6166

BIOST 518

HW06

1a. I modeled the relationship between serum LDL and time death using a proportional hazards regression with robust standard error. The outcome of interest was time to death or censoring. Serum LDL was modelled as a continuous linear term and as a stepwise function with breakpoints at 70, 100, 130, and 160 mg/dL. The slopes for the stepwise function were tested for their difference from zero, two-sided p=0.361, and provided no evidence for nonlinearity.

1b. I modeled the relationship between serum LDL and time death using a proportional hazards regression with robust standard error. The outcome of interest was time to death or censoring. Serum LDL was modelled as a continuous linear term and as a quadratic term. The regression parameter for the quadratic term was tested for a difference from zero, two-sided p=0.055, and provided no evidence for nonlinearity.

1c. I modeled the relationship between serum LDL and time death using a proportional hazards regression with robust standard error. The outcome of interest was time to death or censoring. Serum LDL was modelled as a continuous linear term and as a cubic term. The regression parameter for the cubic term was tested for a difference from zero, two-sided p=0.112, and provided no evidence for nonlinearity.

1d. I modeled the relationship between serum LDL and time death using a proportional hazards regression with robust standard error. The outcome of interest was time to death or censoring. Serum LDL was modelled as a continuous linear term and as a linear spline with knots at 70, 100, 130, and 160 mg/dL. The slopes for the linear spline term was tested simultaneously for a difference from zero, two-sided p=0.119, and provided no evidence for nonlinearity.

1e. I modeled the relationship between serum LDL and time death using a proportional hazards regression with robust standard error. The outcome of interest was time to death or censoring. Serum LDL was modelled as a continuous linear term and as a continuous, natural log-transformed term. The regression parameter for the log transformed term was tested for a difference from zero, two-sided p=0.004, and provided evidence for nonlinearity of the untransformed data.

1f. The models all have similar predicted values at the high end of the serum LDL range, but very different predictions at the lower range. The stepwise function suggests that individuals at the around 50 mg/dl have a much higher relative hazard than the rest of the cohort. This graph clearly suggest that unless the serum LDL is log transformed, a linear model is not the best as when the model allows for flexibility in linearity, the lines are curvilinear.



2a. The regression parameters are now the slopes in each region, not overall. Linear splines allow some flexibility in non-linearity, however they must connect and the knots so information is borrowed across the whole model, not just within the region set by the slope parameters.

2b. There is evidence that there is a u-shaped curve. The parameters for the extremes of the linear spline are opposite signs.

3a. White is the reference group for the dummy variable race. Among white subjects, .461 increase in instanteous risk of death was estimated per 2.718-fold increase in serum LDL levels. We estimate the increase in instanteous risk of death at 0 mg/dl LDL for black subjects was .154 higher than whites, for Asians 305 higher, and for other 3.33X10^8 higher. These parameters are meaningless because no one has no ldl.

3b. Race does modify the serum LDL and time to death association. THe simultaneous test of the race and death interaction parameters for difference from zero gives a statistically significant result, two-sided p<0.001.

3c. There is an association between serum LDL and time to death. When I test the log-transformed serum LDL parameter the two-sided p=0.0004.

3d. There is an association between race and time to death. When I simultaneously test the race parameters for a difference from zero the two-sided p=0.0231.

3e. There is a difference in the distribution of time to death between white and black subjects. When we test the parameter corresponding the black subjects, with white as the reference, the two-sided p<0.001.

4a. There is no evidence of sex bias in mean salary given when holding the year constant, p=0.456. However, the interaction between sex and the year is statistically significant, p=0.0009. When accounting for the differential hiring of men in women in different years, we do see a difference in mean salary given.

4b. There is no evidence of sex bias in mean geometric mean salary given when holding the year constant, p=0.382. However, when simultaneously testing sex and the interaction between sex and the year the parameters are statistically significant, p=0.003. When accounting for the differential hiring of men in women in different years, we do see a difference in mean geometric mean salary given.

4c. Mean salary is easier for other to interpret, but the geometric mean is more precise and more accurately represent how salaries are considered on a multiplicative scale.

4d. If the data had been treated as independent the inference would have been anti-conservative with the p-values too low and the 95% CI to small.