JARAMOGI OGINGA ODINGA UNIVERSITY OF SCIENCE AND TECHNOLOGY

SCHOOL OF HEALTH SCIENCES

UNIVERSITY EXAMINATION FOR DEGREE OF MASTER OF PUBLIC HEALTH

1ST YEAR 2ND SEMESTER 2024/2025 ACADEMIC YEAR

KISUMU CAMPUS

COURSE CODE: **HES 5136**

COURSE TITLE: **STATISTICAL METHODS IN EPIDEMIOLOGY**

EXAM VENUE: **STREAM: MASTERS PUBLIC HEALTH**

DATE:13/1/25 **EXAM SESSION: 9-12.00 NOON**

TIME:3.00HOURS

Instructions:

1. **Answer question 1(Compulsory) and any other three questions**
2. **Candidates are advised not to write on the question paper.**
3. **Candidates must hand in their answer booklets to the invigilator while in the examination room.**

SECTION A

Answer question one(Compulsary)

1. Question one (10 marks).

Sixty-four pregnant women at high risk of pregnancy-induced hypertension participated in a randomized controlled clinical trial comparing 100mg of aspirin daily and a matching placebo during the 3rd trimester of pregnancy. The observed numbers with hypertension are shown in the following table.

Treatment * Hypertension Crosstabulation

			Hypertension yes	Hypertension no	Total
Treatment	Aspirine	Count	5	29	34
		Expected Count	7.8	26.2	34.0
	Placebo	Count	10	21	31
		Expected Count	7.2	23.8	31.0
Total		Count	15	50	65
		Expected Count	15.0	50.0	65.0

i. Give the **estimate** and **approximate 95% confidence interval** for the following of hypertension between aspirin and placebo treated women
   a. Difference in risk (2 marks)
   b. Risk ratio (3 marks)
   c. Odds ratio (3 marks)

ii. Suppose a new study is planned. What sample size is approximately needed in order to have a power of 80% if the risk of hypertension is 0.05 lower in aspirin treated women ($\alpha = 0.05$)? (2 marks)

SECTION B

Answer any three Questions

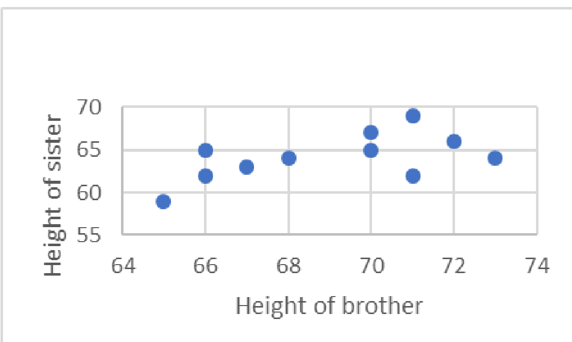
2. Question two (20 marks).

The table below contains an historic dataset on the height (in inches) of brother and sister, published in the second volume of Biometrika by Karl Pearson (indeed the man of the correlation coefficient)

(Pearson K, Lee A. On the laws of inheritance in man. Biometrika 1902; 2:p357. In this case, we look at: - the correlation between the height of brother and si
ster; how to predict height of the brother based on height of the sister or vice versa; - the difference in mean height between brothers and sisters.

family	**Height of brother**	**Height of sister**	**difference**
1	71	69	2
2	68	64	4
3	66	65	1
4	67	63	4
5	70	67	3
6	71	62	9
7	70	65	5
8	73	64	9
9	72	66	6
10	65	59	6
11	66	62	4

Study the SPSS output given below (the SPSS names of variables are: BROTHER, SISTER)



Correlations

		Height of brother	Height of sister
Height of brother	Pearson Correlation	1	.555
	Sig. (2-tailed)		.076
	Sum of Squares and Cross-products	74.000	41.000
	Covariance	7.400	4.100
	N	11	11
Height of sister	Pearson Correlation	.555	1
	Sig. (2-tailed)	.076	
	Sum of Squares and Cross-products	41.000	73.636
	Covariance	4.100	7.364
	N	11	11

Correlations

Descriptive Statistics

	Mean	Std. Deviation	N
Height of brother	69.00	2.720	11
Height of sister	64.18	2.714	11

a. Is Pearson's correlation between height of sister and brother significantly different from zero? Answer this question by solving the following questions (i-iii).
   i. Give bounds for the p-value (3mark)
   ii. Compute also an approximate 90% confidence interval (3 marks)
   iii. Comment on the appropriateness of these assumptions for these data(1mark)

b. Compute the Spearman's correlation coefficient. (3 marks)
   i. Is it statistically different from zero? (1mark)
   ii. Give the upper bound for the p-value (1mark)

Compute the regression line of height of the brother on the height of the sister. (Hint: do not use the original observations, but use the relationship between the correlation coefficient and the slope of a regression line i.e. $r = \frac{s_x}{s_y} b$ or equivalent: $b = \frac{s_y}{s_x} r$)

   iii. What is the best prediction of the height of a brother of a sister with height 70 inches? (3 marks)
   iv. What is the estimated amount of variability in height of brother explained by height of sister? (1mark)

c. Test with a parametric method the hypothesis that the mean height of brother and sister is equal (2marks)
   i. Give the upper bound for the p-value (1mark)
   ii. Give the 95% confidence interval for the mean difference (1mark)

3. Question three (20 marks).

Human beta-endorphin (HBE) is a hormone secreted by the pituitary gland under the condition of stress. An exercise physiologist measured the resting (unstressed) blood concentration of HBE in three groups of men: 15 who had just entered a physical fitness program, 11 who had been jogging regularly for some time, and 10 sedentary people. The mean and standard deviations of the HBE levels (pg/ml) are shown in the table below.

	Fitness program entrants	Joggers	Sedentary
Mean	38.7	35.7	42.5
SD	16.1	13.4	12.8
N	15	11	10

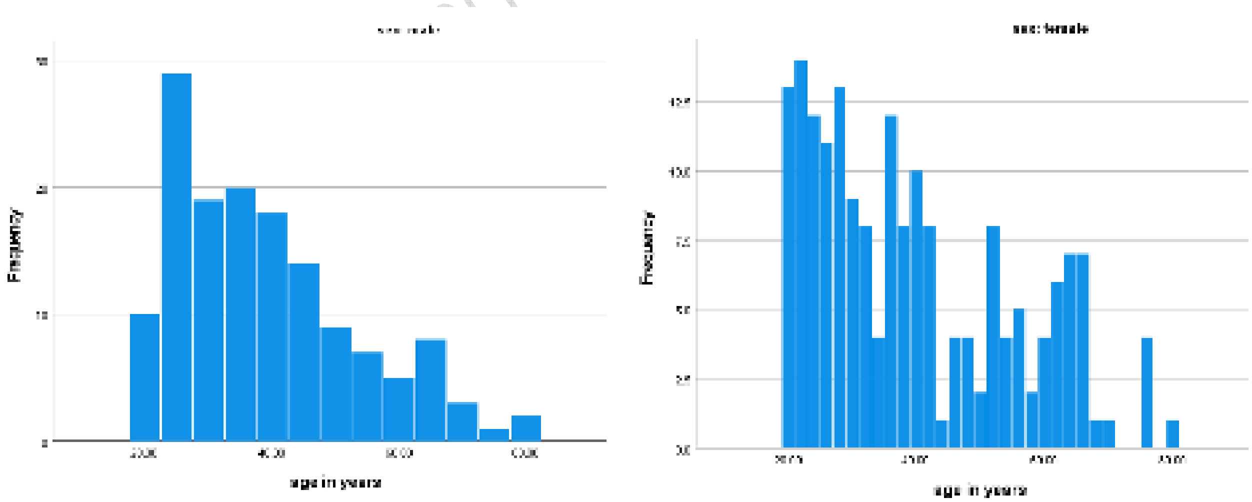
a. Make and complete the ANOVA table (12 marks)
b. Test the null hypothesis that there is no difference in mean HBE levels between the three groups (3 marks)
c. What is the pooled standard deviation (5 marks)

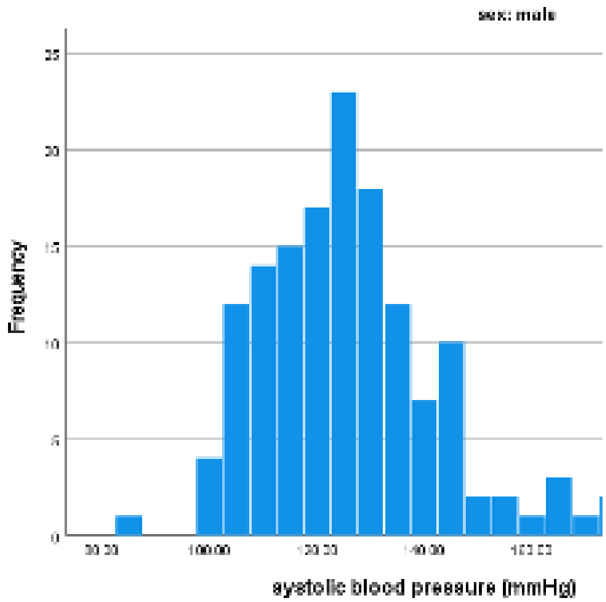
4. Question four (20 marks).

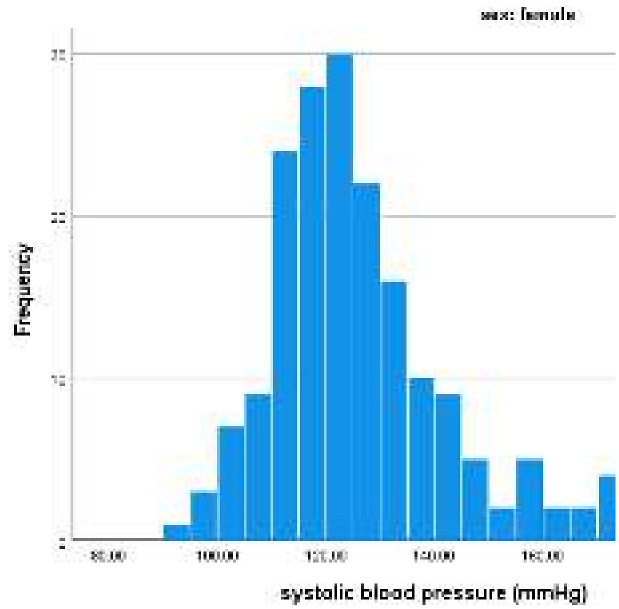
A medical investigator selected from the population of some rural villages in a certain developing country 328 people for his study. Among other variables, systolic and diastolic blood pressure, body weight and pulse frequency were measured. Age and sex were also registered. In the accompanying SPSS output you will find some descriptive statistics and the results of the simple regression analyses of systolic blood pressure on age for males (sex=1, n=145) and females (sex=2, n=183) separately. Use this SPSS output to answer the following questions. First study the results of the analysis for the <u>females</u>. Questions (a) to (f) refer to this analysis.

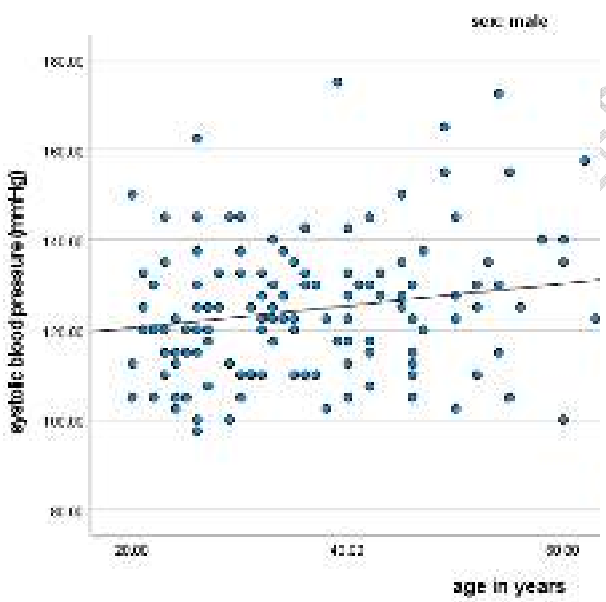
Descriptive Statistics

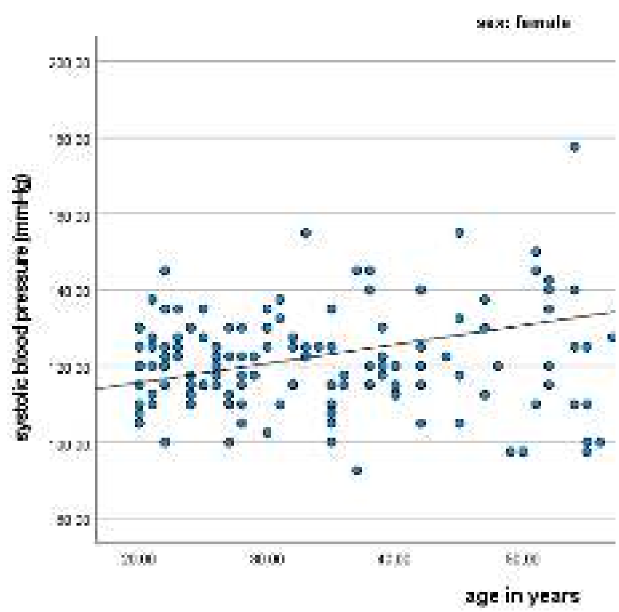
sex		N	Minimum	Maximum	Mean	Std. Deviation
male	age in years	145	20.00	81.00	39.1586	14.40760
	systolic blood pressure (mmHg)	145	85.00	180.00	125.1379	16.64785
	Valid N (listwise)	145				
female	age in years	183	20.00	80.00	39.0984	15.64514
	systolic blood pressure (mmHg)	183	92.50	195.00	125.1913	17.97332
	Valid N (listwise)	183				











Model Summary

sex	Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
male	1	.211[a]	.045	.038	16.32811
female	1	.435[a]	.189	.185	16.22763

a. Predictors: (Constant), age in years

Coefficients[a]

sex	Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
male	1	(Constant)	115.569	3.939		29.340	<.001
		age in years	.244	.094	.211	2.587	.011
female	1	(Constant)	105.649	3.237		32.642	<.001
		age in years	.500	.077	.435	6.501	<.001

a. Dependent Variable: systolic blood pressure (mmHg)

a. Give the estimate for the mean systolic blood pressure of sixty year old women (2marks)

b. Give an estimate of the mean increase per age **decade** for the systolic blood pressure. (1mark)

    i. Give 95% confidence interval for it. (2marks)

c. Give the 90% confidence interval for the mean systolic blood pressure of 30 year old women. (4marks)

d. From the histogram of the systolic blood pressure one can conclude that the distribution is not normal (the distribution is somewhat skewed to the right). Does this imply that the normality assumption underlying linear regression analysis is not fulfilled in this case? (1mark)

Now study also the results of the regression analysis for the male, and answer the following questions.

e. Test whether the difference in mean **yearly** increase of the systolic blood pressure is significantly different between men and women. (the numbers are large, so use a simple and straightforward test). (5marks)

f. The difference in systolic blood pressure between men and women could be studied with the following multiple regression model.

$$systolic\,blood\,pressure = \beta_0 + \beta_1 age + \beta_2 Sex + \beta_3 Sex * age + residual$$

Using the accompanying regression analyses for men and women, give estimates of the β's in this model and their interpretations. Will age play the role of a confounder or effect modifier? (5marks)

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	115.569	3.925		29.441	<.001	107.846	123.292
	age in years	.244	.094	.212	2.596	.010	.059	.430
	sex	-9.920	5.093	-.284	-1.948	.052	-19.940	.100
	Age*Sex	.256	.122	.334	2.100	.037	.016	.495

a. Dependent Variable: systolic blood pressure (mmHg)

5. Question five (20 marks).

The data in the table were collected in Bradford, England, between 1968 and 1977, and relate to 13,384 women giving birth to their first child. The women were classified according to social class (five categories on the Registrar General's scale I-V) and according to the number of cigarettes smoked per day during pregnancy (on a three level categorization: 1 means no smoking, 2 means 1-19 cigarettes per day, and 3 means 20 or more cigarettes per day). The data for each category consist of counts of women showing toxaemic signs (hypertension and/or proteinuria) during pregnancy. The question of interest is how the toxaemic signs vary with smoking status, adjusted for social class. Some SPSS output is given below.

Social Class	Smoking category	No. of women with toxaemic signs	No. of women without toxaemic signs
1	1	131	286
1	2	34	71
1	3	4	13
2	1	350	785
2	2	122	284
2	3	18	34
3	1	1543	3160
3	2	754	2300
3	3	140	383
4	1	328	656
4	2	210	649
4	3	59	163
5	1	121	245
5	2	130	321
5	3	25	65

*Toxaemic signs * Smoking category Crosstabulation*

		Toxaemic signs				Total	
		no		yes			
		Count	% within Smoking category	Count	% within Smoking category	Count	% within Smoking category
Smoking category	non smoking	5132	67.5%	2473	32.5%	7605	100.0%
	1-19 cigarettes	3625	74.4%	1250	25.6%	4875	100.0%
	20 or more cigarettes	658	72.8%	246	27.2%	904	100.0%
Total		9415	70.3%	3969	29.7%	13384	100.0%

Smoking category * Social Class Crosstabulation

		Smoking category						Total	
		non smoking		1-19 cigarettes		20 or more cigarettes			
		Count	% within Social Class	Count	% within Social Class	Count	% within Social Class	Count	% within Social Class
Social Class	1	417	77.4%	105	19.5%	17	3.2%	539	100.0%
	2	1135	71.2%	406	25.5%	52	3.3%	1593	100.0%
	3	4703	56.8%	3054	36.9%	523	6.3%	8280	100.0%
	4	984	47.7%	859	41.6%	222	10.8%	2065	100.0%
	5	366	40.4%	451	49.7%	90	9.9%	907	100.0%
Total		7605	56.8%	4875	36.4%	904	6.8%	13384	100.0%

Toxaemic signs * Social Class * Smoking Category Crosstabulation

Count

		Smoking Category								
		non-smoker			smoker			Total		
		Toxaemic signs			Toxaemic signs			Toxaemic signs		
		no	yes	Total	no	yes	Total	no	yes	Total
Social Class	1	286	131	417	84	38	122	370	169	539
	2	785	350	1135	318	140	458	1103	490	1593
	3	3160	1543	4703	2683	894	3577	5843	2437	8280
	4	656	328	984	812	269	1081	1468	597	2065
	5	245	121	366	386	155	541	631	276	907
Total		5132	2473	7605	4283	1496	5779	9415	3969	13384

Risk Estimate

Social Class		Value	95% Confidence Interval Lower	Upper
1	Odds Ratio for Smoking Category (non-smoker / smoker)	0.988	0.639	1.526
	For cohort Toxaemic signs = no	0.996	0.870	1.141
	For cohort Toxaemic signs = yes	1.009	0.748	1.361
	N of Valid Cases	539		
2	Odds Ratio for Smoking Category (non-smoker / smoker)	0.987	0.780	1.249
	For cohort Toxaemic signs = no	0.996	0.927	1.071
	For cohort Toxaemic signs = yes	1.009	0.857	1.188
	N of Valid Cases	1593		
3	Odds Ratio for Smoking Category (non-smoker / smoker)	0.682	0.619	0.752
	For cohort Toxaemic signs = no	0.896	0.871	0.921
	For cohort Toxaemic signs = yes	1.313	1.224	1.408
	N of Valid Cases	8280		
4	Odds Ratio for Smoking Category (non-smoker / smoker)	0.663	0.547	0.802
	For cohort Toxaemic signs = no	0.888	0.839	0.939
	For cohort Toxaemic signs = yes	1.340	1.169	1.535
	N of Valid Cases	2065		
5	Odds Ratio for Smoking Category (non-smoker / smoker)	0.813	0.610	1.083
	For cohort Toxaemic signs = no	0.938	0.858	1.026
	For cohort Toxaemic signs = yes	1.154	0.947	1.406
	N of Valid Cases	907		
Total	Odds Ratio for Smoking Category (non-smoker / smoker)	0.725	0.672	0.782
	For cohort Toxaemic signs = no	0.911	0.891	0.931
	For cohort Toxaemic signs = yes	1.256	1.190	1.326
	N of Valid Cases	13384		

Tests of Homogeneity of the Odds Ratio

	Chi-Squared	df	Asymptotic Significance (2-sided)
Breslow-Day	11.551	4	.021
Tarone's	11.550	4	.021

Mantel-Haenszel Common Odds Ratio Estimate

Mantel-Haenszel Common Odds Ratio Estimate			
Estimate			0.724
ln(Estimate)			-0.324
Standard Error of ln(Estimate)			0.039
Asymptotic Significance (2-sided)			????
Asymptotic 95% Confidence Interval	Common Odds Ratio	Lower Bound	????
		Upper Bound	????
	ln(Common Odds Ratio)	Lower Bound	????
		Upper Bound	????

a. First look at the cross table of toxaemic signs against smoking category. What test would you choose for the null hypothesis that there is no association between smoking and toxaemic signs?
   i. What are the estimated risks of toxaemic signs for the three smoking categories? (1mark)
   ii. Give for the first smoking category also the corresponding standard error and 95% confidence interval. (3mark)

b. Compute the odds ratio of toxaemic signs of the combined second and third smoking category relative to the non-smokers. (1mark)
   i. Give also a 95% confidence interval (3mark)

c. To correct for possible confounding by social class, a stratified analysis was carried out using SPSS. Since the stratified analysis in SPSS cannot handle larger than 2X2-tables, the smoking categories were combined. Study the output and answer the following questions.
   i. Give the odds ratio per class and compare them with the overall unadjusted one. (1marks)
   ii. What is the most appropriate test for testing the hypothesis that there is no association between toxaemic signs and smoking, adjusted for social class? Give the value of the test statistic and p-value. (2marks)
   iii. Does possible heterogeneity of odds ratio across strata invalidate this test? Is there statistical evidence for the odds ratios being not homogeneous across strata? (1mark)
   iv. Give the results of the tests and comment on it. (1mark)
   v. What is the Mantel-Haenzel estimated odds ratio? (1mark)
   vi. What is its interpretation in this case? (1mark)
   vii. Fill in the question marks in the last table. (5 marks)