

R Assignment One

Katherine Wolf

January 24, 2020

Questions

1. With mathematical notation, write out the statistical model you used to simulate the data for the power analysis given the assumptions listed above. Include: 1) The expression for the log-hazard function in terms of the covariates and parameters, and 2) specify the distribution of failure times given the hazard (general is fine). Clearly define all terms and parameters in the model. (15 points)

The expression for the log-hazard function is

$\log(h(t|x, \beta)) = \log[h_0(t)] + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$, where

$\log(h(t|x, \beta))$ is the natural logarithm of the instantaneous rate (hazard) of the event at time t for a participant who has survived up to that time t , given the covariates in x and coefficients in β ;

$\log[h_0(t)]$ is the baseline instantaneous rate (hazard) of the event at time t for a participant who had neither exposure, i.e., $x_1 = 0$ and $x_2 = 0$, and survived up to time t ;

x is a vector of covariates containing

- x_1 , an indicator variable for the first binary exposure, with $x_1 = 1$ indicating that the participant was exposed and $x_1 = 0$ indicating that the participant was not exposed; and
- x_2 , an indicator variable for the second binary exposure, with $x_2 = 1$ indicating that the participant was exposed and $x_2 = 0$ indicating that the participant was not exposed;

β is a vector of coefficients including $+\beta_1$ is the log hazard ratio comparing the hazard of the event for a participant who had the first exposure, i.e., $x_1 = 1$, to the hazard of the event for a participant who did not, i.e., $x_1 = 0$, holding the values of all other variables constant; $+\beta_2$ is the log hazard ratio comparing the hazard of the event for a participant who had the second exposure, i.e., $x_2 = 1$, to the hazard of the event for a participant who did not, i.e., $x_2 = 0$, holding the values of all other variables constant; and $+\beta_3$ is the log of the ratio of hazard ratios, where - the hazard ratio in the numerator compares the hazard of the event for a participant who had both exposures in its numerator, i.e., $x_1 = 1$ and $x_2 = 1$, to the hazard of the event for a participant who did not have one but still had the other in its denominator, i.e., either $x_1 = 1$ and $x_2 = 0$ or $x_1 = 0$ and $x_2 = 1$, and - the hazard ratio in the denominator compares the hazard of the event for a participant who only had the other exposure not held constant above, i.e., $x_1 = 0$ and $x_2 = 1$ if both numerator and denominator above had the first exposure $x_1 = 1$, or $x_1 = 1$ and $x_2 = 0$ if both numerator and denominator above had the second exposure $x_2 = 1$, to the hazard of the event for a participant who had neither exposure in the denominator, i.e., $x_1 = 0$ and $x_2 = 0$.

The distribution of failure times is here assumed to be exponential, i.e., $f(t) = \lambda e^{-\lambda t}$, where $\lambda = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2}$.

2. In terms of parameter(s) from the model outlined above, specify the null (H_0) and alternative H_A hypotheses for a test of multiplicative interaction. (5 points)

$H_0 : \beta_3 = 0$ (or, equivalently, $e^{\beta_3} = 1$), i.e., the ratio of the hazard ratios described above to define β_3 equals one.

$H_1 : \beta_3 \neq 0$ (or, equivalently, $e^{\beta_3} \neq 1$), i.e., the ratio of the hazard ratios described above to define β_3 is not equal to one.

3. Briefly (1 sentence) interpret the HR for the interaction term in the above model (generally, referencing X_1 and X_2). (5 points)

The ratio of hazard ratios for the interaction term in the above model (e^{β_3}) compares the hazard ratio comparing a participant with both exposures in the numerator to only one in the denominator to the hazard ratio for a participant with only the other exposure in the numerator to neither exposure in the denominator.

4. Referring to the results from the simulation study, what is the estimated probability of rejecting the null if the interaction HR is equal to 1.0? *A priori* what would you expect the probability of rejecting the null to be when the true state is $HR = 1$? Why? (10 points)

The estimated probability of rejecting the null when the interaction ratio was equal to 1.0 and we set the type I error rate to be $\alpha = 0.05$ was 0.05, or 5%.

We expect the power, or the probability of rejecting the null, to equal α (here, 0.05) when the hazard ratio for the interaction equals one. Why? When the hazard ratio for the interaction equals one, the null hypothesis is true. The type I error rate, α , is by definition the probability of rejecting the null hypothesis when it is true. Power, meanwhile, is by definition the probability of rejecting the null hypothesis at a particular type I error rate under a particular alternative hypothesis. If the alternative hypothesis for which the estimated power calculation is done *is* the null hypothesis, then the power calculation should return the probability of rejecting the null hypothesis when it is true, which is the *alpha* we set to estimate the power in the first place.

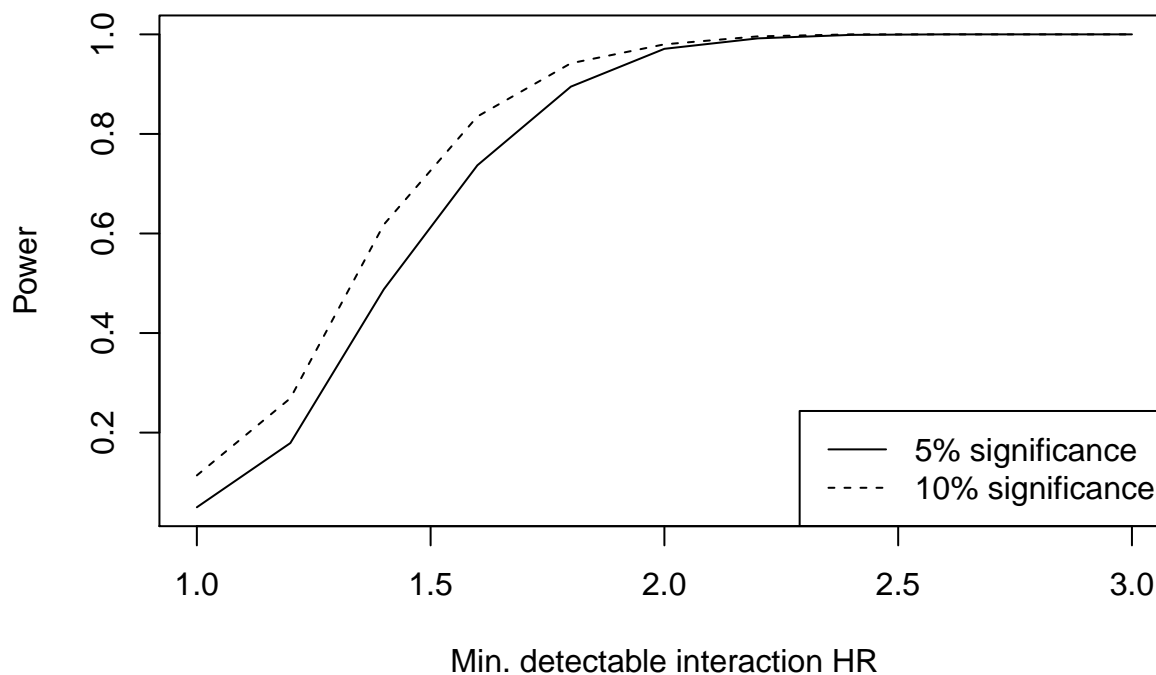
5. What is the HR for interaction that you would need to observe in order to have at least 80% power? If you wanted to be more precise in your estimate of the minimum effect size, explain briefly how you would modify the code to find the HR that gave you an answer closer to 80% (don't provide code—just briefly describe what you would do in a few sentences). (10 points)

These power calculations, done using data simulated for ratios for interaction e^{β_3} ranging from 1 to 3 at intervals of 0.2, estimate that we would need to observe a ratio for interaction of at least 1.8, which would give us 89.5% power. To more precisely estimate the minimum effect size, we could rewrite the code generating the vector of possible ratios for the interaction coefficient e^{β_3} to make it sequence more finely between 1.6 and 1.8 (e.g., we could use intervals of 0.05, or even smaller). If we wanted a certain precision and liked coding, we could also write a loop that would keep dividing the intervals more finely and rerunning the power calculation until the power was within some desired range of 80%.

6. For this question, present the simultaneous plot of the two power curves. Compare the pattern of power for significance tests at the 5% and 10% level and explain what you see in a few sentences. Suggest a reason why you observe what you do. (5 points)

Interaction HR	Power @ 5% significance	Power at 10% significance
1.0	0.050	0.114
1.2	0.179	0.269
1.4	0.488	0.618
1.6	0.737	0.835
1.8	0.895	0.942
2.0	0.971	0.980
2.2	0.992	0.996
2.4	0.999	1.000
2.6	1.000	1.000
2.8	1.000	1.000
3.0	1.000	1.000

Estimated Power for Interaction from Cox Model



```
{r ref.label=knitr::all_labels(), echo = T, eval = F}
```