Question 3 Saufi

Ahmad Firdaus Mohamed

2024-02-11

3. The dataset “stroke\_outcome.dta” comes from a short prospective survival study. The variable outcome is labelled as dead (1) or alive (0) and the variable days is the duration from the start of follow-up untul the last follow-up. (SAUFI)

## a. Present and interpret the result for

- Overall Kaplan-Meier analysis

KM <-survfit(Surv(time = dur\_days, event = event=='dead')~1, type="kaplan-meier", data = Data1)  
summary(KM)

## Call: survfit(formula = Surv(time = dur\_days, event = event == "dead") ~   
## 1, data = Data1, type = "kaplan-meier")  
##   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 0 262 4 0.985 0.00758 0.9700 1.000  
## 1 256 13 0.935 0.01531 0.9052 0.965  
## 2 224 4 0.918 0.01716 0.8850 0.952  
## 3 193 7 0.885 0.02064 0.8452 0.926  
## 4 147 4 0.861 0.02333 0.8161 0.908  
## 5 105 5 0.820 0.02852 0.7656 0.878  
## 6 79 4 0.778 0.03379 0.7147 0.847  
## 7 68 4 0.732 0.03879 0.6602 0.813  
## 9 51 3 0.689 0.04376 0.6087 0.781  
## 10 44 2 0.658 0.04705 0.5719 0.757  
## 12 38 4 0.589 0.05334 0.4929 0.703  
## 14 29 2 0.548 0.05687 0.4473 0.672  
## 18 23 1 0.524 0.05918 0.4202 0.654  
## 22 18 1 0.495 0.06265 0.3864 0.635  
## 25 12 2 0.413 0.07459 0.2895 0.588  
## 28 7 1 0.354 0.08406 0.2220 0.564  
## 29 6 1 0.295 0.08833 0.1638 0.530  
## 41 3 1 0.196 0.09951 0.0728 0.530

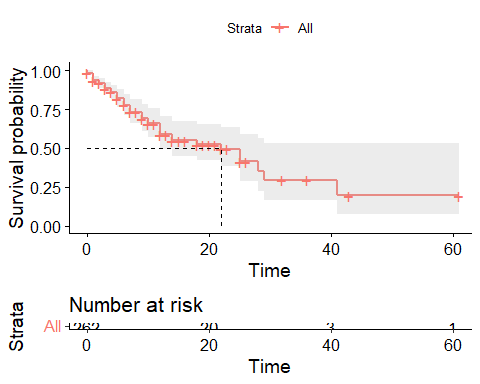
manually calculate for 10th observation

(1-(2/44))\*0.689

## [1] 0.6576818

ggsurvplot(KM, surv.median.line = "hv", pval = TRUE, conf.int = TRUE,risk.table = TRUE)

## Warning in .pvalue(fit, data = data, method = method, pval = pval, pval.coord = pval.coord, : There are no survival curves to be compared.   
## This is a null model.



quantile(KM, probs = c(0.25,0.5,0.75))

## $quantile  
## 25 50 75   
## 7 22 41   
##   
## $lower  
## 25 50 75   
## 6 12 28   
##   
## $upper  
## 25 50 75   
## 12 NA NA

The median time for survival overall is 22 days (95% CI : 50, -)

summary(KM,times=c(0,22))

## Call: survfit(formula = Surv(time = dur\_days, event = event == "dead") ~   
## 1, data = Data1, type = "kaplan-meier")  
##   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 0 262 4 0.985 0.00758 0.970 1.000  
## 22 18 54 0.495 0.06265 0.386 0.635

At time 0, 4 events occured with the survival probability is 0.985 (95% CI : 0.970, 1.0)

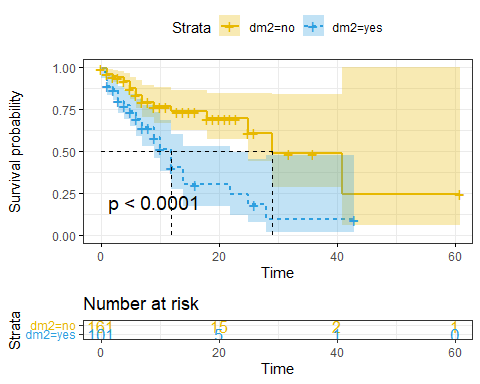
At time 22, which is the median survival time, 58 event occurred. The survival probability is 0.495 (95% 0.386, 0.635).

- Kaplan-Meier analysis on variable dm2

KM.dm2 <-survfit(Surv(time = dur\_days, event = event=='dead')~dm2, type="kaplan-meier", data = Data1)  
summary(KM.dm2)

## Call: survfit(formula = Surv(time = dur\_days, event = event == "dead") ~   
## dm2, data = Data1, type = "kaplan-meier")  
##   
## dm2=no   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 0 161 1 0.994 0.00619 0.9817 1.000  
## 1 158 5 0.962 0.01508 0.9332 0.992  
## 2 137 2 0.948 0.01784 0.9140 0.984  
## 3 117 1 0.940 0.01944 0.9028 0.979  
## 4 90 2 0.919 0.02397 0.8735 0.967  
## 5 62 3 0.875 0.03388 0.8109 0.944  
## 6 47 2 0.838 0.04142 0.7602 0.923  
## 7 40 2 0.796 0.04880 0.7056 0.897  
## 9 30 1 0.769 0.05390 0.6705 0.882  
## 12 24 1 0.737 0.06044 0.6277 0.866  
## 18 18 1 0.696 0.06959 0.5723 0.847  
## 25 8 1 0.609 0.10165 0.4392 0.845  
## 29 5 1 0.487 0.13597 0.2820 0.842  
## 41 2 1 0.244 0.18522 0.0549 1.000  
##   
## dm2=yes   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 0 101 3 0.9703 0.0169 0.9377 1.000  
## 1 98 8 0.8911 0.0310 0.8324 0.954  
## 2 87 2 0.8706 0.0335 0.8074 0.939  
## 3 76 6 0.8019 0.0410 0.7255 0.886  
## 4 57 2 0.7737 0.0441 0.6920 0.865  
## 5 43 2 0.7377 0.0488 0.6480 0.840  
## 6 32 2 0.6916 0.0556 0.5908 0.810  
## 7 28 2 0.6422 0.0616 0.5321 0.775  
## 9 21 2 0.5811 0.0693 0.4600 0.734  
## 10 17 2 0.5127 0.0762 0.3832 0.686  
## 12 14 3 0.4028 0.0821 0.2702 0.601  
## 14 8 2 0.3021 0.0872 0.1717 0.532  
## 22 5 1 0.2417 0.0882 0.1182 0.494  
## 25 4 1 0.1813 0.0844 0.0728 0.451  
## 28 2 1 0.0906 0.0767 0.0172 0.476

ggsurvplot(KM.dm2, pval = TRUE, conf.int = TRUE,  
 risk.table = TRUE, # Add risk table  
 risk.table.col = "strata", # Change risk table color by groups  
 linetype = "strata", # Change line type by groups  
 surv.median.line = "hv", # Specify median survival  
 ggtheme = theme\_bw(), # Change ggplot2 theme  
 palette = c("#E7B800", "#2E9FDF"))



quantile(KM.dm2, probs = c(0.25,0.5,0.75))

## $quantile  
## 25 50 75  
## dm2=no 12 29 41  
## dm2=yes 5 12 22  
##   
## $lower  
## 25 50 75  
## dm2=no 7 25 41  
## dm2=yes 3 9 14  
##   
## $upper  
## 25 50 75  
## dm2=no NA NA NA  
## dm2=yes 9 22 NA

summary(KM.dm2,times=c(0,12,29))

## Call: survfit(formula = Surv(time = dur\_days, event = event == "dead") ~   
## dm2, data = Data1, type = "kaplan-meier")  
##   
## dm2=no   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 0 161 1 0.994 0.00619 0.982 1.000  
## 12 24 19 0.737 0.06044 0.628 0.866  
## 29 5 3 0.487 0.13597 0.282 0.842  
##   
## dm2=yes   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 0 101 3 0.9703 0.0169 0.9377 1.000  
## 12 14 31 0.4028 0.0821 0.2702 0.601  
## 29 1 5 0.0906 0.0767 0.0172 0.476

Comment on the time, number of event and probability of the event for both analyses. (10 marks)

For no DM group, at time zero, 1 event occured with the survival probabilities of 99.4% (95% cI 98.1%, 100%). At time 29, which is the median survival time, 23 events had occurred, with the survival probabilities of 48.7% (95% CI 28.2, 84.2%)

For Dm group, at time zero, 3 event occured with the survival probabilities of 97.4% (95% cI 93.7%, 100%). At time 12, which is the median survival time, 34 events had occurred, with the survival probabilities of 40.2% (95% CI 8.2%, 27.0%)

The median survival is approximately 29 days for no DM group and 12 days for dm group, suggesting a good survival for no DM group compared to DM group.

## b. The analyst proposed to model the log hazard using the Cox proportional hazard (PH) model.

Comment on his proposal. (5 marks)

It is reasonable for the analyst to proceed with COX proportionate hazard because of the advantages over Kaplan-meier which include:

1) it’s multivariable analysis

1. able to analyse both numerical and categorical independent variables
2. it’s semi-parametric analysis which is better than non-parametric KM
3. KM unable to calculate log hazard, but COX PH is able to
4. COX PH also able to provide instantaneous risk.
5. COX PH able to assess confounder, and interaction in the model

## c. Estimate the Cox PH model with covariate: sex, gcs, dm2, sbp, dbp. Present the results in a table. Interpret the result for sbp and dm2. (10 marks)

Cox1 <- coxph(Surv(time = dur\_days, event= event == 'dead')~ sex + gcs + dm2 + sbp + dbp, data = Data1)  
summary(Cox1)

## Call:  
## coxph(formula = Surv(time = dur\_days, event = event == "dead") ~   
## sex + gcs + dm2 + sbp + dbp, data = Data1)  
##   
## n= 262, number of events= 63   
##   
## coef exp(coef) se(coef) z Pr(>|z|)   
## sexfemale 0.186970 1.205591 0.284123 0.658 0.510499   
## gcs -0.215582 0.806072 0.032872 -6.558 5.44e-11 \*\*\*  
## dm2yes 0.908952 2.481720 0.272826 3.332 0.000863 \*\*\*  
## sbp 0.001297 1.001298 0.006324 0.205 0.837529   
## dbp -0.001965 0.998037 0.009824 -0.200 0.841509   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## exp(coef) exp(-coef) lower .95 upper .95  
## sexfemale 1.2056 0.8295 0.6908 2.1040  
## gcs 0.8061 1.2406 0.7558 0.8597  
## dm2yes 2.4817 0.4029 1.4539 4.2363  
## sbp 1.0013 0.9987 0.9890 1.0138  
## dbp 0.9980 1.0020 0.9790 1.0174  
##   
## Concordance= 0.837 (se = 0.026 )  
## Likelihood ratio test= 66.93 on 5 df, p=4e-13  
## Wald test = 64.37 on 5 df, p=2e-12  
## Score (logrank) test = 79.49 on 5 df, p=1e-15

tidy(Cox1, conf.int = T)

## # A tibble: 5 × 7  
## term estimate std.error statistic p.value conf.low conf.high  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 sexfemale 0.187 0.284 0.658 5.10e- 1 -0.370 0.744   
## 2 gcs -0.216 0.0329 -6.56 5.44e-11 -0.280 -0.151   
## 3 dm2yes 0.909 0.273 3.33 8.63e- 4 0.374 1.44   
## 4 sbp 0.00130 0.00632 0.205 8.38e- 1 -0.0111 0.0137  
## 5 dbp -0.00196 0.00982 -0.200 8.42e- 1 -0.0212 0.0173

tblCox1 <- tbl\_regression(Cox1)  
tblCox1

## Table printed with {flextable}, not {gt}. Learn why at  
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html  
## To suppress this message, include `message = FALSE` in the code chunk header.

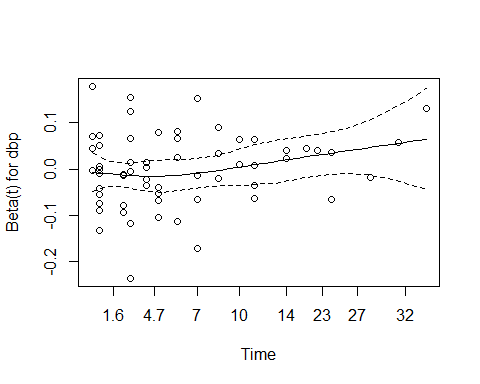
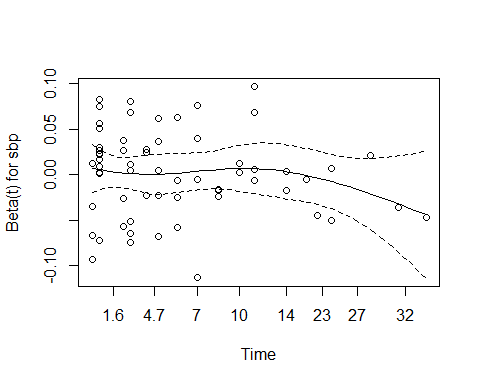
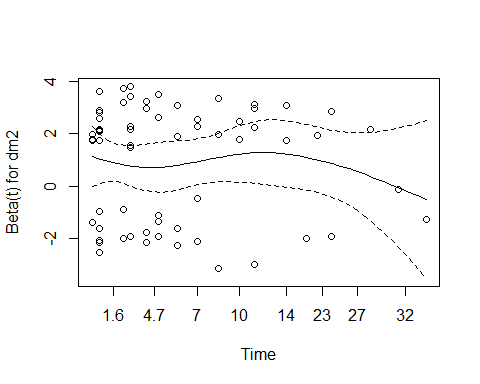
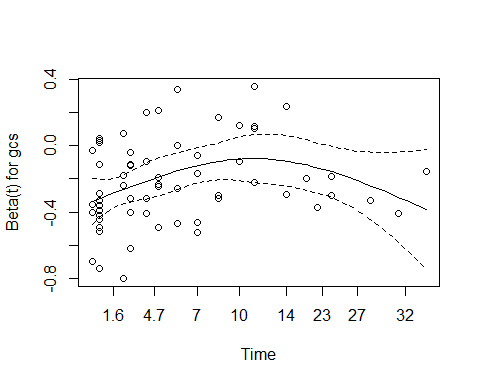
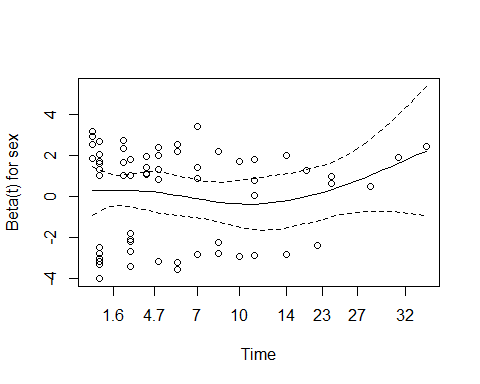
Table 1 : Cox Proportional Hazard Analysis (n=262)

| **Characteristic** | **log(HR)**1 | **95% CI**1 | **p-value** |
| --- | --- | --- | --- |
| sex |  |  |  |
| male | — | — |  |
| female | 0.19 | -0.37, 0.74 | 0.5 |
| earliest Glasgow Coma Scale | -0.22 | -0.28, -0.15 | <0.001 |
| diabetes? |  |  |  |
| no | — | — |  |
| yes | 0.91 | 0.37, 1.4 | <0.001 |
| earliest systolic BP (mmHg) | 0.00 | -0.01, 0.01 | 0.8 |
| earliest diastolic BP (mmHg) | 0.00 | -0.02, 0.02 | 0.8 |
| 1HR = Hazard Ratio, CI = Confidence Interval  Global constant regression coefficient = 0.306 | | | |

1. With every increase of 1 mmhg sbp, the log hazard changes by 0.0001297 (95% CI -0.01109, 0.01369), when adjusted for sex, gcs, dm and dbp.
2. Being diabetic, changes the log hazard by 0.9089 (95% Ci : 0.3742, 1.4436) compared to non-diabetic when adjusted for sbp, sex, gcs and dbp.

## d. Test the assumption for proportional hazard. Save the plot in the thumb-drive and write the name of the plot in the answer sheet. (5 marks)

phm <- cox.zph(Cox1, transform = 'km', global= T)  
plot(phm)



### Additional assumption test

cox.zph (Cox1, transform = 'rank')

## chisq df p  
## sex 7.30e-02 1 0.787  
## gcs 4.26e+00 1 0.039  
## dm2 6.36e-02 1 0.801  
## sbp 4.71e-06 1 0.998  
## dbp 1.69e-01 1 0.681  
## GLOBAL 6.00e+00 5 0.306

Except the gcs, all other individual covariate and global constant regression coefficient are more than 0.05. Hence when the significance level is set at 0.05, the proportional hazard is assumed