**Part 1**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Immunoperoxide negative** | | | |  |  | |  | |
| **Months** | **nk** | | **dk** | **ck** | **(nk-dk)/nk** | | **S(t)** | |
| 19 | 36 | | 1 | 0 | 0.972222222 | | 0.972222 | |
| 25 | 35 | | 1 | 0 | 0.971428571 | | 0.944444 | |
| 30 | 34 | | 1 | 0 | 0.970588235 | | 0.916667 | |
| 34 | 33 | | 1 | 0 | 0.96969697 | | 0.888889 | |
| 37 | 32 | | 1 | 0 | 0.96875 | | 0.861111 | |
| 46 | 31 | | 1 | 0 | 0.967741935 | | 0.833333 | |
| 47 | 30 | | 1 | 0 | 0.966666667 | | 0.805556 | |
| 51 | 29 | | 1 | 0 | 0.965517241 | | 0.777778 | |
| 56 | 28 | | 1 | 0 | 0.964285714 | | 0.75 | |
| 57 | 27 | | 1 | 0 | 0.962962963 | | 0.722222 | |
| 61 | 26 | | 1 | 0 | 0.961538462 | | 0.694444 | |
| 66 | 25 | | 1 | 0 | 0.96 | | 0.666667 | |
| 67 | 24 | | 1 | 0 | 0.958333333 | | 0.638889 | |
| 74 | 23 | | 1 | 0 | 0.956521739 | | 0.611111 | |
| 78 | 22 | | 1 | 0 | 0.954545455 | | 0.583333 | |
| 86 | 21 | | 1 | 0 | 0.952380952 | | 0.555556 | |
| 122 | 20 | | 0 | 1 | 1 | | 0.555556 | |
| 123 | 19 | | 0 | 1 | 1 | | 0.555556 | |
| 130 | 18 | | 0 | 2 | 1 | | 0.555556 | |
| 133 | 16 | | 0 | 1 | 1 | | 0.555556 | |
| 134 | 15 | | 0 | 1 | 1 | | 0.555556 | |
| 136 | 14 | | 0 | 1 | 1 | | 0.555556 | |
| 141 | 13 | | 0 | 1 | 1 | | 0.555556 | |
| 143 | 12 | | 0 | 1 | 1 | | 0.555556 | |
| 148 | 11 | | 0 | 1 | 1 | | 0.555556 | |
| 151 | 10 | | 0 | 1 | 1 | | 0.555556 | |
| 152 | 9 | | 0 | 1 | 1 | | 0.555556 | |
| 153 | 8 | | 0 | 1 | 1 | | 0.555556 | |
| 154 | 7 | | 0 | 1 | 1 | | 0.555556 | |
| 156 | 6 | | 0 | 1 | 1 | | 0.555556 | |
| 162 | 5 | | 0 | 1 | 1 | | 0.555556 | |
| 164 | 4 | | 0 | 1 | 1 | | 0.555556 | |
| 165 | 3 | | 0 | 1 | 1 | | 0.555556 | |
| 182 | 2 | | 0 | 1 | 1 | | 0.555556 | |
| 189 | 1 | | 0 | 1 | 1 | | 0.555556 | |
| **Immunoperoxide Positive** | | | |  |  |  | |
|  | | | |  |  |  | |
| **Months** | | **nk** | **dk** | **ck** | **(nk-dk)/nk** | **S(t)** | |
| 22 | | 9 | 1 | 0 | 0.8888889 | 0.888889 | |
| 23 | | 8 | 1 | 0 | 0.875 | 0.777778 | |
| 38 | | 7 | 1 | 0 | 0.8571429 | 0.666667 | |
| 42 | | 6 | 1 | 0 | 0.8333333 | 0.555556 | |
| 73 | | 5 | 1 | 0 | 0.8 | 0.444444 | |
| 77 | | 4 | 1 | 0 | 0.75 | 0.333333 | |
| 89 | | 3 | 1 | 0 | 0.6666667 | 0.222222 | |
| 115 | | 2 | 1 | 0 | 0.5 | 0.111111 | |
| 144 | | 1 | 0 | 1 | 1 | 0.111111 | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Log Rank |  |  |  |  |  |  |  |  |
| Time | n1 | d1 | n2 | d2 | n= n1+n2 | d = d1+d2 | d1-n1\*d/n | [n1\*n2/n^2][(n-d)/(n-1)]\*d |
| 19 | 36 | 1 | 9 | 0 | 45 | 1 | 0.2 | 0.16 |
| 22 | 35 | 0 | 9 | 1 | 44 | 1 | -0.795455 | 0.162706612 |
| 23 | 35 | 0 | 8 | 1 | 43 | 1 | -0.813953 | 0.151433207 |
| 25 | 35 | 1 | 7 | 0 | 42 | 1 | 0.1666667 | 0.138888889 |
| 30 | 34 | 1 | 7 | 0 | 41 | 1 | 0.1707317 | 0.141582391 |
| 34 | 33 | 1 | 7 | 0 | 40 | 1 | 0.175 | 0.144375 |
| 37 | 32 | 1 | 7 | 0 | 39 | 1 | 0.1794872 | 0.147271532 |
| 38 | 31 | 0 | 7 | 1 | 38 | 1 | -0.815789 | 0.150277008 |
| 42 | 31 | 0 | 6 | 1 | 37 | 1 | -0.837838 | 0.135865595 |
| 46 | 31 | 1 | 5 | 0 | 36 | 1 | 0.1388889 | 0.119598765 |
| 47 | 30 | 1 | 5 | 0 | 35 | 1 | 0.1428571 | 0.12244898 |
| 51 | 29 | 1 | 5 | 0 | 34 | 1 | 0.1470588 | 0.125432526 |
| 56 | 28 | 1 | 5 | 0 | 33 | 1 | 0.1515152 | 0.12855831 |
| 57 | 27 | 1 | 5 | 0 | 32 | 1 | 0.15625 | 0.131835938 |
| 61 | 26 | 1 | 5 | 0 | 31 | 1 | 0.1612903 | 0.135275754 |
| 66 | 25 | 1 | 5 | 0 | 30 | 1 | 0.1666667 | 0.138888889 |
| 67 | 24 | 1 | 5 | 0 | 29 | 1 | 0.1724138 | 0.142687277 |
| 73 | 23 | 0 | 5 | 1 | 28 | 1 | -0.821429 | 0.146683673 |
| 74 | 23 | 1 | 4 | 0 | 27 | 1 | 0.1481481 | 0.126200274 |
| 77 | 22 | 0 | 4 | 1 | 26 | 1 | -0.846154 | 0.130177515 |
| 78 | 22 | 1 | 3 | 0 | 25 | 1 | 0.12 | 0.1056 |
| 86 | 21 | 1 | 3 | 0 | 24 | 1 | 0.125 | 0.109375 |
| 89 | 20 | 0 | 3 | 1 | 23 | 1 | -0.869565 | 0.11342155 |
| 115 | 20 | 0 | 2 | 1 | 22 | 1 | -0.909091 | 0.082644628 |
|  |  |  |  |  |  | SUM | -4.187299 | 3.191229315 |
|  |  |  |  |  |  |  | **Z** | **Var(z)** |
|  |  |  |  |  | Test Statistic | | **5.4942702** |  |

With the test statistic of 5.49, which is greater than 3.84, we reject the null hypothesis. Breast cancer survival differs between immunoperoxidase positive and negative patients.

**Part II**

1. Done
2. 
   1. **H(t|x)** = h0(t) exp(β1gender+ β2bmi+ β3g\_b) = **h0(t) exp[-0.803(gender)+ -.117(bmi)+ 0.041(g\_b)]**
      1. e(β1+β3(20)) = e(-.8033496+.041009(20)) = e(0.0168304) = **1.017**

At any point in time during the study period, the instantaneous risk of death among females with a BMI of 20 is 1.017 times that of men with a BMI of 20.

* + 1. e(β1+β3(30)) = e(-.8033496+.041009(30)) = e(0.4269204) = **1.533**

At any point in time during the study period, the instantaneous risk of death among females with a BMI of 30 is 1.533 times that of men with a BMI of 30.

* + 1. e(β2(10)) = e(-.1170119(10)) = e(-1.170119) = **0.310**

At any point in time during the study period, the instantaneous risk of death among males that have a 10 kg/m2 increase in BMI is .310 times that of men without a 10 kg/m2 increase in BMI.

* + 1. e(β2(10)+β3(10)) = e(-.1170119(10)+0.0410099(10)) = e(-.76002) = **0.468**

At any point in time during the study period, the instantaneous risk of death among females that have a 10 kg/m2 increase in BMI is .468 times that of women without a 10 kg/m2 increase in BMI.

* 1. Bmi is most likely a modifier of Gender, since there was a decrease in Risk by more than 50% in both cases of male and female with an increase in BMI. Whereas, gender is most likely not a modifier of BMI since there was a case where a change in gender only resulted in a 1.7% increase in risk of death for women than men.

1. 

Full-model log likelihood = -1201.3811

Reduced model log likelihood = -1202.3064

LRtest = 2\*-1201.3811 – 2\*-1202.3064 = **1.8506 < 3.84 p-value = 0.1737**

**Due to a failure to reject the null hypothesis, Gender does not significantly modify the effect of BMI**

* 1. Gender



The plot suggests that proportional hazards is satisfied for Gender. The predictor doesn’t seem to be modified by time, no convergence, divergence or cross.

* 1. CVD



The plot suggests that proportional hazards is not satisfied by CVD. There seems to be convergence in the middle and then divergence at the end.

* 1. MIORD



This plot suggests that proportional hazards is satisfied by MIORD, there doesn’t seem to be any modification of MIORD by time. Both lines seem to follow their respective curves without convergence, divergence. There is the cross at the end, but overall, the chart makes it seem that they aren’t changed by time.

* 1. MITYPE

This plot suggests that proportional hazards is not satisfied by MITYPE, there seems to be a divergence in the plot until the end, where there seems to be a convergence if we draw a line from 200 to the last points.

* 1. Variable 2 - **age**



Log likelihood test not significant **p-value = 0.6736**. This indicates that proportional hazards is not violated. The predictor, age, is **not modified by time**.

* 1. Variable 3 – **gender**

Log likelihood test not significant **p-value = 0.4496**. This indicates that proportional hazards is not violated. The predictor, gender, is **not modified by time**.

* 1. Variable 4 – **hr**

Log likelihood test not significant **p-value = 0.2650**. This indicates that proportional hazards is not violated. The predictor, heart rate, is **not modified by time**.

* 1. Variable 5 – **bmi**

Log likelihood test not significant **p-value = 0.2211**. This indicates that proportional hazards is not violated. The predictor, bmi, is **not modified by time**.

* 1. Variable 6 – **cvd**

Log likelihood test not significant **p-value = 0.4351**. This indicates that proportional hazards is not violated. The predictor, cvd, is **not modified by time**.

* 1. Variable 7 – **miord**



Log likelihood test not significant **p-value = 0.1375**. This indicates that proportional hazards is not violated. The predictor, miord, is **not modified by time**.

* 1. Variable 8 – **mitype**

Log likelihood test is significant **p-value = 0.0264**. This indicates that proportional hazards is violated. The predictor, mitype, **is modified by time**.

1. Only, g, or MITYPE from Q4 doesn’t satisfy the PH assumption.

HR @ 1 day = exp(β1mitype + β2log(1)) = exp(.1285006 - .1690017(log(1))) = 1.137

HR @ 365 day = exp(β1mitype + β2log(365)) = exp(.1285006 - .1690017(log(365))) = .420



Based off the HR calculated at day 1 and day 365 here, we see a decrease in HR, indicating a converging trend. In the graph in 3, if we draw a line for trends and include the last points, we can see that the graph also displays a converging trend closer to the end.

1. 

**H(t|x)** = h0(t) exp(β1age+ β2gender+ β3hr + β4bmi + β5cvd + β6miord + β7mitype – β8mitype\*log(n)) =

**h0(t) exp[.0585112(age)-.1683112(gender)+.0119293(hr)-.0427541(bmi)+.0052424(cvd)+.0628722(miord)+0.7680942(mitype)-0.198949{mitype x log(t)}]**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Step** | **Predictor Removed** | **LR p-value** | **Log likelihood** | **AIC** | **BIC\*** |
| 0 | - | - | -1139.2833 | 2294.567 | 2321.532 |
| 1 | CVD | 0.9762 | -1139.2838 | 2292.568 | 2316.162 |
| 2 | MIORD | .6634 | -1139.3785 | 2290.757 | 2310.981 |
| 3 | Gender | .2287 | -1140.1031 | 2290.206 | 2307.059 |
| 4 |  |  |  |  |  |
| 5 |  |  |  |  |  |
| 6 |  |  |  |  |  |
| 7 |  |  |  |  |  |

\*BIC used log(d) instead of log(n) => d = 215 for Volinsky/Raftery

1. Out of all the models above, the last one, with gender removed in addition to the other predictors CVD and MIORD, is the best in terms of AIC and BIC (lowest numbers). The 3rd step backwards selection has the lowest AIC and BIC. (Model that contains, age, hr, bmi, mitype, and the mitype (time interaction)
   1. After performing the lrtest for mitype, we obtained a p-value of 0.0096. This means that the interaction is still necessary to be in the model.
   2. Via the global test of PH for the other terms, we obtained a p-value of 0.4845, this means that the PH was not violated.



* 1. Due to B not being significant, there was no need to perform backward selection elimination to select for the necessary interactions with log(t), and therefore only MITYPE has the interaction.

1. Mathematical Model

**H(t|x)** = h0(t) exp(β1age+ β2HR+ β3BMI + β4MItype + β5MItype\*log(n)) =

**h0(t) exp[.0572666(age)+..0117882(HR)-.0419281(BMI) +.7486013(mitype) -0.19633(mitypex log(n))**

HR @ 1 day: exp(β1mitype + β2log(1)) = exp(.7486013 - .19633(log(1)) = **2.114**

HR @ 1 year (365 days): exp(β1mitype + β2log(365)) = exp(.7486013 - .19633(log(365)) = **0.664**

HR @ 5 years (1825 days): exp(β1mitype + β2log(1825)) = exp(.7486013 - .19633(log(1825)) = **0.484**

With 2.114 as the HR at day one and 0.664 at day 365 and -1.973 for day 1825, we see a decreasing change in HR overtime for MITYPE.

Controlling for heart rate, bmi, and MITYPE, there is a **5.89% increase in instantaneous risk due to a 1 year increase in age**. Controlling for age, bmi, and MITYPE, there an **1.19 % increase in instantaneous risk due to a 1 beat/min increase in heart rate**. Controlling for age, heart rate, and MITYPE, there is a **4.11% decrease in instantaneous risk due to a 1 kg/m2 increase in BMI.** Controlling for age, heart rate, and bmi, there seems to be a 111% increase in instantaneous risk comparing Q-wave to non-Q-wave MITYPE. However, it must be noted that this relationship is not constant over time, in fact, it decreases over time (days). All of the above variables, with the exception of MITYPE have a constant effect over the time of the study.