Applied Survival Analysis
Lab 7: Assessing the PH Assumption

In today's lab, we are going to evaluate the assumption of proportional hazards using several graphical approaches. We will use the same example as in the lecture, the nursing home dataset (*nurshome.dta).*

First we are going to *stset* our data:

```
stset los, f(fail)
```

We want to assess the proportional hazards for gender and marital status. One way is to produce the plot of *log[-log(S(t))]* versus *log(t).* If the lines of the subgroups are parallel then the assumption is satisfied. STATA has already a command to produce this graph, the only difference is that it gives us *-log[-log(S(t))]* instead of *log[-log(S(t))]*. The way to produce this graph for gender and marital status is the following:
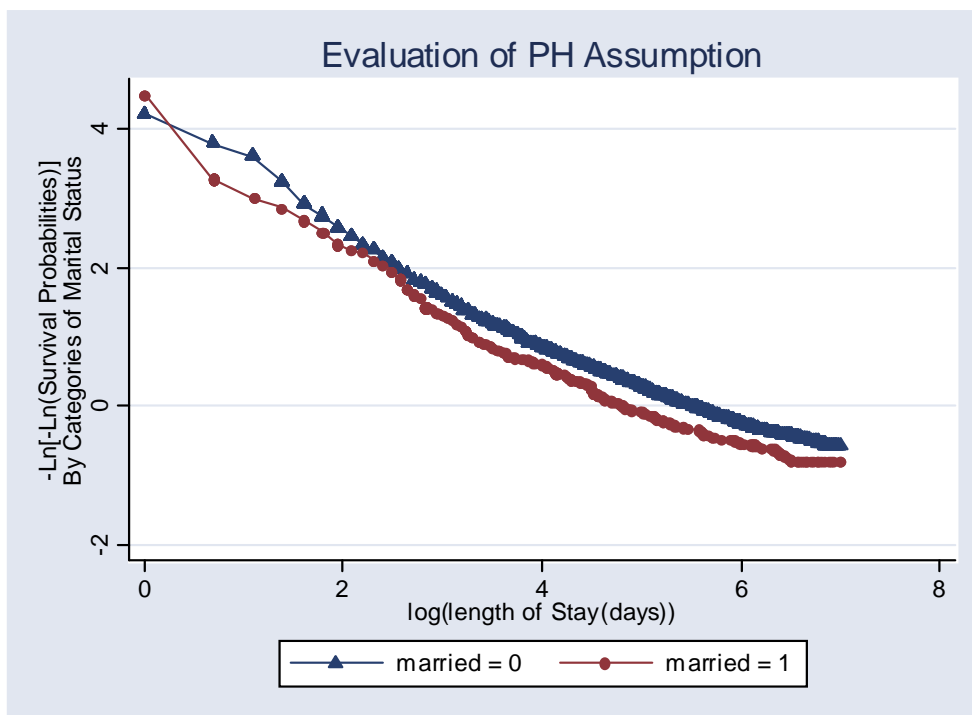For Gender:

```
stphplot, by(gender) c(l l) s(T o) title(Evaluation of PH Assumption)
xtitle(log(length of Stay(days))) ytitle("-Ln[-Ln(Survival
Probabilities)]" "By Categories of Gender")
```
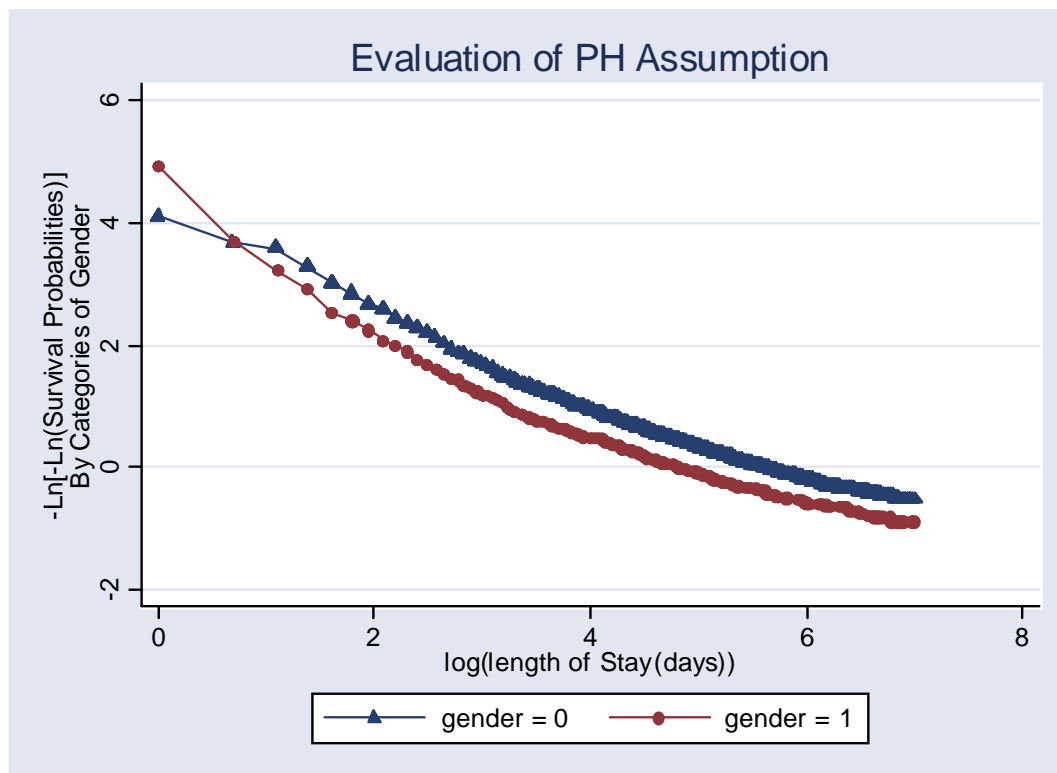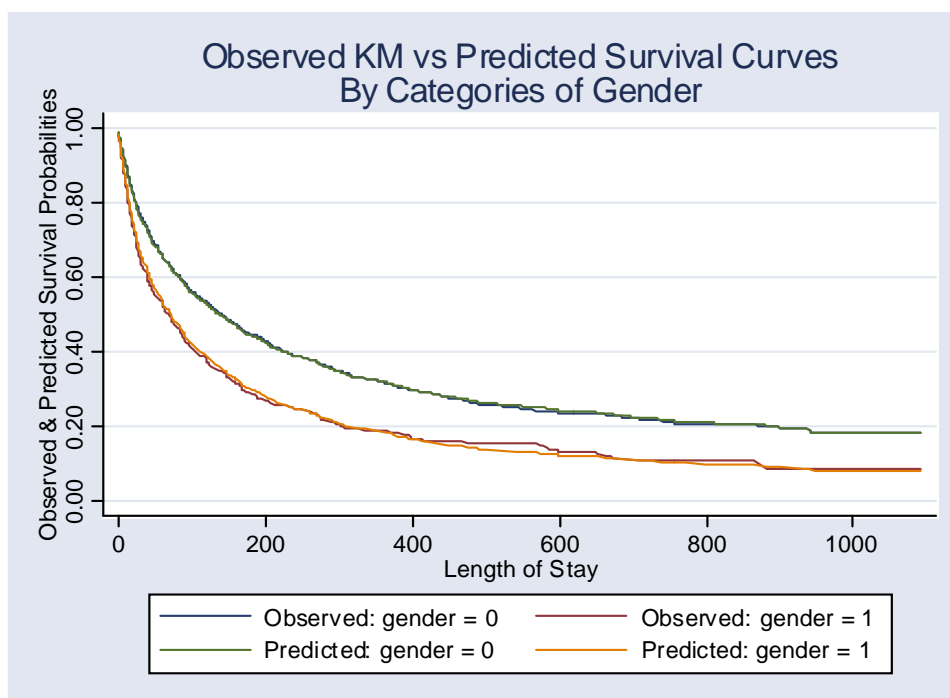
For Marital Status:

```
stphplot, by(married) c(l l) s(T o) title(Evaluation of PH Assumption)
xtitle(log(length of Stay(days)))ytitle("-Ln[-Ln(Survival
Probabilities)]" "By Categories of Marital Status")
```


STATA can also give you another informative graph with the command **stcoxkm** where it plots Kaplan-Meier observed survival curves and compares them to the Cox predicted curves for the same variable. The closer the observed values are to the predicted, the less likely the proportional hazards assumption has been violated.
So for gender we have:

```
stcoxkm, by(gender) c(l l l l) s(i i i i) title("Observed KM vs
Predicted Survival Curves" "By Categories of Gender") xtitle(Length of
Stay) ytitle("Observed & Predicted Survival Probabilities")
```

```
stcoxkm, by(married) c(l l l l) s(i i i i) title("Observed KM vs
Predicted Survival Curves" "By Categories of Marital Status")
xtitle(Length of Stay) ytitle("Observed & Predicted Survival
Probabilities")
```
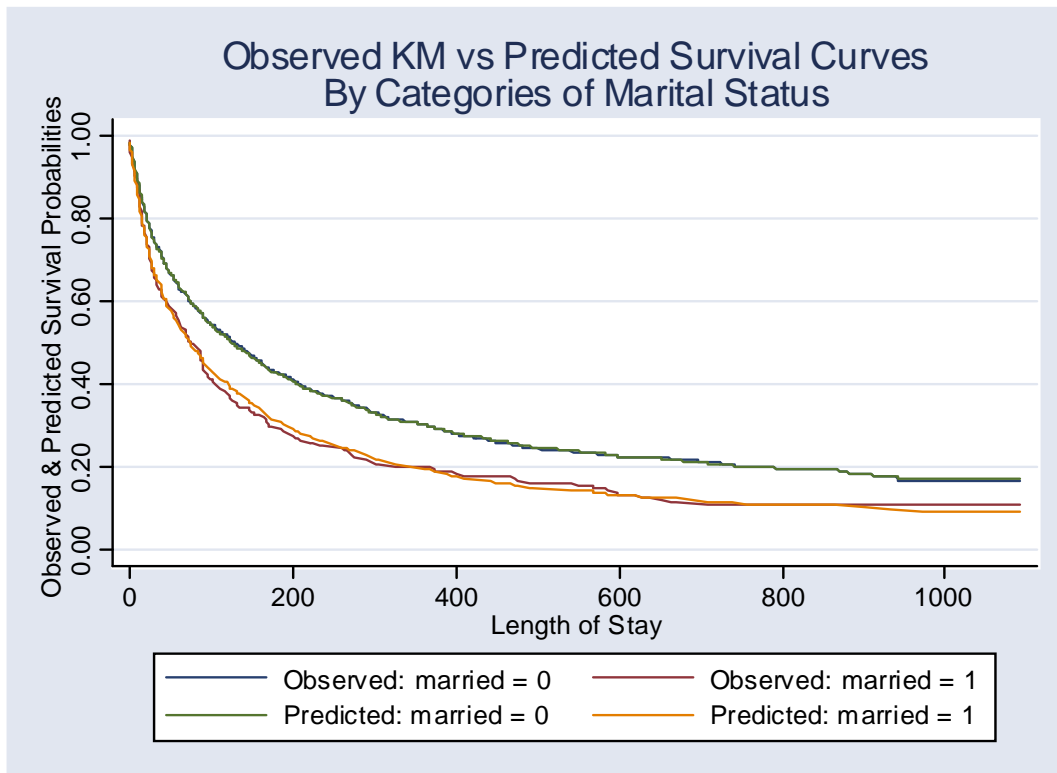

What can you say about the assumption of proportional hazards for gender and marital status?

Next if we have several covariates and want to assess the proportional hazards assumption, we can generate a new covariate that takes a different value for every combination of covariate values. This can be done only if there is enough data and a couple of covariates. In this case we want to generate another covariate that will include the combination of healthy and unhealthy men and women:

```
gen hlthsex=1 if gender==0 & health==2
(1331 missing values generated)

replace hlthsex=2 if gender==1 & health==2
(81 real changes made)

replace hlthsex=3 if gender==0 & health==5
(126 real changes made)

replace hlthsex=4 if gender==1 & health==5
(42 real changes made)

lab define hsfmt 1"Healthier Women" 2"Healthier Men" 3"Sicker Women"
4"Sicker Men"

lab val hlthsex hsfmt
```
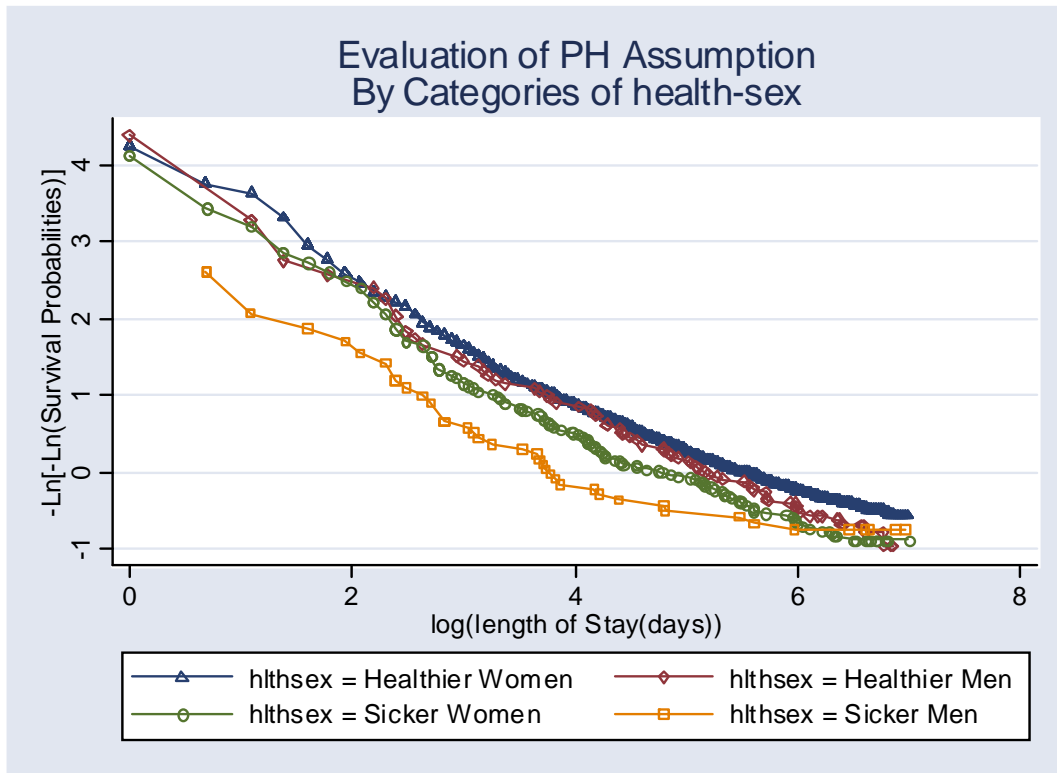
```
stphplot, by(hlthsex ) c(l l l l) s(th dh oh sh) title("Evaluation of
PH Assumption" "By Categories of health-sex") xtitle(log(length of
Stay(days))) ytitle("-Ln[-Ln(Survival Probabilities)]"))
```


Evaluation of PH Assumption
By Categories of health-sex

And the corresponding `stcoxkm` plot:

```
stcoxkm, by(hlthsex)  c(J J J J l l l l) s(i i i i i i i i)
title("Observed KM vs Predicted Survival Curves"  "By Categories of
health-sex") xtitle(Length of Stay)  ytitle(Observed & Predicted
Survival Probabilies) xscale(titlegap(3))  yscale(titlegap(3))
legend(size(small))
```


Observed KM vs Predicted Survival Curves
By Categories of health-sex
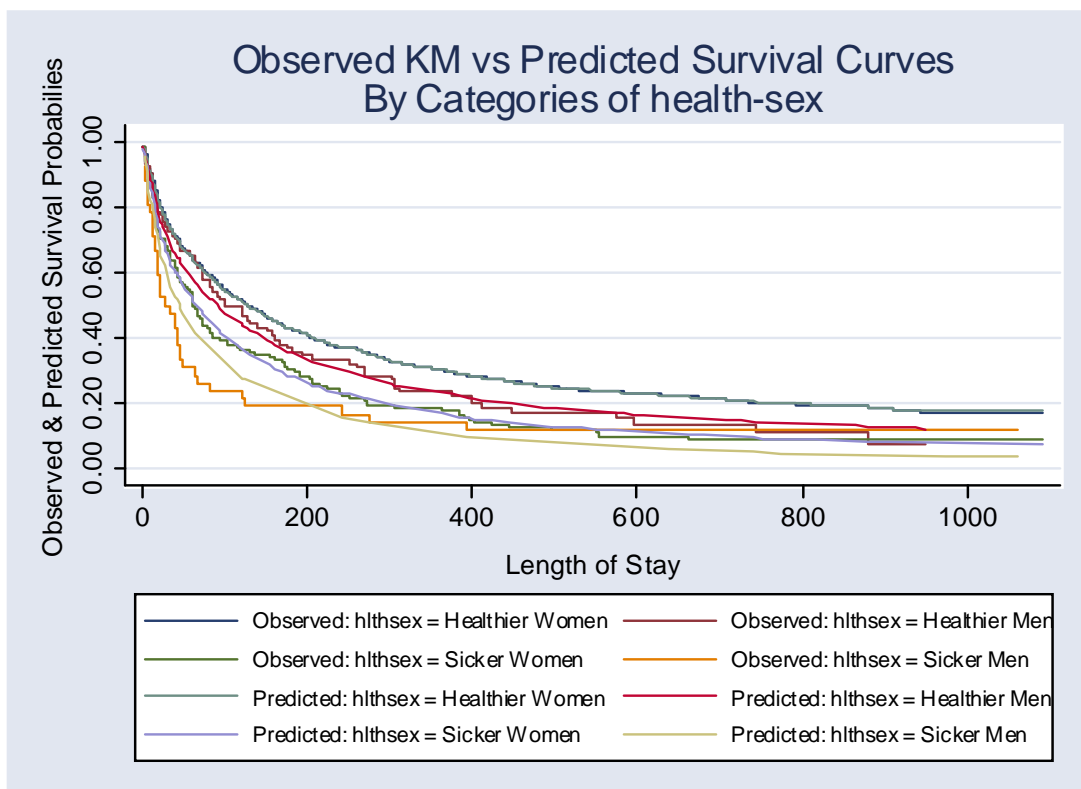
Now we want to assess the proportional hazards assumption within a stratified model. For example we want to stratify for gender and adjust for marital and health status. The **stphplot** for this model is the following:

```
stphplot, strata(gender) adjust(married health) c(l l) s(T o)
title("Evaluation of PH Assumption" "Stratification by Gender")
xtitle(log(length of Stay(days))) ytitle("-Ln[-Ln(Survival
Probabilities)]")
```