Analysis of Student Performance in Mathematics Courses

Jeremy Meyer

STAT 651 Project

December 13, 2019

Overview

Background

- Media is full of explanations on what makes students excel
- Student math performance was collected from Portuguese school questionnaires
- Analysis can be used by school to find areas of improvement
- DISCLAIMER:
    - Correlation does not imply causation.
    - These are results collected from ONE school.

The Data

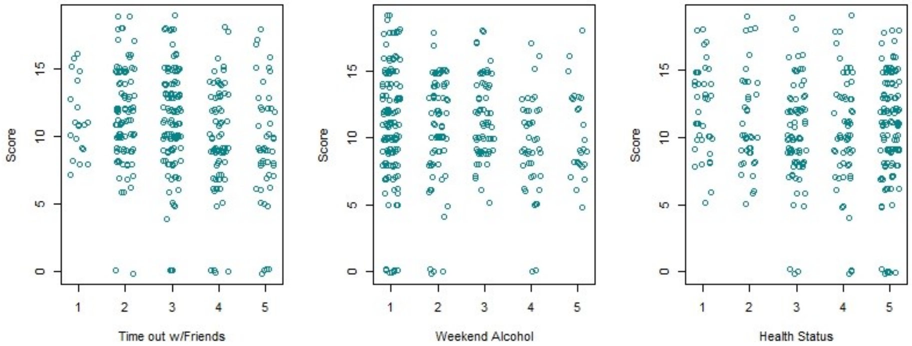
- 349 secondary students aged 15-22
- Response is final score in math courses (0 to 20)
- Goal: Interest is in what factors contributes to a high final grade
    - Socioeconomic factors like study time, family relations, etc.

Socioeconomic variables of interest

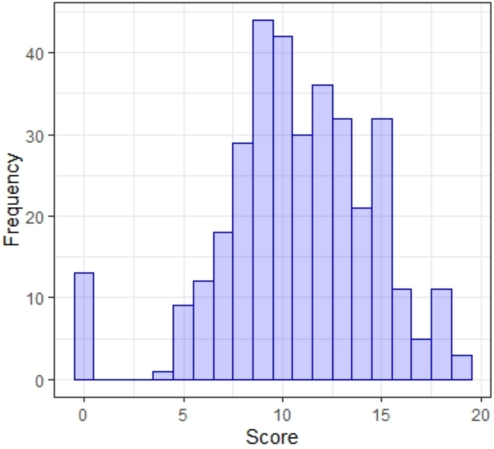
All 10 are discrete

1. Mother's Education level (0 None - 4 College level)
2. Father's Education level (0 None - 4 College level)
3. Study time (1: $<$ 2 hours - 4: $>$ 10 hours)
4. Extra-Carriculars? (Yes / No)
5. Romantic relationship (Yes / No)
6. Family relations (1 Very Bad - 5 Excellent)
7. Free Time after school (1 Very Low - 5 Very High)
8. Time out with friends (1 Very Low - 5 Very High)
9. Weekend Alchohol consumption (1 Very Low - 5 Very High)
10. Health (1 Very Poor - 5 Excellent)

Some Covariates

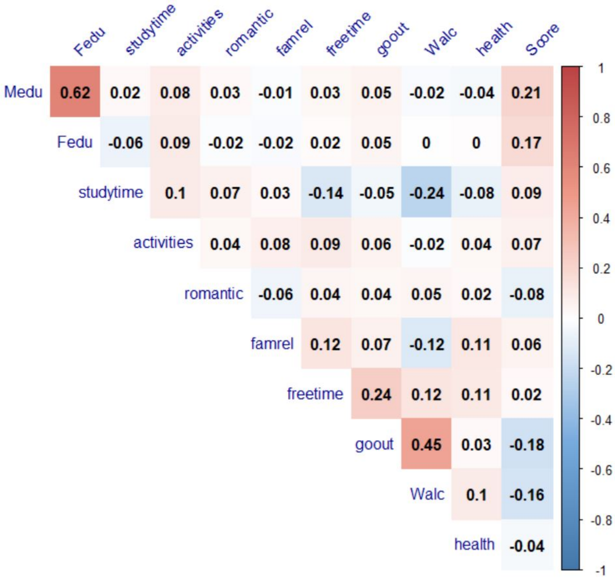


Scores - Outliers



- 13 Students scored zeros. Dropouts?
- These actually greatly affect results (more later)
- Will throw them out for interest in students who finished

Correlations



Model Construction

- Scale response to (0,1) $\rightarrow y_i \sim$ Beta
- Evaluate significant covariates $\rightarrow$ regression
- Logit link function so $0 < \mu_i < 1$
- Noninformative priors

Beta reparameterization

Why not in terms of mean and variance?

$$s^2 = \frac{\alpha\left[\frac{\alpha(1-\bar{x})}{\bar{x}}\right]}{(\alpha+\left[\frac{\alpha(1-\bar{x})}{\bar{x}}\right])^2\left(\alpha+\left[\frac{\alpha(1-\bar{x})}{\bar{x}}\right]+1\right)} = \frac{\alpha^2\bar{x}^2(1-\bar{x})}{(\alpha\bar{x}+\alpha(1-\bar{x}))^2(\alpha\bar{x}+\alpha(1-\bar{x})+\bar{x})}$$

$$= \frac{\alpha^2\bar{x}^2(1-\bar{x})}{\alpha^2(a+\bar{x})} = \frac{\bar{x}^2(1-\bar{x})}{a+\bar{x}}$$

$$\beta = \frac{\alpha(1-\bar{x})}{\bar{x}} \qquad \alpha = \frac{\bar{x}^2(1-\bar{x})-s^2\bar{x}}{s^2} = \bar{x}\left[\frac{\bar{x}(1-\bar{x})}{s^2}-1\right]$$

- Beta variance has upper bound: $\frac{\mu(1-\mu)}{1+\alpha+\beta} < \mu(1-\mu) < 0.25$
- For a fixed μ and variance, MME for α, β may be outside support.
- ex. If $\mu = .9$, var$=.15 \rightarrow \quad \hat{\alpha} = -.2, \quad \hat{\beta} = -.04$
- Makes proposals difficult for MCMC.

Beta reparameterization

Instead let $\phi = \alpha + \beta$

- ϕ is unconstrained for a given μ
- Variance at a given μ becomes $\frac{\mu(1-\mu)}{1+\phi}$
- ϕ is a "dispersion parameter"

Thus if $\mu = \frac{\alpha}{\alpha+\beta}$ $\phi = \alpha + \beta \rightarrow$

- $\alpha = \phi\mu$
- $\beta = \phi(1 - \mu)$

Proposed Model

$$
\begin{gathered}
y_i|\phi, \mu_i \sim \text{Beta}(\phi\mu_i,\ \phi(1-\mu_i)),\ \text{where} \\
\mu_i = \text{logit}^{-1}(\beta_0 + \beta_1(x_{1i}) + \beta_2(x_{2i}) + \ldots + \beta_9(x_{9i}) + \beta_{10}(x_{10i})) \\
\phi \sim \text{Gamma}(.1, .1), \quad \text{each } \beta_j \sim t_4
\end{gathered} \tag{1}
$$

Where

- y_i is the scaled i^{th} student score
- x_{ki} is the k^{th} covariate for student i
- β_k represents the effect of the k^{th} covariate on score.

$$
E(y_i) = \mu_i \qquad V(y_i) = \frac{\mu_i(1-\mu_i)}{1+\phi} \tag{2}
$$

Prior Choice

ϕ Prior	DIC	β_j prior	DIC
G(.1, .1)	-288.32	t_4	-288.32
G(.1, 1)	-286.62	t_2	-287.86
G(1,1)	-286.48	N(0,1)	-287.15
Unif(0, 100)	-287.59	N(0, 100)	-286.74

Note: Top row is baseline level.

- In Beta regression, the coefficients are very small
- Uninformative prior with positive support on ϕ

Computational Approach (MCMC)

- Gibbs sampler had problems mixing
  - For 20,000 draws $\rightarrow$ β_0 had effective sample size 190
- Idea: Sample correlated betas jointly, uncorrelated βs univariately
  - Use Metropolis Algorithim
- Use correlations from Gibbs sampler to determine groupings
- Many variables were negatively correlated with β_0
- β_1 and β_2 correlated.

Sampling algorithm: Multivariate Updates

1. Choose starting values. $(0 \to \beta_i,\ 10 \to \phi)$
2. We sample $(\beta_0, \beta_3, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10})$ together and (β_1, β_2) together.
    - Use a MVN proposal distribution, centered on the current state.
    - Let $n_1 = 7$ and $n_2 = 2$ be the dimension of update
    - Conditioning on all other parameters, use the inverse hessian normal approximation to generate a covariance matrix $\hat{\Sigma}$
    - Variance of the proposal is a scaled version of $\hat{\Sigma}$
    - Thus, only 1 tuning parameter per group.
    - Don't regenerate a new $\hat{\Sigma}$ each iteration, just generate once at start or infrequently.
    - Update first group and current state, (β_1, β_2)

Sampling Algorithim: Univariate updates

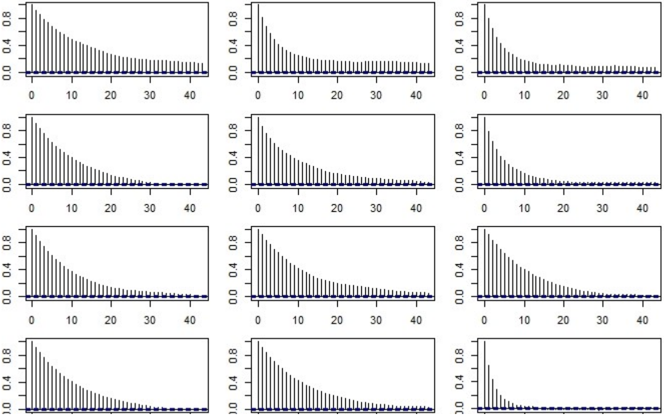
1. For β_4, β_5, and ϕ, use a univariate Metropolis update. For these betas:
   - Use a normal proposal.
   - Update current state with draws from multivariate updates
2. Iterate back and forth between all multivariate and univariate updates until we obtain 20,000 draws.

Acceptance rates / Effective Sample size / $\hat{R}$

	Acc Rate	Eff Size	$\hat{R}$
β_0	.258	697	1.0029
β_1	.369	748	1.0056
β_2	.369	1192	1.0015
$\boldsymbol{\beta_3}$	.258	899	1.0023
$\boldsymbol{\beta_4}$	.277	910	1.0011
β_5	.258	1861	1.0014
β_6	.258	971	1.0010
β_7	.258	817	1.0012
β_8	.258	856	1.0017
β_9	.258	861	1.0013
β_{10}	.258	850	1.0007
ϕ	.439	3848	1.0002

- Eased some mixing problems
- Used 5 different starting locations (see slide 19)

ACF plots



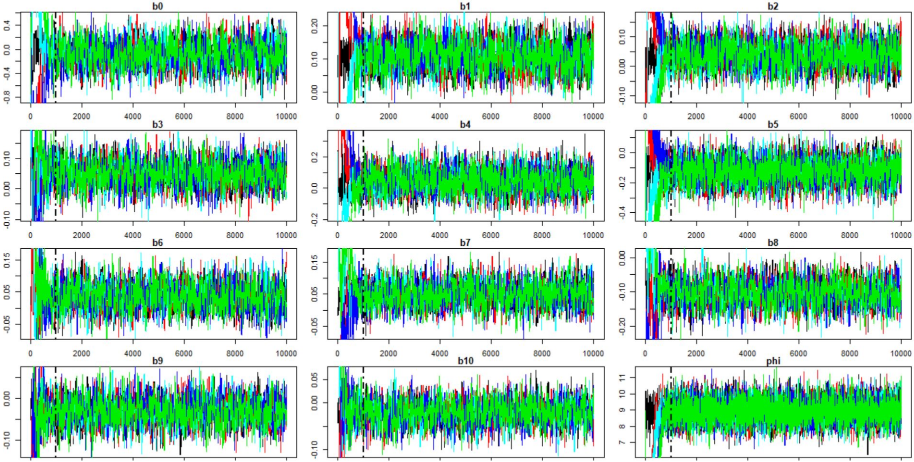
Arranged from left to right in increasing order of β_i (ϕ last)

Traceplots in various locations

Recall $(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 10)$ (BLACK) was the starting value.
Suppose we started

- (-.25, -.25, ..., 9) RED
- (.25, .25, .25, ..., 11) CYAN
- (.5, .5, ..., -.5, -.5, ... 7) BLUE
- (-.5, -.5, ..., +.5, +.5, ... 12) GREEN

Trace Plots



All traceplots show strong signs of convergence!

Results: Significant coefficients

	2.5%	mean	97.5%
Intercept	-0.52	-0.08	0.35
Medu(+)	0.03	0.11	0.19
Fedu	-0.04	0.04	0.12
studytime	-0.03	0.05	0.13
activities	-0.09	0.05	0.19
romantic	-0.27	-0.12	0.02
famrel	-0.04	0.04	0.11
freetime	-0.02	0.05	0.12
goout(-)	-0.18	-0.10	-0.03
Walc	-0.09	-0.03	0.03
health	-0.07	-0.02	0.03
phi	7.64	8.93	10.24

Note: This does NOT change with other priors

What if we had left the zeros in?

	2.5%	mean	97.5%
Intercept	-0.45	0.11	0.67
Medu(+)	0.04	0.15	0.27
Fedu	-0.11	-0.00	0.11
studytime(+)	0.04	0.15	0.26
activities	-0.08	0.11	0.30
romantic(-)	-0.57	-0.37	-0.17
famrel	-0.14	-0.03	0.07
freetime	-0.05	0.06	0.16
goout(-)	-0.34	-0.24	-0.14
Walc(+)	0.02	0.11	0.19
health(-)	-0.16	-0.09	-0.02
phi	3.33	3.81	4.35

Worse health and weekends out on alcohol are correleted with better grades!

Why did this happen?

Table: Group Means

	Zeros	Not Zeros	diff
Medu	2.62	2.81	-0.19
Fedu	2.54	2.55	-0.01
studytime	1.77	2.07	-0.30
activities	0.46	0.53	x0.9
romantic	0.61	0.31	x2.0
famrel	4.38	3.94	0.44
freetime	3.46	3.21	0.25
goout	3.77	3.09	0.68
Walc	1.77	2.28	-0.51
health	4.31	3.55	0.76

The 13 students who scored 0 on average have much better health and drink less alcohol.

These differences had a high influence on the βs.

Comparison to Frequentist methods

betareg package in R does ML frequentist beta regression

- Parameterization is the same. Same conclusions.
- Estimates are almost identical

Covariate	P-value	Coef.F	Coef.B	Width.F	Width.B
(Intercept)	0.666	-0.11	-0.08	1.00	0.87
Medu	**0.007***	0.11	0.11	0.16	0.15
Fedu	0.315	0.04	0.04	0.16	0.16
studytime	0.212	0.05	0.05	0.17	0.16
activities	0.478	0.05	0.05	0.28	0.28
romantic	0.082	-0.13	-0.12	0.29	0.29
famrel	0.348	0.04	0.04	0.16	0.15
freetime	0.165	0.05	0.05	0.14	0.14
goout	**0.004***	-0.10	-0.10	0.14	0.15
Walc	0.360	-0.03	-0.03	0.12	0.12
health	0.419	-0.02	-0.02	0.10	0.10
ϕ	—	9.19	8.93	2.64	2.59

Conclusions

- Educated mothers tend to have kids do well in math class.
- Spending too much time out with friends can be detrimental to grades
- Future work:
    1. Find out why some students scored 0
    2. Trends across schools / subjects
    3. Causal analysis