60080079 Introduction to Statistical Methods
Semester 2 2023-2024
Homework Assignment 2

21 CST H3Art

1. The scores of a reference population on the Wechsler Intelligence Scale for Children (WISC) are normally distributed with $\mu = 100$ and $\sigma = 15$.

1.1 __% of this population have WISC scores below 100.
1) 50
2) 60
3) 70
4) 80

1.2 __% of this population have WISC scores above 140.
1) 0.35
2) 0.36
3) 0.37
4) 0.38

1.3 __.82% of this population have WISC scores between 100 and 120.
1) 30
2) 40
3) 50
4) 60

Write your answer as a three-digit number.

Answer: 142

2. Here are data for a British Ford Escort car. Speed is measured in kilometers per hour; and fuel consumption is measured in liters of gasoline used per 100 kilometers traveled.

Speed (km/h)	Fuel Used (liters/100 km)	Speed (km/h)	Fuel Used (liters/100 km)
10	21.00	90	7.57
20	13.00	100	8.27
30	10.00	110	9.03
40	8.00	120	9.87
50	7.00	130	10.79
60	5.90	140	11.77
70	6.30	150	12.83
80	6.95		

Draw a scatterplot for the data and answer the following question.

2.1 How does the fuel consumption of a car change as its speed increases?

1) Always increases
2) Always decreases
3) Decreases when the speed is lower than 60, but increases when it is higher than 60
4) There is no relationship between the fuel consumption and the speed

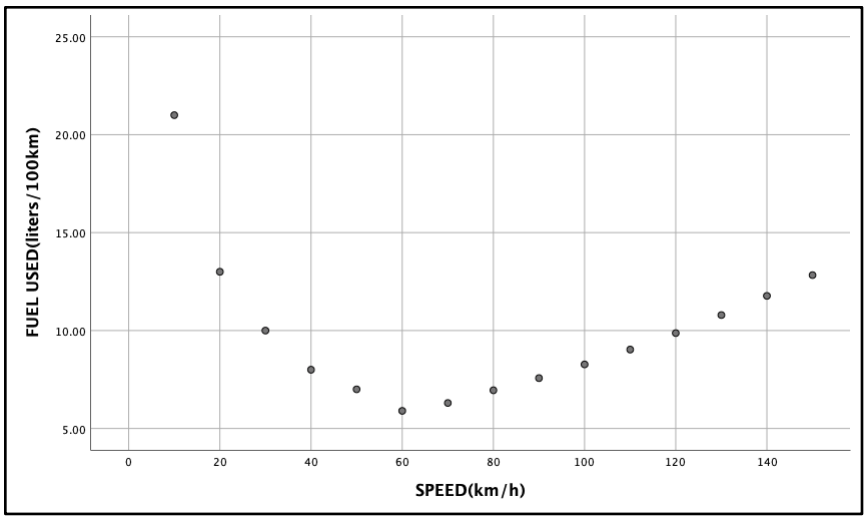
2.2 The direction of the relationship between the fuel consumption and the speed is __.

1) positive
2) negative
3) negative before speed = 60, but positive after speed = 60
4) positive before speed = 60, but negative after speed = 60

2.3 The strength of the relationship (not necessary linear) between the fuel consumption and the speed is __.

1) strong
2) weak

Write your answer as a three-digit number.



Answer: 331

3. Changing the units of measurement of both explanatory and response variables can alter __ and __, but not __ nor __.

1) the correlation between the two variables
2) the intercept of the linear regression
3) the slope of the linear regression
4) the variance in the response variable explained by the explanatory variable

Write your answer as a four-digit number.

Answer: 2314

4. Each of the following statements contains an error. Explain in each case what is wrong.

4.1 "There is a high correlation between the gender of American workers and their income."

1) The unit of income is unknown
2) Correlation can only be applied to quantitative variables
3) The two variables have to have the same unit to calculate their correlation

4.2 "We found a high correlation (r = 1.09) between students' ratings of faculty teaching and ratings made by other faculty members."

1) Correlation can only be applied to quantitative variables
2) Correlation cannot greater than 1
3) The two variables have to have the same unit to calculate their correlation
4) Correlation should have a unit

4.3 "The correlation between planting rate and yield of corn was found to be r = 0.23 bushel."

1) Correlation can only be applied to quantitative variables
2) The two variables have to have the same unit to calculate their correlation
3) Correlation does not have a unit

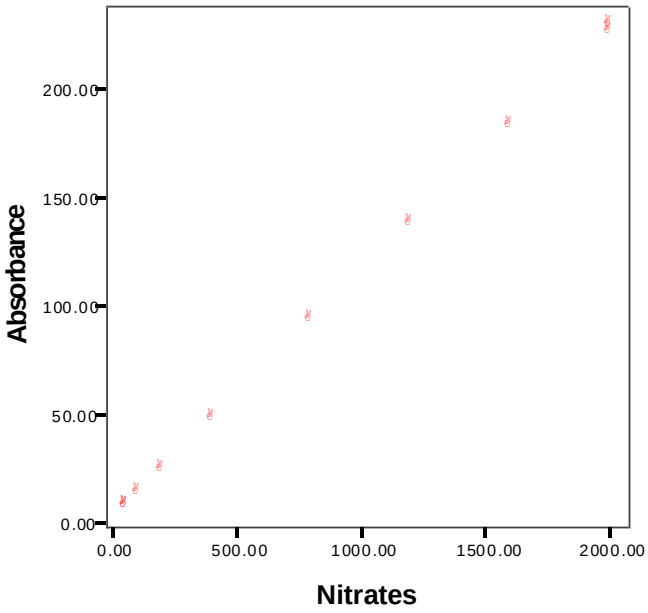
Write your answer as a three-digit number.

Answer: 223

5. Keeping water supplies clean requires regular measurement of levels of pollutants. The measurements are indirect—a typical analysis involves forming a dye by a chemical reaction with the dissolved pollutant, then passing light through the solution and measuring its "absorbance." To calibrate such measurements, the laboratory measures known standard solutions and uses regression to relate absorbance to pollutant concentration. This is usually done every day. Here is one series of data on the absorbance for different levels of nitrates. Nitrates are measured in milligrams per liter of water.

Nitrates	50	50	100	200	400	800	1200	1600	2000	2000
Absorbance	7	7.5	12.8	24	47	93	138	183	230	226

The following is the scatterplot of the data.



5.1 The explanatory variable is __, and the response variable is __.
1) Nitrates
2) Absorbance

5.2 The correlation is __.
1) very close to 1
2) 1
3) 0.8
4) 1.2

Part I. Write your answer as a three-digit number.

Answer: 121

5.3 The intercept of the least-squares line for predicting absorbance from concentration is __, and the slope is __.
1) 1.13
2) 1.66
3) 0.113
4) 16.6

5.4 If the lab analyzed a specimen with 500 milligrams of nitrates per liter, the predicted absorbance is 5_.2.
1) 5
2) 6
3) 7
4) 8

Part II. Write your answer as a three-digit number.

Answer: 234

6. The number of people living on American farms has declined steadily during the past century. Here are data on the farm population (millions of persons) from 1935 to 1980.

Year	1935	1940	1945	1950	1955	1960	1965	1970	1975	1980
Population	32.1	30.5	24.4	23.0	19.1	15.6	12.4	9.7	8.9	7.2

6.1 The least-squares regression line of farm population on year is given by: predicted Population = 11_7 – 0._87 × Year.

1) 5
2) 6
3) 7
4) 8

6.2 According to the regression line, there are __ people declining each year on the average during this period.

1) 117,000
2) 687,000
3) 587,000
4) 687

6.3 About __.7 percent of the observed variation in farm population is accounted for by linear change over time.

1) 56
2) 88
3) 75
4) 97

Write your answer as a four-digit number.

Answer: 2134

7. The following table gives the calories and sodium content for each of 22 brands of meat hot dogs. Brand 15, with just 107 calories, is a low outlier in the distribution of calories.

Brand	Calories	Sodium	Brand	Calories	Sodium
1	149	322	12	139	386
2	158	370	13	136	393
3	139	322	14	153	372
4	148	375	15	107	144
5	152	330	16	135	405
6	141	386	17	140	428

7	153	401	18	138	339
8	157	440	19	129	430
9	149	319	20	132	375
10	147	360	21	135	426
11	146	387	22	142	513

7.1 Use calories as the explanatory variable, and sodium as the response variable to calculate the least-squares regression line WITH the outlier. The intercept is __.604, and the slope is __.536.

1) 18
2) 13
3) 5
4) 2

7.2 Use calories as the explanatory variable, and sodium as the response variable to calculate the least-squares regression line WITHOUT the outlier. The intercept is _9.754, and the slope is _.218.

1) -1
2) -2
3) 55
4) 60

Part I. Write your answer as a four-digit number.

Answer: 2431

7.3 Does a comparison of the two regression lines show that the Brand 15 is influential?

1) yes
2) no

7.4 For this question, we adopt the regression line without the outlier. A new brand of meat hot dog has 170 calories per frank. How many milligrams of sodium do you estimate that one of these hot dogs contains?

1) 352.694
2) 444.724
3) The prediction will not be reliable because of extrapolation.
4) The prediction will not be reliable because we deleted an observation.

Part II. Write your answer as a two-digit number.

Answer: 13

8. A study shows that there is a positive correlation between the size of a hospital (measured by its number of beds *x*) and the median number of days *y* that patients remain in the hospital.

Which of the following conclusions best explains the association between the size of a hospital and the median number of days that patients remain in the hospital?

1) This means that patients can shorten a hospital stay by choosing a small hospital because there is a causation relationship between the two variables.
2) The choice of hospital size and length of stay may be the common response due to seriousness of the illness.
3) Seriousness of the illness may play a role of a confounding factor.

Write your answer as a single-digit number.

Answer: 2