Corso di Dottorato in Oncologia e Immunologia

Course of Biostatistics

Final Assessment AY 2023-24

The file prost.csv was from a randomized trial comparing two treatment groups (placebo and three doses of estrogen) for stage 3 and 4 prostate cancer.

The file prost.csv includes the following variables:

```
status: 0 censor; 1 event
treatment:
      0 - placebo,
      1 - treatment: 0.2 mg or 1.0 mg or 5.0 mg estrogen
stage: cancer stage 3 and 4
dtime: time to event - Months of Follow-up
hg : Serum Hemoglobin (g/100ml)
```

Using the data in file prost.csv address the following questions. Select one exercise of type *SV* and one exercise of type *Stat* below.

I do not mind if you make the assessment in group (of no more than three people).

Your assessment will not be graded, and it is not a pass-fail test.

Please include the names of all participants in your assessment paper.

Papers should be concise, nor exceeding four pages in total length (size 11pt, 1.5 spaced), including plots and tables. Figures should have a maximum height of 8cm.

The assessment paper is due by June 7, 2024.

Exercise type	
Stat1	It is claimed that the patients were randomised, hence there should no difference amongst patients across different groups. Conduct a t-test to assess whether the 'hg' variable exhibits the same mean for patients enrolled in the placebo and treatment group. Prior to conducting the t-test, perform an exploratory analysis to ensure the appropriateness of the test for the given data. After conducting the test, provide a brief discussion of the results considering specifically the randomisation issue mentioned above.
Stat2	It is claimed that the patients were randomised, hence there should no difference amongst patients across different groups. Conduct a Wilcoxon-Mann-Withney test to assess whether the 'hg' variable exhibits the same median for patients enrolled in the placebo or treatment groups. Perform an appropriate explorative analysis based on graphical diagnostics to

further support your conclusions. After conducting the analysis, provide a brief discussion of the results considering specifically the randomisation issue mentioned above.

SV1 Calculate the Kaplan-Meyer survival function by *treatment* for individuals who died from prostatic cancer.

- Represent the estimated function using an appropriate plot.
- Calculate the median survival time by *treatment* and report the results. Interpret the findings.
- Perform a log-rank test and discuss the implications of the test results.

SV2 Calculate the Kaplan-Meyer survival function for individuals who died for prostatic cancer by `stage`.

- Represent the estimated function using an appropriate plot.
- Add the 95% confidence interval to the plot and interpret the result.
- Calculate the median survival time by `stage` and report the finding along with their interpretation.

Bibliographic Reference

D. P. Byar and S. B. Green. The choice of treatment for cancer patients based on covariate information: Application to prostate cancer. *Bulletin Cancer, Paris*, 67:477–488, 1980. _161, 275, 521