Jonathan Le and Vamsi Chavali

2/4/22

QBIO-490

Clinical Data Partner Activity

Written Questions

1. A categorical variable is a variable that can take on one of a limited number of fixed values. For example, in a patient survey assessing mood, their mood can be quantified as a categorical variable and can be measured as either "happy", "neutral", or "sad".

   A discrete variable is a quantitative variable that is countable or measurable in unique units. An example of a discrete variable is the number of heads you get after flipping a coin ten times.

   A continuous variable is a quantitative variable that can be measured along a continuous scale. For example, someone's height is a continuous variable, as it can be measured continuously.

2. We have chosen the "stage_event_pathologic_stage" variable, which describes the stage of cancer that each patient is in. This variable has no NAs.

3. This variable is a categorical variable, as it divides patients into different stages such as "Stage I", "Stage II", "Stage III", or "Stage IV". A patient's pathological stage is determined using a physical exam of the patient, imaging tests, biopsies, and lab tests. Pathological stage also utilizes information learned during and after surgery. Because of this, it is more precise than clinical stage.

4. First Article: https://pubmed.ncbi.nlm.nih.gov/32583567/
   Perioperative chemotherapy showed a significant improvement in R0 resection rates and prognosis in advanced gastric cancer patients with higher safety rates. Volume measurement, repeated laparoscopic exploration combined with exfoliative cytology can be used as a supplementary method in the clinical staging and efficacy evaluation of advanced gastric cancer.

   Second Article: https://pubmed.ncbi.nlm.nih.gov/32955607/
   For patients without neoadjuvant chemoradiotherapy (CRT), a tendency to preoperative overstaging was observed. Lymph node size alone did not reliably predict metastasis. According to current guidelines, nearly 1/3 of these patients would have been overtreated by using CRT. On the background of relevant side effects, complications, and the limited benefit of CRT on overall survival, we suggest that primary surgical resection should be recommended more liberally for stages II and III rectal cancer.

5. The second variable we have chosen is the "BMI" variable. This variable is a continuous variable that measures a patient's body mass index based on their weight and height. BMI

helps separate patients into different categories of whether they are a healthy weight for their height. Because the clinic dataset does not provide us the BMI variable, we created the variable by dividing each patient's weight in kilograms by their height in meters squared. We also created another column with categorical variables for each BMI to help with the survival plot. Patients that had BMI >= 25 were classified as "overweight, while patients that had BMI < 25 and > 18.5 were "normal", and those < 18.5 were "underweight".

6. Hypothesis 1: Patients that have a BMI may be more likely to be diagnosed at later stages of cancer.
   Hypothesis 2: Patients at a higher stage likely have lower survival rates.
   Hypothesis 3: Patients with higher BMI have lower survival rates.

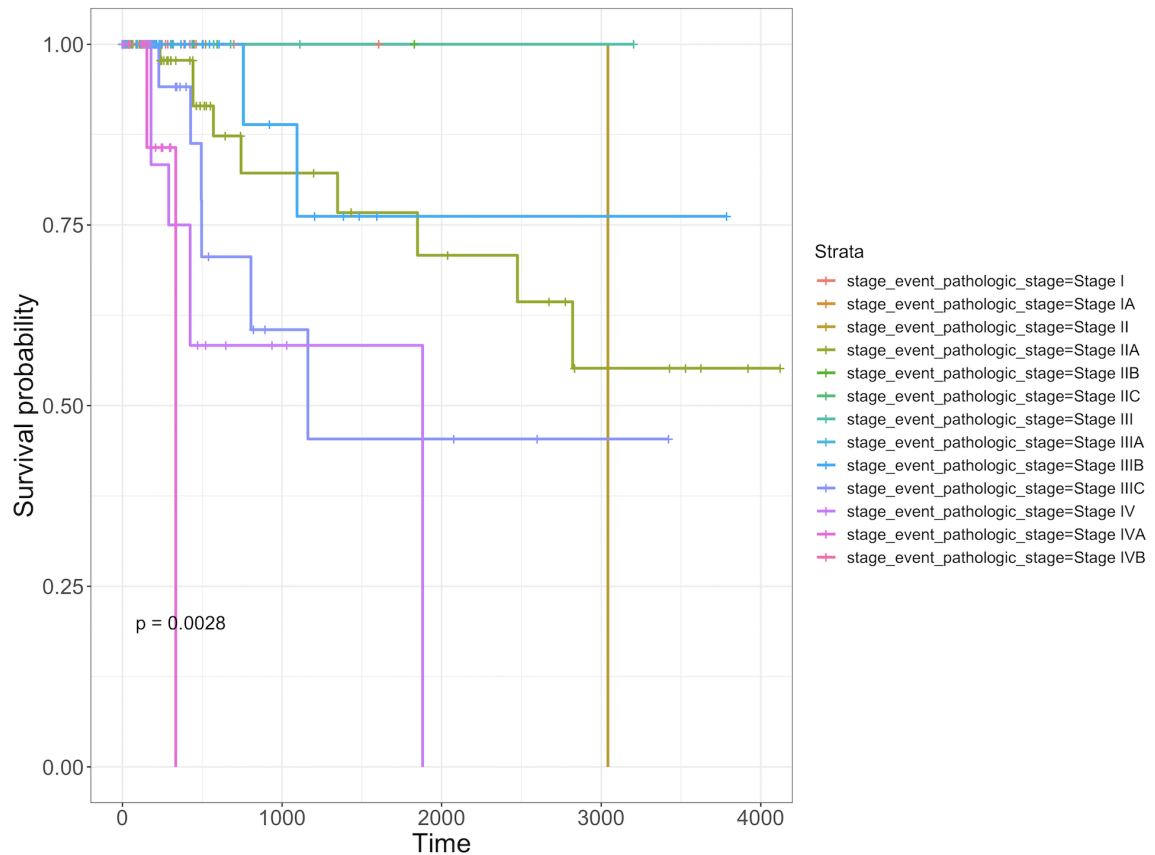7. Figure 1. Boxplot of "BMI" vs "pathological stage"

.
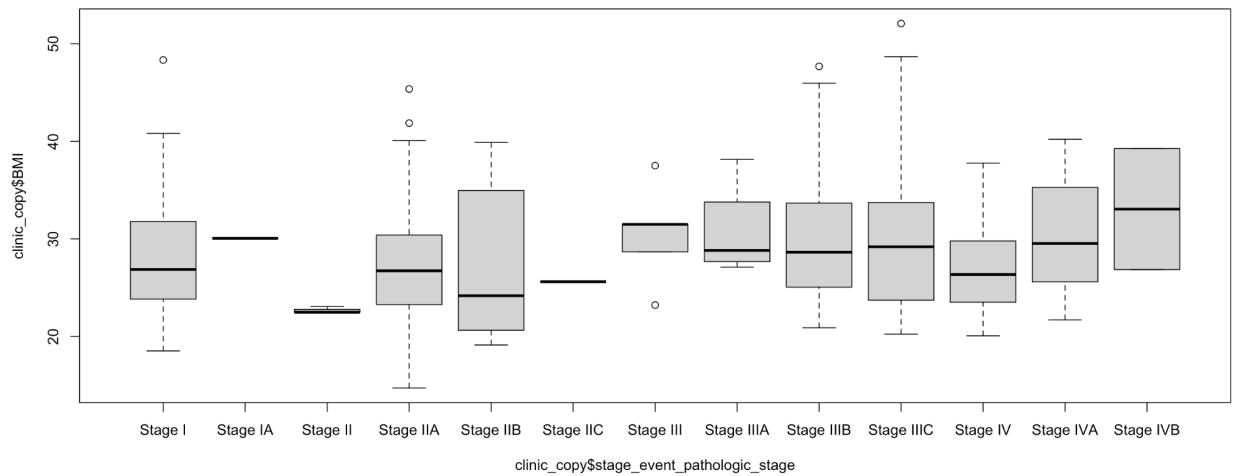
Figure 2. Survival Plot Analysis with Pathological Stage for CRC patients.
It appears that patients with higher stages of CRC tend to have worse survival rates than those at lower stages of disease. For example, patients at Stage IVB have the shortest survival time while those at Stage I or II have higher survival rates at later points in time.
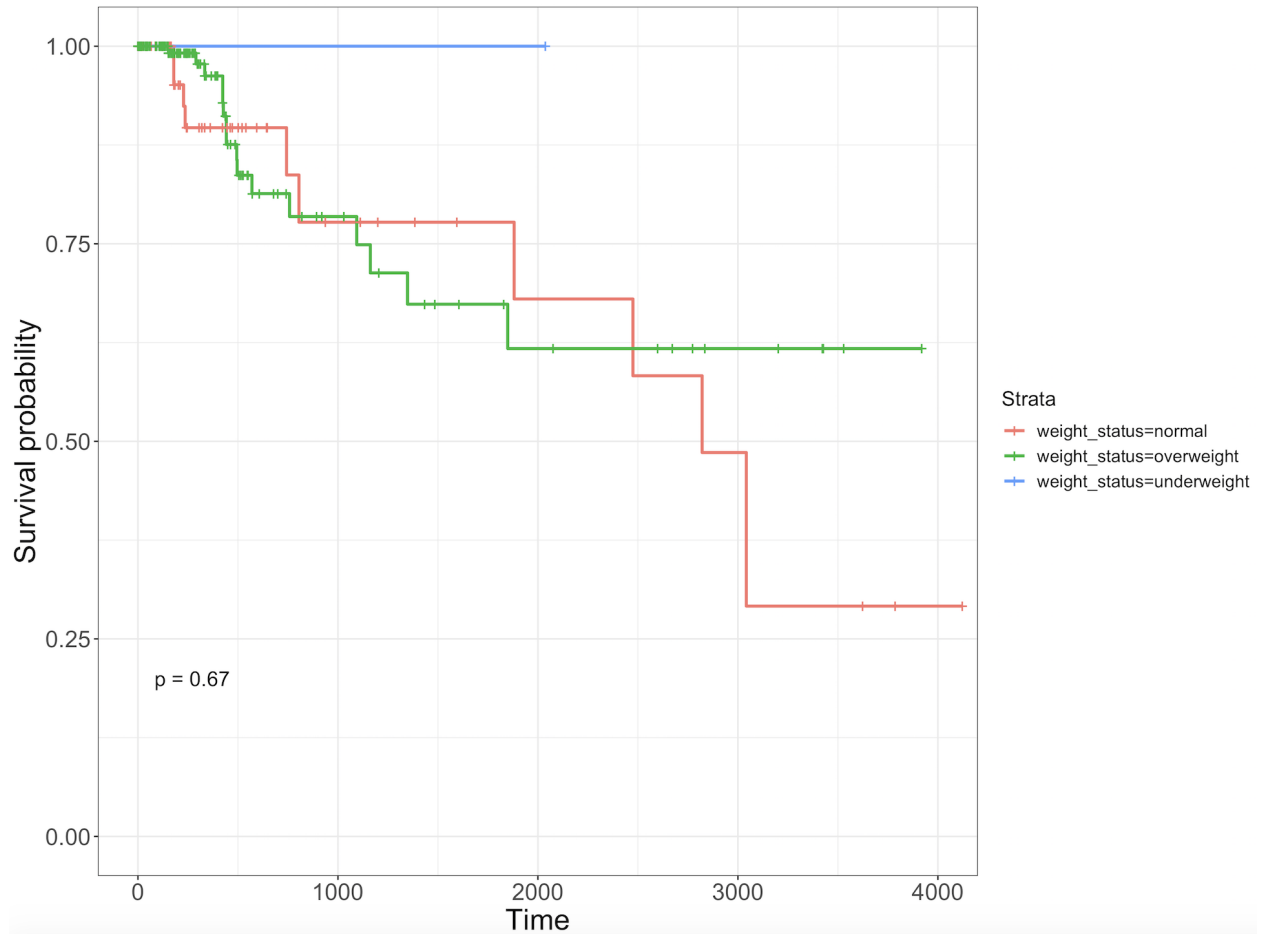
Figure 3. Survival Plot Analysis with BIM for CRC patients.

Interestingly, patients classified as "normal" have the lowest survival rate over time, while those that are underweight have the best survival rate. However, this analysis may be limited by sample size.