Faculty of Medicine
Department of Cancer Research and Molecular Medicine

Exam KLMED8005

Medical Statistics - II

Thuesday June 5th 2012, 9.00 am – 1:00 pm

ECTS credits: 7.5
Allowed examination support (code A):
Calculator, all written and printet aids.
No. pages (including front page): 8

Contact person during the exam:
Grethe Albrektsen, mob. 954 98 743

Exam results: June 26th 2012
Examination results are announced on http://studweb.ntnu.no/

IMPORTANT

- **Read the exercises carefully before answering.**
- **Your answers should be short and precise.**
- **Remember to mark your answers with ordered numbers corresponding to exercise. You do not need to repeat the text to exercises in your answers.**

EXERCISE 1.

Figure 1 below show mean values (given by height of bar) of arterial stiffness by age (3 groups; young, middle-aged, older) and amount of light physical activity (PA) groups (2 groups; high, low). The result is obtained from a recent study (Gando et al, Hypertension, 2012) that examined associations between arterial stiffness and amount of light physical activity in 538 healthy adults (men and women).

Arterial stiffness is assumed to be a risk factor for cardiovascular disease, and is in this study measured by carotid-femoral pulse wave velocity (cfPWV; in cm/s; increasing values means increasing arterial stiffness).

Duration and intensity of physical activity (PA) was assessed by a triaxial accelerometer, something that registers movements (acceleration) in different directions. Data were recalculated into intensity levels. A previous validation study showed high correlation with measurements of oxygen uptake (VO_2). The device was worn at day-time during a 14-day period. Daily time (minutes) spent with light physical activity was recorded for each person (continuous scale), and categorizing into high/low PA group were defined according to whether a person had value higher or lower than median value for persons at the same age and sex.

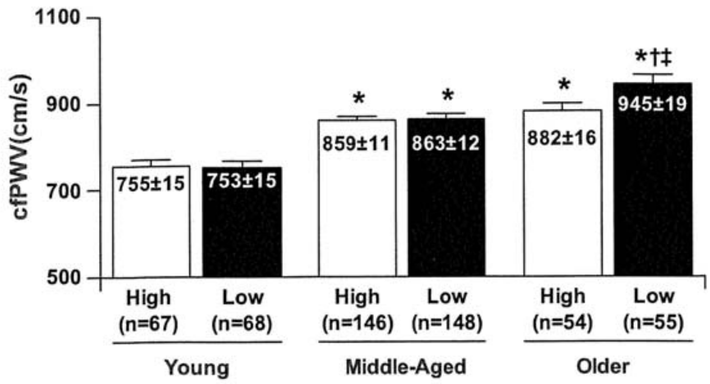


Figure 1. Arterial stiffness in high-light PA and low-light PA groups.

a) A two-way analysis of variance (ANOVA) was applied to examine whether mean level of arterial stiffness differed by age and/or amount of light physical activity.

   According to Figure 1

   – does it seem to be any association between age and arterial stiffness?

   – does it seem to be any association between amount of light PA and arterial stiffness?

   (give a short explanation for your answers)

b) Two alternative ANOVA models can be defined to examine, and formally test, associations between age, light physical activity and arterial stiffness.

   – define the two alternative models in terms of parameters to be tested

   – give a short description of what the different parameters in the model represent (interpretation of it).

   – describe the difference in interpretation of results from the two different models.

c) How can you choose between the two alternative models, and what do you think would be the most suitable model in this case?

d) What is the assumption(s) of the ANOVA model (in general).

e) What can you do if the assumptions on the model are not met?

EXERCISE 2.

In the same study (Gando et al, 2012), a simple linear regression analyses was performed for each age group to examine association between amount of light physical activity (minutes per day, continuous scale) and arterial stiffness (measured by cfPWV, cm/s) according to age.

Figure 2 below show a scatter-plot between the variables in the regression model (amount of light physical activity and arterial stiffness). A significant association was found, but only among the oldest (> 60 years). Results from the linear regression analysis (estimated regression coefficients, regression line), together with Pearson's correlation coefficient (r), is shown for this group.

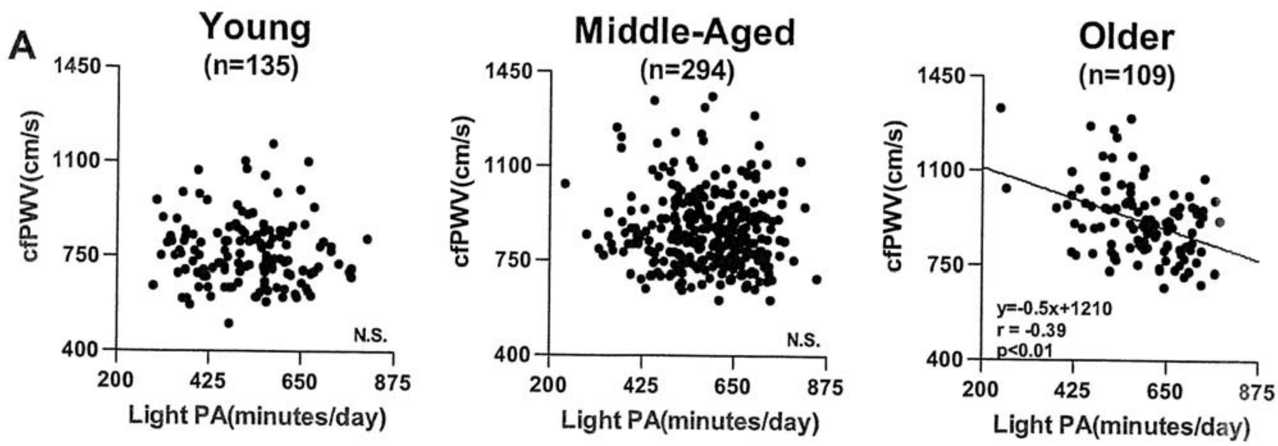


Figure 2. Relationships between daily time spent in each PA intensity and cfPWV.

a) How do you interpret the estimated values of the (unstandardized) regression coefficients (describe in terms of variables in the model)?

b) Is it a strong linear association between amount of light physical activity and arterial stiffness among the oldest? Give an explanation for your answer.

c) What are the assumptions on the linear regression model?
   - describe in terms of the variables in the model.

d) A multiple linear regression analysis was performed to evaluate whether gender was a potential confounder for the association between amount of light physical activity and arterial stiffness.
   Gender did not seem to be a confounder.
   - how could the author come to this conclusion, based on results from the linear regression analysis?

EXERCISE 3.

In another study, the main aim was to evaluate whether prognosis of colorectal cancer differed between men and women. Kaplan-Meier survival plot and Cox proportional hazard (PH) regression model was applied. The study comprised a total of 247 patients (105 men and 142 women).

The data analysis (Kaplan-Meier method) gave the following results (selected parts of SPSS-output):

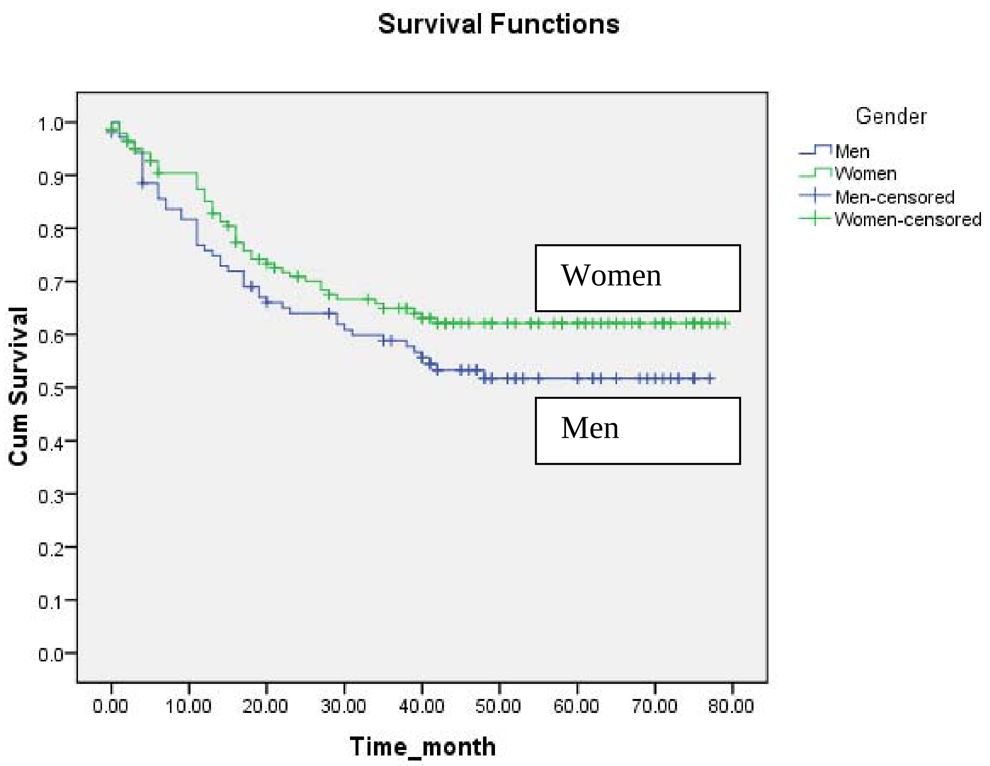
Case Processing Summary

Gender	Total N	N of Events	Censored	
			N	Percent
Men	105	48	57	54.3%
Women	142	48	94	66.2%
Overall	247	96	151	61.1%

Overall Comparisons

	Chi-Square	Df	Sig.
Log Rank (Mantel-Cox)	2.276	1	.131

Test of equality of survival distributions for the different levels of Gender.



a) Based on the Kaplan-Meier survival plot
   - what is the estimated 50% (median) survival time (approximately) in men and women, respectively
   - what is the estimated 75% survival time (approximately) in men and women, respectively
   - what is the estimated proportion of survivors (approximately) after 2.5 years (30 months) in men and women, respectively

b) Specify the null hypothesis and the alternative hypothesis for the log-rank test.

c) Describe result from the survival analysis, based on information in the SPSS output above.

EXERCISE 4

The colorectal cancer data was further analysed in a Cox proportional hazard (PH) regression model.

The SPSS output below show results from an unadjusted analyses (corresponding to Kaplan-Meier survival plot).

Categorical Variable Codings[b]

		Frequency	(1)
Gender[a]	1=Men	105	1
	2=Women	142	0

a. Indicator Parameter Coding

b. Category variable: Gender

Variables in the Equation

	B	SE	Wald	Df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
Gender	.305	.204	2.225	1	.136	1.356	.909	2.023

An analysis adjusted for age at diagnosis (1-year interval, continuous variable) and stage of disease (Duke's classification system, categorical variable), gave the following result (SPSS output, Model 1)

Variables in the Equation

	B	SE	Wald	Df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
Gender	.434	.209	4.326	1	.038	1.544	1.025	2.325
Age_diag	.001	.009	.009	1	.922	1.001	.983	1.019
Duke			140.374	4	.000			
duke(1)	-2.570	.472	29.603	1	.000	.077	.030	.193
duke(2)	-2.078	.343	36.771	1	.000	.125	.064	.245
duke(3)	-.843	.320	6.939	1	.008	.430	.230	.806
duke(4)	1.636	.320	26.066	1	.000	5.135	2.740	9.624

a) Describe results (for gender only) from the Cox PH regression analyses, based on information in the SPSS output. Report

– estimated value (point- and interval estimate) of hazard ratio

– results from the statistical test

b) What is meant by (implies) the proportionality assumption in the Cox PH regression model?

c) How can you check whether the proportionality assumption is met?

Additional analyses with age included as a categorical variable (10-yr intervals, <60, 60-69, 70-79, 80+ years; youngest patients as reference group) was also performed. This model gave the following result (SPSS output, Model 2)

Variables in the Equation

	B	SE	Wald	Df	Sig.	Exp(B)	95,0% CI for Exp(B) Lower	Upper
Gender	.407	.212	3.699	1	.054	1.502	.992	2.275
Duke			138.384	4	.000			
duke(1)	-2.655	.474	31.379	1	.000	.070	.028	.178
duke(2)	-2.234	.352	40.210	1	.000	.107	.054	.214
duke(3)	-.989	.327	9.123	1	.003	.372	.196	.707
duke(4)	1.562	.330	22.410	1	.000	4.770	2.498	9.108
Age_10yr			5.687	3	.128			
Age_10yr(1)	-.171	.285	.359	1	.549	.843	.482	1.474
Age_10yr(2)	-.328	.281	1.358	1	.244	.721	.415	1.250
Age_10yr(3)	.446	.338	1.744	1	.187	1.562	.806	3.027

d) How do you interpret the value of the hazard ratio(s) for age in Model 1 (age as continuous variable) and Model 2 (age as categorical variable), respectively?

e) What do you think would be the best model in this case; age included as a continuous variable (1-year interval, Model 1) or as a categorical variable (10-year interval, Model 2)? Give an explanation for your answer.

NB:
In the exercise above, assume that it is necessary to adjust for age, despite the lack of significant prognostic impact, and assume that the proportionality assumption is met for all variables in the model.