Generalized Linear Models (ST425A)
(Advanced level course, 7.5 hec, Aut. 2019)

Examnation (Part 2)

Gebrenegus Ghilagaber (Professor & Head)
Department of Statistics, Stockholm University

- **Date and time:** Monday 30 September 2019, 13:00 - 15:00
- **Permitted facilities:** Pocket calculator and attached R- and SAS-codes.
- **Return of exam:** Not yet decided (information will be sent via e-mail or athena).
- **Instructions:**
  - The total amount of points for this part of the exam is 20.
  - The minimum requirement to pass this part of examination is 10 points.
  - Solutions to each question should be detailed enough and well-motivated in order to get full points.

The table below presents the results of a survey on ideal number of children (grouped into 3 increasing categories) among 3663 women cross classified by their education (3 levels) and residence (2 levels).

		Ideal # children		
Education	**Residence**	≤ 2	$3-5$	>5
None	Urban	218	200	248
	Rural	731	643	532
Primary	Urban	220	143	121
	Rural	134	44	18
Secondary	Urban	250	94	35
or above	Rural	22	6	4

- a) Fit an appropriate model with *Ideal # children* as response variable and *Education* as explanatory variable.
- b) Repeat (a) with *Education* and *Residence* as explanatory variables.
- c) Use the model-deviances in (a) and (b) to suggest which model is more adequate.
- d) Do the results in (a) and (b) indicate you need to add an interaction term (between *Education* and *Residence*) to the model in (b)?
- e) Interpret your final results on the effect of *Education* and *Residence* on ideal children ever born

Summarize your results in a form of a report that includes choice of a model (with justification), the fitted model, and overall comments on your results (estimates and test statistics). Attach relevant SAS or R codes, tables and figures as appendices.

Generalized Linear Models 2019: R- and SAS Commands

Packages Used
car
ResourceSelection
aod
nnet
MASS
VGAM
vcd
mcprofile
lmtest

Description	Command
Simple/multiple linear regression	lm(formula, data)
One-way / Two-way ANOVA	aov(formula , data)
Type I SS	anova(model)
Type II or III SS	Anova(model , type="II") Anova(model , type="III")
Binary regression	glm(cbind(y,n-y)~ , family=binomial(link="logit"),data) link=c("logit ","probit ","cloglog")
Odds ratio estimates (model with interaction)	predict.glm(model, newdata=data.frame(x1=c(),x2=factor()),se.fit=TRUE)
Multinomial logistic regression	multinom(formula,data,weights)
Proportional odds models	polr(formula,data,weights,Hess=TRUE,method="logistic") method = c("logistic", "probit", "cloglog", "cauchit") vglm(formula=formula, family=cumulative (link="logitlink",parallel=T, reverse=F), weights,data)
Adjacent-Category model (with proportional odds assumption)	vglm(formula, family=acat (link="loglink", parallel=T, reverse=F), weights,data)

Continuation ratio (with proportional odds assumption)	vglm(formula,family=sratio(link="loglink",parallel=T,reverse=F), weights,data)
Log-linear models	glm(formula, family=poisson(link="log"), data)
Rate ratio + CI	exp(confint(mcprofile(model , CM=matrix(C,1))) where C is the contrast matrix
Standardized residuals	rstandard(model)
Jack-knife/studentized residuals	rstudent(model)
Cooks distance	cooks.distance(model)
Leverages	hatvalues(model)
Pearson residuals	residuals(model,type="pearson")
Deviance residuals	residuals(model,type="deviance")
Akaike's information criterion	AIC(model)
Bayesian information criterion	BIC(model)
Log likelihood	logLik(model)
deviance	deviance(model)
Likelihood ratio test	anova(model1,model2, test="Chisq")
Shapiro-Wilk Test	shapiro.test()
Breush Pagan Test	bptest(model)
Tests for homogeneity of variance across groups	fligner.test(y~group, data) leveneTest(y~group, data)
Hosmer-Lemeshow's goodness of fit	hoslem.test(y, fitted(model), g=)
Wald test	wald.test(b = coef(model), Sigma = vcov(model), Terms =)
Testing for single/joint effects of covariates/factors	linearHypothesis(M1,c("coef1"," coef2"),test="Chisq") linearHypothesis(M1,c("coef1=0"),test="Chisq") linearHypothesis(M1,c("coef1"=" coef2"),test="Chisq")
Pearson chi-Squared test for independence	assocstats(H)
Confidence Intervals for parameter estimates	confint() confint.default()

Relevel/order factors	relevel(x, ref="") factor(x,levels=unique(x)) factor(x,ordered=T)

Description	Command

Simple/multiple linear regression
 	
anova
	proc glm data = data plots=diagnostics;
	model formula/ SS1 SS2 SS3 solution; means x;
	lsmeans x / adjust = tukey cl;
	output out = estim p= phat r=resid student = stresid
	cookd=cooks dffits = DFIT h = hat_matrix;
	
	class A (ref = 'A1') B (ref = 'B1');

Binary regression
 	
Odds ratio estimates (model with interaction)
	proc logistic data = *dataset;*
	model *formula* /link=LOGIT (orPROBIT or CLOGLOG)
	aggregate scale = none influence;
	effectplot fit/obs;
	output out=*tablename* pred=*logit (or probit or pcloglog)*
	resdev = resdev reschi= Pearson stdresdev = stdresdev
	stdreschi = stdPearson; run;
	
	EFFECTPLOT FIT (PLOTBY = *name of CLASS variable*)/obs;
	ODDSRATIO name of class var;

Multinomial logistic regression
 	
cumulative logit model
 	
Adjacent-Category model
(with proportional odds assumption)
 	
Continuation ratio (with proportional odds assumption)
	ods graphics on;
	proc logistic data=data;
	freq frequency;
	class var1(ref='' param=ref) var2(ref='' param=ref);
	model response (ref='baseline')=var1var2/ link = glogit;
	effectplot interaction (plotby=var1)/polybar;
	oddsratio var;
	run;
	
	model / LINK = clogit scale=none aggregate;
	effectplot interaction(plotby=covariate);
	output out=pred predprobs= i predprobs = c;

Contingency table
 	
Log-linear models
	proc freq data = dataset order = data;
	weight n;
	tables var1*var2 / chisq expected nopercent nocol nocum norow; run;
	proc genmod data = data;

Rate ratio + CI	class var1 (ref = 'reference group name') var2 (ref = ' reference group name");
	model formula / dist = poisson obstats type3;
	ods output obstats = test1 (keep = frequency type site pred Xbeta std Reschi Resdev); run;
	
	estimate 'RR' var1 1 0 -1 0 var1*var2 1 0...0 /e ;

Standardized residuals Jack-knife/studentized residuals Cooks distance Leverages Pearson residuals Deviance residuals	
Akaike's information criterion Bayesian information criterion Log likelihood deviance	
Likelihood ratio test Shapiro-Wilk Test Breush Pagan Test Tests for homogeneity of variance across groups Hosmer-Lemeshow's goodness of fit Wald test Testing for single/joint effects of covariates/factors Pearson chi-Squared test for independence	proc logistic data = data; model formula/ link=logit lackfit; run;
Confidence Intervals for parameter estimates Relevel/order factors	