1. Running a Cox PH model
2. Cox PH model (Intercept & predictors), (Model 3)

|  |
| --- |
| R code  # Running a COX PH model  Y=Surv(addicts$survt,addicts$status==1)  coxph(Y~prison + dose + clinic,data=addicts) |

Call:

coxph(formula = Y ~ prison + dose + clinic, data = addicts)

coef exp(coef) se(coef) z p

prison 0.32655 1.38618 0.16722 1.95 0.051

dose -0.03537 0.96525 0.00638 -5.54 2.9e-08

clinic -1.00990 0.36426 0.21489 -4.70 2.6e-06

Likelihood ratio test=64.6 on 3 df, p=6.23e-14

n= 238, number of events= 150

The above Cox PH model (model 3), is modelling time to relapse of drug addicts with the predictor variables prison, (i.e. whether the drug addict had a prison record or not), dose of methadone (the dose of methadone used to maintain abstinence from drug taking after drug withdrawal has been achieved) and whether the drug addict attended clinic 1 or clinic 2.

A brief glance at the direction and magnitude of the beta coefficients and hazard ratios for each variable indicate the following.

1. The drug addict having a prison record increases the risk of drug relapse. The hazard ratio for drug addicts with prison record, controlling for the variables dose and clinic indicates that the drug addicts are 38% more likely to relapse on drugs than those without a prison record, but the result is marginally non-significant
2. The drug addict receiving the maintenance drug methadone has a slightly negative effect on the outcome time to drug relapse. The dose of methadone the drug addict is on, controlling for the effects of clinic and prison reduces the risk of relapse by roughly 3.5%, and the result is highly significant,
3. The drug addict attending clinic 2 over clinic 1 has the largest negative effect on drug relapse. The drug addict attending clinic 2 over clinic 1 adjusting for the effects of prison record and dose of methadone reduces the risk of relapse by again the result is highly significant

The likelihood ratio statistic compares observed frequencies to those predicted from the model constructed the statistic here is highly significant with a . This highly significant result suggests the model has good predictive power.

1. Detailed Cox PH model with CI and descriptive statistics

|  |
| --- |
| R code  # Detailed COX PH model  summary(coxph(Y~prison+dose+clinic,data=addicts)) |

Call:

coxph(formula = Y ~ prison + dose + clinic, data = addicts)

n= 238, number of events= 150

coef exp(coef) se(coef) z Pr(>|z|)

prison 0.326555 1.386184 0.167225 1.953 0.0508 .

dose -0.035369 0.965249 0.006379 -5.545 2.94e-08 \*\*\*

clinic -1.009896 0.364257 0.214889 -4.700 2.61e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

exp(coef) exp(-coef) lower .95 upper .95

prison 1.3862 0.7214 0.9988 1.9238

dose 0.9652 1.0360 0.9533 0.9774

clinic 0.3643 2.7453 0.2391 0.5550

Concordance= 0.665 (se = 0.026 )

Rsquare= 0.238 (max possible= 0.997 )

Likelihood ratio test= 64.56 on 3 df, p=6.228e-14

Wald test = 54.12 on 3 df, p=1.056e-11

Score (logrank) test = 56.32 on 3 df, p=3.598e-12

The detailed output for model 3 confirms through the confidence intervals that the variables of dose and clinic are significant (absence of “unity value” of 1, no effect), but having a prison record is not.

The detailed output gives the estimated hazard ratio for comparing the different levels of variable either way.

Estimated hazard ratio for clinic 2 versus clinic 1 at A drug addict attending clinic 2 reduces the chances of drug relapse over clinic 1 by between 44.50 – 76.09%

Estimated hazard ratio for clinic 1 versus clinic 2 at 2.7453 (the reciprocal of 0.3643). A drug addict attending clinic 1 increases the chances of drug relapse over clinic 2 by 174.53%.

1. Handling tied events in the same time interval

|  |
| --- |
| R code  # Handling ties (i.e. more than one event in a defined time interval) - several methods  coxph(Y~prison + dose + clinic,data=addicts, method="efron")  coxph(Y~prison + dose + clinic,data=addicts, method="breslow")  coxph(Y~prison + dose + clinic,data=addicts, method="exact") |

1. Efron method

Call:

coxph(formula = Y ~ prison + dose + clinic, data = addicts, method = "efron")

coef exp(coef) se(coef) z p

prison 0.32655 1.38618 0.16722 1.95 0.051

dose -0.03537 0.96525 0.00638 -5.54 2.9e-08

clinic -1.00990 0.36426 0.21489 -4.70 2.6e-06

Likelihood ratio test=64.6 on 3 df, p=6.23e-14

n= 238, number of events= 150

1. Breslow method

Call:

coxph(formula = Y ~ prison + dose + clinic, data = addicts, method = "breslow")

coef exp(coef) se(coef) z p

prison 0.32651 1.38612 0.16722 1.95 0.051

dose -0.03540 0.96522 0.00638 -5.55 2.9e-08

clinic -1.00887 0.36463 0.21487 -4.70 2.7e-06

Likelihood ratio test=64.5 on 3 df, p=6.36e-14

n= 238, number of events= 150

1. Exact method

Call:

coxph(formula = Y ~ prison + dose + clinic, data = addicts, method = "exact")

coef exp(coef) se(coef) z p

prison 0.32697 1.38676 0.16742 1.95 0.051

dose -0.03547 0.96516 0.00639 -5.55 2.8e-08

clinic -1.01080 0.36393 0.21507 -4.70 2.6e-06

Likelihood ratio test=64.6 on 3 df, p=6.02e-14

n= 238, number of events= 150

The three different methods for handling tied observations in the same time interval has not produced any significant differences between the original model 3 and the models above using the tied methods.

1. Cox PH mod el (Intercept, predictors & interactions), (Model 4)

|  |
| --- |
| R code  # Testing the significance of interactions  # - mod 1 - the reduced model (MODEL 3)  # - mod 2 - the full model (MODEL 4)  mod1=coxph(Y~prison + dose + clinic,data=addicts)  mod2=coxph(Y~prison + dose + clinic + clinic\*prison + clinic\*dose, data=addicts)  mod1  mod2 |

mod1

Call:

coxph(formula = Y ~ prison + dose + clinic, data = addicts)

coef exp(coef) se(coef) z p

prison 0.32655 1.38618 0.16722 1.95 0.051

dose -0.03537 0.96525 0.00638 -5.54 2.9e-08

clinic -1.00990 0.36426 0.21489 -4.70 2.6e-06

Likelihood ratio test=64.6 on 3 df, p=6.23e-14

n= 238, number of events= 150

> mod2

Call:

coxph(formula = Y ~ prison + dose + clinic + clinic \* prison +

clinic \* dose, data = addicts)

coef exp(coef) se(coef) z p

prison 1.1924 3.2949 0.5414 2.20 0.028

dose -0.0192 0.9810 0.0194 -0.99 0.322

clinic 0.1796 1.1967 0.8933 0.20 0.841

prison:clinic -0.7383 0.4779 0.4315 -1.71 0.087

dose:clinic -0.0140 0.9861 0.0143 -0.97 0.330

Likelihood ratio test=68.2 on 5 df, p=2.45e-13

n= 238, number of events= 150

1. Testing the significance of an interaction

|  |
| --- |
| R code  # Utilising information from mod 1 and mod 2  names(mod2) |

[1] "coefficients" "var" "loglik" "score"

[5] "iter" "linear.predictors" "residuals" "means"

[9] "concordance" "method" "n" "nevent"

[13] "terms" "assign" "wald.test" "y"

[17] "formula" "call"

Use the Loglik function, (log-likelihood) to discern the difference between the reduced model mod 1, and the full model mod 2.

|  |
| --- |
| R code  # Call up the log-likelihood function  # first term = model with no predictors, second term model with predictors  mod2$loglik  # or  mod2[[3]] |

mod2$loglik

[1] -705.5393 -671.4500

> # or

> mod2[[3]]

[1] -705.5393 -671.4500

1. Compare full model with reduced model using Chi-square statistic

|  |
| --- |
| R code  # calculate the difference in -2 Log-likelihood (deviance) of the two models  (-2)\*(mod1$loglik[2]-mod2$loglik[2]) |

[1] 3.618211

The 2 in square brackets indicates the difference in degrees of freedom between the reduced model and the full model, hence the difference in degrees of freedom is due to the two interactions included in the full model.

1. Obtaining p-values from Chi-square statistic

|  |
| --- |
| R code  # Determine the significance of the difference between the models using Chi-square statistic  LRT=(-2)\*(mod1$loglik[2]-mod2$loglik[2])  Pvalue = 1 - pchisq(LRT,2)  Pvalue |

[1] 0.1638006

Result – no significant difference between the two models, therefore the interactions do not add explanatory power to the model.

1. Determining significance using a self-defined function

|  |
| --- |
| R code  # Determine the significance of the difference between models using a self-defined function.  lrt.surv=function(mod.full,mod.reduced,df){  lrts=(-2)\*(mod.full$loglik[2]-mod.reduced$loglik[2])  pvalue=1-pchisq(lrts,df)  return(pvalue)  }  lrt.surv(mod1,mod2, 2) |

[1] 0.1638006

The non-significant difference between model 3 (no interactions, reduced model) and model 4 (interactions, full model) prove the interactions do not add to the explanatory power of this model. This observation is also shown when comparing the initial computer output of the two models. Model 4 has a whereas model 3 which is smaller and therefore more significant as a model overall.