Cox_Models & Model Evaluation

Ixtaccihuatl Obregon, Jane Ma

2024-12-06

COX MODELS

Using the stepwise selection model: age + edema + bilirubin + albumin + copper + sgot + prothrombin + stage will be used

```
cirrhosis = read_csv("data/cirrhosis.csv")|>
  janitor::clean_names() |>
  mutate(age = round(age / 365),
         sex = if_else(sex == "M", "Male", "Female"),
         ascites = if_else(ascites == "N", "No", "Yes"),
         hepatomegaly = if_else(hepatomegaly == "N", "No", "Yes"),
         spiders = if_else(spiders == "N", "No", "Yes"),
         edema = if_else(edema == "N", "No", "Yes"),
         stage = factor(stage),
         drug = factor(drug, levels = c("Placebo", "D-penicillamine"), order = T)) |>
  na.omit()
```

```
## Rows: 418 Columns: 20
## -- Column specification -----------------------------------------------------
## Delimiter: ","
## chr  (7): Status, Drug, Sex, Ascites, Hepatomegaly, Spiders, Edema
## dbl (13): ID, N_Days, Age, Bilirubin, Cholesterol, Albumin, Copper, Alk_Phos...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
cirrhosis = cirrhosis |>
  mutate(
    status = case_when(
      status == "D" ~ 1,  # Event of interest (death)
      status == "C" | status == "CL" ~ 0,  # Censored data
      TRUE ~ as.numeric(status)))
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `status = case_when(...)`.
## Caused by warning:
## ! NAs introduced by coercion
```

Table 1: Multivariate Cox Proportional Hazards Analysis

| Characteristic	HR[1]	95% CI[1]	p-value
drug			
drug.L	1.05	0.79, 1.39	0.7
age	1.03	1.01, 1.05	0.004
edema			
No	—	—	
Yes	1.47	0.88, 2.47	0.14
bilirubin	1.09	1.05, 1.13	<0.001
albumin	0.47	0.28, 0.82	0.007
copper	1.00	1.00, 1.00	0.002
sgot	1.00	1.00, 1.01	0.015
prothrombin	1.33	1.07, 1.64	0.010
stage			
1	—	—	
2	3.88	0.47, 32.1	0.2
3	5.29	0.68, 41.1	0.11
4	8.02	1.04, 61.8	0.046

[1]HR = Hazard Ratio, CI = Confidence Interval

```r
# cox model based on stepwise selection varibales above (ixta)
cox_model_a = coxph(Surv(n_days, status) ~ drug + age + edema +
                        bilirubin + albumin + copper + sgot +
                        prothrombin + stage,
                id=id,
            data = cirrhosis)

# Summarize the results
cox_summary_a = tbl_regression(cox_model_a, exponentiate = TRUE) |>
  modify_caption("Multivariate Cox Proportional Hazards Analysis")
cox_summary_a
```

```r
surv_fit_cox_a = survfit(cox_model_a)
plot(surv_fit_cox_a, xlab = "Days", ylab = "Survival Probability", main = "Survival Curves - Unadjusted
```
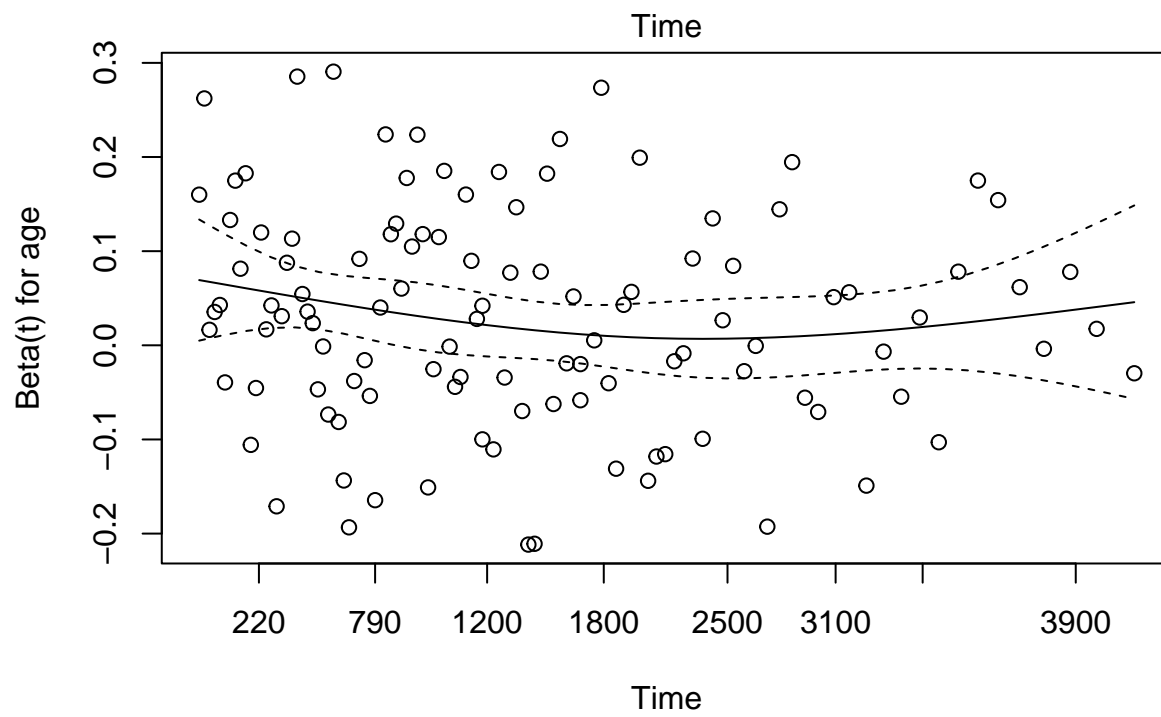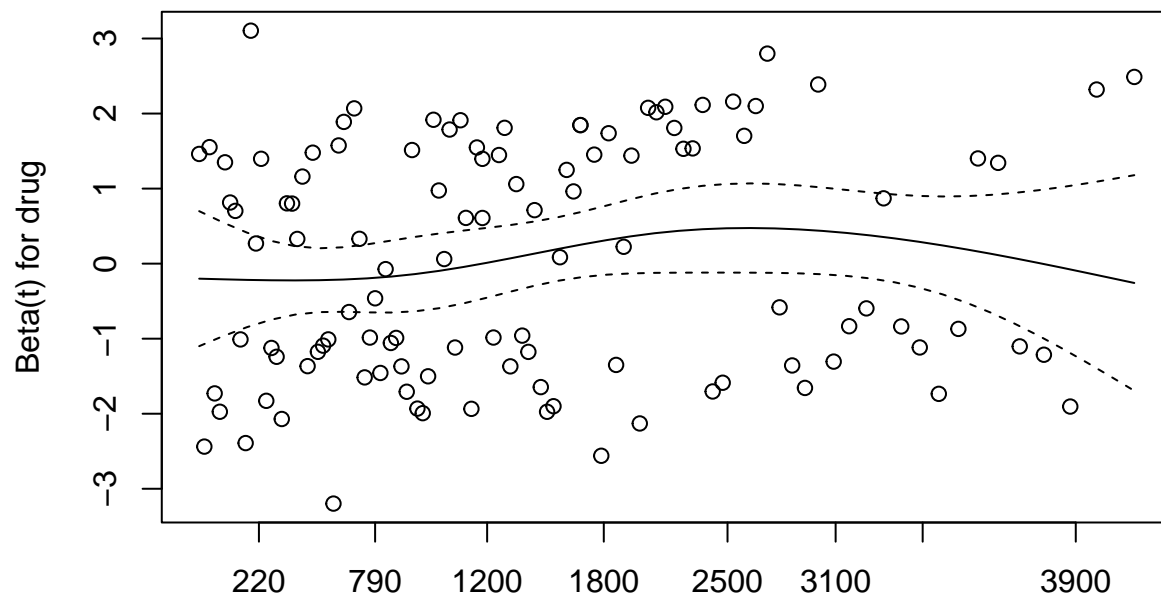
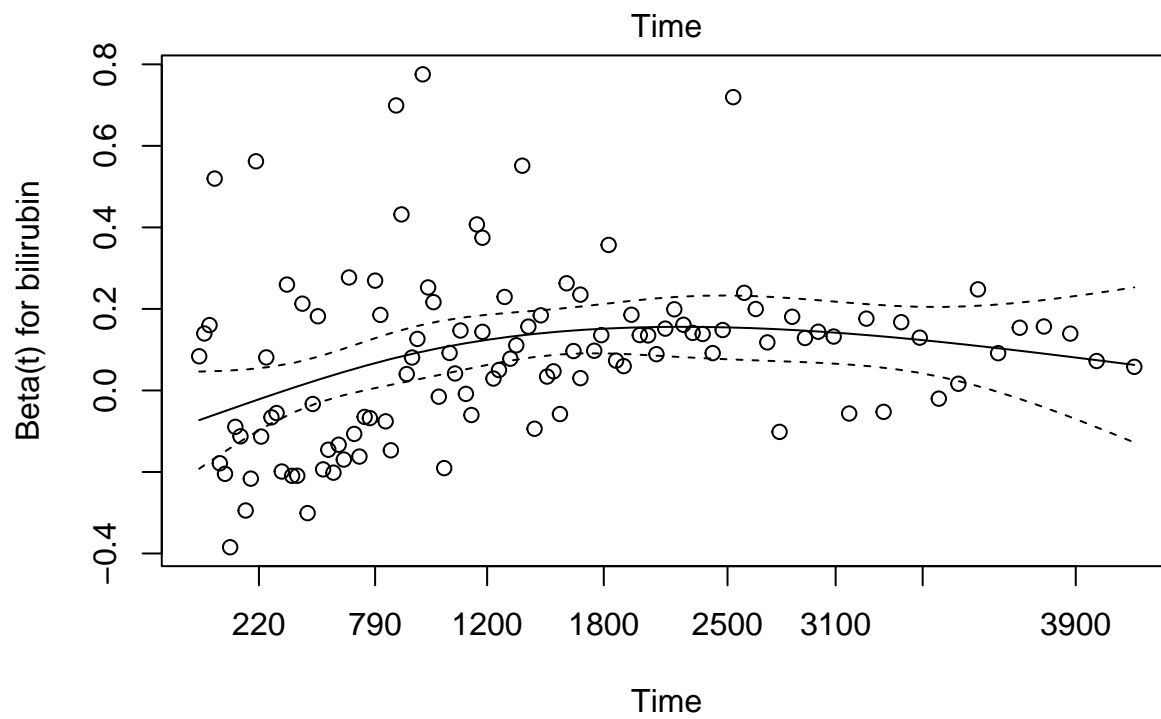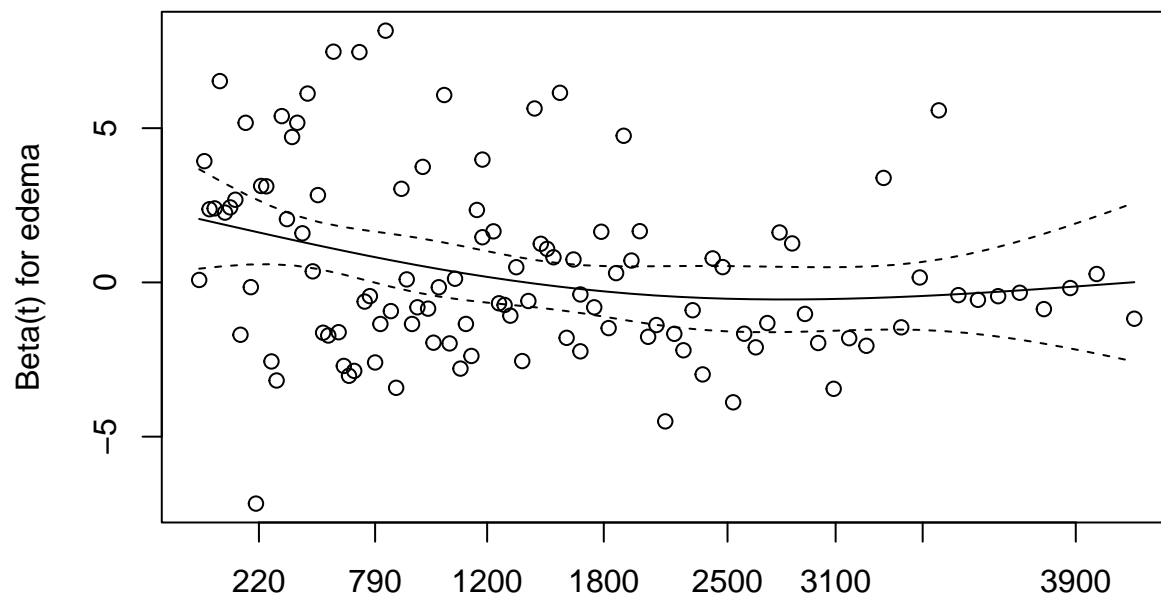## Survival Curves – Unadjusted
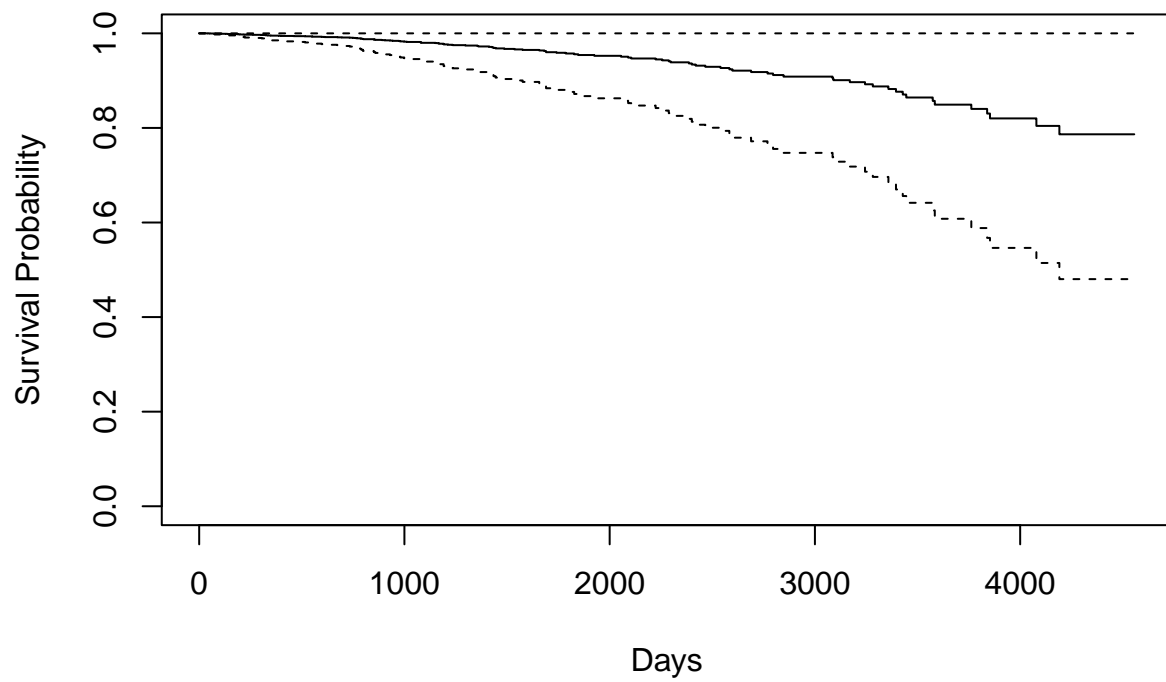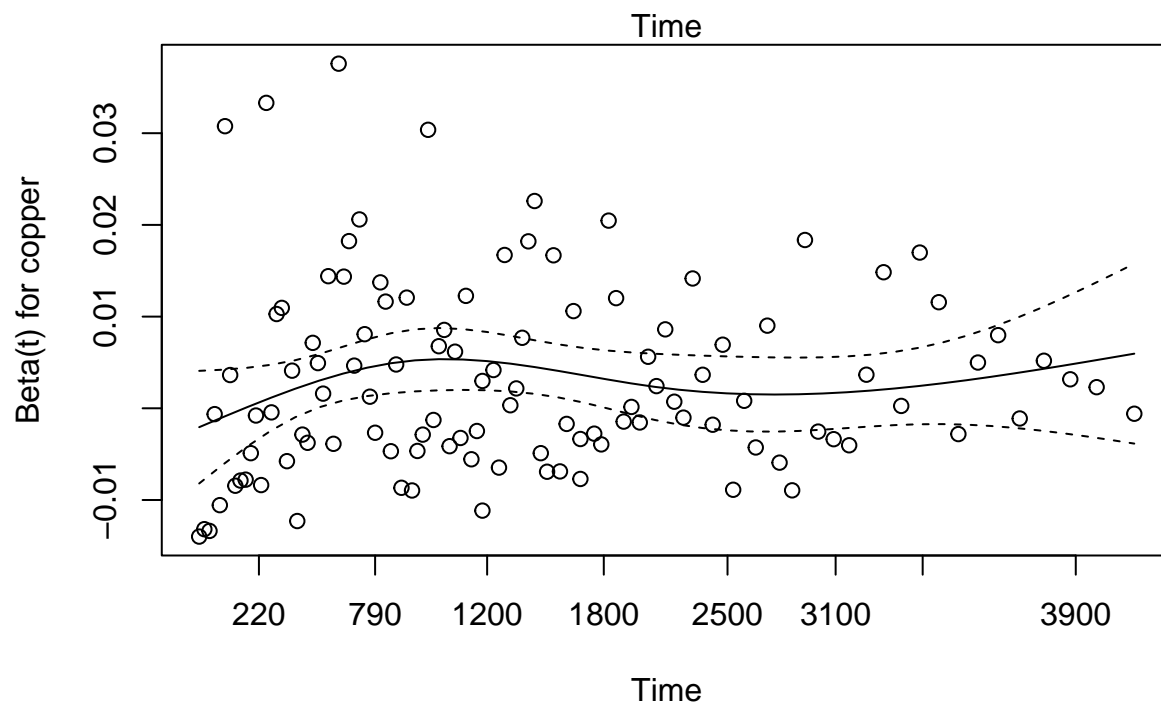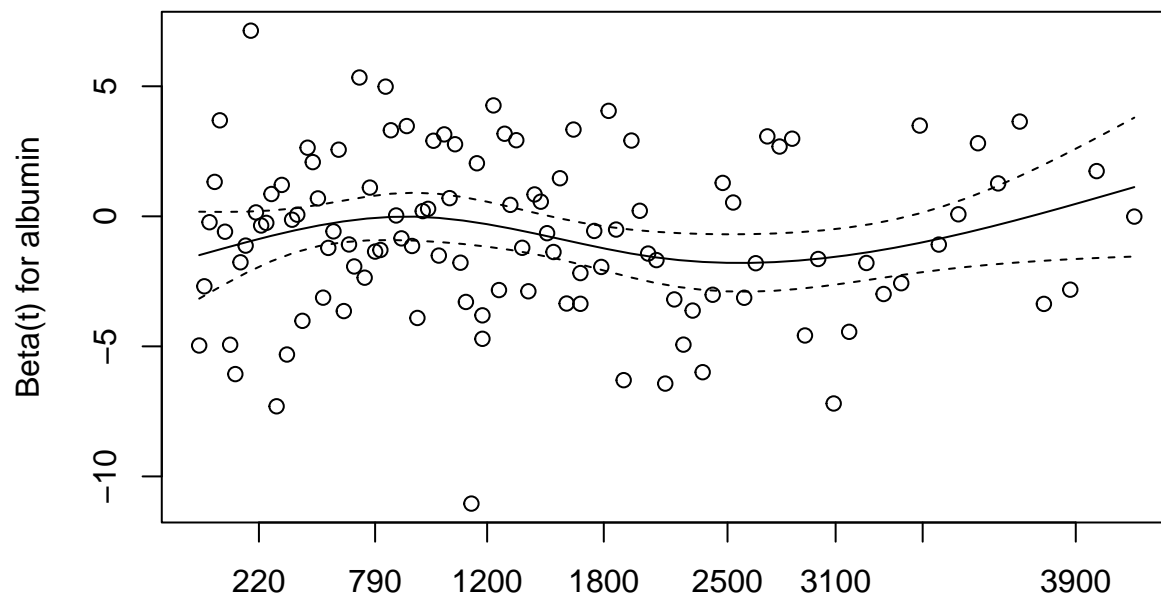


```
#ggforest(cox_model_a, data = cirrhosis, main = "Hazard Ratios for Cox Model")
```
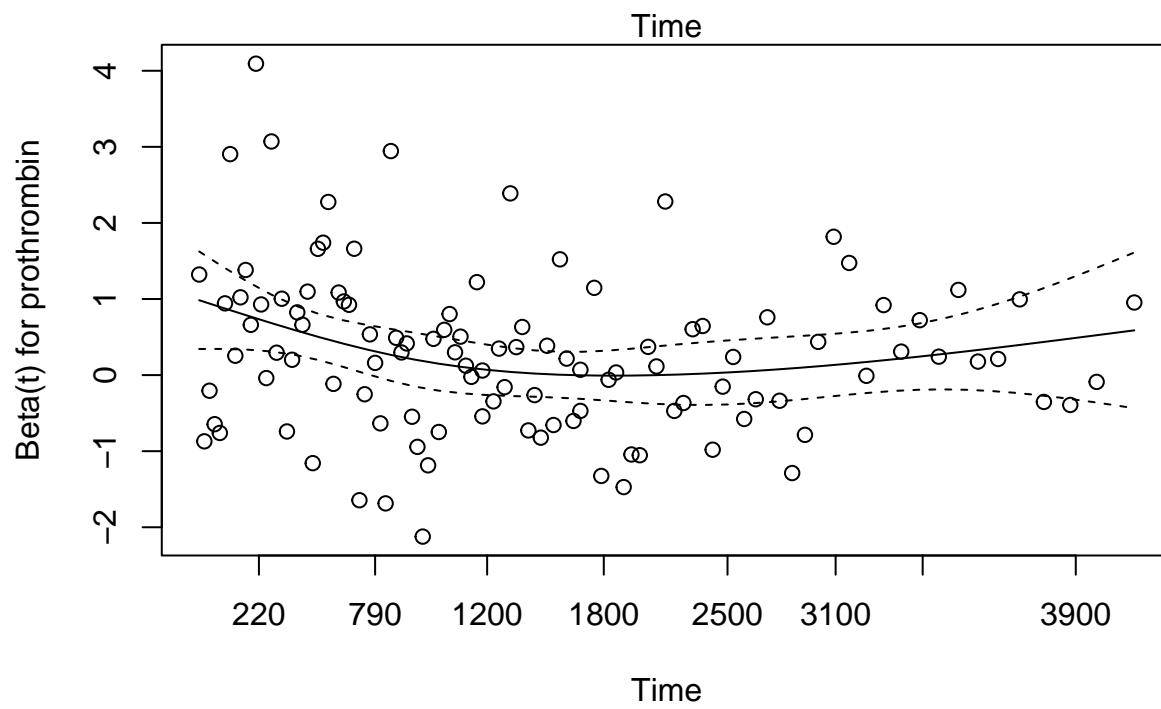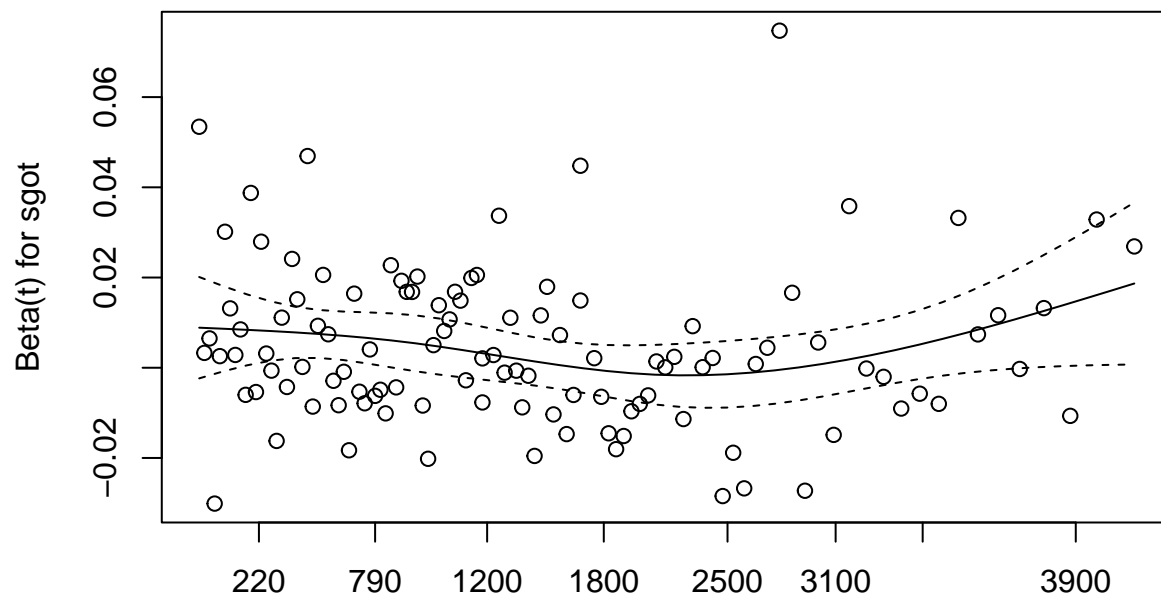
Check PH Assumption

```
ph_assumption_a = cox.zph(cox_model_a)
plot(ph_assumption_a)
```

5
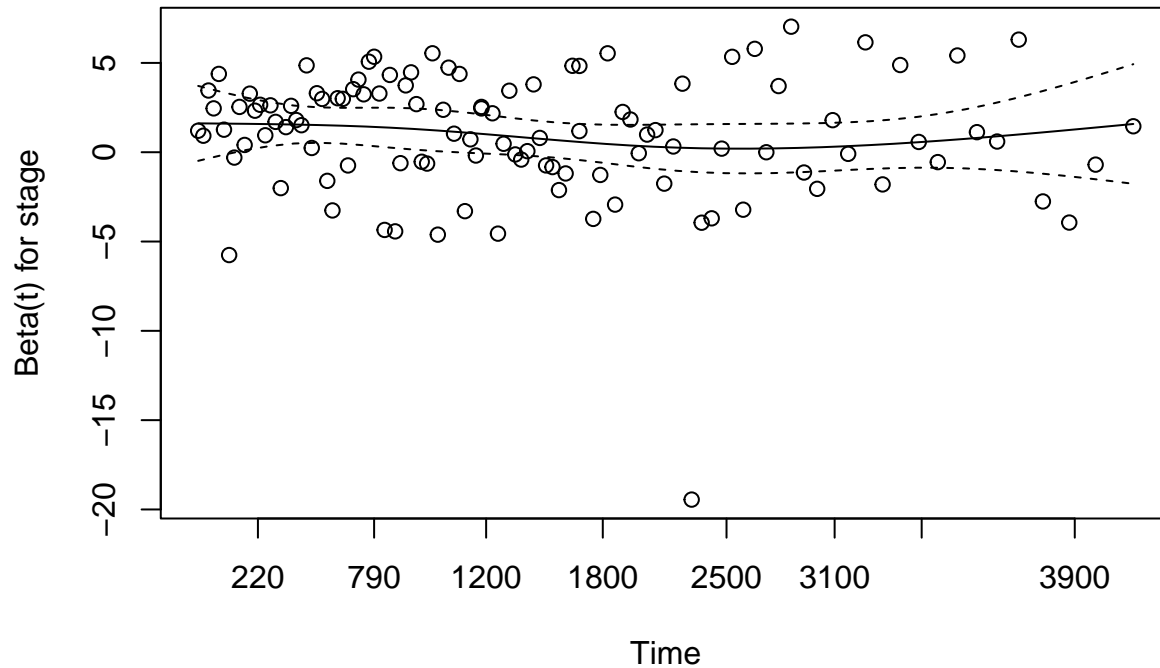
```
ph_assumption_df = as.data.frame(ph_assumption_a$table)
knitr::kable(ph_assumption_df, caption = "Proportional Hazards Assumption Test for Cox PH Model Unadjus
```

Table 2: Proportional Hazards Assumption Test for Cox PH Model Unadjusted

	chisq	df	p
drug	0.1600772	1	0.6890854
age	2.6909476	1	0.1009198
edema	6.1134319	1	0.0134158
bilirubin	8.3071868	1	0.0039489
albumin	0.6258766	1	0.4288719
copper	0.1021024	1	0.7493211
sgot	1.3384725	1	0.2473035
prothrombin	5.0189196	1	0.0250718
stage	4.5185052	3	0.2106456
GLOBAL	24.6087203	11	0.0103973

edema pval=0.0165; varies over time violating the proportional hazards assumption. bilirubin pval=0.0038; varies over time violating the proportional hazards assumption. prothrombin pval=0.0211; varies over time violating the proportional hazards assumption. global pval=0.0037; varies over time violating the proportional hazards assumption for overall model.

If assumptions not met: - do a stratified analysis - include a time-varying covariate to allow changing hazard ratios over time - include interactions with time

Checking Drug Variation

```
#checking if drug is actually useful for the model
summary(cirrhosis$drug)
```

```
##          Placebo D-penicillamine
##              140            136
```

```
table(cirrhosis$drug)
```

```
##
##          Placebo D-penicillamine
##              140            136
```

```
cox_model_drug = coxph(Surv(n_days, status) ~ drug, data = cirrhosis, id=id)
summary(cox_model_drug)
```

```
## Call:
## coxph(formula = Surv(n_days, status) ~ drug, data = cirrhosis,
##     id = id)
##
##   n= 276, number of events= 111
##
##             coef exp(coef) se(coef)      z Pr(>|z|)
## drug.L 0.08564   1.08942  0.13445 0.637    0.524
##
##        exp(coef) exp(-coef) lower .95 upper .95
## drug.L     1.089     0.9179     0.837     1.418
##
## Concordance= 0.508  (se = 0.026 )
## Likelihood ratio test= 0.41  on 1 df,    p=0.5
## Wald test            = 0.41  on 1 df,    p=0.5
## Score (logrank) test = 0.41  on 1 df,    p=0.5
```

pval=0.5 drug does not have a significant impact on survival outcomes, and we might consider removing it.

Stratification of Edema

```
cox_model_edema_strat_simple = coxph(Surv(n_days, status) ~ strata(edema), id = id, data = cirrhosis)

cox_model_edema_strat = coxph(Surv(n_days, status) ~ drug + age + strata(edema) +
                              bilirubin + albumin + copper + sgot +
                              prothrombin + stage,
                              id = id,
                              data = cirrhosis)
ph_assumption_edema_strat = cox.zph(cox_model_edema_strat)
ph_assumption_edema = as.data.frame(ph_assumption_edema_strat$table)
knitr::kable(ph_assumption_edema, caption = "Proportional Hazards Assumption Test COX PH Model Stratific
```

Table 3: Proportional Hazards Assumption Test COX PH Model Stratified for Edema

	chisq	df	p
drug	1.4344763	1	0.2310353
age	1.8937409	1	0.1687806
bilirubin	11.5075689	1	0.0006931
albumin	0.0049180	1	0.9440915
copper	0.3914633	1	0.5315312
sgot	1.0484233	1	0.3058705
prothrombin	2.6117691	1	0.1060734
stage	3.0850010	3	0.3787045
GLOBAL	18.9514038	10	0.0408843

```r
surv_fit_edema_strat = survfit(cox_model_edema_strat)
plot(surv_fit_edema_strat,
    xlab = "Days",
    ylab = "Survival Probability",
    main = "Survival Curves - Edema Stratified",
    col = c("blue", "red"),   # Assigning colors to the strata
    lty = 1,                  # Line type (solid)
    lwd = 2)                  # Line width
```

Survival Curves – Edema Stratified
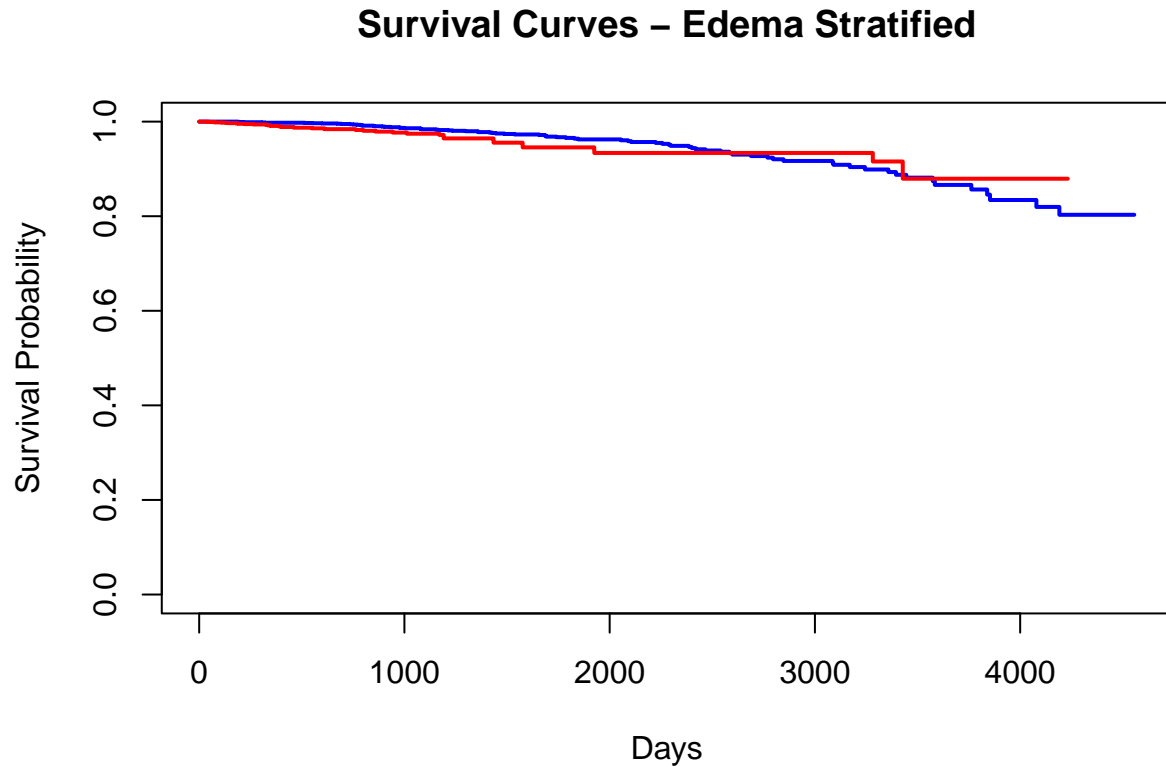

```r
#ggforest(cox_model_edema_strat, data = cirrhosis, main = "Hazard Ratios for Cox Model Edema Stratified
```

bilirubin pval=0.0006; varies over time violating the proportional hazards assumption. global pval=0.0153; varies over time violating the proportional hazards assumption for overall model.

Time-varying covariate to allow changing hazard ratios over time

```
# checking if the dataset allows for time-varying covariates
table(cirrhosis$id)
```

```
##
##   1   2   3   4   5   7   8   9  10  11  12  13  15  16  17  18  19  20  21  22
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
##  23  24  25  26  27  28  29  30  31  32  33  34  35  36  37  38  39  43  44  46
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
##  47  48  50  51  52  54  55  56  57  59  60  61  62  63  64  65  66  67  68  69
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
##  71  72  73  74  75  76  77  78  79  80  81  82  83  84  85  86  87  88  89  90
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
##  91  92  93  94  97  98  99 100 101 102 103 104 105 107 108 109 110 111 112 113
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
## 114 115 116 117 118 119 120 121 122 124 125 127 130 131 132 133 134 135 136 137
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
## 138 139 140 141 142 143 144 145 147 148 149 151 152 153 154 155 156 157 158 159
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
## 160 161 162 163 165 166 167 169 170 172 173 175 177 179 180 181 183 184 185 186
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
## 187 188 189 191 192 193 194 195 196 197 198 199 200 201 202 203 204 206 208 209
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
## 210 212 213 214 215 217 219 220 221 222 223 224 225 226 227 228 229 230 231 232
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
## 233 234 235 236 237 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
## 254 255 256 257 258 259 260 262 263 264 265 266 267 268 269 270 271 272 273 275
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
## 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
## 296 297 298 299 301 302 303 304 305 306 307 308 309 310 311 312
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
```

```
sum(table(cirrhosis$id) > 1)
```

```
## [1] 0
```

Only one observation per id therefore can not to time-varying model.

Rescaling bilirubin and prothrombin

Try to see if rescaling helps.

```
cox_model_logs = coxph(Surv(n_days, status) ~ age + edema +
                          log(bilirubin) + albumin + copper + sgot +
                          log(prothrombin) + stage,
                       data = cirrhosis, id=id)
```

```
#ggforest(cox_model_logs, data = cirrhosis, main = "Hazard Ratios for Cox Model Log(bilirubin) and log(
cox.zph(cox_model_logs) # violation in  edema, log(pro)
```

```
##                    chisq df       p
## age                2.042  1 0.1530
## edema              6.869  1 0.0088
## log(bilirubin)     0.920  1 0.3374
## albumin            0.577  1 0.4477
## copper             0.147  1 0.7019
## sgot               0.679  1 0.4100
## log(prothrombin)   5.613  1 0.0178
## stage              3.817  3 0.2819
## GLOBAL            14.920 10 0.1350
```

```
cox_model_logpro = coxph(Surv(n_days, status) ~ age + edema +
                         bilirubin + albumin + copper + sgot +
                         log(prothrombin) + stage,
                       data = cirrhosis, id=id)
```

```
#ggforest(cox_model_logpro, data = cirrhosis, main = "Hazard Ratios for Cox Mode log(prothrombin)")
cox.zph(cox_model_logpro) # violation in  edema, bili, log(pro)
```

```
##                     chisq df       p
## age                2.6154  1 0.1058
## edema              5.7796  1 0.0162
## bilirubin          8.2124  1 0.0042
## albumin            0.6331  1 0.4262
## copper             0.0983  1 0.7538
## sgot               1.3957  1 0.2374
## log(prothrombin)   5.2746  1 0.0216
## stage              4.6281  3 0.2011
## GLOBAL            21.7343 10 0.0165
```

```
cox_model_logbili = coxph(Surv(n_days, status) ~ age + edema +
                          log(bilirubin) + albumin + copper + sgot +
                          prothrombin + stage,
                        data = cirrhosis, id=id)
```

```
#ggforest(cox_model_logbili, data = cirrhosis, main = "Hazard Ratios for Cox Mode log(bilirubin)")
cox.zph(cox_model_logbili) # violation in  edema,  log(pro)
```

```
##                    chisq df       p
## age                2.092  1 0.1481
## edema              6.967  1 0.0083
## log(bilirubin)     0.938  1 0.3328
## albumin            0.612  1 0.4342
## copper             0.137  1 0.7110
## sgot               0.678  1 0.4102
## prothrombin        5.538  1 0.0186
## stage              3.814  3 0.2823
## GLOBAL            15.125 10 0.1276
```

Violations still occur with re scaling. But for log(bilirubin) and log(prothrombin) global pval=0.09 so ok model?

Interactions with Time

```
cox_model_timeint_all = coxph(Surv(n_days, status) ~ drug + age + strata(edema) +
                           bilirubin + albumin + copper + sgot +
                           prothrombin + stage + bilirubin:n_days + prothrombin : n_days,
                           id=id,
                       data = cirrhosis |> na.omit())
```

```
## Warning in coxph(Surv(n_days, status) ~ drug + age + strata(edema) + bilirubin
## + : a variable appears on both the left and right sides of the formula
```

```
#ggforest(cox_model_timeint_all, data = cirrhosis, main = "Hazard Ratios for Cox Model Time Interaction
ph_assumption_bili_pro = cox.zph(cox_model_timeint_all) # violation in bilirubin, but global pval > 0.0
ph_assumption_bp= as.data.frame(ph_assumption_bili_pro$table)
knitr::kable(ph_assumption_bp, caption = "Proportional Hazards Assumption Test COX PH Model - Time Inte
```

Table 4: Proportional Hazards Assumption Test COX PH Model - Time Interaction for Bilirubin and Prohrombin

	chisq	df	p
drug	0.0000954	1	0.9922070
age	5.6872539	1	0.0170886
bilirubin	33.8750626	1	0.0000000
albumin	4.4234600	1	0.0354481
copper	4.5102704	1	0.0336919
sgot	0.6776265	1	0.4104052
prothrombin	223.2379810	1	0.0000000
stage	3.9806497	3	0.2635612
bilirubin:n_days	7.0520026	1	0.0079177
prothrombin:n_days	94.3126489	1	0.0000000
GLOBAL	241.8746701	12	0.0000000

```
cox_model_timeint_bil = coxph(Surv(n_days, status) ~ age + strata(edema) +
                           bilirubin + albumin + copper + sgot +
                           prothrombin + stage + bilirubin : n_days ,
                       id=id,
                        data = cirrhosis)
```

```
## Warning in coxph(Surv(n_days, status) ~ age + strata(edema) + bilirubin + : a
## variable appears on both the left and right sides of the formula
```

```
#ggforest(cox_model_timeint_bil, data = cirrhosis, main = "Hazard Ratios for Cox Model Time Interaction
ph_assumption_bili = cox.zph(cox_model_timeint_bil) # violation in bilirubin, but global pval > 0.05
ph_assumption_b = as.data.frame(ph_assumption_bili$table)
knitr::kable(ph_assumption_b, caption = "Proportional Hazards Assumption Test COX PH Model - Time Intera
```

Table 5: Proportional Hazards Assumption Test COX PH Model - Time Interaction for Bilirubin

	chisq	df	p
age	2.1727589	1	0.1404740
bilirubin	156.4489525	1	0.0000000
albumin	0.2287249	1	0.6324709
copper	7.8506692	1	0.0050802
sgot	3.9551455	1	0.0467283
prothrombin	0.2988300	1	0.5846169
stage	1.7545923	3	0.6248657
bilirubin:n_days	66.5972171	1	0.0000000
GLOBAL	160.5633738	10	0.0000000

```r
cox_model_timeint_pro = coxph(Surv(n_days, status) ~ age + strata(edema) +
                    bilirubin + albumin + copper + sgot +
                    prothrombin + stage + prothrombin : n_days,
                 id=id,
               data = cirrhosis)
```

```
## Warning in coxph(Surv(n_days, status) ~ age + strata(edema) + bilirubin + : a
## variable appears on both the left and right sides of the formula
```

```r
#ggforest(cox_model_timeint_pro, data = cirrhosis, main = "Hazard Ratios for Cox Model Time Interaction
ph_assumption_pro = cox.zph(cox_model_timeint_pro) # no violation
ph_assumption_p = as.data.frame(ph_assumption_pro$table)
knitr::kable(ph_assumption_p, caption = "Proportional Hazards Assumption Test COX PH Model - Time Inter
```

Table 6: Proportional Hazards Assumption Test COX PH Model - Time Interaction for Prothrombin

	chisq	df	p
age	4.4256842	1	0.0354019
bilirubin	9.3485413	1	0.0022316
albumin	3.7384028	1	0.0531753
copper	6.7691484	1	0.0092747
sgot	0.0015555	1	0.9685394
prothrombin	220.7119921	1	0.0000000
stage	3.7539572	3	0.2892873
prothrombin:n_days	90.8223488	1	0.0000000
GLOBAL	238.1570909	10	0.0000000

```r
tibble(
  Model = c(
    "Plain",
    "Stratified by Edema",
    "Time Interaction Bilirubin and Prothrombin",
    "Time Interaction Bilirubin",
    "Time Interaction Prothrombin"
  ),
  AIC = c(
    AIC(cox_model_a),
```

Table 8: Multivariate Cox Proportional Hazards Analysis

Characteristic	HR[1]	95% CI[1]	p-value
drug			
drug.L	1.22	0.86, 1.72	0.3
age	1.02	0.99, 1.05	0.15
bilirubin	0.98	0.89, 1.07	0.7
albumin	1.14	0.58, 2.24	0.7
copper	1.00	1.00, 1.01	0.066
sgot	1.00	1.00, 1.01	0.2
prothrombin	6.48	4.21, 9.97	<0.001
stage			
1	—	—	
2	2.24	0.27, 18.5	0.5
3	4.48	0.53, 37.7	0.2
4	6.74	0.79, 57.5	0.081
bilirubin * n_days	1.00	1.00, 1.00	0.083
prothrombin * n_days	1.00	1.00, 1.00	<0.001

[1]HR = Hazard Ratio, CI = Confidence Interval

```
    AIC(cox_model_edema_strat),
    AIC(cox_model_timeint_all),
    AIC(cox_model_timeint_bil),
    AIC(cox_model_timeint_pro)
  )
) |> knitr::kable(caption = "AIC Comparison between Different Models")
```

Table 7: AIC Comparison between Different Models

Model	AIC
Plain	960.1851
Stratified by Edema	849.5708
Time Interaction Bilirubin and Prothrombin	415.2000
Time Interaction Bilirubin	818.6991
Time Interaction Prothrombin	415.6481

We use the model stratified by Edema and with the interaction with time for bilirubin and prothrombin. The coefficients are:

```
tbl_regression(cox_model_timeint_all, exponentiate = TRUE) |>
  modify_caption("Multivariate Cox Proportional Hazards Analysis")
```

It can be concluded that prothrombin and stage are significantly associated with survival risk. As we stratified by edema, edema is also a potential significant factor. The result complies with what was discovered in the KM model.

Interaction between Convariates

```
cox_init = coxph(Surv(n_days, status) ~ drug + age + strata(edema) +
                               bilirubin + albumin + copper + sgot +
                               prothrombin + stage + bilirubin : n_days,
                       id = id,
                       data = cirrhosis |> na.omit())
variables = c("drug", "age", "albumin", "copper", "sgot",
              "prothrombin", "stage")
vars_df = tibble()
for(var in variables[1 : (length(variables) - 1)])
{
  left_vars = variables[(which(variables == var) + 1) : length(variables)]
  for(var2 in left_vars)
  {
    cox_fit = coxph(Surv(n_days, status) ~ drug + age + strata(edema) +
                                 bilirubin + albumin + copper + sgot +
                                 prothrombin + stage + bilirubin : n_days +
                       eval(parse(text = var2)) : eval(parse(text = var)),
                         id = id,
                         data = cirrhosis |> na.omit())
    # aic_vec= c(aic_vec, AIC(model_four))
    chisq_stat=-2 * (logLik(cox_init)-logLik(cox_fit))
    p_val = 1 - pchisq(chisq_stat,
                       attr(logLik(cox_fit), "df") -
                         attr(logLik(cox_init),"df"))
    if(p_val < 0.05)
    {
      vars_df = vars_df |> rbind(c(round(p_val, 4), var, var2))
    }
  }
}

colnames(vars_df) = c("p_value", "variable1", "variable2")
vars_df |>
  mutate(interaction = paste0(variable1, " * ", variable2)) |>
  select(interaction, p_value) |>
  knitr::kable(col.names = c("Interaction Term", "P Value"),
    caption = "Siginificant Interaction term")
```

Table 9: Siginificant Interaction term

Interaction Term	P Value
age * copper	0.031
albumin * copper	5e-04
copper * sgot	0.0082
copper * prothrombin	0.0238
copper * stage	0.0015

We first add the albumin*copper term into the model and evaluate again.

```
cox_fit2 = coxph(Surv(n_days, status) ~ drug + age + strata(edema) +
                     bilirubin + albumin + copper + sgot + prothrombin + stage +
                     bilirubin : n_days + albumin * copper,
                 id = id, data = cirrhosis)
vars_df = tibble()
for(var in variables[1 : (length(variables) - 1)])
{
  left_vars = variables[(which(variables == var) + 1) : length(variables)]
  for(var2 in left_vars)
  {
    cox_fit = coxph(Surv(n_days, status) ~ drug + age + strata(edema) +
                        bilirubin + albumin + copper + sgot + prothrombin +
                        stage + bilirubin : n_days + albumin * copper +
                        eval(parse(text = var2)) : eval(parse(text = var)),
                      id = id,
                      data = cirrhosis)
    # aic_vec= c(aic_vec, AIC(model_four))
    chisq_stat=-2 * (logLik(cox_fit2)-logLik(cox_fit))
    p_val = 1 - pchisq(chisq_stat,
                     attr(logLik(cox_fit), "df") -
                       attr(logLik(cox_fit2),"df"))
    if(p_val < 0.05)
    {
      vars_df = vars_df |> rbind(c(round(p_val, 4), var, var2))
    }
  }
}
```

This is our final model.

```
cox_final = cox_fit2
summary(cox_final)$coefficient %>% .[, c(1, 2, 5)] |>
  data.frame() |> mutate(significance = c("", "**", "***", "***", "**", "***", "*",
                                     "", "", "*", "***", "***")) |>
  knitr::kable(col.names = c(" ", "Estimate", "Hazard Ratio", "p value", "Sig."),
               digits = 4, caption = "Final Model Parameter Results")
```

Table 10: Final Model Parameter Results

	Estimate	Hazard Ratio	p value	Sig.
drug.L	0.2406	1.2720	0.1124	
age	0.0337	1.0343	0.0025	**
bilirubin	0.2467	1.2798	0.0000	***
albumin	-1.4627	0.2316	0.0001	***
copper	-0.0224	0.9779	0.0020	**
sgot	0.0065	1.0065	0.0010	***
prothrombin	0.2819	1.3257	0.0168	*
stage2	1.3093	3.7034	0.2228	
stage3	1.6975	5.4604	0.1058	
stage4	2.0812	8.0139	0.0467	*
bilirubin:n_days	-0.0002	0.9998	0.0000	***
albumin:copper	0.0076	1.0076	0.0004	***

Model Evaluation

```r
deviance_res = residuals(cox_final, type = "deviance")

plot(deviance_res, ylab = "Deviance Residuals", xlab = "Index",
     main = "Deviance Residuals Scatterplot")
abline(h = c(-3, 3), col = "red", lty = 2) # Flag large residuals
```
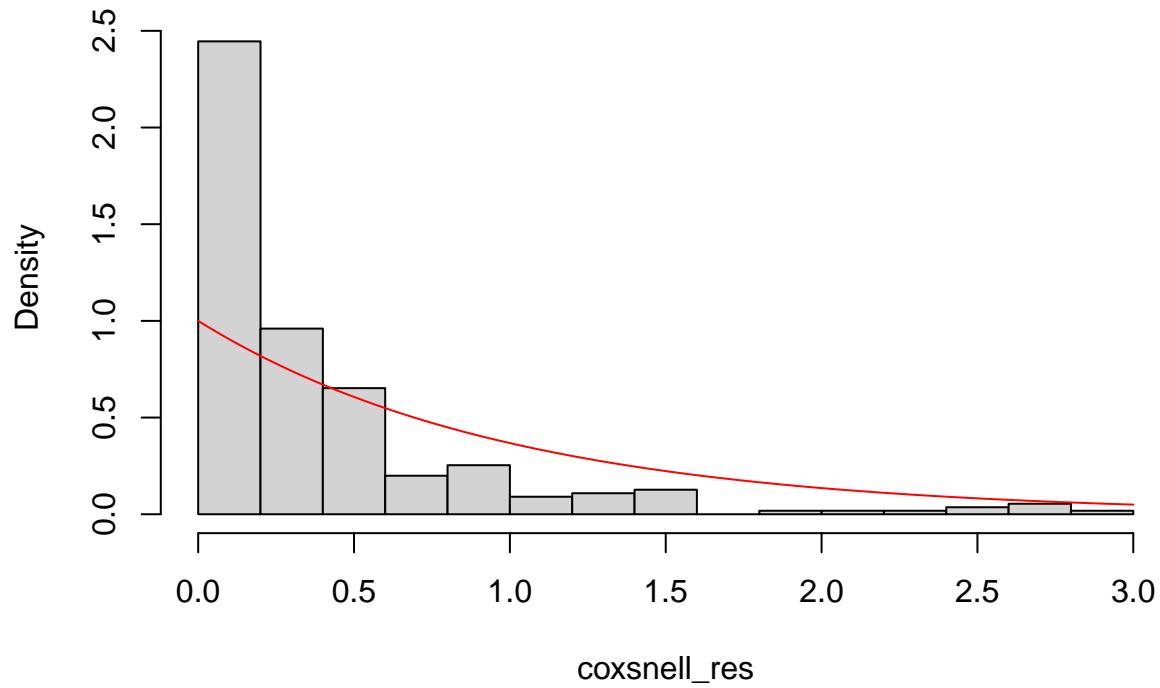
Deviance Residuals Scatterplot
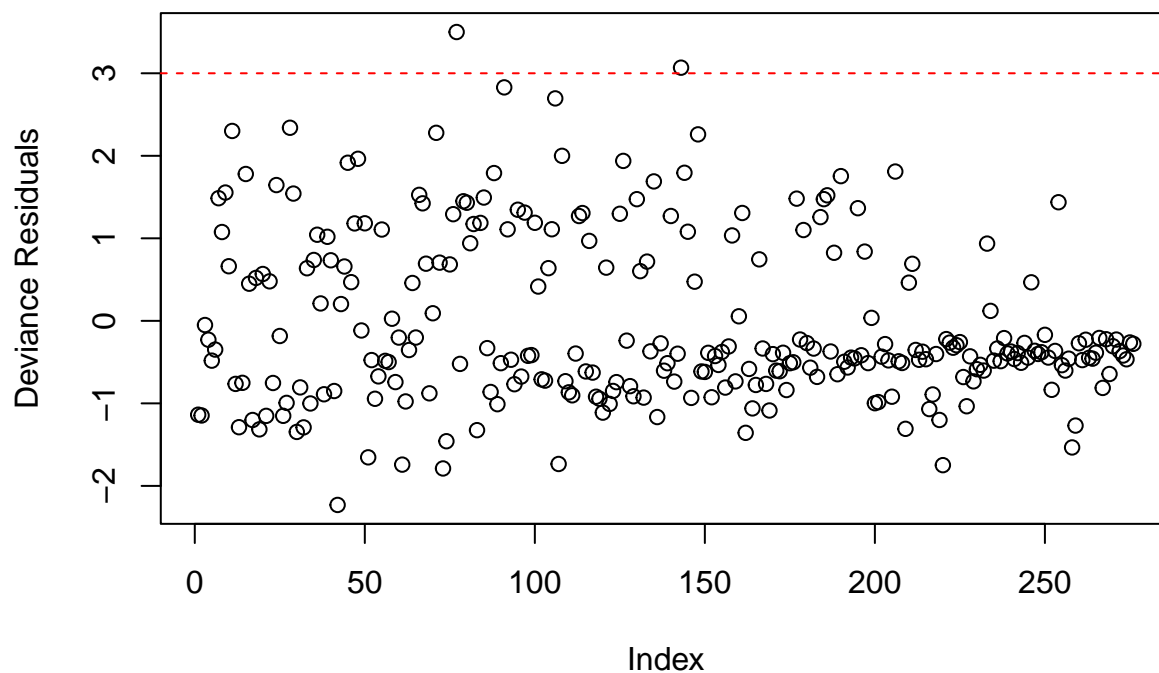

```r
which(deviance_res > 3)
```
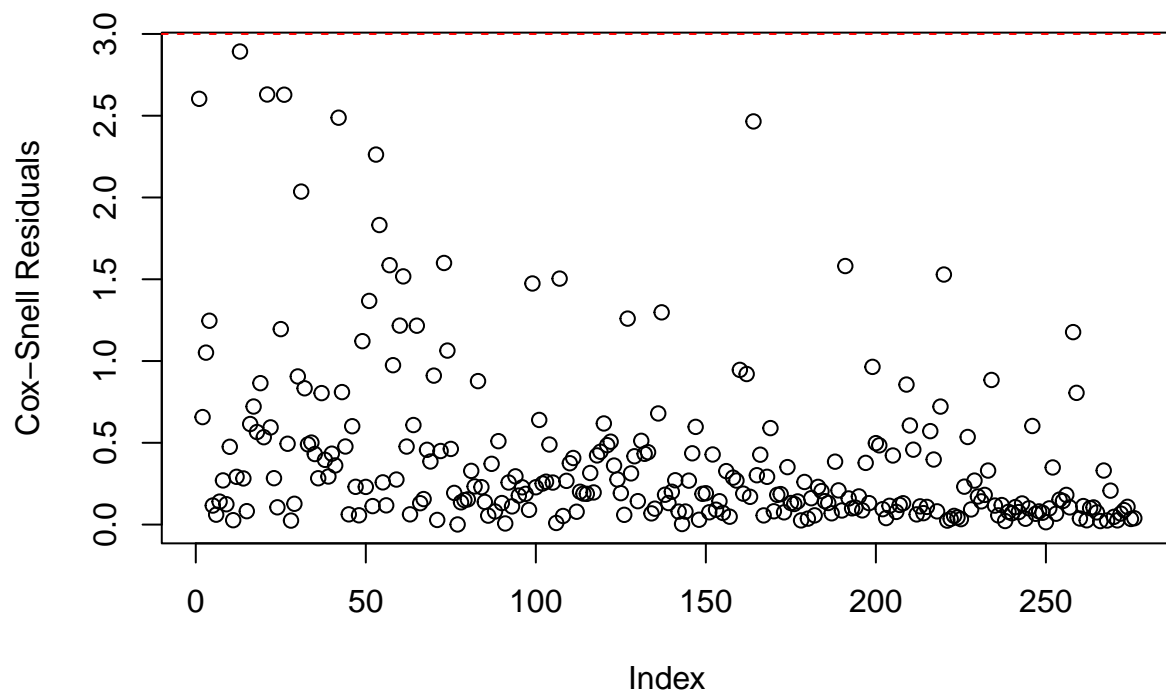
```
##  77 143
##  77 143
```

```r
coxsnell_res = - (predict(cox_final, type = "survival") |> log())
hist(coxsnell_res, main = "Cox-Snell Residuals Histogram", freq = F, breaks = 15)
curve(exp(- x), add = T, col = "red")
```

Cox−Snell Residuals Histogram


```
plot(coxsnell_res, ylab = "Cox-Snell Residuals", xlab = "Index",
    main = "Cox-Snell Residuals Scatterplot")
abline(h = 3, col= "red", lty = 2)
```

Cox−Snell Residuals Scatterplot

Based on the plot, the Cox-Snell residuals approximately follow exponential (1) distribution, which indicates

good fit of the model.

For influence diagnostics, we use the LD option. We calculate it manually and select the individuals that provide the 5 largest absolute differences.

```
ld_res = c()
for(i in 1 : nrow(cirrhosis))
{
  dat = cirrhosis |> slice(- i)
  model_ld = coxph(Surv(n_days, status) ~ drug + age + strata(edema) +
                      bilirubin + albumin + copper + sgot + prothrombin + stage +
                      bilirubin : n_days + albumin * copper,
                  id = id, data = dat)
  ld_res = c(ld_res, 2 * abs(logLik(model_ld) - logLik(cox_final)))
}
hist(ld_res)
```

Histogram of ld_res
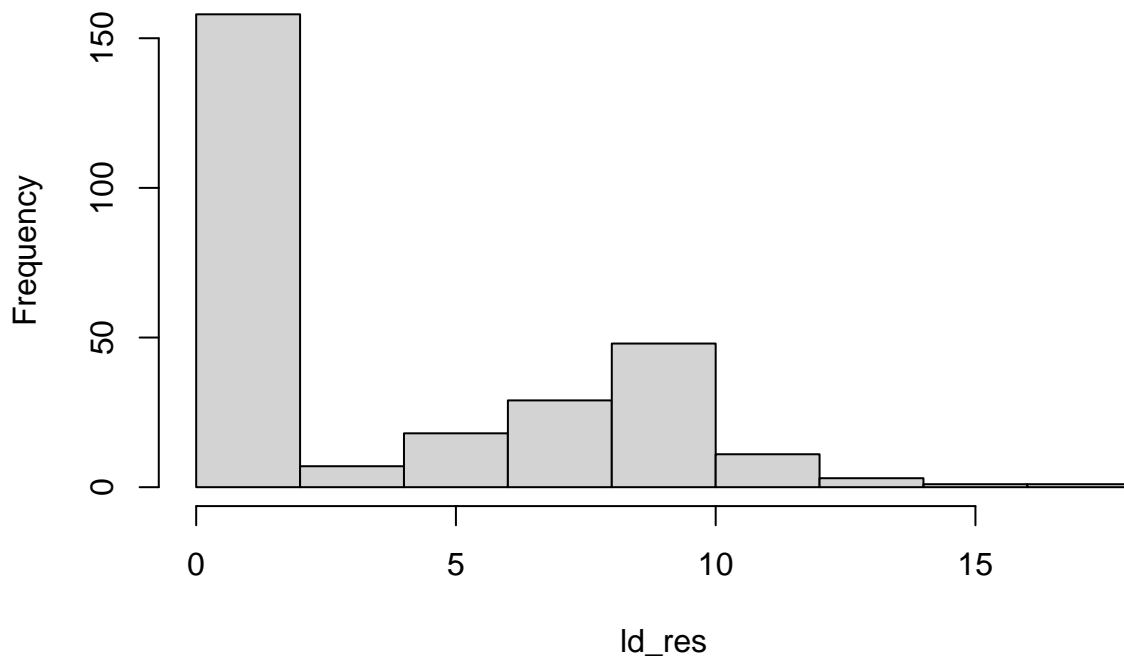
The indices for the patients who seem to have a lot of influence on the model based on the LD option are 82, 100, 108, 129, 210.

```
cox_after = cirrhosis |>
  slice(c(- 77, - 143, - 82, - 100, - 108, - 129, - 210)) |>
  coxph(Surv(n_days, status) ~ drug + age + strata(edema) +
                      bilirubin + albumin + copper + sgot + prothrombin + stage +
                      bilirubin : n_days + albumin * copper,
                  id = id, data = _)
summary(cox_final)$coefficient %>% .[, c(1, 2, 5)] |>
  cbind(summary(cox_after)$coefficient %>% .[, c(1, 2, 5)]) |>
  knitr::kable(col.names = c(" ", rep(c("Estimate", "Hazard Ratio", "p value"), 2)),
```

```
                digits = 4, caption = "Model Parameter Results Comparison") |>
  add_header_above(header = c(" " = 1, "Original Model" = 3, "New Model" = 3))
```

Table 11: Model Parameter Results Comparison

	Original Model			New Model		
	Estimate	Hazard Ratio	p value	Estimate	Hazard Ratio	p value
drug.L	0.2406	1.2720	0.1124	0.2907	1.3374	0.0662
age	0.0337	1.0343	0.0025	0.0303	1.0308	0.0098
bilirubin	0.2467	1.2798	0.0000	0.3033	1.3543	0.0000
albumin	-1.4627	0.2316	0.0001	-1.5639	0.2093	0.0001
copper	-0.0224	0.9779	0.0020	-0.0226	0.9777	0.0021
sgot	0.0065	1.0065	0.0010	0.0063	1.0063	0.0024
prothrombin	0.2819	1.3257	0.0168	0.3428	1.4089	0.0063
stage2	1.3093	3.7034	0.2228	1.4232	4.1505	0.1921
stage3	1.6975	5.4604	0.1058	1.7935	6.0104	0.0930
stage4	2.0812	8.0139	0.0467	2.1404	8.5030	0.0440
bilirubin:n_days	-0.0002	0.9998	0.0000	-0.0002	0.9998	0.0000
albumin:copper	0.0076	1.0076	0.0004	0.0078	1.0078	0.0003