Lab classes, Biostatistics, Biomedical Engineering

Sara Pérez Jaume

Josep Lluís Carrasco Jordan

August 2023

Contents

Lab class 1

Question 1

Use R to answer the following questions:

Create a vector with 1 ones, 1 twos, 1 threes, 5 eights, 8 nines, 2 tens, 3 twelves, 4 fifteens and 5 sixteens. Call this vector x.

Create a second vector with the consecutive integer numbers between 1 and 38. Call this vector y.

Create a third vector with the consecutive odd integer numbers between 1 and 43. Call this vector z.

- The sum of all the values in x is …
- If all the values in y are multiplied, and the natural logarithm is applied to the result, we obtain the following value: (round the result to 1 decimal digit) …
- Create a vector (called x1) that contains the first nine elements of x. Create another vector (called y1) that contains the last nine elements of y. Compute the element-wise division of x1 by y1. Now sum all the values obtained. The result is (rounded to 2 decimal digits) …

Now create a 2-column matrix by merging the first 18 elements of x (first column) and the last 18 elements of z (second column). Call this matrix X. What is the value located at row 3 and column 2 of X? …

Create a vector with 42 A and 25 B. Call it w. Create a 3-column object with the first 30 elements of x, the first 30 elements of y, and the first 30 elements of w. To create this object you should use the function … . Call this new object dades.

Assign the following names to the columns of dades: Var1, Var2 and Grup. To assign the names you have to use the function …

Save the dades object to a .txt file using a comma as the column separator. The file should contain no row names. In this case you must use the function …

Question 2

The file pumps.txt contains information on the operation of the pumps in three clinical services (A, B and C), pumps in stock (S) and pumps of a new design

(N). The pumps can work with two doses (dose 1 or 2). To analyse some of the characteristics of the pumps, they were running for 12 hours, and some outcomes were collected. The collected information was:

- Dose (**dosi** variable)
- The origin of the pump (**origen** variable)
- If the alarm sounded, 1=yes, 0=no (**alarma** variable)
- Number of times that the alarm sounded (**nalarma** variable)
- Time until the first alarm sounded (**temps.alarma** variable)
- Maximum pressure of the system expressed in psi (**pressio** variable). The system is blocked and the maximum pressure reached is registered.
- System flow expressed in ml/h (**flux** variable). A certain amount of fluid is injected into the system and we measured how long it takes for the fluid to be expelled.

- Read the file containing the data. Which R function should you apply? … What is the column separator character? … Once the file has been read we observe that it contains … observations and … variables.
- Define the **dosi** variable as qualitative. The appropriate function to do this is …
- Generate two data frames, one for each dose. You should use the following R function: …
- Consider the data.frame containing only data at dose 1. Save this data.frame in a .txt file. What R function should you apply? …
- Using the data at dose 1, create a new variable by categorizing the variable **temps.alarma**. What R function can you apply? … The new variable should aggregate the time to alarm data according to the following intervals:

Interval		**Category**
[0,1)	→	Short
[1,6)	→	Medium
[6, ∞)	→	Large

- The new variable values have to be the categories in the table. To achieve this, we can use the argument … . How many subjects are there in the category Medium? …

Lab class 2

Load the file pumps.txt and answer the questions. The file contains information on the operation of some pumps. Remember the description of the variables that are collected:

- Dose (**dosi** variable)
- The origin of the pump (**origen** variable)
- If the alarm sounded, 1=yes, 0=no (**alarma** variable)

- Number of times that the alarm sounded (**nalarma** variable)
- Time until the first alarm sounded (**temps.alarma** variable)
- Maximum pressure of the system expressed in psi (**pressio** variable). The system is blocked and the maximum pressure reached is registered.
- System flow expressed in ml/h (**flux** variable). A certain amount of fluid is injected into the system and we measured how long it takes for the fluid to be expelled.

Round the result to 3 decimal digits (at least) for numerical answers unless otherwise indicated.

- We want to know the number of pumps whose alarm sounded (at least once) during the 12-hour run in clinical services A, B and C. So we need to compute …

  In these services (A, B and C) during the 12-hour run, the alarm has sounded in … pumps at low dose (dose = 1); that is, a percentage of …% of total pumps in the clinical services A, B and C at low dose.

  Moreover, if you look at high dose results (dose = 2), the alarm sounded in … pumps; that is, a percentage of …% of total pumps in the clinical services A, B and C at high dose.

From now on, we will only use the data at low dose (dose = 1) in the clinical services A, B and C.

- The mean number of times that the alarm sounded during the 12-hour follow-up is …
- In 25% of the pumps, the alarm sounded … or less times during the 12-hour monitoring.
- Now select the pumps whose alarm sounded (at low dose in the clinical services A, B and C). The median of the variable **temps.alarma** is … and the standard deviation is …
- Use the appropriate chart to describe the variable **temps.alarma** for every service separately (in those pumps whose alarm sounded). Which service shows outliers for this variable? …

Now use the data from all the pumps in services A, B and C and dose = 1 (include again the pumps whose alarm did not sound).

- Categorize the variable **number of times that the alarm sounded** in the following levels: Never, Between 1 and 3 (both included), 4 or more times.
- The number of pumps whose alarm sounded 4 or more times is equal to … out of a total of … infusion pumps, which as a percentage is …%.
- The mode of the new variable is equal to … (if there is a tie, choose the lowest category).
- If we want to describe this new variable as a function of the clinical service, we should use …

- In service A, the alarm sounded 4 or more times in a …% of the infusion pumps, whereas this percentage was …% in service C.

From now on, the questions continue refer to the pumps in services A, B and C at low dose (dose = 1). Note that variable **pressio** is expressed in psi, but we want to change the units to atmospheres (1 atm = 15 psi, approximately).

- The mean is equal to … atm while the interquartile range is … atm.

Lab class 3

Question 1

Load the file pumps.txt corresponding to the example of the infusion pumps.

Only use data at a low dose (dosi = 1).

Generate a contingency table showing the frequencies of the variables **Number of alarms** (nalarma) and **Pump origin** (origen).

Answer the following questions using the table. Round the results to 3 decimal digits (at least).

- Probability that a pump comes from origin C AND the alarm sounds: …
- Probability that a pump does not sound independently of its origin: …
- Probability that the alarm sounds fewer than 5 times if the pump is of origin S: …
- If the alarm of a pump sounded more than twice, the most probable origin is …, with probability …

Question 2

The usual treatment of breast cancer consists of several cycles of chemotherapy before surgical treatment. The final outcome of the chemotherapic treatment is classified as a success (complete or partial reduction of the tumour) or failure (no reduction of the tumour). Given that the treatment consists of several cycles, it is important to early predict the treatment success or failure in the initial cycles. For this purpose, positron emission tomograpy (PET) (a non-invasive image technique) is applied. Using PET, the uptake of the tumoral mass is quantified as the percentage of the standardized uptake value (SUV). The difference in SUV between the second and first cycles has been proposed as a marker of the final result of the chemotherapy. A

difference in percentage of SUV equal to or greater than 58% is considered to be a positive result.

Data from the study are stored in the file SUV.txt, which contains 3 variables: subject identifier (ID); chemotherapy treatment outcome (Status: E=success, F=failure) and the difference in uptake percentage (SUV).

Compute the following quantities and answer the questions. Remember that the event to predict is the success of the treatment.

Round to 4 decimals for the probabilities and to 2 for the remaining quantities.

The TPF is … and the FPF is … . Thus, this test gives more correct results in subjects that at the end of the treatment have …

The likelihood ratios are LR+= … and LR-= …; a … result is therefore more informative.

If a subject has a percentage difference in SUV greater than or equal to 58%, the probability of complete or partial reduction is …

If the difference in the percentage SUV is lower than 58%, the probability of complete or partial reduction is …

Now let us assess the diagnostic ability of the continuous marker "difference of % SUV". The value of the area under the ROC curve is (use 2 decimals) … . If we use the following rating:

- AUC $\in$ [0.5,0.7), diagnostic ability is poor.
- AUC $\in$ [0.7,0.9), diagnostic ability is good.
- AUC $\geq$ 0.9, diagnostic ability is excellent.

...we can affirm that the diagnostic ability of the difference in % SUV is …

Lab class 4

Question 1

Let us continue with the injection pump assay example. Round to (at least) 4 decimals for probabilities and to 2 for the remaining quantities. To answer the following questions you DO NOT NEED to load any data files.

- The engineer who designed the pumps is convinced that in 12 hours of working, the alarm will sound in 12 out of 25 pumps. Additionally, it is reasonable to assume that pumps are independent in relation to the

behaviour of their alarms. Thus, the variable **number of pumps whose alarm sounds** follows a … distribution.

- Hence, concerning the 50 injection pumps in stock it is expected that the alarm will sound in … pumps after 12 hours of working.
- The probability that the alarm sounds in 28 or more pumps in stock is …
- However, the probability that the alarm sounds in 28 or more pumps in stock but also in 37 or fewer pumps is …
- With a probability of 91%, we could affirm that … alarms will sound in the pumps in stock as much.

- Now let us suppose that after 12 hours working, we revise the pumps one after another to check whether the alarm sounded. The variable **number of pumps to revise before finding the first pump with an alarm that sounded** follows a … distribution. Therefore, the expected number of pumps to revise is (include the pump with the alarm that sounded in the count) …
- The probability that one alarm sounded in any of the first 3 pumps that were revised is …
- However, the probability of finding the first alarm that sounded exactly in the 5th revised pump is …
- Otherwise, the variable **number of pumps with no sounding alarm before finding 2 pumps with an alarm that sounded** follows a … distribution. Therefore, the expected number of total pumps to revise before finding 2 pumps whose alarm sounded is (include the 2 pumps in the count) …
- The probability of revising 6 or less pumps before finding 2 pumps whose alarm sounded is …
- Furthermore, with a probability of 98%, it will be necessary to review … pumps as much to find 2 pumps whose alarm sounded.

- Now let us focus on the variable **number of sounding alarms per pump**. The manufacturer thinks that pumps in stock have a rate of 0.219 alarms/hour. Assuming independence between the alarms, this variable follows a … distribution.
- The probability of no alarms sounding in an hour is … . However, the probability of no alarms sounding in 2 hours is …
- In any case, the expected number of observed alarms in 2 hours is …
- Now let us consider a period of 6 hours. We may affirm that, in 96% of the pumps, … alarms will sound as much.

Question 2

Let us continue with the injection pump assay example. Round to (at least) 4 decimals for probabilities and to 2 for the remaining quantities. To answer the questions you DO NOT NEED to load any data files.

- Regarding the maximum pressure of the pump system, the engineer believes that pressure values follow a Normal distribution with a mean equal to 21.6 psi and standard deviation equal to 2 psi. Thus, the probability that a pump has a pressure lower than 23 psi is … . Moreover, there is a probability of … that the pressure reaches a value greater than 21.6 psi.
- Nevertheless, 71% of the pumps will reach a pressure superior to … psi.
- Yet approximately 99% of the pumps will have pressure values (interval symmetric in relation to the mean) between … and … psi.
- Furthermore, the engineer is convinced that the number of sounding alarms per pump follows a Poisson process. Therefore, in a specific time interval, the rate of alarms …
- Thus, …
- Following this hypothesis, the manufacturer thinks that the pumps in stock have a rate of 0.143 alarms/hour. The probability that an alarm (at least) sounds in an hour is …
- The expected time needed to observe one alarm sounding is … hours.
- Hence, the probability that one alarm takes more than 13 hours to sound is … . Additionally, in 85% of cases one alarm will sound before … hours.
- However, the probability that one alarm sounds twice before 5 hours is …

Lab class 5

In this exercise we will use the pump infusion data: pumps.txt (TO DOWNLOAD: RIGHT CLICK + DOWNLOAD).

Remember that pumps from origin A are those from the service where some complaints were received. We are going to use these data to find out if the complaints were justified. Use only data at low dose (DOSI=1), which we consider a sample of the potential population of pumps.

Round to a precision of (at least) 4 decimals for the proportions/probabilities and 2 decimals for the remaining values.

Firstly, let us assess the proportion of pumps whose alarm sounded. To estimate a confidence interval for a proportion, the applicability conditions are: … . Do they hold in this case? …

So, the 95% confidence interval for the proportion of pumps whose alarm sounded is defined by the values (lower limit) … and (upper limit) …

In a period of 12 hours under standard working conditions, the manufacturer expects that 80% of the pumps as much will detect some anomaly and the alarm will sound. If the frequency is higher, this must be caused by pump malfunction and the pumps should be replaced.

In order to demonstrate that they are non-conforming (using the probability of a sounding alarm), it is therefore necessary to carry out a test of …

In this test, the alternative hypothesis (which we aim to demonstrate to decide that the pumps should be replaced), with a predefined small risk of error, is …

To solve this hypothesis, the test statistic, under the null hypothesis, is distributed as … but only if the following is fulfilled: …

In this case, using the appropriate procedure in relation to the tested hypothesis, we get a p-value of … . If the probability of type I error (alpha) is set to 5%, the decision is …

Thus, the conclusion is: …

Now let us assess the variable **time to the alarm sounds** considering only those pumps from origin A at dose 1 whose alarm has sounded.

To estimate a confidence interval for the mean **time to the alarm sounds**, the applicability conditions are: …

The mean **time to the alarm sounds** is … hours with a 95% confidence interval of (lower limit) … and (upper limit) …

The manufacturers believe that given the sample results, they can affirm that the mean of the time to the alarm sounds is lower than 4 hours. Let us find out if this hypothesis is right.

We are therefore dealing with a hypothesis test of …

In this test, if \(\mu\) is the mean of the time to the alarm sounds, the alternative hypothesis (the affirmation of the manufacturers) is …

In this case (given that applicability conditions hold), the distribution of the test statistic under the null hypothesis is …

In this case, and using the appropriate procedure in relation to the hypothesis tested, we get a p-value of …

So the conclusion is: …

If a pump from origin A at low dose whose alarm has sounded is chosen at random, we may affirm with 95% confidence that the time to the alarm sounds will be within (lower limit) … and (upper limit) … hours.

However, we may say with 95% confidence that 80% of the pumps with origin A will take to sound between (lower limit) … and (upper limit) … hours.

Lab class 6

To answer the questions use the blood pressure data: bloodpressure.txt (RIGHT CLICK + DOWNLOAD). Read the document containing a description of the variables. Briefly, data correspond to blood pressure readings obtained with automatic (METODE=1) and hand-operated (METODE=2) sphygmomanometers. Both readings were simultaneously obtained. The measurement process was repeated so every subject was measured twice by each device. The reading order is in **NM** variable, where 1 stands for the first reading while 2 indicates the second reading. At the same time, subject information such as age (EDAD), gender (SEXO), weight (PESO) and height (ALTURA) was also recorded. Note that every row in the data file represents one blood pressure reading. So, subject information is repeated four times.

We aim to analyse the weight and height structure of the population that the sample comes from. The body mass index (BMI) will be used as a measure. The BMI is defined as the weight divided by the squared height, where weight must be measured in kg and height in metres. Based on their BMI, subjects are classified in the following categories:

- A) BMI ≤ 20 (low weight)
- B) 20 < BMI ≤ 25 (normal weight)
- C) BMI > 25 (overweight)

It is believed that the proportions of BMI in this population are 15% low weight, 30% normal weight and 55% overweight. What is the null hypothesis in this case? …

What is the appropriate procedure to test this hypothesis? …

If the appropriate procedure is applied to test this hypotesis, what is the probability distribution of the test statistic if the null hypothesis is true? …

What are the applicability conditions of this test? …

In this case, the lowest expected count is (round to 2 decimals at least) …, and the p-value is equal to (round to 4 decimals at least) …

If the probability of type I error (alpha) is set to 5%, the decision is …

So the conclusion is: …

We are also interested in finding out whether the BMI proportions vary depending on gender. What is the null hypothesis now? …

Now what is the appropriate procedure to test this hypothesis? …

Now the lowest expected count is (round to 2 decimals at least) …, and the appropriate p-value is equal to (round to 4 decimals at least) …

If the probability of type I error (alpha) is set to 5%, the decision is …

So the conclusion is: …

Now let us assess whether the two devices work similarly when classifying subjects into low systolic blood pressure (lower than or equal to 140 mmHg) and high systolic blood pressure (greater than 140 mmHg). Use only the first reading of each device (NM=1).

In this case we are facing …

The sample proportion of subjects with low systolic blood pressure when device 1 is used is (round to 3 decimals at least) … while the proportion is … when device 2 is used.

What is the appropriate procedure to test this hypothesis? …

Using this methodology, we obtain a p-value of (round to 3 decimals at least) … . If the probability of type I error (alpha) is set to 5%, the decision is …

So the conclusion is: …

Lab class 7

To answer the questions use the blood pressure data: bloodpressure.txt (RIGHT CLICK + DOWNLOAD). Read the document containing a description of the variables. Briefly, data correspond to blood pressure readings obtained with automatic (METODE=1) and hand-operated (METODE=2) sphygmomanometers. Both readings were obtained simultaneously. The measurement process was repeated so that every subject was measured twice by each device. The reading order is in NM variable, where 1 stands for the first reading while 2 indicates the second reading. At the same time, subject information such as age (EDAD), gender (SEXO), weight (PESO) and height (ALTURA) was also recorded. Note that every row in the data file represents one blood pressure reading. Therefore, subject information is repeated four times.

Only use data from the first method (METODE=1) and the first reading (NM=1).

We want to find out whether there are systolic blood pressure (SIS) differences (on average) between subjects under treatment for hypertension (TNSI_MED=1) and those who do not receive any treatment (TNSI_MED=2) (subjects with no information about treatment, coded as 8, should be excluded from the analysis). It is reasonable to assume that systolic blood pressure is distributed as a Normal model in the population under study.

Here we are facing a …

What is the systolic blood pressure mean in those patients who receive treatment? (round to 2 decimals) …

What is the systolic blood pressure mean in those patients who do not receive treatment? (round to 2 decimals) …

What is the null hypothesis? …

What is the appropriate statistic to test the hypothesis? …

What is the probability distribution of the test statistic if the null hypothesis is true? …

Which of the following is an applicability condition of the test statistic? …

By applying the appropriate procedure, the following p-value is obtained (round to 4 decimals): …

Furthermore, the 95% confidence interval for the difference of means (treatment - no treatment) is defined by (round to 2 decimals) … and …

If the type-I error rate (alpha) is set to 5%, what should be the decision about the hypothesis that is tested? …

Therefore, concerning the research hypothesis, the conclusion is …

Next, we want to assess whether there are differences between the systolic blood pressure means of the two devices. Let us use only data from the first reading (NM=1) and from the two devices. As before, subjects with no information about treatment (coded as 8) should be excluded from the analysis. Now, we are facing …

What is the null hypothesis? …

What is the appropriate statistic to test the hypothesis? …

What is the probability distribution of the test statistic if the null hypothesis is true? …

Which of the following is an applicability condition of the test statistic? …

Application of the appropriate procedure results in the following p-value (round to 4 decimals): …

If the type-I error rate (alpha) is set to 5%, what should be the decision about the hypothesis that is tested? …

Concerning the research hypothesis, the conclusion is …

Lab class 8

To answer the questions, use the blood pressure data: bloodpressure.txt (read the document containing a description of the variables).

Only use data from the first measure (NM=1) and the first method (METODE=1).

Note: Throughout the exercise you may assume normality wherever it is needed.

We aim to assess the level of linear relationship between the systolic blood pressure (SIS) and the body mass index (BMI). Remember that the BMI is defined as $BMI = w/h^2$ where w is the weight in kg and h is the height in metres.

- The Pearson correlation coefficient estimate between SIS and BMI is (round to 3 decimal digits) … with a 95% confidence interval with limits (round to 3 decimal digits) … and …
- Answer the following question using this scale regarding the correlation coefficient estimate:
    - ≤0.2: Almost independent
    - (0.2, 0.4]: Poor
    - (0.4, 0.6]: Moderate
    - (0.6, 0.8]: Good
    - >0.8: Excellent

  With regards to the independence hypothesis between SIS and BMI, using the p-value related to hypothesis testing, we may conclude that …

Let us now model the relationship between SIS and BMI.

- What is the change in SIS if BMI is increased by one unit? (round to 2 decimal digits) … . The corresponding 95% confidence interval has limits (round to 2 decimal digits) … and …
- What is the percentage of variability of SIS that is explained by the BMI? (round to 0 decimal digits) …%.
- Let us suppose that a subject has a BMI of 35.9 kg/m^2. What is its SIS prediction (mean)? (round to 2 decimal digits) …, with a 95% confidence interval with limits (round to 2 decimal digits) … and …
- On the other hand, 95% of subjects who have a BMI of 35.9 kg/m^2 will have a SIS between (round to 2 decimal digits) … and …

Now let us introduce the variables gender (SEXO) and age (EDAD) into the previous model. Remember that SEXO must be codified as a factor (Hint: dades$SEXO=as.factor(dades$SEXO)).

- What is now the percentage of variability of SIS that is explained by the model? (round to 0 decimal digits) %.

- What is now the change in SIS if BMI is increased by one unit and the remaining variables are kept constant? (round to 2 decimal digits) …, with a 95% confidence interval with limits (round to 2 decimal digits) … and …
- Let us suppose now that a woman is 48 years old and has a BMI of 23.9 kg/m^2. What is her SIS prediction (mean)? (round to 2 decimal digits) …

Finally, the aim is to assess whether the BMI slope varies according to the gender, i.e., whether there is a BMI-gender interaction (in the presence of BMI, age and gender).

- What is the p-value associated with this hypothesis testing? (round to 3 decimal digits) … . Thus, …

Lab class 9

Question 1

The ICU study dataset, which is stored in the file ICU.txt (RIGHT CLICK TO DOWNLOAD), consists of a sample of subjects who were part of a much larger study on patients following admission to an adult intensive care unit (ICU). The main goal of this study was to find factors related to the vital status of these patients at hospital discharge. The data used here are adapted from: aplore3: Datasets from Hosmer, Lemeshow and Sturdivant, Applied Logistic Regression (3rd Ed., 2013). R package version 0.9.

The variables included in the ICU study dataset are the following:

id Identification code of the patient

sta Vital status of the patient at hospital discharge (0=Lived, 1=Died)

age Age at ICU admission (years)

gender Gender (Male/Female)

race Ethnicity (White/Black/Other)

ser Service at ICU admission (Medical/Surgical)

can Is cancer part of the present problem? (Yes/No)

crn History of chronic renal failure (Yes/No)

inf Infection probable at ICU admission (Yes/No)

sys Systolic blood pressure at ICU admission (mmHg)

hra Heart rate at ICU admission (beats/min)

type Type of ICU admission (Elective/Emergency)

fra Bone fracture at ICU admission (Yes/No)

loc Level of consciousness at ICU admission (Conscious/Stupor or Coma)

When you perform hypothesis tests, use alpha=0.05.

Given that the aim of the study is to model the probability of being dead at hospital discharge, the outcome variable is …, so we need to use …

The percentage of patients who were discharged from the hospital alive is …% (round to 1 decimal, at least).

We suspect that vital status at hospital discharge is related with age. Fit a model with age as predictor (X) to explore this relationship.

The parameter estimate corresponding to age is … (round to 3 decimals, at least) and the odds ratio (OR) is … (round to 3 decimals, at least). The limits of the 95% confidence interval for the OR are … (lower limit) and … (upper limit) (round to 3 decimals, at least). According to this model, the odds of being dead at discharge … when age increases, and this effect … statistically significant.

Now let us consider a second model with age and service at ICU admission. According to this model, which is the OR for a 1-year increase in age? … (round to 3 decimals, at least). In comparisons of patients who were admitted to he surgical service versus those admitted to the medical service, the OR of being dead at discharge is … (round to 3 decimals, at least).

Question 2

Bone marrow transplants are a standard treatment for acute leukaemia. Recovery following bone marrow transplantation is a complex process and may depend on risk factors known at the time of transplantation. The file bonemarrow.txt (RIGHT CLICK TO DOWNLOAD) contains a subset of the data obtained in a multicentre trial of patients diagnosed with acute leukaemia who underwent bone marrow transplantation (data adapted from: KMsurv: Data sets from Klein and Moeschberger (1997), Klein, Moeschberger and Yan, 2012). Several potential risk factors were measured at the time of transplantation.

The variables included in the dataset are the following:

id: Patient identifier

group: Disease group (ALL=Acute lymphoblastic leukaemia, LR-AML=Low-risk acute myeloid leukaemia, HR-AML=High-risk acute myeloid leukaemia)

time: Time from transplantation to death or last contact (days)

status: Censoring status (1=dead, 0=alive)

plat: Platelet recovery indicator. Did platelets ever return to normal levels? (Y=yes, N=no)

patient.age: Patient age (years)

donor.age: Donor age (years)

patient.gender: Gender of the patient (F=female, M=male)

donor.gender: Gender of the donor (F=female, M=male)

hospital: Hospital where the bone marrow transplantation was done (OSU=The Ohio State University Hospital in Columbus, AH=Alfred Hospital in Melbourne, SVH=St. Vincent Hospital in Sydney, HU=Hahnemann University in Philadelphia)

mtx: Was methotrexate (a chemotherapy agent) used after the transplantation? (Y=yes, N=no)

When you perform hypothesis tests, use alpha=0.05.

- The aim of this study is to model the time to death. Thus, the outcome variable is … and we need to use …
- The follow-up time was between … (minimum) days and … (maximum) days.

  At the end of the follow-up period, … patients were alive.

- Plot the survival curve using the Kaplan-Meier method for the entire sample.

  The probability of surviving 1400 days is … (round to 3 decimals, at least), with a 95% confidence interval equal to … (lower limit) and … (upper limit) (round to 3 decimals, at least).

- Now plot the survival curve according to disease group.

  Median survival time for the ALL group is … days (round to 1 decimal, at least).

For those diagnosed with LR-AML, how many deaths occurred on day 48? … And on day 508? …

The p-value corresponding to the comparison of survival times between disease groups is … (round to 3 decimals, at least), so …