Question 1. I have provided a data set (“exam 01”) in two formats: Stata and Excel. The excel file consists of two worksheets, one with the data, and another describing what the variables are. The data are from a study of advanced prostate cancer. The investigators are studying the “bone scan index” a new way of quantifying the extent of metastatic disease. They want to see whether the index predicts survival (I didn’t include survival data in the data set because we aren’t going to analyze that). What I’d like you to do is create a “table 1” to describe the study cohort. You can either do this by analyzing real data or by explaining what you’d do. As a hypothetical illustration, if I’d sent you data from a study on pain, you could either send me:

Table 1. Characteristics of sample. Data are mode (range) or percentage

|  |  |
| --- | --- |
| Mean baseline pain | 4.5 (1.4 - 45) |
| Mean post-treatment pain | 2.3 (1.9 -234) |
| Women | 52% |

Or

Table 1. Characteristics of sample. Data are mode (range) or percentage

|  |  |
| --- | --- |
| Mean baseline pain | ?? (??-??) |
| Mean post-treatment pain | ?? (??-??) |
| Women | ?% |

Note that both of these tables are rather silly, I am just doing this for illustration. Your table should be in a format suitable for publication in a journal.

Question 2. Some colleagues of yours are working on a project to predict preoperatively which patients will be found to have positive lymph nodes. They send you the following print-out from Stata describing a logistic regression and give you the following: “nodes is coded 1 for positive, 0 for negative; yos is year of surgery; age is in years; CA125 is in units of 10 μg/mL; histol is coded as 0, 1 or 2 whether the tumor was well differentiated, moderately differentiated or poorly differentiated on biopsy”.

logistic nodes yos age CA125 histol

Logistic regression Number of obs = 4701

LR chi2(4) = 401.26

Prob > chi2 = 0.0000

Log likelihood = -2619.9091 Pseudo R2 = 0.0711

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Odds Ratio | Std. Err. | z | P>z | [95% Conf. | Interval] |
|  |  |  |  |  |  |  |
| yos | 1.026621 | .0168909 | 1.60 | 0.110 | .9940435 | 1.060266 |
| age | 1.039007 | .0054137 | 7.34 | 0.000 | 1.02845 | 1.049672 |
| CA125 | 1.04026 | .0051741 | 7.94 | 0.000 | 1.030169 | 1.050451 |
| histol | .4306271 | .0292454 | -12.41 | 0.000 | .3769584 | .4919369 |

Put this information in a format suitable for reporting in a journal article. Briefly state (in no more than 2 – 3 sentences) any criticisms, comments or questions you might you offer your colleagues on their analysis.

Question 3. You read the following in a journal article.

“There was no difference in progression-free survival (p=0.1433) between the group given platinum only therapy compared to those on platinum plus immunotherapy. Immunotherapy therefore does not improve progression-free survival for patients receiving platinum therapy for advanced lung cancer. In the correlative analysis, the serum marker YLK44 (p=0.0132) but not ELCA (p=0.0622), EPLA (p=0.7764), LDH (p=0.6475) nor PPR-3 (p=0.2150) was associated with response, defined as a 50% or greater reduction in tumor size. We conclude that there is a statistically significant association between YLK44 and response rates in second-line therapy. Patients who responded to platinum (mean age 69, 95% C.I. 62 to 76) were on average no younger than non-responders (mean age 73, 95% C.I. 70 to 76).”

What errors of statistical analysis or interpretation can you find in this paragraph? Write out your answers *briefly*, in bullet point form.

Question 4. Look through the PDF of the study on childhood cancer survivors. You don’t have to read the discussion, or the second half of the results if you don’t want, because the questions focus on the earlier part of the paper.

1. Ignoring the first section of the results on patient characteristics, and the associated table 1, what is the null hypothesis associated with the first statistical test reported in the paper? You don’t have worry too much about working out exactly which null hypothesis is the first one, a null hypothesis somewhere near the beginning will do.
2. Look at table 2. Write two *brief* bullet points commenting on the statistical methods used in the table and / or the conclusions that the authors draw from the data presented in the table.