2017 First Year Exam – Methods
Statistics 210-211-212
June 26, 2017
9:00 – 12:00

Instructions

- There are 4 questions on the examination, each with multiple parts. Select any 3 of them to solve.
- Your solutions to each of the 3 problems you solve should be written on separate sheets of paper. Label *each sheet* with your student id number, the problem number, and the page for that problem written in the upper right hand corner. For example, the labeling on a page might be:

  ID# 912346378
  Problem 2, page 3
- You have 3 hours to complete your solution. Please be prepared to turn in your exam at 12:00 noon.

1. A nationwide study was conducted with the aim of improving patient satisfaction at urban hospitals. A total of n hospitals were enrolled in this study between 2011 – 2013. Each hospital received a score at the start of the study in 2011 and at the end of the study in 2013. At the start of the study in 2011, all the hospitals received satisfaction scores between 70 – 75 which is considered to be good or average. The hospital-specific score was the average of all the scores recorded from each patient immediately upon discharge through a phone or a face-to-face interview conducted by a research company. The lowest possible score given by a patient is 0 (highly dissatisfied) and the highest possible score is 100 (highly satisfied).

   As noted above, all the hospitals have comparable satisfaction scores at the start of the study. As part of the effort to improve satisfaction, the hospitals were randomly assigned to participate in one of two workshops focused on improving communication by nurses and technicians. The approach of the first workshop was heavily based on practical exercises while the style of the second workshop was online self-training. The variables used in this study are defined as follows: Y_i is the patient satisfaction score for hospital i at the end of the study in 2013; W_i is the workshop training to which hospital i was assigned (here W_i which is either 1 or 2) and x_i is the average number of nurses assigned to 5 patients.

   (a) Formulate a model for patient satisfaction where the mean score takes into account potential different outcomes for the two workshops and potential interaction between the workshop method and the average number of nurses per 5 patients. Write the model in the form $Y_i = \mu_i + \epsilon_i$. Specify the mean component μ_i and the random component ϵ_i. For this problem it will be considered valid to consider the distribution of the scores to be approximately Gaussian, assuming standard classical assumptions for the linear regression model.

   (b) Write the model in matrix notation. Make sure that you specify the components of all vectors and matrices used in the model.

   (c) Explain how you could conduct a test for no interaction between the workshop method and the average number of nurses per 5 patients using the concept of nested models, where the full model $\mathcal{M}$ contains the main effects of training type and the average number of nurses per 5 patients and the interaction of the two main effects, and the reduced model $\mathcal{M}_0$ does not contain the interaction effect. A complete answer should include the null and alternative hypotheses; the linear models for $\mathcal{M}$ and $\mathcal{M}_0$; the test statistic; the distribution of the test statistic under the the null hypothesis; and the rejection region.

   (d) Define $\delta(a)$ to be the difference in the expectation of the distributions of satisfaction scores for the two training types where the average number of nurses per 5 patients is a. Under the no-interaction model, suppose that a 95% confidence interval for $\delta(a)$ (assuming equal probability tails) is given by (L_a, U_a). Now denote a 95% confidence interval (again assuming equal probability tails) for the difference $\delta(b)$, where $a < b$, to be (L_b, U_b). Which of the following is true? Explain.

      i. $U_a - L_a < U_b - L_b$
      ii. $U_a - L_a = U_b - L_b$
      iii. $U_a = U_b$ and $L_a = L_b$
      iv. No conclusion due to incomplete information.

   (e) Following the above notation, consider the following two parameterizations of the mean function μ_i:

$$\begin{aligned} \mu_i^A &= \alpha_0 + \alpha_1 W_{1i} + \alpha_2 x_i + \alpha_{12} W_{1i} x_i \\ \mu_i^B &= \beta_0 + \delta_0 W_{1i} + (\beta_1 + \delta_1 W_{1i}) x_i \end{aligned}$$

   where $W_{1i} = 1$ if the i-th hospital adopted the first workshop and $W_{1i} = 0$ if it adopted the second workshop.

i. Using the first parameterization μ_i^A, derive the mean function for the hospital population that used the first workshop and then derive the mean function for the hospital population that used the second workshop.

ii. Derive the mean functions from (i) using the second parameterization μ_i^B.

iii. Show that these two parameterizations are equivalent, i.e., there is a one-to-one function between the slopes and intercepts in the two parameterizations.

2. To study predictors that are associated with elevated fasting blood glucose Y_i (in mg/dl units) among African-American seniors, a geriatric doctor recorded the body mass index (BMI) x_{1i} (in kg/m^2), total daily average calorie intake x_{2i} (averaged over the previous 6 months) and gender. This dataset consists of $n = 27$ subjects randomly selected from the male senior African-American population and $n = 27$ from the female senior African-American population. Gender is encoded through the indicator variables G_{1i} for the male group and G_{2i} for the female group. We will assume that the distribution of the morning blood glucose, Y_i, for any level of BMI and total daily average calorie intake and for both males and females to be Gaussian. Consider the model

$$Y_i = (\beta_0 + \delta_0 G_{2i}) + (\beta_1 + \delta_1 G_{2i})x_{1i} + (\beta_2 + \delta_2 G_{2i})x_{2i} + \epsilon_i, \quad i = 1, \ldots, 54 \tag{1}$$

where the ϵ_i's are iid $N(0, \sigma^2)$. Moreover, the male subjects are indexed by $i = 1, \ldots, 27$ and the females by $i = 28, \ldots, 54$.

(a) Denote the response vector to be $\mathbf{Y} = [Y_1, \ldots, Y_{54}]'$; the error vector to be $\epsilon = [\epsilon_1, \ldots, \epsilon_{54}]'$; and the parameter vector to be $\underline{\beta} = [\beta_0, \beta_1, \beta_2, \delta_0, \delta_1, \delta_2]'$. Let's formulate the regression model in matrix notation to be

$$\mathbf{Y} \ = \ \mathbf{X}\underline{\beta} \ + \ \epsilon.$$

Give the elements of the design matrix $\mathbf{X}$.

(b) Denote the least squares estimator of $\underline{\beta}$ to be $\widehat{\underline{\beta}}$; the vector of predicted values to be $\widehat{\underline{\mathbf{Y}}} = \mathbf{X}\widehat{\underline{\beta}}$ and the residuals to be $\underline{R} = \underline{\mathbf{Y}} - \widehat{\underline{\mathbf{Y}}}$. Suppose that the squared norm of the observed residual vector is $\underline{R}'\underline{R} = 50$. Give an unbiased estimate of the error variance σ^2.

(c) In the following questions, you will need to perform statistical inference. You are given the following calculations from the data:

$$\widehat{\underline{\beta}} \ = \ \begin{pmatrix} 100.00 & 1.00 & .01 & 10 & 0.10 & 0.01 \end{pmatrix}'$$

$$(\mathbf{X}'\mathbf{X})^{-1} \ = \ \begin{pmatrix} 1.0 & 0.10 & 0.001 & 0.02 & 0.01 & 0.01 \\ 0.10 & 0.10 & 0.5 & 0.2 & 0.01 & 0.01 \\ 0.001 & 0.5 & 0.001 & 0.1 & 0 & 0 \\ 0.02 & 0.02 & 0.01 & 1.0 & 0 & 0 \\ 0.01 & 0.01 & 0 & 0 & 1.0 & 0.01 \\ 0.01 & 0.01 & 0 & 0 & 0.01 & 1.0 \end{pmatrix}.$$

i. Consider only the population with total daily calorie intake of 2000 cal. Use ANOVA to test the null hypothesis that the male and female regression lines are parallel across BMI. That is, test the null hypothesis that there is no interaction between gender and BMI for the population with total daily calorie intake of 2000 cal.

ii. Using the calculations above, give a 95% confidence interval for the difference in the expected morning fasting blood glucose between male vs female African-Americans for the subpopulation with daily calorie intake of 2000 cal and BMI of 30 kg/m^2. Note: (i.) a complete answer should include an expression of the true unknown value in terms of the parameter vector $\underline{\beta}$; (ii.) you should specify the percentiles used in calculating the confidence interval; (iii.) you do not need to carry out any matrix calculations.

(d) Let $\underline{Q} = [1, \ldots, 1, 2, \ldots, 2]$ be a vector whose first 27 elements are all 1's and whose last 27 elements are all 2's. Show that the vector of residuals $\underline{R}$ and the vector $\underline{Q}$ are orthogonal, i.e., $\sum_{i=1}^{54} R_i Q_i = 0$.

3. The data in Table 1, taken from Wakefield et al. (1994), were collected following the administration of a single 30 mg dose of the drug cadralazine to a cardiac failure patient.

Table 1: Concentrations, y_i, of the drug cadralazine as a function of time, x_i.

Observation (i)	Time (hrs) (x_i)	Concentration (mg/ltr) (y_i)
1	2	1.63
2	4	1.01
3	6	0.73
4	8	0.55
5	10	0.41
6	24	0.01
7	28	0.06
8	32	0.02

The response y_i represents the drug concentration at time x_i, $i = 1, \ldots, 8$. The most straightforward model for these data is to assume

$$\log(y_i) = \mu_i(\beta) + \epsilon_i = \log\left\{\frac{D}{V}\exp(-\kappa_e x_i)\right\} + \epsilon_i \tag{2}$$

where $\epsilon_i \sim_{iid} \mathcal{N}(0, \sigma^2)$, $\beta = (-\log(V), \kappa_e)$ and the dose is $D = 30$. The parameters of scientific interest are the volume of distribution $V > 0$ and the elimination rate κ_e.

(a) In order to obtain parameter estimates and draw inference for the above regression model, one could turn to the theory of generalized linear models provided that the probability distribution of the outcome is a member of the exponential dispersion family. Write down the form of the probability density function (pdf) for a member of the exponential dispersion family with canonical location parameter θ, dispersion parameter $a(\phi)$ and mean $b'(\theta)$.

(b) Consider a regression model of the form $g(\mu_i) \equiv \eta_i = \mathbf{X}_i\beta$, where μ_i denotes the mean of the response variable of interest, $\mathbf{X}_i$ is the i-th row of the design matrix, β is a vector of regression parameters, and $g(\cdot)$ is a differentiable function linking μ_i to the linear predictor, η_i. Using a generic likelihood pertaining to a member of the exponential dispersion family (in the form provided for (a)), derive the score equation used to obtain maximum likelihood estimate (MLE) of β.

(c) For the general case in (b), derive Fisher's expected information.

(d) Let $\widehat{\beta}$ denote the MLE obtained from solving the score equations you derived in part (b). What is the asymptotic distribution of $\widehat{\beta}$?

(e) Show that the probability distribution for the model defined in (1) is a member of the exponential dispersion family and identify each of the parts of the pdf. Use this and your results from (b) and (c) to provide expressions for the score equations and Fisher's expected information in the setting of model (1).

(f) Based upon the data in Table 1 and assuming model (1), the MLE for β, $\widehat{\beta}$, and the estimated variance-covariance matrix for $\widehat{\beta}$ are

```
> fit$coef
      beta0       beta1
   -2.81321    -0.15211

> vcov(fit)
              beta0         beta1
beta0    0.1825338 -0.00799111
beta1   -0.0079911  0.00056078
```

Using these estimates, obtain a 95% confidence interval for the volume of distribution, V.(Note: You may leave your numerical expressions unevaluated, but you explicitly state how each element is computed.)

(g) The clearance, $Cl = V \times \kappa_e$ and elimination half-life $x_{1/2} = (\log 2)/\kappa_e$ are also parameters of scientific interest in this experiment. Find the MLEs of these parameters along with asymptotic 95% confidence intervals. (Note: You may leave your numerical expressions unevaluated, but you explicitly state how each element is computed.)

(h) State explicitly what residuals plots you would consider in order to assess the assumptions made by the model in (1). For each plot, you should specify (i) the assumption you are assessing, (ii) what would be plotted on the x- and y-axis, and (iii) what would indicate a violation of the model assumption.

(i) Suppose that in reality, $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ with ϵ_i independent of ϵ_i, $i \neq j$. What would be the implication for the inference you provide in parts (f) and (g). If you stated that your inference would be invalid, explicitly state how you would attempt to correct your inference and discuss any potential limitations to your approach.

4. Equity among faculty salaries at US universities is of interest. To assess whether or not inequity exists between male and female faculty members, a longitudinal study of faculty salaries for faculty employed at a single US university was conducted. Annual salaries for all faculty employed at anytime between 1998 and

2008 were collected. Let Y_{ij} denote the annual salary for faculty member i during year j. Beyond gender and year, data on faculty race/ethnicity (Caucasian, Asian, African-American, Latin/Mexican, Other), type of degree (PhD, Professional, Other), year highest degree was earned, academic school (ICS/Engineering, Arts/Education/Humanities, Business, Biological Sciences, Medicine, Physical Sciences, Social Ecology, and Social Sciences), and an indicator of whether or not the faculty member had an administrative position in a given year were also collected.

Primary interest was in determining whether the linear rate of change of salary differs between male and female faculty members. To answer this question, an initial mean model of the following form was specified:

$$\mathrm{E}[\log(Y_{ij})|\vec{X}_i] = \mu_{ij} = \beta_0 + \beta_1 MALE_i + \beta_2(YEAR_{ij} - 2008) + \beta_3 MALE_i \times (YEAR_{ij} - 2008) + \vec{\gamma}\vec{Z}_{ij},$$

$i = 1, 2, \ldots, N,\;\; j = 1, 2, \ldots, n_i$, with $\vec{Z}_{ij}$ denoting a vector of adjustment covariates including an indicator for faculty race/ethnicity, an indicator for type of degree , year highest degree was earned, an indicator for academic school, and an indicator of whether or not the faculty member had an administrative position in the given year. Note that $MALE$ is an indicator of male gender and $YEAR$ can take on the values 1998, 1999, ..., 2008. Also note that in the above model, salary has been log-transformed and that the distribution of log-transformed salary is roughly symmetric.

(a) Provide precise interpretations of β_1 and β_2 in terms that can be easily understood by a statistical layman. (Hint: A suitable transformation may help.)

(b) Consider the model that is specified in the call for `fit1` in APPENDIX 1. Based upon this specification, write down the full probability model that is being assumed. That is write down all random components along with the distributional assumptions for the random components being assumed in the model.

(c) Under the distributional assumptions for the random components being assumed in model `fit1`, give the general form for the covariance matrix $\Sigma_i \equiv \mathrm{Cov}(\vec{Y}_i^*)$, where $Y_{ij}^* = \log Y_{ij}$.

(d) Suppose that the assumptions made in `fit1` where true, but we use ordinary least squares (OLS) to estimate $\vec{\theta} = (\beta_0, \beta_1, \beta_2, \beta_3, \vec{\gamma})$. Call this estimator $\widehat{\vec{\theta}}_I$.

  i. What is the approximate (large sample) variance of $\widehat{\vec{\theta}}_I$? (Matrix notation would be perfect. I'm looking for the *true* variance here, not what would be returned from a software package under the usual OLS assumptions. If you could not answer (c), you can still answer this question by leaving your solution in terms of the Σ_i's.)

  ii. Provide an expression for the robust (empirical sandwich) estimator of the variance of $\widehat{\vec{\theta}}_I$ and list any asymptotic properties of this estimator.

  iii. Suppose we wish to test the null hypothesis that the first order rate of change of salaries for males is 0 using $\widehat{\vec{\theta}}_I$ as the estimate of $\vec{\theta}$, and using the robust variance estimate of $\mathrm{Var}[\widehat{\vec{\theta}}_I]$ (call it $\mathbf{C}$). Provide a test statistic and critical value to yield an asymptotic level $\alpha = .05$ test of the null hypothesis. (Note: You may leave the test statistic in matrix form and you need not provide a numeric critical value, but should say how the critical value would be obtained by specifying the referent distribution for your test statistic and what quantile of that distribution you would use. If you did not provide an answer for (d) you may just assume a the robust variance estimate is given by $\mathbf{C}$.)

(e) Now return to model `fit1` in APPENDIX 1. Using the summary output for model `fit1`, provide an estimate of the relative difference in salaries between male and female faculty members in the year

2000, controlling for all other adjustment variables in the model.

(f) Using the summary output for model `fit1`, provide an estimate of the relative difference in the first-order rate of change of salaries comparing males to females after adjustment for all other adjustment variables in the model, along with a 95% confidence interval.

(g) It was also interest to determine if the rate of change of salaries between males and females differs by school appointment. As such a 3-way interaction between $MALE$, $(YEAR - 2008)$, and the factored indicator for school affiliation was fit in model `fit3`.

i. A colleague suggests that in order to test the null hypothesis that the rate of change of salaries between males and females does not differ by school appointment, the model in `fit3` should be compared to the reduced model given in `fit1`. Explain why this is incorrect.

ii. Using the output provided in APPENDIX 1, carry out a level $\alpha = .05$ test of the null hypothesis that the rate of change of salaries between males and females does not differ by school appointment. (Note: You need not provide a numeric critical value, but should say how the critical value would be obtained by specifying the referent distribution for your test statistic and what quantile of that distribution you would use.)

APPENDIX 1

```
> fit1 <- lme( lsalary ~ male*I(year-2008) + ethgrp + deg + yrdeg + factor(school.grp) + admin,
+             method = "ML",
+             random = reStruct( ~ 1| uniqueid, pdClass="pdSymm"),
+             correlation = corExp( form = ~ I(year-2008) | uniqueid, nugget=TRUE),
+             data = temp )
> summary( fit1 )
Linear mixed-effects model fit by maximum likelihood
 Data: temp
      AIC    BIC logLik
  -27792 -27625  13919

Random effects:
 Formula: ~1 | uniqueid
        (Intercept) Residual
StdDev:  7.4116e-05  0.18493

Correlation Structure: Exponential spatial correlation
 Formula: ~I(year - 2008) | uniqueid
 Parameter estimate(s):
    range    nugget
33.194239  0.011678
Fixed effects:
                                   Value Std.Error   DF t-value p-value
(Intercept)                       51.455   0.78295 8791  65.719  0.0000
maleMale                           0.012   0.01151 1455   1.004  0.3157
I(year - 2008)                     0.044   0.00088 8791  49.972  0.0000
ethgrpAsian                       -0.008   0.01387 1455  -0.565  0.5720
ethgrpAfrican-American             0.024   0.02984 1455   0.809  0.4184
ethgrpLatin/Mexican               -0.004   0.02079 1455  -0.190  0.8491
ethgrpOther                       -0.003   0.02390 1455  -0.118  0.9062
degProfessional                   -0.118   0.02110 1455  -5.594  0.0000
degOther                          -0.177   0.02232 1455  -7.936  0.0000
yrdeg                             -0.020   0.00039 1455 -50.971  0.0000
factor(school.grp)Arts/Ed/Human   -0.199   0.01509 8791 -13.190  0.0000
factor(school.grp)Business         0.314   0.02441 1455  12.885  0.0000
factor(school.grp)Bio Sci         -0.215   0.01938 8791 -11.092  0.0000
factor(school.grp)Medicine        -0.120   0.01867 8791  -6.438  0.0000
factor(school.grp)Phys Sci        -0.133   0.01791 1455  -7.413  0.0000
factor(school.grp)Social Ecology  -0.204   0.01767 8791 -11.543  0.0000
factor(school.grp)Social Sciences -0.146   0.01639 8791  -8.886  0.0000
adminYes                          -0.009   0.00396 8791  -2.232  0.0256
maleMale:I(year - 2008)           -0.004   0.00103 8791  -3.828  0.0001
```

```
> fit2 <- lme( lsalary ~ male*I(year-2008) + ethgrp + deg + yrdeg + male*factor(school.grp) +
                       I(year-2008)*factor(school.grp) + admin,
+             method = "ML",
+             random = reStruct( ~ 1| uniqueid, pdClass="pdSymm"),
+             correlation = corExp( form = ~ I(year-2008) | uniqueid, nugget=TRUE),
+             data = temp )

> summary(fit2)
Linear mixed-effects model fit by maximum likelihood
 Data: temp
     AIC    BIC logLik
  -27828 -27561  13951

Random effects:
 Formula: ~1 | uniqueid
        (Intercept) Residual
StdDev:  7.8021e-05  0.18317

Correlation Structure: Exponential spatial correlation
 Formula: ~I(year - 2008) | uniqueid
 Parameter estimate(s):
    range    nugget
32.981382  0.012175
Fixed effects:
                                                 Value Std.Error   DF t-value p-value
(Intercept)                                     51.565   0.77676 8779  66.384  0.0000
maleMale                                         0.003   0.02859 1453   0.100  0.9200
I(year - 2008)                                   0.042   0.00151 8779  27.581  0.0000
ethgrpAsian                                     -0.008   0.01376 1453  -0.598  0.5498
ethgrpAfrican-American                           0.030   0.02962 1453   1.010  0.3127
ethgrpLatin/Mexican                             -0.004   0.02065 1453  -0.200  0.8411
ethgrpOther                                      0.003   0.02377 1453   0.138  0.8899
degProfessional                                 -0.122   0.02100 1453  -5.831  0.0000
degOther                                        -0.178   0.02216 1453  -8.011  0.0000
yrdeg                                           -0.020   0.00039 1453 -51.503  0.0000
factor(school.grp)Arts/Ed/Human                 -0.222   0.02923 8779  -7.599  0.0000
factor(school.grp)Business                       0.416   0.04671 1453   8.914  0.0000
factor(school.grp)Bio Sci                       -0.210   0.04171 8779  -5.035  0.0000
factor(school.grp)Medicine                      -0.162   0.03618 8779  -4.489  0.0000
factor(school.grp)Phys Sci                      -0.121   0.04261 1453  -2.847  0.0045
factor(school.grp)Social Ecology                -0.156   0.02841 8779  -5.477  0.0000
factor(school.grp)Social Sciences               -0.119   0.03367 8779  -3.527  0.0004
adminYes                                        -0.009   0.00395 8779  -2.217  0.0267
maleMale:I(year - 2008)                         -0.004   0.00106 8779  -3.840  0.0001
maleMale:factor(school.grp)Arts/Ed/Human         0.032   0.03325 8779   0.973  0.3308
maleMale:factor(school.grp)Business             -0.097   0.05415 1453  -1.797  0.0726
maleMale:factor(school.grp)Bio Sci               0.016   0.04638 8779   0.348  0.7276
maleMale:factor(school.grp)Medicine              0.041   0.03909 8779   1.055  0.2913
maleMale:factor(school.grp)Phys Sci              0.016   0.04655 1453   0.354  0.7234
maleMale:factor(school.grp)Social Ecology       -0.081   0.03595 8779  -2.260  0.0239
maleMale:factor(school.grp)Social Sciences      -0.004   0.03777 8779  -0.097  0.9225
I(year - 2008):factor(school.grp)Arts/Ed/Human   0.001   0.00155 8779   0.854  0.3933
I(year - 2008):factor(school.grp)Business        0.009   0.00246 8779   3.660  0.0003
I(year - 2008):factor(school.grp)Bio Sci         0.005   0.00190 8779   2.594  0.0095
I(year - 2008):factor(school.grp)Medicine       -0.001   0.00160 8779  -0.594  0.5524
I(year - 2008):factor(school.grp)Phys Sci        0.007   0.00176 8779   3.812  0.0001
I(year - 2008):factor(school.grp)Social Ecology  0.002   0.00217 8779   0.982  0.3262
I(year - 2008):factor(school.grp)Social Sciences 0.006   0.00175 8779   3.245  0.0012
```

```
> fit3 <- lme( lsalary ~ male*I(year-2008)*factor(school.grp) + ethgrp + deg + yrdeg + admin,
+            method = "ML", random = reStruct( ~ 1| uniqueid, pdClass="pdSymm"),
+            correlation = corExp( form = ~ I(year-2008) | uniqueid, nugget=TRUE), data = temp )
> summary(fit3)
     AIC    BIC logLik
  -27823 -27504  13955

Random effects:
 Formula: ~1 | uniqueid
        (Intercept) Residual
StdDev:    6.07e-05  0.18317

Correlation Structure: Exponential spatial correlation
 Formula: ~I(year - 2008) | uniqueid
 Parameter estimate(s):
    range    nugget
33.101130  0.012255
Fixed effects:
                                                          Value Std.Error   DF t-value p-value
(Intercept)                                              51.577   0.77717 8772  66.365  0.0000
maleMale                                                  0.015   0.03044 1453   0.477  0.6337
I(year - 2008)                                            0.038   0.00316 8772  12.133  0.0000
factor(school.grp)Arts/Ed/Human                          -0.217   0.03097 8772  -7.015  0.0000
factor(school.grp)Business                                0.435   0.04919 1453   8.851  0.0000
factor(school.grp)Bio Sci                                -0.201   0.04403 8772  -4.574  0.0000
factor(school.grp)Medicine                               -0.144   0.03841 8772  -3.754  0.0002
factor(school.grp)Phys Sci                               -0.089   0.04525 1453  -1.972  0.0488
factor(school.grp)Social Ecology                         -0.143   0.03037 8772  -4.700  0.0000
factor(school.grp)Social Sciences                        -0.114   0.03602 8772  -3.165  0.0016
ethgrpAsian                                              -0.008   0.01377 1453  -0.597  0.5504
ethgrpAfrican-American                                    0.030   0.02963 1453   1.023  0.3065
ethgrpLatin/Mexican                                      -0.004   0.02067 1453  -0.180  0.8571
ethgrpOther                                               0.003   0.02379 1453   0.134  0.8938
degProfessional                                          -0.123   0.02101 1453  -5.846  0.0000
degOther                                                 -0.178   0.02217 1453  -8.026  0.0000
yrdeg                                                    -0.020   0.00039 1453 -51.502  0.0000
adminYes                                                 -0.009   0.00395 8772  -2.239  0.0252
maleMale:I(year - 2008)                                   0.000   0.00342 8772  -0.090  0.9280
maleMale:factor(school.grp)Arts/Ed/Human                  0.030   0.03609 8772   0.823  0.4103
maleMale:factor(school.grp)Business                      -0.123   0.05847 1453  -2.109  0.0352
maleMale:factor(school.grp)Bio Sci                        0.006   0.04979 8772   0.114  0.9094
maleMale:factor(school.grp)Medicine                       0.019   0.04217 8772   0.444  0.6572
maleMale:factor(school.grp)Phys Sci                      -0.022   0.04997 1453  -0.437  0.6620
maleMale:factor(school.grp)Social Ecology                -0.100   0.03962 8772  -2.526  0.0116
maleMale:factor(school.grp)Social Sciences               -0.009   0.04148 8772  -0.205  0.8375
I(year - 2008):factor(school.grp)Arts/Ed/Human            0.003   0.00347 8772   0.959  0.3378
I(year - 2008):factor(school.grp)Business                 0.015   0.00501 8772   2.926  0.0034
I(year - 2008):factor(school.grp)Bio Sci                  0.008   0.00451 8772   1.754  0.0795
I(year - 2008):factor(school.grp)Medicine                 0.004   0.00398 8772   1.091  0.2755
I(year - 2008):factor(school.grp)Phys Sci                 0.016   0.00481 8772   3.369  0.0008
I(year - 2008):factor(school.grp)Social Ecology           0.007   0.00417 8772   1.576  0.1150
I(year - 2008):factor(school.grp)Social Sciences          0.008   0.00393 8772   1.930  0.0537
maleMale:I(year - 2008):factor(school.grp)Arts/Ed/Human  -0.002   0.00390 8772  -0.427  0.6692
maleMale:I(year - 2008):factor(school.grp)Business       -0.007   0.00578 8772  -1.258  0.2083
maleMale:I(year - 2008):factor(school.grp)Bio Sci        -0.003   0.00498 8772  -0.692  0.4890
maleMale:I(year - 2008):factor(school.grp)Medicine       -0.006   0.00435 8772  -1.455  0.1457
maleMale:I(year - 2008):factor(school.grp)Phys Sci       -0.011   0.00517 8772  -2.116  0.0344
maleMale:I(year - 2008):factor(school.grp)Social Ecology -0.006   0.00496 8772  -1.191  0.2335
maleMale:I(year - 2008):factor(school.grp)Social Sciences -0.002  0.00441 8772  -0.427  0.6696
```