1 Local Tests

Often, one is interested in testing a hypothesis about a subset of the $\boldsymbol{\beta}$. The hypothesis is then $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10}$, where $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^t, \boldsymbol{\beta}_2^t\right)^t$. Here $\boldsymbol{\beta}_1$ is a $q \times 1$ vector of the $\boldsymbol{\beta}$'s of interest and $\boldsymbol{\beta}_2$ is the vector of the remaining $p - q\boldsymbol{\beta}$'s.

- Wald test -MPLE
- LR test
- score test -score

Example: On stage of cancer in a study of 90 males diagnosed with cancer of the larynx. Here we shall test the hypothesis that there is no difference in survival between patients with different stages of disease, adjusting for the age of the patient. Our test is based on the model with covariates : indicators of stage II, III, and IV disease, and patient's age at diagnosis.

- The local hypothesis of interest is $H_0 : \beta_2 = 0, \beta_3 = 0\ \beta_4 = 0$ against the alternative hypothesis that, at least, one of these β 's is nonzero.

```
proc phreg data = ex1;
class stage(desc);
  model time*death(0) =  age stage ;
  contrast "beta2=beta3=beta4=0" stage 1 0 0,
                                 stage 0 1 0,
                                 stage 0 0 1/TEST(WALD LR SCORE);
run;
```

- The CONTRAST statement enables you to specify a matrix, $\mathbf{L}$, for testing the hypothesis $\mathbf{L}\boldsymbol{\beta} = \mathbf{0}$.

Class Level Information				
Class	Value	Design Variables		
stage	4	1	0	0
	3	0	1	0
	2	0	0	1
	1	0	0	0

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
age		1	0.01890	0.01425	1.7589	0.1848	1.019	
stage	4	1	1.69333	0.42218	16.0876	<.0001	5.438	stage 4
stage	3	1	0.63815	0.35609	3.2116	0.0731	1.893	stage 3
stage	2	1	0.13842	0.46232	0.0896	0.7646	1.148	stage 2

Contrast Test Results									
	LR Statistics			Score Statistics			Wald Statistics		
Contrast	DF	Chi-Square	Pr > ChiSq	DF	Chi-Square	Pr > ChiSq	DF	Wald Chi-Square	Pr > ChiSq
beta2=beta3=beta4=0	3	15.4529	0.0015	3	20.5766	0.0001	3	17.6371	0.0005

- all tests lead to the same conclusion that there is significant differences in survival between patients with different stages of disease.

Q: Do different stages have different relative risk for different age, ie interaction between age and stage.

```
proc phreg data = ex1;
class stage(desc);
  model time*death(0) =  age stage age*stage;
  contrast "beta5=beta6=0" age*stage 1 0 0,
                                age*stage 0 1 0/TEST(WALD LR SCORE);
run;
```

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
age		1	-0.00256	0.02605	0.0096	0.9218	.	
stage	4	1	0.84702	2.42571	0.1219	0.7270	.	stage 4
stage	3	1	-0.12250	2.46833	0.0025	0.9604	.	stage 3
stage	2	1	-7.94614	3.67821	4.6670	0.0307	.	stage 2
age*stage	4	1	0.01367	0.03597	0.1445	0.7038	.	stage 4 * age
age*stage	3	1	0.01135	0.03745	0.0919	0.7618	.	stage 3 * age
age*stage	2	1	0.12025	0.05231	5.2853	0.0215	.	stage 2 * age

Contrast Test Results									
	LR Statistics			Score Statistics			Wald Statistics		
Contrast	DF	Chi-Square	Pr > ChiSq	DF	Chi-Square	Pr > ChiSq	DF	Wald Chi-Square	Pr > ChiSq
beta5=beta6=0	2	0.1612	0.9226	2	0.1614	0.9224	2	0.1613	0.9225

- The effect of stage II on survival may be different for different ages because a local test of $\beta_7 = 0$ may be rejected.
- Furthermore, it is suggested by the local tests of $\beta_5 = 0$ and $\beta_6 = 0$ that the effects of stages III and IV on survival may not be different for different ages. To test the hypothesis that $\beta_6 = \beta_7 = 0$: effects of stages III and IV are the same for different age.
- Other local tests can be carried out similarly.

2 Model Building/Selection

A major question when we have collected many covariates/risk factors, is how to incorporate these covariates in the modeling procedure.

- to adjust for potential confounding (or explanatory) variables when one has a specific hypothesis in mind and the desire is to compare two or more groups with respect to survival times
  - if one has a particular hypothesis in mind, then interest centers upon that particular hypothesis and any model building will be done to adjust that particular comparison (or comparisons) for other noncontrollable factors. Often, the other explanatory factors are simply viewed as confounders and interest in them matters only insofar as they affect the assessment of the basic hypothesis.
- to predict the distribution of the time to some event from a list of explanatory variables with no particular prior hypothesis in mind. Utilizing the proportional hazards model introduced

Model selection criterion

- p-value associated with tests
- AIC $= -2 \log L + kp$: balance model fitting + parsimonuous

Use Bone Marrow Transplant from last Lab to illustrate forward selection:

- Acute leukemia patients being given a bone marrow transplant: To adjust the basic comparisons of the three risk groups, acute lymphoblastic leukemia (ALL), low-risk acute myeloctic leukemia (AML low-risk), and high- risk acute myeloctic leukemia (AML high-risk), so as to reduce the possible bias which may exist in making those comparisons.
  - Step 1: global test of the hypothesis of no difference in disease-free survival between three groups, last lab: there is significant difference.
  - Step 2:
    - Always contain group indicator in the model.
    - fit a model with 1 covariate + group,do this for all remaining covariates
    - find the covariate with smallest AIC, add it to the model
  - Step 3: Repeat step 2 until no significance is found, then we stop and base our inference about the primary hypothesis on the last model.

```
proc phreg data = bone_marrow1;
  model t2*dfree(0)= g2 g3 z7 ; /*waiting time is z7*/
  z7: test z7 = 0;
  ods output  FitStatistics = z7aic;
  ods output TestStmts = z7stat;
run;
proc phreg data = bone_marrow1;
  model t2*dfree(0)= g2 g3 z8 ; /*fab is z8*/
```

```
  z8: test z8 = 0;
  ods output  FitStatistics = z8aic;
  ods output TestStmts = z8stat;
run;
proc phreg data = bone_marrow1;
  model t2*dfree(0)= g2 g3  z10 ; /*MTX is z10*/
  z10: test z10 = 0;
  ods output  FitStatistics = z10aic;
  ods output TestStmts = z10stat;
run;
proc phreg data = bone_marrow1;
  model t2*dfree(0)= g2 g3 z3 z4 z3xz4 ; /*SEX is z3 and z4*/
  sex: test z3 = z4 = z3xz4 =0;
  ods output  FitStatistics = sexaic;
  ods output TestStmts = sexstat;
run;
proc phreg data = bone_marrow1;
  model t2*dfree(0)= g2 g3 z5 z6 z5xz6 ; /*CMV status is z5 and z6*/
  cmv: test z5 = z6 = z5xz6 =0;
  ods output  FitStatistics = cmvaic;
  ods output TestStmts = cmvstat;
run;
proc phreg data = bone_marrow1;
  model t2*dfree(0)= g2 g3 z1 z2 z1xz2 ; /*age is z1 and z2*/
  age: test z1 = z2 = z1xz2 =0;
  ods output  FitStatistics = ageaic;
  ods output TestStmts = agestat;
run;

data aic_table8_4;
  set z7aic z8aic z10aic sexaic cmvaic ageaic;
run;
proc print data = aic_table8_4 noobs;
where criterion = "AIC";
var criterion withcovariates;
run;
data stat_table8_4;
  set z7stat z8stat z10stat sexstat cmvstat agestat;
run;
proc print data = stat_table8_4 (drop=status) noobs;
run;
```

Criterion	WithCovariates
AIC	737.947
AIC	731.020
AIC	737.348
AIC	741.440
AIC	743.105
AIC	733.181

Label	WaldChiSq	DF	ProbChiSq
z7	1.1816	1	0.2770
z8	8.0815	1	0.0045
z1	2.0256	1	0.1547
se	1.9097	3	0.5914
cm	0.1861	3	0.9798
ag	11.9778	3	0.0075

- Pick z8, repeat

Criterion	WithCovariates
AIC	731.678
AIC	731.057
AIC	736.114
AIC	736.999
AIC	725.982

Label	WaldChiSq	DF	ProbChiSq
z7	1.1811	1	0.2771
z1	2.0492	1	0.1523
se	0.9240	3	0.8196
cm	0.0218	3	0.9992
ag	13.0528	3	0.0045

- Pick age, repeat

Criterion	WithCovariates
AIC	727.480
AIC	726.580
AIC	730.610
AIC	731.420

Label	WaldChiSq	DF	ProbChiSq
z7	0.4648	1	0.4954
z1	1.4438	1	0.2295
se	1.3675	3	0.7132
cm	0.5772	3	0.9016

- All the local tests are nonsignificant and that the AIC is larger than the last best one, the model building process stops.
- Lastly, re-run and analyze the final model selected to finish the model building process.

Only discussed forward selection procedure for when we have a scientific hypothesis in mind. Backward selection procedure which starts with the model with all factors, and, at each step, removes the least significant factor from the model. A stepwise model selection procedure combines the forward and backward procedures. All automatic selection is available in SAS.