Solutions to tutorial 4c

[1] Here are the R commands to do the step-wise model building.

```r
library(survival);library(MASS)
attach(D)
S=Surv(week,arrest)
Scope=list(upper=~(age+factor(educ)+fin+mar+paro+prio+race+wexp),
   lower=~1)
phm_e = coxph(S~1)
phm_f = stepAIC(phm_e,Scope,direction="both")
detach(D)

The final part of the output is

Step:  AIC=1327.35
S ~ age + prio + fin + mar

<table>
<thead>
<tr>
<th>Df</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;none&gt;</td>
<td>1327.3</td>
</tr>
<tr>
<td>- mar</td>
<td>1 1327.7</td>
</tr>
<tr>
<td>+ race</td>
<td>1 1328.2</td>
</tr>
<tr>
<td>+ wexp</td>
<td>1 1328.8</td>
</tr>
<tr>
<td>- fin</td>
<td>1 1329.0</td>
</tr>
<tr>
<td>+ paro</td>
<td>1 1329.2</td>
</tr>
<tr>
<td>+ factor(educ)</td>
<td>4 1330.2</td>
</tr>
<tr>
<td>- age</td>
<td>1 1335.4</td>
</tr>
<tr>
<td>- prio</td>
<td>1 1336.2</td>
</tr>
</tbody>
</table>
```

so the variables that were selected were age, prio, fin and mar. These were not the same variables as obtained in the purposeful approach: important covariates have been missed out, such as education and work experience.
Here is one possible version of the purposeful approach. Note it is very possible that you obtained a different model by making your own decisions along the way. This is perfectly fine.

```r
D=read.table("http://www.sph.emory.edu/~dkleinb/surv2datasets/addicts.dat", header=TRUE)
attach(D)
#clinic, prison, dose
S=Surv(t,depart)
m1=coxph(S~clinic)
m2=coxph(S~prison)
m3=coxph(S~dose)

Prison is not significant, but clinic and dose are. Try the model with both clinic and dose:

m4=coxph(S~clinic+dose)

Both terms are significant. Try to include prison in this model:

m5=coxph(S~clinic+dose+prison)

Prison has become significant at the 20% level so we retain it.

Only dose needs to have its scale checked. This can be done using the following commands:

```r
m6 =coxph(S~clinic +I(dose^2) +prison)
m7 =coxph(S~clinic +dose+I(dose^2) +prison)
m8 =coxph(S~clinic +I(1/dose) +prison)
m9 =coxph(S~clinic +I(log(dose)) +prison)
m10=coxph(S~clinic +I(sqrt(dose)) +prison)
aic5 =2*(length(m5$coeff)-m5$logl[2])
```
A change of scale does nothing to improve the fit. Now attempt to add interactions:

\[
\begin{align*}
\text{m11} &= \text{coxph}(S \sim \text{clinic} \cdot \text{dose} + \text{prison}) \\
\text{m12} &= \text{coxph}(S \sim \text{clinic} + \text{dose} \cdot \text{prison}) \\
\text{m13} &= \text{coxph}(S \sim \text{clinic} \cdot \text{prison} + \text{dose})
\end{align*}
\]

None of these interactions are significant. The final model is model m5. We now assess the PHA:

\[
> \text{cox.zph}(m5, \text{transform}='\text{rank}', \text{global}=\text{FALSE})
\]

<table>
<thead>
<tr>
<th></th>
<th>rho</th>
<th>chisq</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>clinic</td>
<td>-0.2498</td>
<td>10.495</td>
<td>0.00120</td>
</tr>
<tr>
<td>dose</td>
<td>0.0905</td>
<td>1.096</td>
<td>0.29521</td>
</tr>
<tr>
<td>prison</td>
<td>-0.0462</td>
<td>0.322</td>
<td>0.57068</td>
</tr>
</tbody>
</table>

Clinic does not satisfy the PHA. We would extend the model or stratify over clinic to address this.