

Statistics 2000, Section 001, Quiz 2 (200 Points)

March 24, 2000, Dr. Jürgen Symanzik

Your Name: _____

First look at all 5 questions. Then start with the question that looks easiest to you. Continue with a more difficult question. Try to answer as many questions as possible in these 50 minutes.

Note that you will obtain at least partial credit if you indicate a correct formula but your final result is incorrect. If you just rely on your calculator without indicating the formula that should be used and your result is incorrect, you will obtain no credit at all for this part of a question.

Question 1: Sums and Straight Lines (30 Points)

As promised, here is one question from our homework assignments:

1. For $x_1 = 5, x_2 = 3, x_3 = 4, x_4 = -2, x_5 = 20$, and $n = 5$, determine the following sums: (18 Points)

$$\sum_{i=2}^{n-1} x_i =$$

$$\sum_{i=2}^{n-1} x^{(i)} =$$

$$\sum_{i=2}^{n-1} x_1 =$$

$$\sum_{i=\frac{n+1}{2}}^{\frac{n^2-21}{2}} \frac{x_{n-i}}{x^{(i+1)}} =$$

2. Determine the slope and the y-intercept of the lines whose equations are given as:
(8 Points)

(a) $5x - 4y = 20$

(b) $x + 6y = 12$

3. Indicate the slope and the y-intercept of the line whose graph looks as follows:
(4 Points)

Question 2: Samples and Populations (60 Points)

Below are the ages of all 40 blood donors from a computer company in Logan, Utah, that visited a bloodmobile unit at their company on a particular day. We consider these 40 blood donors as our population of interest for the remainder of this question. And here is the data:

35 53 61 31 21 42 23 29 35 37

39 58 27 64 27 31 36 48 41 22

37 35 42 32 43 34 59 50 38 43

31 30 41 37 29 45 23 56 46 41

1. Obviously, **35%** of the blood donors who visit the bloodmobile at this company this particular day fall into the 30 to 39 years age group. Is this number (i.e., **35%**) a parameter or a statistic? (**5 Points**)

2. Based on this data, do you think it is correct to state that about 35% of all blood donors anywhere in the United States fall into the 30 to 39 years age group? Explain, using the appropriate statistical vocabulary. (**15 Points**)

3. A local newspaper reporter from the Utah Statesman wants to learn more about some of these blood donors and their motivation to come to the bloodmobile. Since the reporter once took a Stat 2000 class, she thinks that it might be best to randomly select people for the interview rather than taking the first few people that showed up at the bloodmobile. The reporter wants to talk to 5 people and the bloodmobile operator hands her a list of those 40 people that participated.

Explain how you would proceed if you were the reporter, using the appropriate statistical vocabulary. Use the attached random digits table, starting at line 120, to draw your sample. Which persons (indicate the numbers and circle the corresponding age on the previous page) do you finally select for your interview? **(20 Points)**

4. Based on your 5 persons from (3.) above, calculate the sample proportion \hat{p} of blood donors at this company that falls into the 30 to 39 years age group. Is this number a parameter or a statistic? **(8 Points)**
5. Unfortunately, our reporter does not have the time to calculate the mean age for those 5 persons from (3.) above. Instead, she asks her assistant (a student who dropped out of Stat 2000 after the first week) to calculate the mean age of these 5 persons. The student comes up with 22 years. Should the reporter be suspicious about this number or is this just a possible effect of sampling? Explain, using the appropriate statistical vocabulary. Note that the reporter may be using different persons than the 5 persons you used in the previous parts of this question. **(12 Points)**

Question 3: Linear Regression & Correlation (58 Points)

The average prices (in dollars) per ounce of gold and silver for the years 1986 through 1994 are given below. (Source: U.S. Bureau of Mines.)

Year	Gold	Silver
1986	368	5.47
1987	478	7.01
1988	438	6.53
1989	383	5.50
1990	385	4.82
1991	363	4.04
1992	345	3.94
1993	361	4.30
1994	389	5.30

1. Draw a scatterplot where x is the silver price and y is the gold price. Does the appearance of the scatterplot suggest a linear relationship between x and y ? (10 Points)

2. Fit a least squares (linear regression) line to the data. Indicate what your variables stand for. It might help to know that:

$$\sum_{i=1}^n x_i = 46.91, \sum_{i=1}^n x_i^2 = 253.6095, \sum_{i=1}^n y_i = 3,510, \sum_{i=1}^n y_i^2 = 1,383,102, \sum_{i=1}^n x_i y_i = 18,625.9$$

(38 Points)

3. Calculate Pearson's correlation coefficient r between x and y . How can we interpret this value for our given data set? **(10 Points)**

Question 4: Mosaic Plots (32 Points)

Look at the Mosaic Plot of the Titanic data again.

1. Mark the following regions, using the letters “A”, “B”, and “C”:
“A”: Crew, Adult, Female, Survived.
“B”: Third Class, Child, Female, Not Survived.
“C”: Second Class, Child, Male, Not Survived.
(12 Points)

2. Indicate which of these statements are correct and which are incorrect:
 - (i) There were more Third Class than First Class passengers.
 - (ii) The majority of Third Class, Adult, Female passengers survived.
 - (iii) Among the Adult, Male passengers, the largest percentage that survived was in First Class.
 - (iv) There were more Female than Male children on board of the Titanic.
 - (v) There were more Male Third Class passengers than Male First Class and Male Second Class passengers together.**(20 Points)**

Question 5: Probability (20 Points)

Assume that a person has to taste 4 different types of cereal and rank those cereals from 1 (worst) to 4 (best). Let us further assume that this person thinks that all 4 cereals are equally good and therefore randomly assigns the numbers 1 through 4. A possible sample space Ω that describes this ranking when just one testing person has to taste the cereal is

$$\begin{aligned}\Omega = \{ & 1234, 1243, 1423, 1432, 1324, 1342, \\ & 2134, 2143, 2413, 2431, 2314, 2341, \\ & 3214, 3241, 3421, 3412, 3124, 3142, \\ & 4231, 4213, 4123, 4132, 4321, 4312\}\end{aligned}$$

1. We assume that each sample point from Ω is equally likely to occur. Indicate **(12 Points)**

$$P(\text{testing person indicates ranking } 1342) =$$

$$P(\text{testing person indicates a ranking where 1 occurs in 1st position}) =$$

$$P(\text{testing person indicates a ranking where 2 occurs in 2nd position}) =$$

$$\begin{aligned}P(\text{testing person indicates a ranking where 1 occurs in 1st position} \\ \text{and 2 occurs in 2nd position}) =\end{aligned}$$

2. Based on your answers in (1) above, are the events “1 occurs in 1st position” and “2 occurs in 2nd position” dependent or independent. Explain why. **(8 Points)**