Practical 3
Inferential Statistics 2

Introduction

This is the third practical session. In this session we will do exercises on Pearson correlation and linear regression.

Pearson Correlation

1a. Write down the formula for Pearson Correlation in the boxes below;

Basic Formula for r	(x-mean x)^2	(y-mean y)^2	(x- mean x)(y-mean y)

As you can see from the formulas above, to calculate the correlation coefficient (r), you need to identify the following;

- Total of the first variable ($\sum x$),
- Total of the first variable squared ($\sum x^2$),
- Total of the second variable ($\sum y$),
- Total of the second variable squared ($\sum y^2$) and
- Total of the two variables multiplied ($\sum xy$).

Just imagine the number of calculations that you have to do before you even get to calculate the correlation coefficient (r). If the sample size is 150, you will have to do more than 455 calculations. Since you'll be doing this calculations manually, the chance of error occurring is quite high indeed☺.

For exercise, complete the following table. Measure the time required to complete it. Once done, please note that you may have to do the same thing again for a dataset 5 times larger than this.…….☺

2. A case-control study to identify factors that can cause small for gestational age – SGA was conducted.

   In the past exercise, we have proven that there is an association between the mothers' first trimester weight and SGA.

   Now we want to see whether there is an association between the mothers' first trimester weight (WEIGHT) and the child's birth weight (BIRTHWGT).

   Please complete the following table;

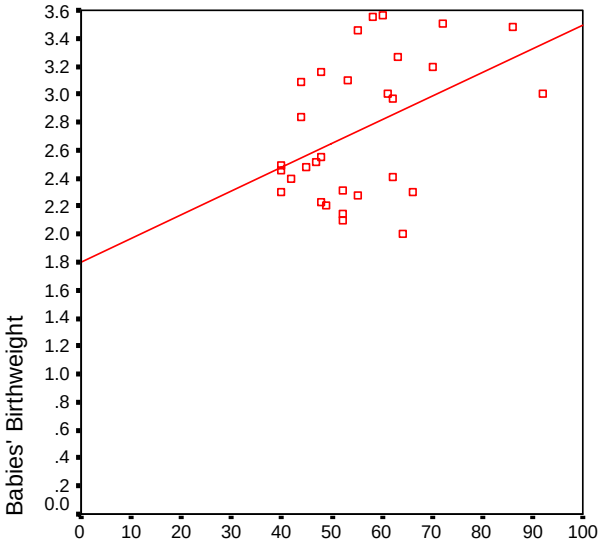
INDEX	WEIGHT	$WEIGHT^2$	BIRTHWGT	$BIRTHWGT^2$	$\sum xy$
9	42.00		2.40		
10	40.00		2.30		
12	66.00		2.30		
20	51.50		2.10		
21	47.50		2.23		
29	39.50		2.49		
31	40.00		2.46		
32	46.50		2.52		
34	55.00		2.28		
43	49.20		2.20		
60	45.00		2.48		
70	63.50		2.00		
72	52.40		2.31		
79	52.30		2.15		
90	47.50		2.55		
97	62.00		2.41		
117	55.10		3.46		
126	72.00		3.50		
131	61.50		2.97		
138	86.00		3.48		
145	60.80		3.00		
146	44.00		2.84		
156	58.00		3.55		
159	70.00		3.19		
171	44.00		3.09		
173	59.50		3.56		
174	47.50		3.16		
175	53.00		3.10		
178	62.50		3.27		
181	92.00		3.00		
TOTAL					

a. State the null hypothesis for correlation test between the two variables.

b. Conduct the correlation test and calculate the r (correlation coefficient). How strong is the relationship between the two variables?

c. Is the r significant? What is the p value? How is it calculated?

If the r is significant, it is best to demonstrate it using a scatter diagram like the one below;



To expect the students to calculate all that during the examination, would be rather cruel. Instead, usually, all the required data will be given, along with some extraneous data, just to confuse the students. It is up to the students to select the appropriate data and use it in the appropriate statistical test.

3. A case-control study to identify factors that can cause small for gestational age – SGA was conducted. Among the factors studied were whether there is an association between the mothers' height in cm (HEIGHT) and the child's birth weight in kilogram (BIRTHWGT).

n = 218	HEIGHT	BIRTHWGT
Mean	151.65	2.79
Standard deviation	5.26	0.54
∑(observation)	33059.00	608.46
∑ (observation^2)	5019291.00	1760.98
∑ (observation 1 x observation 2)	92386.35	

a. Name the appropriate statistical test to test the association between the two variables.

b. State the null hypothesis for the above statistical test.

c. Conduct the statistical test including the test of significance. Discuss the result of the test.

Linear Regression

4a. Write down the formula for linear regression in the boxes below;

Basic Formula	b	a

b. Using the data from Q2, conduct the test for linear regression and calculate the regression co-efficient (b) and constant (a).

c. Write down the final equation of the calculation.

d. Draw a rough diagram of the final equation from the calculation.

Research Project 3

Students will be guided on how to enter the data that they have collected into the computer using Excel or SPSS. Each lab is required to prepare a notebook for the session.

For homework, students are required to complete the data entry for all collected data and bring the completed file to the fourth practical session.