UNIVERSITY OF JOHANNESBURG

PROGRAM : MASTER OF PUBLIC HEALTH

SUBJECT **Environmental Epidemiology and Biostatistics**

CODE : **EEB01X1**

DATE : SUPP EXAMINATION
JULY 2017

DURATION : 3 HOURS

WEIGHT : 50: 50

TOTAL MARKS : 150

EXAMINER : DR V NKOSI,

MODERATOR : MS VAN WYK

NUMBER OF PAGES :14

INSTRUCTIONS TO CANDIDATES:

1. Answer all questions.
2. Read your questions carefully, you will be penalized if your answers are not properly structured.
3. You can start with any question, but do not divide sub-questions of the same question.
4. Please write neatly.

GOOD LUCK

QUESTION 1

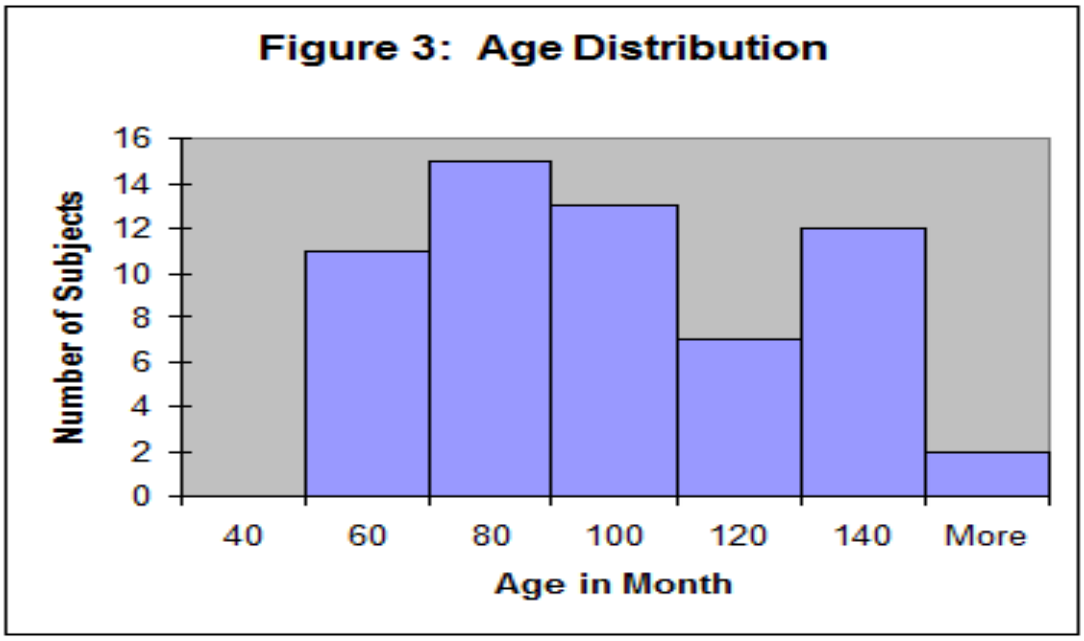
Choose the correct answer:

1.1 Specify at least five key thing that you may consider when reporting the sample size calculation. (10)

[10 Marks]

QUESTION 2

The following histogram shows the distribution of Age (in Months) babies in the National Medicare - Mean: 80; Median: 60; Standard Deviation: 7



2.1 How would you characterize this distribution (symmetric, right skewed, etc.)? (2)

2.2 Suppose you take a random sample of 20 observations from this "population" of over 130 respondents. What shape will the histogram of these sample values likely have? (4)

2.3 Suppose you take a random sample of 100 observations from this "population" of over 130 respondents. What shape will the histogram of these sample values likely have? (4)

[10 Marks]

QUESTION 3

The following data is the annual income (in, or per R10 000) taken from nine randomly selected students in the Johannesburg Environmental Health Program:
66; 89; 41; 98; 76; 77; 68; 60; 60; 67

3.1 Calculate the sample mean score (2)

3.2 Calculate the sample median score (2)

3.3 Calculate the sample mode score (2)

3.4 Calculate the sample standard deviation of these scores (2)

3.5 What population could this sample represent? (2)

[10 Marks]

QUESTION 4

Use the table below to answer the questions.

Blood Pressure Category					
	optimal	**Normal**	**Pre-Htn**	**Htn**	**Total**
Male	20	15	15	30	**80**
Female	2	15	25	25	**70**
Total	**25**	**30**	**40**	**55**	**150**

4.1 What is the probability of selecting a male with optimal blood pressure? (2)

4.2 What is the probability of selecting a patient with Pre-Htn or Htn blood pressure? (2)

Consider the following table which cross-classifies subjects by CVD and gender.

	CVD	**Free of CVD**	**Total**
Men	35	265	300
Women	45	355	400
Total	80	620	700

4.3 What proportion of men has prevalent CVD? (2)

4.4 What proportion of patients with CVD are men? (2)

Consider the following table which cross-classifies subjects by their family history of CVD and current (prevalent) CVD status.

Family History	**Current CVD**		
	No	**Yes**	**Total**
No	215	25	240
Yes	90	15	105
Total	305	40	345

Calculate the following probabilities.

4.5 The probability of Current CVD given that the subject has Family History? (2)

4.6 The probability of Current CVD given that the subject does not have Family History? (2

[12 Marks]

QUESTION 5

5.1 An Randomized Control Trial is planned to show the efficacy of a new drug vs. placebo to lower total cholesterol. What are the hypotheses? Choose the correct one. (2)

A. $H_0: m_P=m_N$ $H_1: m_P \neq m_N$

B. $H_0: m_P=m_N$ $H_1: m_P>m_N$

C. $H_0: m_P=m_N$ $H_1: m_P<m_N$

5.2 The null value of a difference in means is: (2)

A. 0

B. 0.5

C. 1

D. 2

5.3 The null value of a mean difference is: (2)

A. 0

B. 0.5

C. 1

D. 2

5.4 The null value of a relative risk is: (2)

A. 0

B. 0.5

C. 1

D. 2

5.5 The null value of a difference in proportions is: (2)

A. 0

B. 0.5

C. 1

D. 2

5.6 The null value of an odds ratio is: (2)

A. 0

B. 0.5

C. 1

D. 2

5.7. A two-sided test for the equality of means produces a p-value equals 0.20. Reject H_0?

A. Yes

B. No

C. Maybe

5.8 In Framingham Heart Study, we want to assess risk factors for Impaired Glucose. The outcome is Glucose and is grouped into 4 categories (Diabetes (glucose $\geq$ 126), Impaired Fasting, Glucose (glucose 100-125), and Normal Glucose). The associated Risk Factors are Sex (male, female), Age (years), and BMI (normal weight, overweight, obese). What kind of variable is Glucose? (2)

A. Binary
B. Nominal
C. Ordinal
D. Continuous

5.9 What statistical test would be used to assess whether age is associated with Glucose Category? (2)

A. ANOVA

B. Chi-Square GOF

C. Chi-Square test of independence

D. Test for equality of means

E. Other

5.10 Consider a Tertiary Outcome Diabetes Status (Diabetes / No Diabetes). The associated Risk Factors are Sex (male, female), Age (years), and BMI (normal weight, overweight, obese). What test would be used to assess whether AGE is associated with Diabetes? (2)

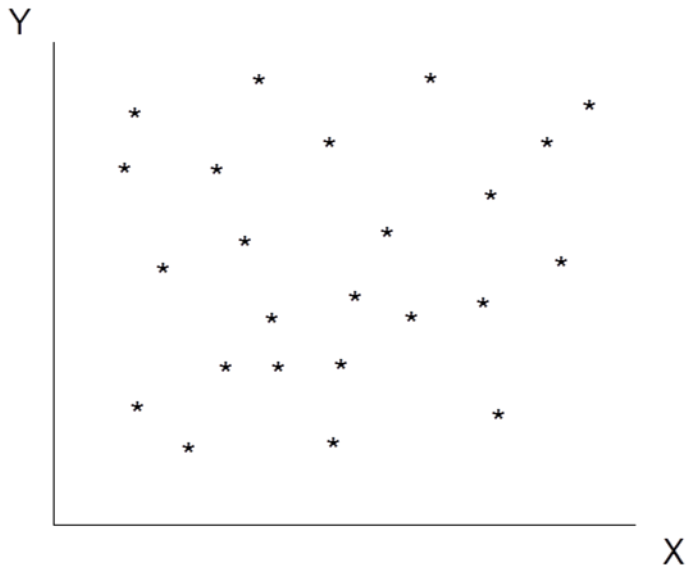
A. ANOVA
B. Chi-Square GOF
C. Chi-Square test of independence
D. Test for equality of means
E. Other

5.11 What test would be used to assess whether BMI is associated with Diabetes? (2)

A. ANOVA
B. Chi-Square GOF
C. Chi-Square test of independence
D. Test for equality of means

E. Other

5.11 Correlation (r) – measures the nature and strength of linear association between two variables at a time. What is the most likely value of r for the data shown below? (2)



A. r=-0.5

B. r=0

C. r=0.5

D. r=1

5.12 What is r (4)

5.13 What is r^2 (4)

[30 Marks]

QUESTION 6

6.1 If the 95% confidence interval for this study is 0.95 to 9. For statistical significance using an alpha level of 0.05, the correct interpretation of these results is that: (2)

A. A statistical significant association exists between family history and an increased risk for chronic otitis media

B. A statistically significant association exists between family history and decreased risk for chronic otitis media

C. It can be concluded with 95% confidence that family history protects against chronic otitis media

D. It can be concluded with 95% confidence that family history increases the risk of chronic otitis media

E. The risk of chronic otitis media is not statistically significantly different between children with family history compared to children without family history of otitis media.

6.2 If a study fails to detect a significant difference between treatment A and treatment B when in fact the treatments do differ, the following type of error has occurred. (2)

A. Type I error
B. Type II error
C. Treatment discrimination error
D. Difference testing error

6.3 Which of the following is true about the *P* value? (2)

A. Indicates the probability of seeing the observed result, and results more extreme, by chance alone (given that the null hypothesis is true)
B. Indicates the probability that the null hypothesis is true
C. Rules out the role of bias and/or confounding
D. Indicates that the results observed are of medical or public health significance

6.4. Which of the following statements is/are true about the 95% confidence interval? (2)

A. If you did the study 100 times and got 100 point estimates and 100 confidence intervals, in 95 of the 100 results, the true point estimate would lie within the given interval.
B. The range within which the true measure of effect lies with a stated probability, or a certain degree of assurance (95%).
C. The confidence interval is calculated around the point estimate and quantifies the variability around the point estimate.
D. The 1st and 2nd answers only
C. All of the above

A paediatrician wished to determine the relationship between chronic otitis media in young children and parental history of such infections at his practice in Monroe County. From the records of a large paediatric practice, he identified 50 children between one and three years of age who had experienced at least three middle ear infections (chronic otitis media) during the preceding year. Fifty children in the same age group treated by the same practice for other illnesses were also identified. The paediatrician interviewed the parents of the subjects in both groups to determine their history of chronic otitis media as young children. Of the

children with recurrent ear infections, 30 had a family history of chronic otitis media, compared to 20 of the children treated for other illnesses.

6.5 What is the most appropriate statistical test for determining whether a significant association exists between chronic otitis media in children between the ages of one and three and a parental history of otitis media? (2)

A. Paired *t* test

B. Chi-square test

C. Correlation analysis

D. Analysis of variance

E. Independent sample (pooled) *t* test

[10 Marks]

QUESTION 7

Consider the SPSS binary logistic regression for HIV survey among adults in South Africa and use them to answer the questions below.

Table 1: HIV status (dependent variable). NB. The largest coded group is identified as the "target".

Dependent Variable Encoding	
Original Value	Internal Value
negative	0
positive	1

Table 2: The coding for all categorical predictors is specified

Categorical Variables Codings						
			Parameter coding			
		Frequency	(1)	(2)	(3)	(4)
educ	no schooling	2146	1.000	.000	.000	.000
	up to std5	431	.000	1.000	.000	.000
	std6 to std10	1680	.000	.000	1.000	.000
	diploma	4028	.000	.000	.000	1.000
	bachelor or higher	807	.000	.000	.000	.000
marstat	married	315	1.000	.000	.000	.000
	single	3225	.000	1.000	.000	.000
	dirvorced	47	.000	.000	1.000	.000

	living together	5278	.000	.000	.000	1.000
	other	227	.000	.000	.000	.000
Race	african	5815	1.000	.000	.000	
	white	1914	.000	1.000	.000	
	coloured	736	.000	.000	1.000	
	indian	627	.000	.000	.000	
Sex	male	5572	1.000			
	female	3520	.000			

Table 3: The contribution of each predictor were it added alone into the equation.

Variables in the Equation

								95% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1[a]	marstat			29.844	4	.000			
	marstat(married)	-.762	.236	10.434	1	.001	.467	.294	.741
	marstat(single)	-.694	.164	17.994	1	.000	.500	.362	.688
	marstat(divorced)	.041	.473	.008	1	.931	1.042	.412	2.636
	marstat(co-habit)	-.383	.171	4.998	1	.025	.682	.487	.954
	age_exact	.028	.004	47.415	1	.000	1.029	1.020	1.037
	Sex(Male)	.504	.071	50.292	1	.000	1.655	1.440	1.902
	Race			322.150	3	.000			
	Race(Black)	2.657	.341	60.753	1	.000	14.252	7.307	27.798
	Race(White)	.893	.359	6.193	1	.013	2.442	1.209	4.932
	Race(Coloured)	-.257	.476	.292	1	.589	.773	.304	1.965
	educ			14.747	4	.005			
	educ(No school)	.519	.160	10.520	1	.001	1.681	1.228	2.300
	educ(up to std5)	.495	.191	6.734	1	.009	1.640	1.129	2.382
	educ(std6-std10)	.580	.158	13.504	1	.000	1.786	1.311	2.433
	educ(diploma)	.559	.151	13.649	1	.000	1.749	1.300	2.352
	Constant	-5.299	.430	151.597	1	.000	.005		

Variable(s) entered on step 1: marstat, age_exact, Sex, Race, educ.

7.1) From table 2 above, Please identify the reference (comparison) group for each categorical variable (4)

7.2). Interpreting the model in table 3: (6)

[10 Marks]

QUESTION 8

The Scenario is based on a paper by Feresu et al

Incidence of stillbirth and perinatal mortality and their associated factors among women delivering at Harare Maternity Hospital, Zimbabwe: a cross-sectional retrospective analysis

Shingairai A Feresu*1,3, Siobán D Harlow2, Kathy Welch4 and
Brenda W Gillespie4

Address: 1Department of Preventive and Societal Medicine, University of Nebraska Medical Center 984350 Nebraska Medical Center, Omaha, NE, 68198-4350, USA,

2Department of Epidemiology, School of Public Health, University of Michigan, 109 Observatory Rd, Ann Arbor, MI 48105,

USA, 3Department of Community Medicine, University of Zimbabwe, Box A178, Avondale, Harare, Zimbabwe and

4Center for Statistical Consultation and Research, University of Michigan, 3554 Rackham Building, 915 E Washington St, Ann Arbor, MI 48109-1070, USA

Email: Shingairai A Feresu* - sferesu@unmc.edu; Siobán D Harlow - harlow@isr.umich.edu; Kathy Welch - kwelch@umich.edu; Brenda W Gillespie - bgillesp@sph.umich.edu
* Corresponding author

Abstract
Background: Death of an infant in utero or at birth has always been a devastating experience for the mother and of concern in clinical practice. Infant mortality remains a challenge in the care of pregnant women worldwide, but particularly for developing countries and the need to understand contributory factors is crucial for addressing appropriate perinatal health.
Methods: Using information available in obstetric records for all deliveries (17,072 births) at Harare Maternity Hospital, Zimbabwe, we conducted a cross-sectional retrospective analysis of a one-year data, (1997–1998) to assess demographic and obstetric risk factors for stillbirth and early neonatal death. We estimated risk of stillbirth and early neonatal death for each potential risk factor.

Results: The annual frequency of stillbirth was 56 per 1,000 total births. Women delivering stillbirths and early neonatal deaths were less likely to receive prenatal care (adjusted relative risk [RR] = 2.54; 95% confidence intervals [CI] 2.19–2.94 and RR = 2.52; 95% CI 1.63–3.91), which for combined stillbirths and early neonatal deaths increased with increasing gestational age (Hazard Ratio [HR] = 3.98, HR = 7.49 at 28 and 40 weeks of gestation, respectively). Rural residence was associated with risk of infant dying in utero, (RR = 1.33; 95% CI 1.12–1.59), and the risk of death increased with increasing gestational age (HR = 1.04, HR = 1.69, at 28 and 40 weeks of gestation, respectively). Older maternal age was associated with risk of death (HR = 1.50; 95% CI 1.21–1.84). Stillbirths were less likely to be delivered by Cesarean section
(RR = 0.64; 95% CI 0.51–0.79), but more likely to be delivered as breech (RR = 4.65; 95% CI 3.88–5.57, as were early neonatal deaths (RR = 3.38; 95% CI 1.64–6.96).

Conclusion: The frequency of stillbirth, especially macerated, is high, 27 per 1000 total births. Early prenatal care could help reduce perinatal death linking the woman to the health

care system, increasing the probability that she would seek timely emergency care that would reduce the likelihood of death of her infant in utero. Improved quality of obstetric care during labor and delivery may help reduce the number of fresh stillbirths and early neonatal deaths.

Interpret the table below and answer the following questions

8.1 Interpret the risk of stillbirth for male vs female infant (1)

8.2 Is this association significant? (1)

8.3 Suppose the crude risk was 1.69 95% confidence interval was 1.69 (0.89- 7.45), interpret these findings (2)

8.4 How do findings in **c** differ from those in **a** above, and what could be the reason (2)

8.5 What does adjusted estimates mean (2)

[8 Marks]

QUESTION 9

9.1 Discuss different approaches that are applied during human exposure assessment (10)

[10 Marks]

QUESTION 10

Chronic respiratory disease among the elderly in South Africa: any association with proximity to mine dumps?

[The following passage has been extracted from an article written by V Nkosi et al and published in Environmental Health 2015;13:33. DOI 10.1186/s12940-015-0018-7]

There is increasing evidence that environmental factors such as air pollution from mine dumps, increase the risk of chronic respiratory symptoms and diseases

Mine dump facilities are the main source of airborne particulate matter pollution, the dust is blown into the surrounding communities and can potentially have adverse health effects on human health and ecology. Communities located close to mine dumps are of lower socio-economic status, often children and the elderly. These communities consist of historically disenfranchised ethnic groups living in government-funded houses, informal settlements and

retired homes. Epidemiological studies have shown that residing near mines is a major risk for exposure to particulate matter and metals such as cadmium, lead, silica, manganese, lead and arsenic. Exposure to mine dump dust that is that is rich in silica has been linked to the development of chronic bronchitis, emphysema and airflow obstruction. Settle-able dust has a negative effect on visibility when it forms dust plumes while its deposition on fabrics, buildings, vehicles and water tanks constitutes a nuisance. The ongoing reclamation of mine dumps for gold recovery observed during the survey, is worsening dust pollution with further deterioration of ambient air quality in the study populations.

The study findings suggest that there is a high prevalence chronic respiratory symptoms and diseases among the elderly in communities located near mine dumps. The significant risk factors are proximity to mine dump, smoking habits, low level of education and domestic use of gas or paraffin.

10.1 Name and describe Bradford Hills Criteria that would be followed to prove that exposure to dust from mine dumps is a cause of chronic respiratory diseases among the elderly? (16)

10.2 Name four routes of exposure to dust from mine dumps (4)

10.3 Name four sources of exposure to dust from mine dumps (4)

10.4 What type of study design was used? (2)

10.5 Name two advantages and disadvantages of the study design used (4)

10.6 How did the researchers control for confounding in this study? (2)

10.6 What is the biomarker (2)

10.7 What type of biomarker can be applied to study exposure to dust from mine dumps (2)

10.8 Name three examples of biomarkers that can be used as samples to study exposure to dust from mine dumps (6)

10.9 Name a limitation of using a questionnaire to collect data of respiratory outcomes (2)

10.10 List and compare differences environmental and occupational epidemiology (6)

[50 Marks]

Total = 150 marks