Quiz 3:

1. A hypothetical Singapore based cohort study on diabetes yielded the following raw data. The odds of diabetes for Malays is ___ and the odds ratio of diabetes for Malays, given the Chinese as the baseline, is ___

	Diabetic	Healthy	Total
Chinese	200	500	700
Malay	70	130	200
Indian	23	77	100
Total	293	707	1000

A) 1.86, 0.74

B) 0.54, 1.35

C) 1.86, 1.35

D) 0.54, 0.74

Explanation:

The odds for an event is equal to the number of events divided by the number of non-events. Thus, the odds of diabetes for Malay is 70/130=0.54, and the odds for Chinese is 200/500=0.40. The odds ratio is (70/130)/(200/500)=1.35. Please refer to chapter 2 unit 3 for more information.

2. Multiple sclerosis (MS) is a rare disease (about 50 cases per 100,000 people) that affects the central nervous system. Suppose you want to study the association between smoking and MS. How would you do that?

A) I would design a cohort study. I would randomly select 1000 smokers and 1000 non-smokers. Then I would follow them over the years to see what percentage of each group develop MS. In this way I can calculate the risk ratio.

B) I would design a cohort study. I would randomly select 1000 smokers and 1000 non-smokers. Then I would follow them over the years to see what percentage of each group develop MS. In this way I can calculate the odds ratio.

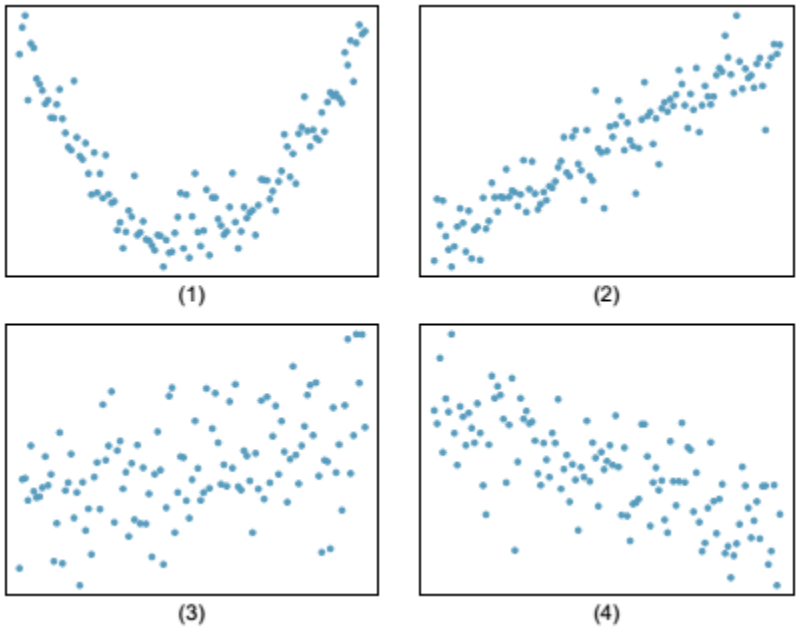
C) I would design a case-control study. I would randomly select 1000 MS patients and 1000 healthy individuals. Then I would look into their smoking habit. In this way I can calculate the risk ratio.

D) I would design a case-control study. I would randomly select 1000 MS patients and 1000 from healthy individuals. Then I would look into their smoking habit. In this way I can calculate the odds ratio.

Explanation:

Since there are about 50 MS cases in 100,000 people, in 1,000 people, we may not observe a single case. Hence, a cohort study is not feasible. For rare diseases, case-control design is more suitable. The risk ratio cannot be calculated based on a case-control study [Chapter 1 Unit 10].

3. Match the calculated correlations to the corresponding scatterplot.



A) r1 = 0.8, r2 = 0.92, r3 = 0.45, r4 = -0.7

B) r1 = 0.06, r2 = 0.92, r3 = 0.9, r4 = -0.7

C) r1 = 0.06, r2 = 0.92, r3 = 0.45, r4 = -0.7

D) r1 = 0.06, r2 = 0.92, r3 = 0.45, r4 = 0.05

Explanation:

For (1), there is a strong non-linear association, but the linear correlation is not strong. (2) shows a relatively strong linear correlation. (3) shows a weak linear correlation and (4) shows a negative correlation. Considering all the above facts, (C) is the best option.

4. The population sizes of a country at various years are given below:

Year	Population (in Millions)
2010	309.3
2000	282.2
1990	248.8
1980	228.5
1970	203.3
1960	179.3
1950	151.3
1940	132.2
1930	123.2
1920	106
1910	92.2
1900	76.2
1890	36
1880	50.2
1870	35.6
1860	31.4
1850	23.2
1840	17.1
1830	12.9
1820	9.6
1810	7.2
1800	5.3
1790	3.9

Draw a scatter plot using EXCEL and choose the best option.

a) The population size increases every 10 years.
b) The correlation coefficient is greater than 0.9, hence a straight line is best for predicting future population sizes.
c) The correlation coefficient is greater than 0.9, but a straight line is not the best for predicting future population sizes.
d) The correlation coefficient is less than 0.9, but a straight line is best for predicting future population sizes.
e) The correlation coefficient is less than 0.9, hence a straight line is not the best for predicting future population sizes.

Explanation:

(A) is not the best option because population in 1890 decreases.

(B) is not the best option because as you can see from the plot, relation is not linear and hence a linear regression may not be the best option.

(C) is the answer because although the association is strong, but it is non-linear

(D) and (E) are not the best options as correlation coefficient is greater than 0.9.

To calculate the value of correlation coefficient, you can use the following online calculator:

http://www.socscistatistics.com/tests/pearson/Default2.aspx

5. Which of the following is true for the correlation coefficient?

A) The larger the range of the data, the closer it is to 1.

B) It cannot be negative.

C) A zero value does not imply absence of association between the variables.

D) It can, although rarely, take a value below -1, but never a value exactly 0.

E) It equals the gradient of the line of best fit.

Explanation:

(A) is not correct. For example, if the x values were multiplied by 2, the value of r is unchanged [chapter 2 unit 6 slide 15].

(B) is not correct because r can be negative.

(C) is the answer because in the case for a non-linear strong association, r can have a value near zero because it shows the strength of linear correlation.

(D) is not true because r is always between -1 and 1.

(E) is not true because in general r is not the gradient of the best fit line.