HW #3 due on 11/25 (Friday)

1. **Data Analysis:** Triple-combination pharmacotherapy for medically ill smokers: A randomized trial.

Dataset: pharmacoSmoking

Variables

id: patient ID number

ttr: Time in days until relapse

relapse: Indicator of relapse (return to smoking)

grp: Randomly assigned treatment group with levels combination (combination of nicotine patch, nicotine oral inhaler, and bupropion ad libitum) or patchOnly (nicotine patch alone)

age: Age in years at time of randomization

gender: Female or Male

race: black, hispanic, white, or other

employment: ft (full-time), pt (part-time), or other

yearsSmoking: Number of years the patient had been a smoker

levelSmoking: heavy or light

ageGroup2: Age group with levels 21-49 or 50+

ageGroup4: Age group with levels 21-34, 35-49, 50-64, or 65+

priorAttempts: The number of prior attempts to quit smoking

longestNoSmoke: The longest period of time, in days, that the patient has previously gone without smoking

pharmacoSmoking data were collected from a randomized clinical trial (Steinberg, 2009), which was conducted to evaluate whether the combination therapy can prolong time to relapse (secondary endpoint). Use the data to answer the following questions.

1. (10%) Use appropriate summary statistics/graphs to summarize survival data by group assignment (i.e. combination or patch only).
2. (5%) Specify both null and alternative the hypotheses for this study using appropriate notation.
3. (5%) Perform an appropriate test for evaluating the hypothesis in Q2 using SAS and draw a conclusion.
4. (5%) The investigator believed age at randomization could affect the intervention effect. Propose a test that can account for age at randomization and then specify the hypothesis using appropriate notation.
5. (5%) Perform the test you propose in Q4 using SAS and draw a conclusion.
6. (5%) Compare results in Q3 and Q5.
7. **Cox PH Model Specification and Estimation**

The following table gives a small data set of survival times and a covariate *z*:

|  |  |  |
| --- | --- | --- |
| ID | Survival time (in years) | Z |
| 1 | 8 | 3 |
| 2 | 7 | 4 |
| 3 | 9+ | 5 |
| 4 | 10 | 6 |

Where + means a right censored observation. Use the data to answer the following questions.

1. (5%) Define a proportional hazard model using appropriate notation to first describe the relationship between hazard and Z (continuous) and then interpret the regression coefficient for Z, i.e. β.
2. (5%) Write down the partial likelihood of β based on the above 4 observations.
3. (10%) Plot the log partial likelihood for β in [-8, 3], and convince yourself that this function is concave.
4. (10%) Find that maximizes the log partial likelihood function and then calculate the second derivative of the log partial likelihood function at .
5. (5%) Use SAS to fit the above proportional hazards model to the data. How do your results compare with the SAS output?

1. **Data Analysis (KM 1.11)**

Effects of ploidy on the prognosis of patients with tongue cancer.

Variables

Time: time to death

Delta: death indicator

Type: type of tumor (1: Aneuploid; 2: Diploid)   
Use the data to answer the following questions.

1. (10%) Summarize the survival data using appropriate summary statistics/graphs by tumor type.
2. (5%) Perform logrank test to compare risk of death overtime between two tumor types and draw a conclusion.
3. (5%) Fit a Cox PH model with tumor type as the only covariate to the data and draw a conclusion.
4. (5%) Calculate the ratio of median survival time derived from Kaplan-Meier survival curve between two tumor types (Aneuploid vs. Diploid) and then compare it with the hazard ratio (Diploid vs. Aneuploid) derived from Q8. Under what distribution, those two would be identical?
5. (5%) Compare logrank test with score test of the Cox PH model.