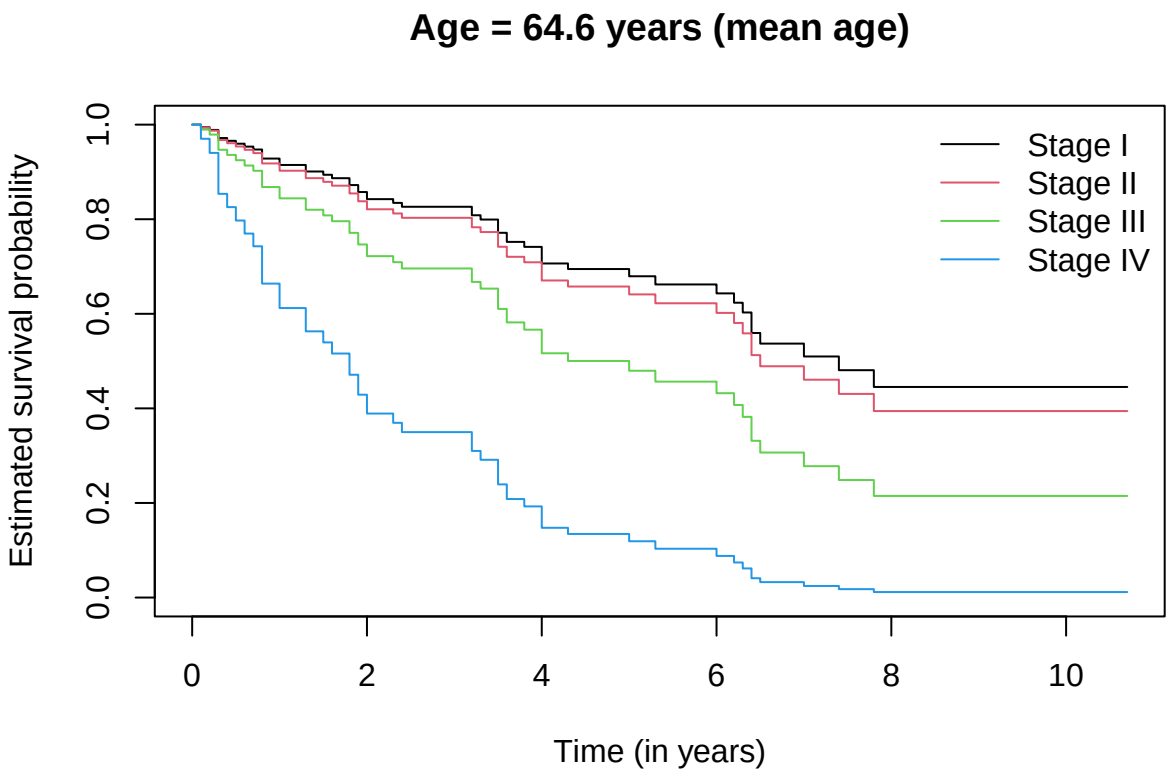
Biostatistics (MATH11230)

Vanda Inácio

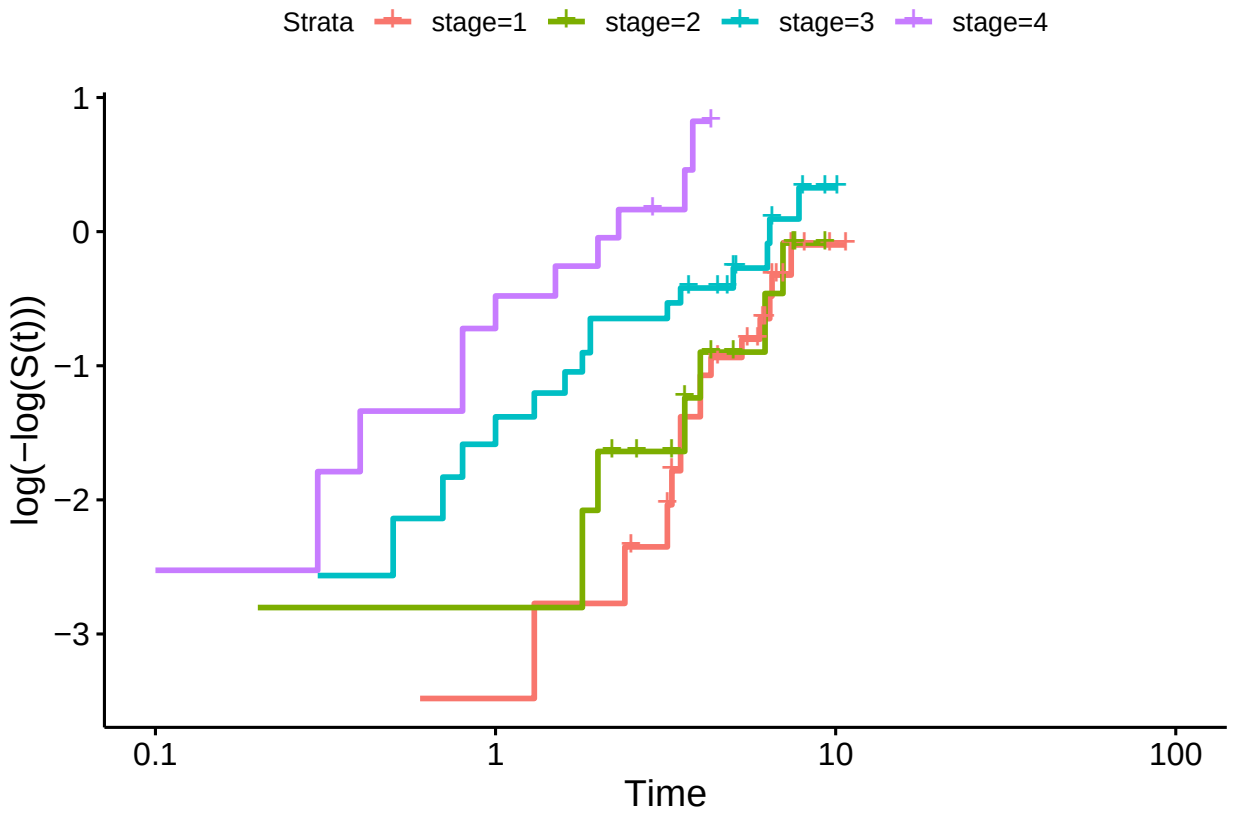
Here we reproduce the results presented in the slides.

```
require(survival)
require(KMsurv)
data(larynx)
res_1 <- coxph(Surv(time, delta) ~ as.factor(stage) + age, data = larynx)

#adjusted survival curve for the mean age
newdf <- data.frame("stage" = levels(as.factor(larynx$stage)),
                    "age" = rep(mean(larynx$age), 4))
fit <- survfit(res_1, newdata = newdf)

plot(fit, conf.int = FALSE, col = c(1:4),
     ylab = "Estimated survival probability",
     xlab = "Time (in years)",
     main = "Age = 64.6 years (mean age)")
legend("topright", legend = c("Stage I", "Stage II", "Stage III", "Stage IV"),
       col = c(1:4), lty = c(1,1,1,1), bty = "n")
```

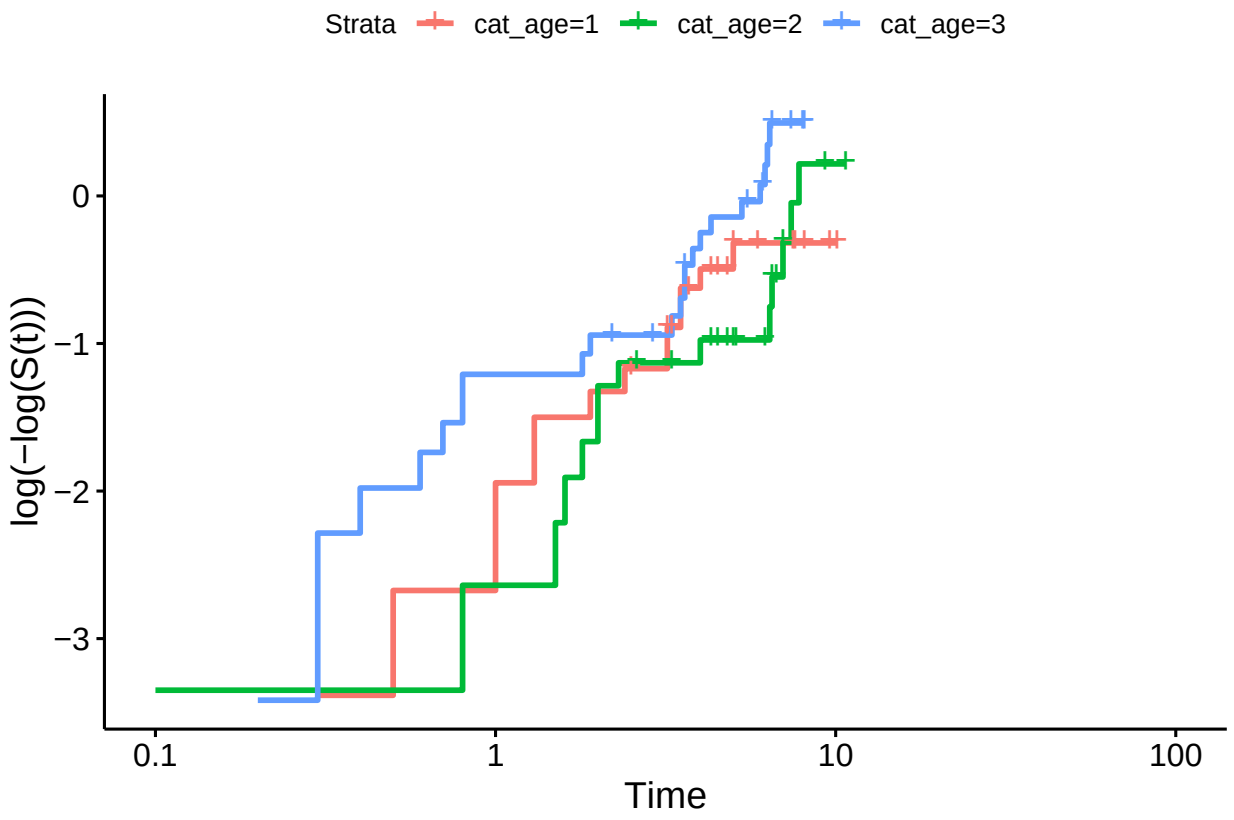


Below the code for the log cumulative hazard plot.

```
km_fit_stage <- survfit(Surv(time, delta) ~ stage, data = larynx)
require(survminer)
ggsurvplot(km_fit_stage, fun = "cloglog")
```

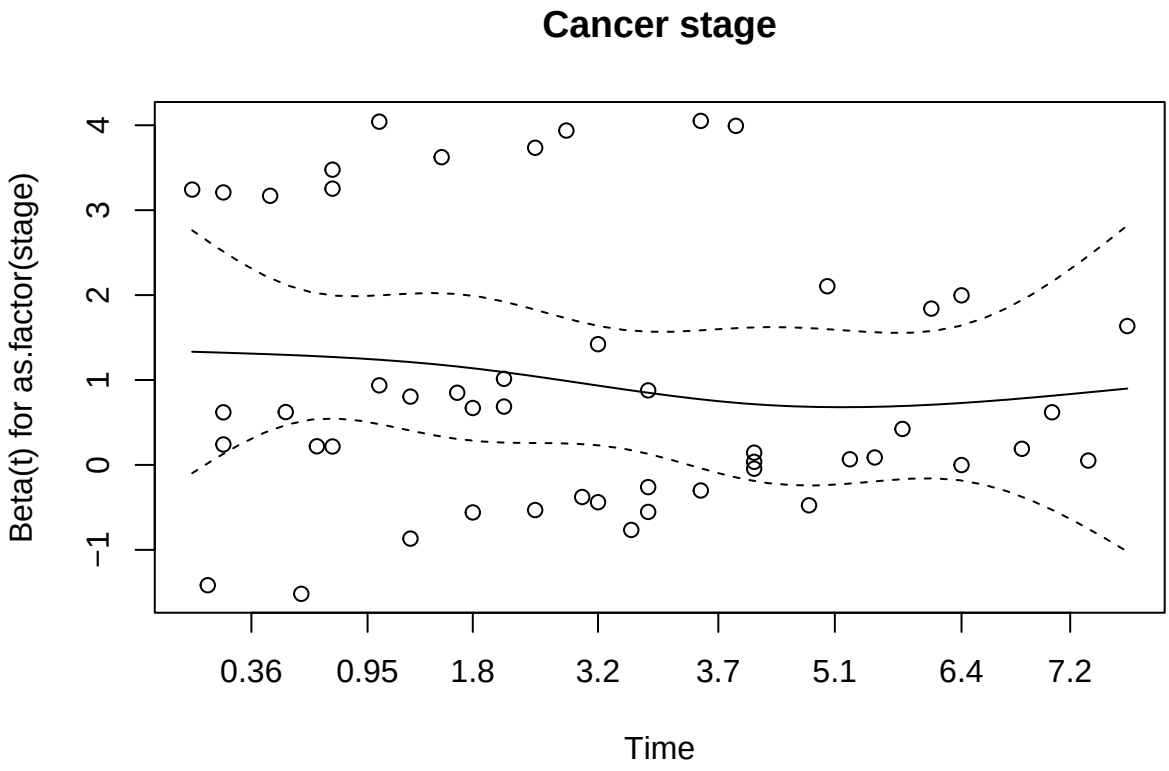



```
larynx$cat_age <- numeric(length(larynx$age))
for(i in 1:length(larynx$age)){
if(larynx$age[i] < quantile(larynx$age, 0.33)) {larynx$cat_age[i] <- 1}
else{
  if(larynx$age[i] >= quantile(larynx$age, 0.33) & larynx$age[i] < quantile(larynx$age, 0.66)) {larynx$
else {larynx$cat_age[i] <- 3}
}
}
larynx$cat_age <- as.factor(larynx$cat_age)
km_fit_age <- survfit(Surv(time, delta) ~ larynx$cat_age, data = larynx)
ggsurvplot(km_fit_age, fun = "cloglog")
```

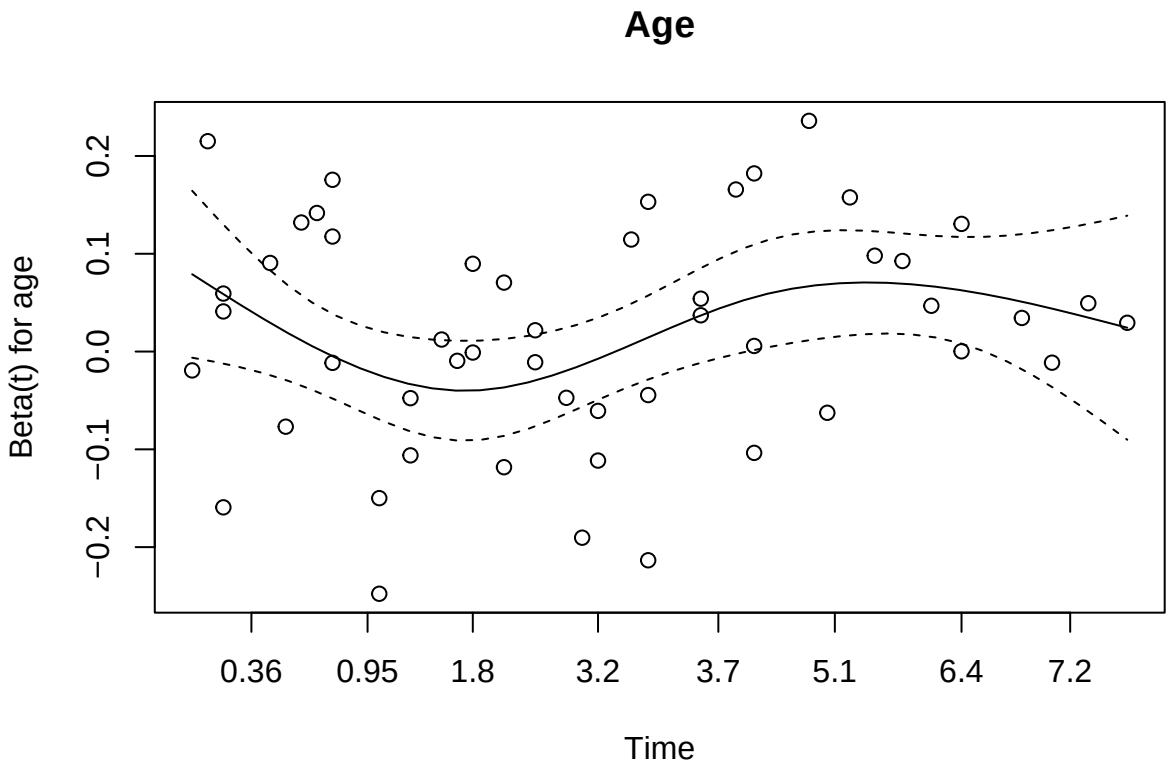


And finally the results for Schoenfeld residuals.

```
zph <- cox.zph(res_1)
plot(zph, var = 1, main = "Cancer stage")
```



```r
plot(zph, var = 2, main = "Age")
```




```r
zph
```

```
##                   chisq df    p
## as.factor(stage)  3.72  3 0.29
## age               1.13  1 0.29
## GLOBAL            5.14  4 0.27
```