Applied Statistics: Sample Exam

Examination of course 2DMT00

A longitudinal data set on the neurological performance of children who are conceived with in-vitro fertilization

Edwin van den Heuvel

Nazanin Nooraee

Data Description

Women who have difficulties in getting pregnant can apply for an in-vitro fertilization treatment. These treatments do increase the probability of getting pregnant, but little is known about its long-term effect on the offsprings. A longitudinal follow-up study of mothers participating in the randomized controlled trial on fertilization was conducted. This longitudinal data set contains information on the neurological performance of $n = 253$ children observed at three different ages (4, 10, and 18 months). The neurological performance in children is quantified with the infant motor profile (IMP) and higher scores means better performance. The data set contains the following variables

1. Design variables
   (a) ID: A unique patient number
   (b) TRT: Treatment indicator (Control = 0, TRT-M = 1, TRT-C = 2)
   (c) PER: Age of child for neurological tests (4, 10, or 18 months)
2. Characteristics of pregnancy and delivery
   (a) GA: Duration of pregnancy measured from last menstrual period (in weeks)
   (b) SEC: Cesarian section (No = 0, Yes = 1)
3. Characteristics of the child
   (a) SEX: Gender of the offspring (Girl = 0, Boy = 1).
   (b) BW: The weight of the child at birth (in grams)
   (c) FB: First born child (No = 0, Yes = 1)
   (d) IMP: The neurological response (an integer$\in [0, 100]$)

All questions in this exam are either related to the analysis of this data set or otherwise to theory that is related to such an analysis. Write down all the answers on the exam paper and provide a well-organized SAS file (or a file from any other statistical software package) with the codes that were used to answer the questions.

All tests are performed at an overall significance level of $\alpha = 0.05$.

Questions

1. In this exercise we will consider gestational age (GA) as the outcome variable of interest and ignore all other variables. To be able to answer the questions, create a data set that contains only data from period 4 months (PER=4) and calculate the transformation $\log(44 - GA)$, with log the natural logarithm.

   (a) Report the first, second, and third quartile, the IQR, the percentiles p_1, p_5, p_{95}, and p_{99}, the minimum, the maximum, and the range for gestational age.

   (b) Provide a 95% confidence interval on the first quartile for gestational age. Provide the necessary calculation details.

   (c) Test if gestational age is normally distributed. Report the value of the test statistic, the p-value, the conclusion, and the reason for choosing this specific test statistic.

   (d) Using the normality assumption of the transformation $\log(44 - GA)$, give a 95% prediction interval for a new observation of gestational age.

   (e) Premature birth is defined as gestational age less than 37 weeks. Literature suggests that premature births occur once in every ten deliveries. Report the proportion of premature births with an appropriate 95% confidence interval. Conclude if the group of women under study have a higher risk of premature births?

2. In this exercise we will consider the neurological response (IMP) of the child at time points 10 and 18 months. Create a data set where the IMP scores at 10 and 18 months are in different columns next to a column that indicates the child identity (ID).

   (a) Apply Grubbs outlier test on the neurological response at 10 months to determine if there exist an extreme lower neurological response. Report the test statistic, the critical value, and discuss the results.

   (b) Calculate Pearson's rho correlation coefficient between the neurological response at 10 and 18 months and provide an appropriate 95% confidence interval on this correlation coefficient.

   (c) It is assumed that the Farlie-Gumbel-Morgenstern copula describes the association between the neurological responses at 10 and 18 months. Estimate the association parameter

   (d) From the age of 10 months it is expected that the neurological score is at least 85. In case it is lower, the child is at risk to develop a neurological deficiency. Test the null hypothesis that the proportion of children with a possible neurological deficiency is the same at 10 and 18 months. Report the null hypothesis, the alternative hypothesis, the test statistic, the result of the test statistic, the p-value, and the conclusion.

   (e) The child ID also indicates the order in which the children were tested at 10 month. Calculate with the conditional random runs test using the median value to investigate if the neurological responses of the children at 10 months are independent. Report the number of runs, the asymptotic test value, the corresponding p-value, and the conclusion. (TIP: eliminate the eight missing values in the neurological response at 10 month from the data set).

3. In this exercise we will investigate differences between subgroups of children (independent samples) on the data that is reported at four months (PER=4). First create this data set or otherwise use a statement that guarantees that the correct data is selected.

   (a) To test the null hypothesis that girls and boys have the same birth weight with a two-sample test, investigate if the assumption of equal variances can be assumed. Report the null hypothesis for this assumption, the alternative hypothesis, the test statistic, the value of the test statistic, the reason for the choice of this test statistic, and the conclusion.

(b) Assuming that the data is normal, perform a two-sample t-test. Report the null hypothesis, the alternative hypothesis, the test statistic, the value of the test statistic, the p-value, and the conclusion.

(c) Test if first born children are more likely to be boys than girls. Report the null hypothesis, the alternative hypothesis, the test statistic, the value of the test statistic, the p-value, the choice of test statistic, and the conclusion.

(d) Test with the Kruskal-Wallis test if treatment affects the neurological response at four months. Report the null hypothesis, the alternative hypothesis, the test statistic, the value of the test statistic, the p-value, and the conclusion.

Good luck!