

STA 2101/442 Assignment 4¹

These questions are practice for the midterm and final exam, and are not to be handed in.

1. It is well known that people who graduate from university have higher lifetime earnings on average than those who do not. Mention at least one confounding variable that could have produced this result.
2. Let $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where \mathbf{X} is an $n \times p$ matrix of known constants, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown constants, and $\boldsymbol{\epsilon}$ is multivariate normal with mean zero and covariance matrix $\sigma^2\mathbf{I}_n$. The constant $\sigma^2 > 0$ is unknown.
 - (a) Show $X^\top \mathbf{e} = \mathbf{0}$.
 - (b) Why does $X^\top \mathbf{e} = \mathbf{0}$ tell you that if a regression model has an intercept, the residuals must add up to zero?
 - (c) Consider a regression model with an intercept, so that the sum of residuals is equal to zero. Prove the following decomposition of sums of squares, also given on the formula sheet: $SST = SSE + SSR$. Hint: Starting with scalar calculations, add and subtract \hat{y}_i . Switch to matrix notation partway through the calculation.
3. High School History classes from across Ontario are randomly assigned to either a discovery-oriented or a memory-oriented curriculum in Canadian history. At the end of the year, the students are given a standardized test and the median score of each class is recorded. Please consider a regression model with these variables:
 - X_1 Equals 1 if the class uses the discovery-oriented curriculum, and equals 0 if the class uses the memory-oriented curriculum.
 - X_2 Average parents' education for the classroom.
 - X_3 Average family income for the classroom.
 - X_4 Number of university History courses taken by the teacher.
 - X_5 Teacher's final cumulative university grade point average.
 - Y Class median score on the standardized history test.

The full regression model (as opposed to the reduced models for various null hypotheses) implies

$$E[Y|X] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5.$$

For each question below, please give

¹This assignment was prepared by [Jerry Brunner](#), Department of Statistics, University of Toronto. It is licensed under a [Creative Commons Attribution - ShareAlike 3.0 Unported License](#). Use any part of it as you like and share the result freely. The L^AT_EX source code is available from the course website: <http://www.utstat.toronto.edu/~brunner/oldclass/appliedf18>

- The null hypothesis in terms of β values.
 - $E[Y|X]$ for the reduced model you would use to answer the question. Don't re-number the variables.
- (a) If you allow for parents' education and income and for teacher's university background, does curriculum type affect test scores? (And why is it okay to use the word "affect?")
 - (b) Controlling for parents' education and income and for curriculum type, is teacher's university background (two variables) related to their students' test performance?
 - (c) Correcting for teacher's university background and for curriculum type, are parents' education and family income (considered simultaneously) related to students' test performance?
 - (d) Taking curriculum type, teacher's university background and parents' education into consideration, is parents' income related to students' test performance?
 - (e) Here is one final question. Assuming that X_1, \dots, X_5 are random variables (and I hope you agree that they are),
 - i. Would you expect X_1 to be related to the other explanatory variables?
 - ii. Would you expect the other explanatory variables to be related to each other?
4. In the United States, admission to university is based partly on high school marks and recommendations, and partly on applicants' performance on a standardized multiple choice test called the Scholastic Aptitude Test (SAT). The SAT has two sub-tests, Verbal and Math. A university administrator selected a random sample of 200 applicants, and recorded the Verbal SAT, the Math SAT and first-year university Grade Point Average (GPA) for each student. The data are available [here](#). We seek to predict GPA from the two test scores. Throughout, please use the usual $\alpha = 0.05$ significance level.
- (a) First, fit a model using just the Math score as a predictor. "Fit" means estimate the model parameters. Does there appear to be a relationship between Math score and grade point average?
 - i. Answer Yes or No.
 - ii. Fill in the blank. Students who did better on the Math test tended to have _____ first-year grade point average.
 - iii. Do you reject $H_0 : \beta_1 = 0$?
 - iv. Are the results statistically significant? Answer Yes or No.
 - v. What is the p -value? The answer can be found in *two* places on your printout.
 - vi. What proportion of the variation in first-year grade point average is explained by score on the SAT Math test? The answer is a number from your printout.

vii. Give a predicted first-year grade point average for a student who got 700 on the Math SAT. The answer is a number you could get with a calculator from your printout.

(b) Now fit a model with both the Math and Verbal sub-tests.

i. Give the test statistic, the degrees of freedom and the p -value for each of the following null hypotheses. The answers are numbers from your printout.

A. $H_0 : \beta_1 = \beta_2 = 0$

B. $H_0 : \beta_1 = 0$

C. $H_0 : \beta_2 = 0$

D. $H_0 : \beta_0 = 0$

ii. Controlling for Math score, is Verbal score related to first-year grade point average?

A. Give the value of the test statistic. The answer is a number from your printout.

B. Give the p -value. The answer is a number from your printout.

C. Do you reject the null hypothesis?

D. Are the results statistically significant? Answer Yes or No.

E. In plain, non-statistical language, what do you conclude? The answer is something about test scores and grade point average.

iii. Allowing for Verbal score, is Math score related to first-year grade point average?

A. Give the value of the test statistic. The answer is a number from your printout.

B. Give the p -value. The answer is a number from your printout.

C. Do you reject the null hypothesis?

D. Are the results statistically significant? Answer Yes or No.

E. In plain, non-statistical language, what do you conclude? The answer is something about test scores and grade point average.

iv. Give a predicted first-year grade point average for a student who got 650 on the Verbal and 700 on the Math SAT.

v. Let's do one more test. We want to know whether expected GPA increases faster as a function of the Verbal SAT, or the Math SAT. That is, we want to compare the regression coefficients, testing $H_0 : \beta_1 = \beta_2$.

A. Express the null hypothesis in matrix form as $\mathbf{L}\boldsymbol{\beta} = \mathbf{h}$.

B. Carry out an F test².

²If you do not remember how to do this with R from your regression course, several packages provide functions to do it — for example, the `linear.hypothesis` function in the `car` (Companion to Applied Regression) package. Or you could just search online. When I did this, I found a useful [example from a regression course I taught several years ago](#). Feel free to use my `ftest` function if you wish.

C. State your conclusion in plain, non-technical language. It's something about first-year grade point average.

5. Ordinary linear regression is often applied to data sets where the independent variables are best modeled as random variables: write $y_i = \mathbf{X}_i^\top \boldsymbol{\beta} + \epsilon_i$. In what way does the usual conditional linear regression model with normal errors imply that random explanatory variables have zero covariance with the error term? Hint: Assume \mathbf{X}_i as well as ϵ_i continuous. What is the conditional distribution of ϵ_i given \mathbf{X}_i ?
6. For a model with just one (random) explanatory variable, show that $E(\epsilon_i | X_i = x_i) = 0$ for all x_i implies $Cov(X_i, \epsilon_i) = 0$, so that a standard regression model without the normality assumption still implies zero covariance, though not necessarily independence, between the error term and explanatory variables.
7. In the following regression model, the explanatory variables X_1 and X_2 are random variables. The true model is

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \epsilon_i,$$

independently for $i = 1, \dots, n$, where $\epsilon_i \sim N(0, \sigma^2)$.

The mean and covariance matrix of the explanatory variables are given by

$$E \begin{pmatrix} X_{i,1} \\ X_{i,2} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and} \quad Var \begin{pmatrix} X_{i,1} \\ X_{i,2} \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{pmatrix}$$

The explanatory variables $X_{i,1}$ and $X_{i,2}$ are independent of ϵ_i .

Unfortunately $X_{i,2}$, which has an impact on Y_i and is correlated with $X_{i,1}$, is not part of the data set. Since $X_{i,2}$ is not observed, it is absorbed by the intercept and error term, as follows.

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \epsilon_i \\ &= (\beta_0 + \beta_2 \mu_2) + \beta_1 X_{i,1} + (\beta_2 X_{i,2} - \beta_2 \mu_2 + \epsilon_i) \\ &= \beta'_0 + \beta_1 X_{i,1} + \epsilon'_i. \end{aligned}$$

The primes just denote a new β_0 and a new ϵ_i . It was necessary to add and subtract $\beta_2 \mu_2$ in order to obtain $E(\epsilon'_i) = 0$. And of course there could be more than one omitted variable. They would all get swallowed by the intercept and error term, the garbage bins of regression analysis.

- (a) What is $Cov(X_{i,1}, \epsilon'_i)$?
- (b) Calculate the variance-covariance matrix of $(X_{i,1}, Y_i)$ under the true model. Is it possible to have non-zero covariance between $X_{i,1}$ and Y_i when $\beta_1 = 0$?

(c) Suppose we want to estimate β_1 . The usual least squares estimator is

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_{i,1} - \bar{X}_1)(Y_i - \bar{Y})}{\sum_{i=1}^n (X_{i,1} - \bar{X}_1)^2}.$$

You may just use this formula; you don't have to derive it. Is $\hat{\beta}_1$ a consistent estimator of β_1 if the true model holds? Answer Yes or no and show your work. You may use the consistency of the sample variance and covariance without proof.

- (d) Are there *any* points in the parameter space for which $\hat{\beta}_1 \xrightarrow{p} \beta_1$ when the true model holds?
8. Independently for $i = 1, \dots, n$, let $Y_i = \beta X_i + \epsilon_i$, where $X_i \sim N(\mu, \sigma_x^2)$ and $\epsilon_i \sim N(0, \sigma_\epsilon^2)$. Because of omitted variables that influence both X_i and Y_i , we have $Cov(X_i, \epsilon_i) = c \neq 0$.
- (a) The least squares estimator of β is $\frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}$. Is this estimator consistent? Answer Yes or No and prove your answer.
- (b) Give the parameter space for this model. There are some constraints on c .
- (c) First consider points in the parameter space where $\mu \neq 0$. Give an estimator of β that converges almost surely to the right answer for that part of the parameter space. If you are not sure how to proceed, try calculating the expected value and covariance matrix of (X_i, Y_i) .
- (d) What happens in the rest of the parameter space — that is, where $\mu = 0$? Is a consistent estimator possible there? So we see that parameters may be identifiable in some parts of the parameter space but not all.

9. We know that omitted explanatory variables are a big problem, because they induce non-zero covariance between the explanatory variables and the error terms ϵ_i . The residuals have a lot in common with the ϵ_i terms in a regression model, though they are not the same thing. A reasonable idea is to check for correlation between explanatory variables and the ϵ_i values by looking at the correlation between the residuals and explanatory variables.

Accordingly, for a multiple regression model with an intercept so that $\sum_{i=1}^n e_i = 0$, calculate the sample correlation r between explanatory variable j and the residuals e_1, \dots, e_n . Use the formula for r from the formula sheet. Simplify. What can the sample correlations between residuals and x variables tell you about the correlation between ϵ and the x variables?

10. Let X_1, \dots, X_n be random sample from a binomial distribution with parameters 4 and θ , where θ is unknown. The prior distribution on θ is beta with parameters α and β .
- Find the posterior density of θ , including the constant that makes it integrate to one.
 - For $n = 20$ observations and $\alpha = \beta = 1$ (the uniform distribution), we obtain $\bar{X} = 2.3$. What is the posterior mean? Hint: The expected value of a beta random variable is $\frac{\alpha}{\alpha+\beta}$.
11. Let X_1, \dots, X_n be random sample from a Poisson distribution with parameter $\lambda > 0$. The prior on λ will be gamma; use the parameterization on the formula sheet. Please derive the posterior density of λ , including the constant that makes it integrate to one.
12. Let X_1, \dots, X_n be random sample from a normal distribution with mean μ and precision τ (the precision is one over the variance – see formula sheet).
- Suppose that the parameter μ is known, while τ is unknown. The prior on τ is gamma, with the parameterization given on the formula sheet. Give the posterior distribution of τ .
 - Suppose that τ is known, while this time μ is unknown. The prior on μ is standard normal. Find the posterior distribution of μ .
13. Suppose the prior is a finite mixture of prior distributions. That is, the parameter θ has prior density

$$\pi(\theta) = \sum_{j=1}^k a_j \pi_j(\theta)$$

The constants a_1, \dots, a_k are called *mixing weights*; they are non-negative and they add up to one.

Show that the posterior distribution is a mixture of the posterior distributions corresponding to $\pi_1(\theta), \dots, \pi_k(\theta)$. What are the mixing weights of the posterior?

This result can be useful if your model has a conjugate prior family, because you can represent virtually any prior opinion by a mixture of conjugate priors. For example, a bimodal prior might be just a mixture of two normals. Thus, you can have essentially any prior you wish, and also the convenience of an exact posterior distribution.

14. Let θ represent the probability that a particular type of cancer, apparently wiped out by chemotherapy, will recur within 2 years. Suppose you *really have no prior idea* about the value of θ . Therefore, you adopt a “non-informative” uniform prior distribution on the interval from zero to one.

Of course, if you have no idea about the probability, you also have no idea about the log odds. The log odds is given by $\ln \frac{\theta}{1-\theta}$. Derive the density of the log odds if θ has a uniform distribution, and use R to plot it. Do you seem to have an idea about what the log odds should be? The moral of this story is that when you adopt a uniform prior, you are still expressing an opinion.