STA248 A2 Question 2

Eric Zhu

April 2020

Part a

i

The null hypothesis is that CRP, which represents the concentration of C-reactive protein in the patients' blood, follows a $Weibull(2, 6)$ distribution.

ii

We will provide 10 equally likely intervals for CRP concentration under the null hypothesis, i.e, CRP concentration follows a $Weibull(2, 6)$ distribution. Thus, we will fix the probability of each interval to be 0.1. By hand we would first fix an arbitrary real number, x, such that x is a possible CRP concentration value from the data, then:

$$\int_a^b f(x; \lambda, k)dx = 0.1, \text{ where } a, b \in \mathbb{R}$$
$$\iff$$
$$\int_a^b \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1}dx = 0.1$$

Since we know the parameters of the Weibull distribution, i.e, $\lambda = 2$ and $k = 6$, then we have:

$$\int_a^b \frac{6}{2}(\frac{x}{2})^5 dx = 0.1$$

Since we know that the above definite integral (CDF) can calculate a left tailed probability given one end point, we actually know that b will be equal to x and a will be equal to 0 as part of the first interval. So to better express the various intervals, we can define $k \in \mathbb{N}$ where k is the interval number. And so for the first interval we'd have $b = x_1$ and $a = 0$, and for the second interval we'd have $b = x_2$ and $a = x_1$, and for the third interval we'd have $b = x_3$ and $a = x_2$ and so on until we have our 10 intervals. Each time we solve for the next interval, we can use the previous upper bound as the new lower bound and set the upper bound of the integral to be x. We can represent this generalized definite integral for the k^{th} interval as follows:

$$\int_{x_{k-1}}^{x_k} \frac{6}{2}(\frac{x_k}{2})^5 dx_k = 0.1$$

We will now use R to solve for these 10 intervals through the qweibull function. And we get the following intervals:

(0, 1.374494), (1.374494, 1.557617), (1.557617, 1.684259), (1.684259, 1.788170), (1.788170, 1.881486), (1.881486, 1.971071), (1.971071, 2.062843), (2.062843, 2.165089), (2.165089, 2.298261), (2.298261, Inf)

iii

Using the 10 intervals, we will generate a table of the number of patients that fall within each interval of CRP concentration:

```
##                    Interval 1 Interval 2 Interval 3 Interval 4 Interval 5
## Number of patients        602        179        121        121        103
##                    Interval 6 Interval 7 Interval 8 Interval 9 Interval 10
## Number of patients        109        100        120        168        6902
```

iv

Done by hand.

Part b

i

Will be evident in the R script PDF

ii

```
##                  Stage 1 Cancer Stage 2 Cancer Stage 3 Cancer Stage 4 Cancer
## Underweight BMI               1              2              0              0
## Normal BMI                  750           1063            534            259
## Overweight BMI              841           1130            557            290
## Obese BMI                   966           1214            614            304
```

iii

Done by hand.

iv

Done by hand.