Homework 07

Spencer Pease

11/27/2019

(Q1) Serum Creatinine Level and 5-Year All-Cause Mortality

(Q1.a)

The model fit to determine the nature of the association between serum creatinine level (*crt*) and 5-year all-cause mortality binary survival status (*surv*) is expressed as:

$$E[Log(crt_i)|surv_i] = \beta_0 + \beta_1 \cdot surv_i$$

With β_0 = 0.138 and β_1 = -0.132.

(Q1.b,c)

In this model, the intercept is interpreted as the log of the geometric mean of the response variable (*crt*) when the predictor (*surv*) is 0. This has a relevant scientific interpretation, since the exponentiated intercept is the geometric mean of creatinine level in the population of subjects who did not survive at least 5 years.

The slope of this model is interpreted as the the log of the ratio of the geometric means of the response between two groups differing by one unit of our predictor. Again, this has a relevant scientific interpretation as the exponentiated slope is the ratio of the geometric mean creatinine levels between the population that survived at least 5 years and the population that didn't. Using this and the intercept, we can find the geometric mean of the subset of individuals who survived at least 5 years.

(Q1.d,e)

Using the regression model estimates:

- $GM(crt)$ for a population surviving $\geq$ 5 years = e^{β_0} = 1.006
- $GM(crt)$ for a population surviving $<$ 5 years = $e^{(\beta_0+\beta_1)}$ = 1.148

Calculating the geometric mean directly:

- $GM(crt)$ for a population surviving $\geq$ 5 years = 1.006
- $GM(crt)$ for a population surviving $<$ 5 years = 1.148

Using either the model estimates or direct calculation of the geometric mean produce the same results, showing that the geometric linear model can be used to determine the population geometric mean for populations set by the predictor value.

(Q1.f)

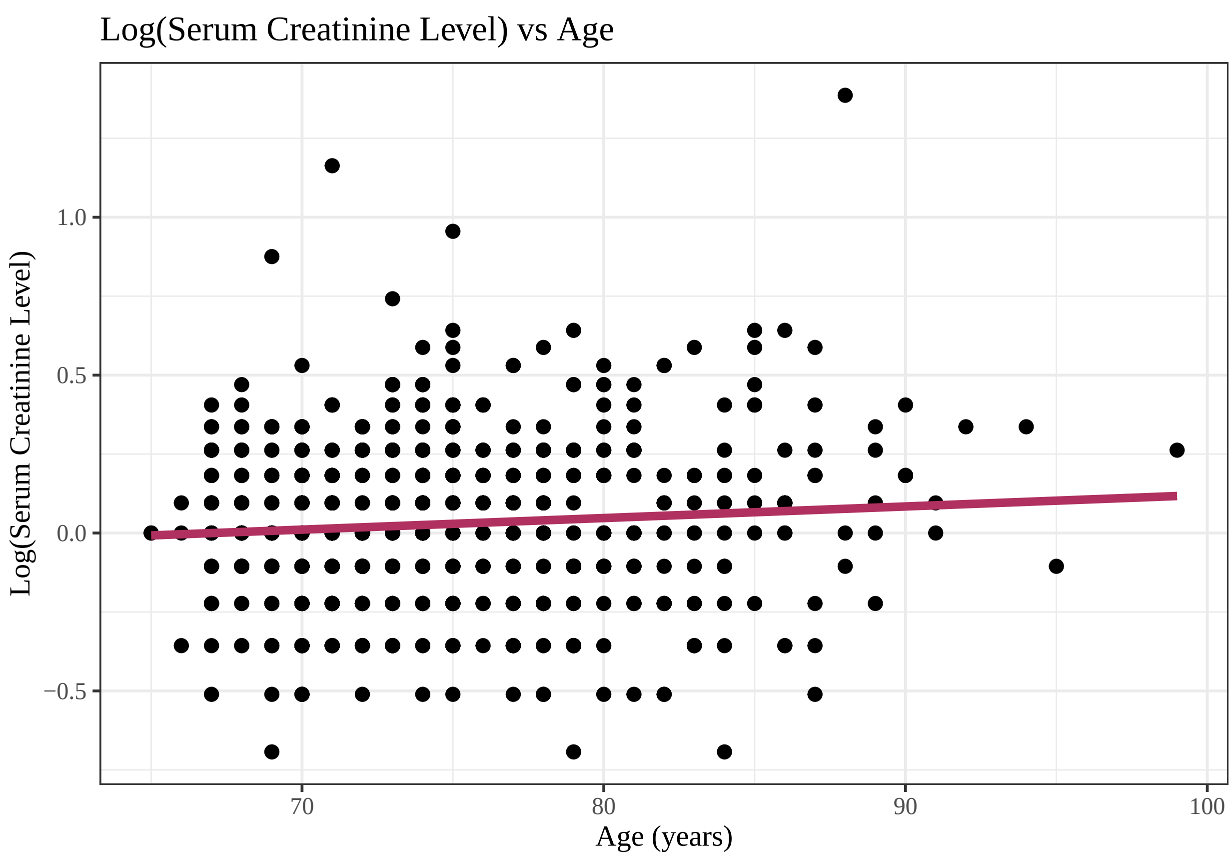
Table 1: Statistical inference for ratio of geometric means between survivorship groups

Estimate	2.5%	97.5%	P-val
0.876	0.824	0.931	2.36e-5

From a linear regression analysis of the log-transformed serum creatinine level and binary 5-year all-cause mortality using Huber-White estimates of the standard error, we find that the population of individuals who survive at least five years form their observation date have a geometric mean creatinine level 0.876 times the value of the the geometric mean of the population that survives less than five years. A 95% confidence interval suggests this ratio is not unusual if the true ratio of geometric means is between 0.824 and 0.931. Because this interval does not include 1, and the two-sided P-value is less than .05, we reject the null hypothesis that there is no linear trend in serum creatinine level across survival status groups.

(Q2) Serum Creatinine Level and Age

(Q2.a)



(Q2.b)

The model fit to determine the nature of the association between serum creatinine level (crt) and age (age) is expressed as:

$$E[Log(crt_i)|age_i] = \beta_0 + \beta_1 \cdot age_i$$

With β_0 = -0.246 and β_1 = 0.004.

(Q2.c,d)

The intercept of this model (β_0) represents the log of the geometric mean of our response variable (*crt*) when the predictor (*age*) is 0. This has no scientific interpretation, as a newborn child is far outside of the range of ages for which we have observations, meaning the model doesn't have any power that far from our observations.

The slope of this model (β_1) represents the log of the ratio of the geometric means of the response between two groups separated by one unit of the predictor (1 year). This has a meaningful scientific interpretation, as the exponentiated slope is the ratio of the geometric means between two age groups one year apart. Within the range of our observed data, this can be used to determine the geometric mean for any age group.

(Q2.e,f)

Table 2: Estimated mean serum creatinine levels for both geometric and linear models

Age	Est. geometric mean crt	Est. Arithmetic mean crt
72	1.018	1.050
85	1.068	1.122
95	1.108	1.178

Comparing predictions between our log-transformed linear model and a standard linear model using the same predictor/response variables, we see that while predictions across the models are generally, similar, there is a definite trend of the geometric mean increasing at a slower rate than the arithmetic mean as age increases. This can be explained by the fact that the geometric mean is less biased towards outliers than the arithmetic mean.

(Q2.g)

Table 3: Statistical inference for ratio of geometric means between one year age groups

Estimate	2.5%	97.5%	P-val
1.004	1	1.007	0.054

From a linear regression analysis of the log-transformed serum creatinine level and patient age using Huber-White estimates of the standard error, we estimate that for every one year increase in age an age group's geometric mean creatinine level is 1.004 time higher. A 95% confidence interval suggests this ratio is not unusual if the true ratioo of geometric means is between 1 and 1.007. Becuase this interval includes 1, and the P-value is greater than .05 we do not reject the null hypothesis that there is no linear association between serum creatinine level and age.

(Q2.h)

Table 4: Estimate of the percent change in geometric means between two age groups 10 years apart

Estimate	95% CI (lower)	95% CI (upper)
3.67	-0.06	7.42