4/11 Part I

Ingen anmäld, ingen tenta arkiverad

Generalized Linear Models (ST425A)
(Advanced level course, 7.5 hec, Aut. 2019)
Re-Examnation (Part 2)

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

- **Date and time:** Monday 4 November 2019, 14:00 - 16:00
- **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 values of response variable ($\mathbf{y}$) and two explanatory variables ($\mathbf{x_1}$ and $\mathbf{x_2}$) on 20 observations. The table is also available in text and excel formats.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
y_i	1	10	29	25	29	40	21	0	13	4	0	7	21	9	7	22	6	2	29	11
x_{1i}	46	74	89	77	84	89	68	70	60	55	35	51	87	83	68	84	74	73	84	91
x_{2i}	0	0	16	16	21	15	14	6	13	9	3	7	23	4	0	19	3	0	15	7

- a) Begin by plotting all possible pairs of variables to get an idea on associations between them. Describe your results (plots)
- b) Fit appropriate model to assess the relationship of Y and X_1 and interpret the resulting parameter estimates. What assumption did you make to fit your chosen model?
- c) Extend your model in (b) by adding X_2 and interpret your output (estimated parameters and associated measures of fit)
- d) Compute the Gross and Net Effects of the explanatory variables (X_1 and X_2) on the response variable
- e) Use some form of diagnostics on your model in (c) and examine if any observation(s) looks to be an outlier

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 Rate ratio + CI	glm(formula, family=poisson(link="log"), data) exp(confint(mcprofile(model , CM=matrix(C,1))) where C is the contrast matrix
Standardized residuals Jack-knife/studentized residuals Cooks distance Leverages Pearson residuals Deviance residuals	rstandard(model) rstudent(model) cooks.distance(model) hatvalues(model) residuals(model,type="pearson") residuals(model,type="deviance")
Akaike's information criterion Bayesian information criterion Log likelihood deviance	AIC(model) BIC(model) logLik(model) deviance(model)
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	anova(model1,model2, test="Chisq") shapiro.test() bptest(model) fligner.test(y~group, data) leveneTest(y~group, data) hoslem.test(y, fitted(model), g=) wald.test(b = coef(model), Sigma = vcov(model), Terms =) linearHypothesis(M1,c("coef1"," coef2"),test="Chisq") linearHypothesis(M1,c("coef1=0"),test="Chisq") linearHypothesis(M1,c("coef1"=" coef2"),test="Chisq") 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	