Masters Comprehensive Exam
Stat 505-506 August 2014
100 points

Write your answers on separate sheets of paper with your name on each sheet.

1. Use the linear model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2\mathbf{V})$.
   Do not assume $\boldsymbol{X}$ is of full column rank.

   (a) What is the sampling distribution of the BLUE of $\boldsymbol{\lambda}^\mathsf{T}\boldsymbol{\beta}$?
   – This question might not make sense for all $\boldsymbol{\lambda}$. What qualifier is needed? (10 pts)

   (b) Assuming testability of $\boldsymbol{\lambda}^\mathsf{T}\boldsymbol{\beta} = 0$, write out the variance of $\boldsymbol{\lambda}^\mathsf{T}\widehat{\boldsymbol{\beta}}$ and explain how to use it to test $H_0: \ \boldsymbol{\lambda}^\mathsf{T}\boldsymbol{\beta} = 0$ versus $H_A: \ \boldsymbol{\lambda}^\mathsf{T}\boldsymbol{\beta} \neq 0$. (10 pts)

   (c) In part (a) you gave the sampling distribution for the BLUE using a frequentist perspective. If we instead take a Bayesian approach, what quantity has a distribution (conditional on the observed data) and what is that distribution? (5 pts)

   (d) One inference we might obtain after fitting a model to such data is an interval estimate for the linear combination of interest.

      i. Pick a confidence level and interpret the the interval estimate using a frequentist perspective. (5 pts)

      ii. Now change to the Bayesian mode and interpret the same interval in a Bayesian context. (5 pts)

2. In exploratory analysis when the true model is unknown and we have to decide which variables to include in a linear model, we can err by including variables which are not needed or by omitting variables which are part of the true model. Specifically, assume $[\boldsymbol{X}_1 \ \boldsymbol{X}_2]$ is of full column rank and that $\boldsymbol{X}_1$ is not orthogonal to $\boldsymbol{X}_2$.

   (a) If we overfit and use
   $$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$$
   when really $\boldsymbol{\beta}_2 = \boldsymbol{0}$, then what can you say about the least squares estimator of $\boldsymbol{\beta}_1$? Explain in terms of general properties of the estimator. (6 pts)

   (b) If we underfit and omit $\boldsymbol{X}_2\boldsymbol{\beta}_2$ when it really should be included, what can you say about the least squares estimator of $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\mathsf{T} \ \ \boldsymbol{\beta}_2^\mathsf{T})^\mathsf{T}$? Explain. (6 pts)

   (c) Does the Gauss–Markov theorem apply to either (or both) of the above conclusions? Explain. (9 pts)

3. A Poisson regression model was fit to the number of insurance claims filed per year using (fixed) effects for district, car size, and age of the insured and one interaction term. Additionally, the number of existing policies (within each cross–classified category) was used as an offset. I obtained the following table by using the `anova` function on the single model.

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			63	236.26	
Size	3	88.35	60	147.91	0.0000
District	3	11.62	57	136.29	0.0088
Age	3	84.87	54	51.42	0.0000
District:Age	9	6.56	45	44.86	0.6827

   (a) Explain what the lines labeled District:Age and Age are testing. (6 pts)

   (b) A simpler model was fit to obtain the cofficient estimates below. Use them to compare rates of claims for Districts 1 and 4. (Include proper qualifiers on your inference.)

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.81	0.03	-54.91	0.00
Group.L	0.43	0.05	8.69	0.00
Group.Q	0.00	0.04	0.11	0.91
Group.C	-0.03	0.03	-0.89	0.38
District2	0.03	0.04	0.60	0.55
District3	0.04	0.05	0.76	0.45
District4	0.23	0.06	3.80	0.00
Age.L	-0.39	0.05	-7.98	0.00
Age.Q	-0.00	0.05	-0.01	0.99
Age.C	-0.02	0.05	-0.35	0.73

(4 pts)

   (c) Define overdispersion and explain how an estimated overdispersion parameter of 4 would change the tabled results. (4 pts)

4. Suppose we are fitting a linear mixed model with repeated (over time) measures on each of 16 subjects.

   (a) When would you consider using an AR(1) correlation structure? (5 pts)

   (b) What is the distribution of errors under AR(1)? (5 pts)

   (c) Why is an iterative algorithm needed to estimate the parameters of model? (5 pts)

5. When fitting models with MCMC software, we have to wonder if the Markov chains have converged.

   (a) Explain what we mean by convergence in this context. (5 pts)

   (b) Describe two methods used to assess convergence of Markov chains. One should be graphical (include a sketch). Include a description of results which indicate that the chains probably have converged, and then one for results which indicate more iterations are needed. (10 pts)