood glucose level (mg/dL) among prediabetic patients

District	Number of prediabetic	Mean	Standard deviation
Kajang	200	174.6	30.5
Putrajaya	150	204.1	28.7

a. What is the probability that the difference of mean fasting blood glucose level of prediabetic patients, $\bar{x}_1 - \bar{x}_2$ between this two universities is more than 40.0 (mg/dL)?

(6 marks)

b. Using the data provided in the table above, construct a 95% confidence interval for the true difference in mean fasting blood glucose levels between the two populations $(\mu_1 - \mu_2)$.

(7 marks)

c. Using the data provided in the table above, construct a 90% confidence interval for the true difference in mean fasting blood glucose levels between the two populations $(\mu_1 - \mu_2)$.

(7 marks)

Question 2

A research team investigated the prevalence of anemia among pregnant women in Selangor. They collected blood samples from 400 women attending antenatal clinics and measured their hemoglobin levels (g/dL). The analysis revealed an average blood hemoglobin concentration of 12.8 g/dL with a standard deviation of 1.5 g/dL.

a) Calculate the percentage of pregnant women with hemoglobin level of 10 g/dL and lower.

(4 marks)

b) Calculate the percentage of pregnant women with hemoglobin levels 13 to 17 g/dL.

(9 marks)

c) Determine the central limits for 90 percent of hemoglobin levels present.

(7 marks)

Question 3

Answer all the following questions.

a) A researcher at Putrajaya Health Centre collected data from patients diagnosed with hypertension, cancer and kidney disease. The research tries to understand how health status influences the type of treatment chosen. The data collected were analysed using SPSS and are presented in figures below.

Refer below – Figure 3: C & Figure 4: D

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Treatment type * Health status	83	100.0%	0	0.0%	83	100.0%

Treatment type * Health status Crosstabulation

			Health status			
			Hypertension	Cancer	Kidney disease	Total
Treatment type	Conventional	Count	9	23	15	47
		% within Treatment type	19.1%	48.9%	31.9%	100.0%
	Alternative medicine	Count	15	11	10	36
		% within Treatment type	41.7%	30.6%	27.8%	100.0%
Total		Count	24	34	25	83
		% within Treatment type	28.9%	41.0%	30.1%	100.0%

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	5.372[a]	2	.068
Likelihood Ratio	5.389	2	.068
Linear-by-Linear Association	2.424	1	.119
N of Valid Cases	83		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 10.41.

Figure 3: C

			Health status	Treatment type
Spearman's rho	Health status	Correlation Coefficient	1.000	-.171
		Sig. (2-tailed)	.	.122
		N	83	83
	Treatment type	Correlation Coefficient	-.171	1.000
		Sig. (2-tailed)	.122	.
		N	83	83

Figure 4: D

i. Assuming the data is not normally distributed. From the SPSS Outputs above, select the correct analysis test used to analyse the above situation. Provide justification of the test selection and state the null hypothesis for the analysis.

(3 marks)

ii. Identify the dependent and independent variables used in this data analysis.

(1 marks)

iii. Using the SPSS Output, identify the relevant findings and construct a reporting table to summarize the findings. Provide clear interpretation of the results.

(4 marks)

iv. Determine if there was a statistically significant association between health status and type of treatment chosen among patients in Putrajaya Health Centre. Justify your determination.

(2 marks)

b) A small survey was conducted by the occupational safety officer to assess the relationship between noise-induced hearing impairment and the workers' workstation positions relative to the heavy-machinery noise sources. All 14 workers from Factory MN were examined using an audiometry booth as part of the assessment. The collected data were analysed using SPSS and are presented in the figures below.

Refer below – Figure 5: E

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Workstation Positions * Noise-induced Hearing Impairment	14	100.0%	0	0.0%	14	100.0%

Workstation Positions * Noise-induced Hearing Impairment

			Noise-induced Hearing Impairment		
			No	Yes	Total
Workstation Positions	200 meter away	Count	1	10	11
		% within Workstation Positions	9.1%	90.9%	100.0%
	400 meter away	Count	3	0	3
		% within Workstation Positions	100.0%	0.0%	100.0%
Total		Count	4	10	14
		% within Workstation Positions	28.6%	71.4%	100.0%

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	9.545[a]	1	.002		
Continuity Correction[b]	5.611	1	.018		
Likelihood Ratio	10.050	1	.002		
Fisher's Exact Test				.011	.011
Linear-by-Linear Association	8.864	1	.003		
N of Valid Cases	14				

a. 3 cells (75.0%) have expected count less than 5. The minimum expected count is .86.

b. Computed only for a 2x2 table

Figure 5: E

i. Identify the correct analysis used to analyse the above situation. Provide justification of the test selection and state the null hypothesis for the analysis.

(3 marks)

ii. Identify the dependent and independent variables used in this analysis.

(1 mark)

v. Using the SPSS Output, identify the relevant findings and construct a reporting table to summarize the findings. Provide clear interpretation of the results.

(4 marks)

iii. Determine if there was a statistically significant relationship between noise-induced hearing impairment and the workers' workstation positions among 14 store workers in Factory MN. Justify your determination.

(2 marks)

Question 4

Answer all the following questions.

a) A comparative health study was conducted to determine the high-density lipoprotein (HDL) or "good cholesterol" markers among individuals with two different lifestyle categories (active in regular exercise and those with a sedentary lifestyle.) The measurements were taken after 60 days of monitoring. HDL is recognized as a protective factor against heart disease, with higher HDL levels associated with a lower risk of developing cardiovascular conditions. The collected data were analysed using SPSS and are presented in figure below.

Refer below – Figure 6: F

Group Statistics

	Lifestyle	N	Mean	Std. Deviation	Std. Error Mean
HDL (mg/dL)	Sedentary	30	38.9667	9.67928	1.76719
	Regular exercise	30	44.4000	10.00207	1.82612

		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
HDL (mg/dL)	Equal variances assumed	.325	.571	-2.138	58	.037	-5.43333	2.54119	-10.52008	-.34658
	Equal variances not assumed			-2.138	57.938	.037	-5.43333	2.54119	-10.52020	-.34647

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.270[a]	.073	.057	9.84200

a. Predictors: (Constant), Lifestyle

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	442.817	1	442.817	4.571	.037[b]
	Residual	5618.167	58	96.865		
	Total	6060.983	59			

a. [redacted]
b. [redacted]

[redacted]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	33.533	4.018		8.346	.000
	Lifestyle	5.433	2.541	.270	2.138	.037

Figure 6: F

i. Assuming the data is normally distributed. From the SPSS Outputs, select the correct analysis test used to analyse the above situation. Provide justification of the test selection and state the null hypothesis for the analysis.

(3 marks)

ii. Using the SPSS Output, identify the relevant findings and construct a reporting table to summarize the findings. Provide clear interpretation of the results.

(5 marks)

iii. Determine if there was a statistically significant difference in HDL (mg/dL) between individuals with two different lifestyle categories (regular exercise and sedentary lifestyle). Justify your determination.

(2 marks)

b) An environmental consultant company was appointed to determine the levels of chromium (Cr) in the soil samples collected from Bukit Goh area, Kuantan, near the bauxite mining activity. During the investigation, it was confirmed that villagers living in the vicinity of Bukit Goh were exposed to this carcinogenic heavy metal. The study also included measuring Cr levels in the villagers' blood samples to assess potential health effects resulting from exposure through inhalation, skin contact, and ingestion. The collected data were analysed using SPSS and are presented in figures below.

Refer below – Figure 7: G

			Chromium (ppm) in Soil	Chromium (ppm) in Blood
[redacted]	Chromium (ppm) in Soil	Correlation Coefficient	1.000	.844**
		Sig. (2-tailed)	.	.000
		N	50	50
	Chromium (ppm) in Blood	Correlation Coefficient	.844**	1.000
		Sig. (2-tailed)	.000	.
		N	50	50

** [redacted]

Ranks

		N	Mean Rank	Sum of Ranks
Chromium (ppm) in Blood - Chromium (ppm) in Soil	Negative Ranks	50[a]	25.50	1275.00
	Positive Ranks	0[b]	.00	.00
	Ties	0[c]		
	Total	50		

a. Chromium (ppm) in Blood < Chromium (ppm) in Soil

b. Chromium (ppm) in Blood > Chromium (ppm) in Soil

c. Chromium (ppm) in Blood = Chromium (ppm) in Soil

Test Statistics[a]

	Chromium (ppm) in Blood - Chromium (ppm) in Soil
Z	-6.154[b]
Asymp. Sig. (2-tailed)	.000

a. [redacted]

b. Based on positive ranks.

Descriptives

			Statistic	Std. Error
Chromium (ppm) in Soil	Mean		25.50	2.062
	95% Confidence Interval for Mean	Lower Bound	21.36	
		Upper Bound	29.64	
	5% Trimmed Mean		25.50	
	Median		25.50	
	Variance		212.500	
	Std. Deviation		14.577	
	Minimum		1	
	Maximum		50	
	Range		49	
	Interquartile Range		26	
	Skewness		.000	.337
	Kurtosis		-1.200	.662
Chromium (ppm) in Blood	Mean		4.854	.3541
	95% Confidence Interval for Mean	Lower Bound	4.142	
		Upper Bound	5.566	
	5% Trimmed Mean		4.813	
	Median		4.700	
	Variance		6.269	
	Std. Deviation		2.5038	
	Minimum		.5	
	Maximum		10.1	
	Range		9.6	
	Interquartile Range		3.5	
	Skewness		.155	.337
	Kurtosis		-.640	.662

Figure 7: G

i. Assuming the data is not normally distributed. From the SPSS output, select the correct analysis used to analyse the above situation. Provide justification of the test selection and state the null hypothesis for the analysis.

(3 marks)

ii. Identify the dependent and independent variables used in this analysis.

(1 mark)

iii. Using the SPSS Output, identify the relevant findings and construct a reporting table to summarize the findings. Provide clear interpretation of the results.

(4 marks)

iv. Determine if there was a statistical effect of Chromium (Cr) exposure from soil on the level of Cr in the blood samples of residents living in Bukit Goh areas, Kuantan. Justify your determination.

(2 marks)

END OF EXAMINATION PAPER