

Biometry 711
Summer 2015
Project 1
Due 6pm, Monday 20 July 2015

The data set `CRCData.csv` (posted on the class website) is a subset of SC cancer registry data from colorectal cancer patients with metastatic (advanced stage) disease diagnosed between 1996 and 2006. For this analysis, death within 12 months of diagnosis is considered an event. Significant changes in CRC standard of care occurred in 2000 with recommended 5-Fluorouracil (5FU) chemotherapy, and again in 2004 with the introduction of targeted therapy. For this analysis, the primary objectives are to: 1) examine differences in 12 month survival rates comparing African American (AA) to non-AA subjects; 2) examine differences in 12 month survival rates by tumor location and explore how these vary by race; and 3) examine differences in 12 month survival rates by race and tumor location after adjusting for subject's age, gender, year of diagnosis, tumor histology, and tumor grade.

The variables are defined as follows:

subjectID: Subject ID

surv12mo: 12 month survival, 1 = dead within 12 months of diagnosis, 0 = alive at 12 months

age: Subject's age at diagnosis (years)

race: Subject's race, 1 = non-AA, 2 = AA

sex: Subject's gender, 1 = male, 2 = female

histology: Tumor histology, 0 = adenocarcinoma, 1 = mucinous carcinoma, 2 = signet cell carcinoma

location: Tumor location, 1 = proximal, 2 = distal, 3 = rectum

grade: Tumor grade, 1, 2, 3 or 4

year: Diagnosis year, 1 = 1996 - 1999, 2 = 2000 - 2003, 3 = 2004 - 2006

Conduct an appropriate data analysis to address the stated research objectives. You should write a report in the form of a paper with introduction, materials and methods, results and discussion sections. The paper may not exceed 8 pages in length, must be double spaced, and use 11 or 12 point font. An additional document is provided on the class website providing the rubric detailing how your work will be evaluated.