Biost/Stat 533
Emerson, Spring 99
Midterm Examination Key
May 3, 1999

In this key, the answer given in boldface was considered sufficient to receive full credit.

1. Consider a regression model in which response variables $Y_i, i = 1, \ldots, n$ satisfy

$$Y_i = \beta_0 + Z_i\beta_1 + W_i\beta_2 + \epsilon_i$$

with the ϵ_i's independent and identically distributed according to $E(\epsilon_i) = 0$, $Var(\epsilon_i) = \sigma^2$. We consider the situation in which we are interested in making inference about β_1, and we are trying to decide whether to regress $\vec{Y}$ on the full model including $\vec{Z}$ and $\vec{W}$, or whether to regress $\vec{Y}$ on a reduced model including only $\vec{Z}$ as a predictor (in both cases we will include an intercept). Notationally, the reduced model is

$$Y_i = \gamma_0 + Z_i\gamma_1 + \epsilon_i^*$$

and let $\mathbf{X} = (\vec{1}_n \quad \vec{Z} \quad \vec{W})$ and $\mathbf{U} = (\vec{1}_n \quad \vec{Z})$ be the design matrices for the full and reduced models, respectively, with $\widehat{\vec{\beta}}$ and $\widehat{\vec{\gamma}}$ be the ordinary least squares estimates from the corresponding regression models.

a. Without loss of generality, we may assume $\sum_{i=1}^n Z_i = 0$ and $\sum_{i=1}^n W_i = 0$. Why?

Ans: **Centering the variables will change the estimate of the intercept parameter, but will not affect the estimates of the slope.** (See the key to homework #3.)

b. Under what conditions does $\hat{\gamma}_1 = \hat{\beta}_1$?

Ans: **Equality of the estimates holds if the sample correlation r_{ZW} between $\vec{Z}$ and $\vec{W}$ is zero.**

$\widehat{\vec{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\vec{Y}$, and letting $S_{WW} = \sum_{i=1}^n W_i^2$, $S_{ZZ} = \sum_{i=1}^n Z_i^2$, $S_{ZW} = \sum_{i=1}^n Z_iW_i$, $S_{WY} = \sum_{i=1}^n W_iY_i$, $S_{ZY} = \sum_{i=1}^n Z_iY_i$, and $r_{ZW} = S_{ZW}/\sqrt{S_{ZZ}S_{WW}}$,

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} n & 0 & 0 \\ 0 & S_{ZZ} & S_{ZW} \\ 0 & S_{ZW} & S_{WW} \end{pmatrix}$$

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{pmatrix} \frac{1}{n} & 0 & 0 \\ 0 & \frac{1}{S_{ZZ}(1-r_{ZW}^2)} & -\frac{r_{ZW}^2}{S_{ZW}(1-r_{ZW}^2)} \\ 0 & -\frac{r_{ZW}^2}{S_{ZW}(1-r_{ZW}^2)} & \frac{1}{S_{WW}(1-r_{ZW}^2)} \end{pmatrix} \qquad \mathbf{X}^T\vec{Y} = \begin{pmatrix} n\overline{Y} \\ S_{ZY} \\ S_{WZ} \end{pmatrix}$$

$$\widehat{\vec{\beta}} = \begin{pmatrix} \overline{Y} \\ \frac{S_{ZY}}{S_{ZZ}(1-r_{ZW}^2)} - \frac{S_{WY}r_{ZW}^2}{S_{ZW}(1-r_{ZW}^2)} \\ \frac{S_{WY}}{S_{WW}(1-r_{ZW}^2)} - \frac{S_{ZY}r_{ZW}^2}{S_{ZW}(1-r_{ZW}^2)} \end{pmatrix}$$

$\widehat{\vec{\gamma}} = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\vec{Y}$, and

$$\mathbf{U}^T\mathbf{U} = \begin{pmatrix} n & 0 \\ 0 & S_{ZZ} \end{pmatrix}$$

$$(\mathbf{U}^T\mathbf{U})^{-1} = \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{S_{ZZ}} \end{pmatrix} \qquad \mathbf{U}^T\vec{Y} = \begin{pmatrix} n\overline{Y} \\ S_{ZY} \end{pmatrix}$$

$$\widehat{\vec{\gamma}} = \begin{pmatrix} \overline{Y} \\ \frac{S_{ZY}}{S_{ZZ}} \end{pmatrix}$$

By inspection, then, $\hat{\gamma}_1 = \hat{\beta}_1$ if $S_{ZW} = 0$, which in turn implies $r_{ZW} = 0$.

c. What are the expectations of $\widehat{\vec{\beta}}$ and $\widehat{\vec{\gamma}}$? Under what conditions is $\hat{\gamma}_1$ unbiased for β_1?

Ans: **The slope estimate from the reduced model is unbiased for the true slope if $r_{ZW} = 0$ OR $\beta_2 = 0$.**

$E[\vec{Y}] = \mathbf{X}\vec{\beta}$, and as $\mathbf{X} = (\mathbf{U} \quad \vec{W})$ we have

$$\begin{aligned} E[\widehat{\vec{\beta}}] &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\vec{\beta} = \vec{\beta} \\ E[\widehat{\vec{\gamma}}] &= (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{X}\vec{\beta} \\ &= (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\begin{pmatrix}\mathbf{U} & \vec{W}\end{pmatrix}\vec{\beta} \\ &= (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\left(\mathbf{U}\begin{pmatrix}\beta_0 \\ \beta_1\end{pmatrix} + \vec{W}\beta_2\right) \\ &= (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{U}\begin{pmatrix}\beta_0 \\ \beta_1\end{pmatrix} + (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\vec{W}\beta_2 \\ &= \begin{pmatrix}\beta_0 \\ \beta_1\end{pmatrix} + \begin{pmatrix} 0 \\ \frac{S_{ZW}\beta_2}{S_{ZZ}}\end{pmatrix} \end{aligned}$$

By inspection, $\hat{\gamma}_1$ is unbiased for β_1 when $S_{ZW} = 0$ (no linear association between $\vec{Z}$ and $\vec{W}$) or when $\beta_2 = 0$ (no linear association between $\vec{W}$ and $\vec{Y}$).

d. What is the variance of $\widehat{\vec{\beta}}$ and $\widehat{\vec{\gamma}}$? Under what conditions is $Var(\hat{\gamma}_1) = Var(\hat{\beta}_1)$?

Ans: **For the full model $Var(\widehat{\vec{\beta}}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$, and for the reduced model $Var(\widehat{\vec{\gamma}}) = \tau^2(\mathbf{U}^T\mathbf{U})^{-1}$ where τ^2 is the error variance for the reduced model. The variances of the slope estimates are equal if the sample correlation r_{ZW} between $\vec{Z}$ and $\vec{W}$ is zero AND $\sigma^2 = \tau^2$. Note that $\tau^2 = \sigma^2 + \beta_2^2 Var(W_i|Z_i)$, so $\sigma_2 = \tau^2$ if $\beta_2 = 0$ OR $Var(W_i|Z_i) = 0$. The latter condition would hold when we wish to treat the relationship between $\vec{Z}$ and $\vec{W}$ as fixed by design.** *However, it should be noted that when we regard that the relationship between $\vec{Z}$ and $\vec{W}$ is fixed by design, using the variance estimate from the reduced model will estimate $\widehat{Var}(\hat{\gamma}_1)$ as if $\vec{W}$ were a random variable, and thus the variance estimate will be too large. This means that the true type I error will be smaller than desired (a conservative test) and the statistical power will be smaller than desired. Hence, if we design an experiment to have some covariate uncorrelated with our predictor of interest, then it is imperative that we model that predictor if we are to realize the full statistical efficiency of our design.*

For the full model, $Var(\widehat{\vec{\beta}}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$, and from the derivations given in the answer to part (b)

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{S_{ZZ}(1 - r_{ZW}^2)}$$

For the reduced model,

$$Var(\hat{\vec{\gamma}}) = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T Var(\vec{Y}|\vec{Z})\mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1} = Var(\vec{Y}|\vec{Z})(\mathbf{U}^T\mathbf{U})^{-1}$$

and $Var(\vec{Y}|\vec{Z}) = \beta_2^2 Var(W|Z) + \sigma^2$. So

$$Var(\hat{\gamma}_1) = \frac{\sigma^2 + \beta_2^2 Var(W|Z)}{S_{ZZ}}$$

The results follow by inspection.

e. Suppose $\beta_2 = 0$ and $\sum_{i=1}^n Z_i W_i = 0$. What are the relative advantages and disadvantages of choosing the reduced model over the full model?

Ans: **By the above, the inference based on the asymptotic distribution for the slope parameter for $\vec{Z}$ will be the same in the two models. Thus it is immaterial which model we would choose in large samples. Simplicity would argue for the reduced model.** *(In small samples, we will use one degree of freedom to estimate $\hat{\beta}_2$ in the full model, and that will have a very slight effect on the power to make inference about $\hat{\beta}_1$ using the t distribution. Thus in small samples it is preferable to use the reduced model in this situation.)*

f. Suppose $\beta_2 = 0$ and $\sum_{i=1}^n Z_i W_i \neq 0$. What are the relative advantages and disadvantages of choosing the reduced model over the full model?

Ans: **By the above, the slope estimate in the reduced model is unbiased for β_1. Because $S_{WZ} \neq 0$ there will be some variance inflation if the full model is used, with no advantage gained from modeling more of the unexplained error.**

g. Suppose $\beta_2 \neq 0$ and $\sum_{i=1}^n Z_i W_i = 0$. What are the relative advantages and disadvantages of choosing the reduced model over the full model?

Ans: **By the above, the slope estimate in the reduced model is unbiased for β_1. Because $\beta_2 \neq 0$ there will be some reduction in the error variance (estimates, if not the variance itself) if the full model is used, and that will provide greater precision to estimate the slope for the predictor of interest.**

h. Suppose $\beta_2 \neq 0$ and $\sum_{i=1}^n Z_i W_i \neq 0$. What are the relative advantages and disadvantages of choosing the reduced model over the full model?

Ans: **This represents a situation where the estimate of β_1 will be confounded by the relationship between $\vec{Y}$ and $\vec{Z}$ if $\vec{Z}$ is not included in the model. Hence, in the reduced model, $\hat{\gamma}_1$ is biased for β_1. The variance of $\hat{\gamma}_1$ may be larger or smaller than the variance for $\hat{\beta}_1$ depending upon the amount of precision gained (how much σ^2 is less than τ^2) relative to the variance inflation caused by the correlation between $\vec{W}$ and $\vec{Z}$.**

2. Consider an "error in the variables" model in which there is a true relationship between response Y and predictor W given by $Y_i = \beta_1 + \beta_2 W_i + \epsilon_i$ with $\epsilon_i \sim (0, \sigma^2 > 0)$ totally independent. Suppose that W is unobserved, and we instead have X, an imprecise measurement of W which follows the relation $X_i = \alpha_1 + \alpha_2 W_i + \delta_i$, with $\delta_i \sim (0, \tau^2 > 0)$ totally independent of each other and the ϵ's. We thus fit a regression model $Y_i = \gamma_1 + \gamma_2 Z_i + \eta_i$. Show that the ordinary least squares estimate $\hat{\vec{\gamma}}$ is biased as an estimator of $\vec{\beta}$, even when $\alpha_0 = 0$ and $\alpha_1 = 1$. (Hint: What must $\mathbf{X}^T\mathbf{W}$ equal if $\hat{\vec{\gamma}}$ is to be unbiased for $\vec{\beta}$? What does it equal?)

Ans: **Let $\mathbf{X} = (\vec{1}_n \quad \vec{X})$ and $\mathbf{W} = \vec{1}_n \quad \vec{W}$. Then $E[\vec{Y}] = \mathbf{W}\vec{\beta}$. Now OLSE $\hat{\vec{\gamma}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\vec{Y}$, so $E[\hat{\vec{\gamma}}] = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\vec{\beta}$. This then says that $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}$ must be the identity matrix if $\hat{\vec{\gamma}}$ is to be unbiased for $\vec{\beta}$, and thus $\mathbf{X}^T\mathbf{W}$ would have to be equal to $\mathbf{X}^T\mathbf{X}$. Now**

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{pmatrix} \quad \textbf{and} \quad \mathbf{X}^T\mathbf{W} = \begin{pmatrix} n & \sum W_i \\ \sum X_i & \sum X_i W_i \end{pmatrix}$$

Thus we would need $\sum W_i = \sum X_i$ and $\sum X_i W_i = \sum X_i^2$. But

$$\sum X_i = n\alpha_1 + \alpha_2 \sum W_i + \sum \delta_i$$
$$\sum X_i^2 = n\alpha_1^2 + 2\alpha_1\alpha_2 \sum W_i + \alpha_2^2 \sum W_i^2 + 2\alpha_1 \sum \delta_i + 2\alpha_2 \sum \delta_i W_i + \sum \delta_i^2$$
$$\sum X_i W_i = \alpha_1 \sum W_i + \alpha_2 \sum W_i^2 + \sum \delta_i W_i$$

From this, we see that for totally arbitrary α_1 and α_2, in general $\sum W_i \neq \sum X_i$ and $\sum X_i W_i \neq \sum X_i^2$ and $\widehat{\vec{\gamma}}$ is biased for $\vec{\beta}$. Furthermore, even if $\alpha_1 = 0$ and $\alpha_2 = 1$ (as we might hope for a variable merely measured with additive noise),

$$\sum X_i = \sum W_i + \sum \delta_i$$
$$\sum X_i^2 = \sum W_i^2 + 2 \sum \delta_i W_i + \sum \delta_i^2$$
$$\sum X_i W_i = \sum W_i^2 + \sum \delta_i W_i$$

In order to have $\sum W_i = \sum X_i$, we would need $\sum \delta_i = 0$ (an event that would not necessarily be true, but might sometimes happen). In order to have $\sum X_i W_i = \sum X_i^2$, we would need $\sum \delta_i W_i + \sum \delta_i^2 = 0$. Note that we are uninterested in the case that $\sum \delta_i^2 = 0$ because $\tau^2 > 0$. Thus we have $\sum \delta_i (W_i + \delta_i) = \sum \delta_i X_i = 0$, which is not impossible, but is highly unusual (the δ_i's would have to be dependent upon the W_i's. Hence, in general, if our interest is in measuring the association between a response Y and the underlying variable W, regression with a surrogate variable will produce biased estimates of the true relation.

3. Suppose independent response variables $Y_i \sim \mathcal{E}(\lambda_i)$, $\lambda_i > 0$, for $i = 1, \ldots, n$ are distributed according to an exponential distribution with

$$\text{density} \quad f_i(y_i) = \frac{1}{\lambda_i} e^{-y_i/\lambda_i}$$
$$\text{cdf} \quad F_i(y_i) = 1 - e^{-y_i/\lambda_i}$$
$$\text{mean} \quad E[Y_i] = \lambda_i$$
$$\text{variance} \quad Var(Y_i) = \lambda_i^2$$

Recall that in the exponential, λ is a scale parameter such that if $Y \sim \mathcal{E}(\lambda)$ then for $c > 0$, $cY \sim \mathcal{E}(c\lambda)$.

a. Consider a linear regression model with $\lambda_i = \vec{x}_i^T \vec{\beta}$ for known predictor vectors $\vec{x}_i$. Is inference based on the asymptotic normality of least squares estimators of $\vec{\beta}$ valid in this setting? Justify your answer. If it is not valid, briefly describe a regression analysis that would provide asymptotically valid inference for this model.

Ans: **Because there is a mean variance relationship, OLS based inference would only be valid if the sampling of the predictors and the value of $\vec{\beta}$ were such that $\vec{x}_i^T \vec{\beta}$ were the same for all individuals.**

One approach around this problem would be to iteratively use weighted least squares with the current estimate of $\widehat{\vec{\beta}}$ at each iteration used to estimate the covariance matrix for $\vec{Y}$. (see homework #2)

b. Suppose $Z_i = \mu_i + \delta_i$ where μ_i is an unknown parameter and $e^{\delta_i} \sim \mathcal{E}(1)$ are independent. What is the distribution of e^{Z_i}?

Ans: $e^{Z_i} = e^{\mu_i} e^{\delta_i}$ **so** $e^{Z_i} \sim \mathcal{E}(e^{\mu_i})$**, a scaled exponential random variable.**

c. For independent response variables Y_i as above, consider a linear regression model

$$\log(Y_i) = \vec{x}_i^T \vec{\gamma} + \epsilon_i$$

Is inference based on the asymptotic normality of least squares estimators of $\vec{\gamma}$ valid in this setting? Justify your answer. If it is not valid, briefly describe a regression analysis that would provide asymptotically valid inference for this model.

Ans: **Using the result from part (b), we see that $Z_i = \log(Y_i)$ can be written as $Z_i = \log(\lambda_i) + \epsilon_i$ where the ϵ_i's are independent and identically distributed. This suggests that asymptotic inference for $\vec{\beta}$ based on ordinary least squares estimates would be valid. It should be noted that $E[\epsilon_i] = 1 \neq 0$, so the LSE of the intercept is biased, but that will not affect the distribution of the estimates for the slopes.**

4. Consider response variables $Y_i, i = 1, \ldots, n$ and known predictors x_i. Let $x_i^* = x_i - \overline{x}$ be transformed predictors obtained by centering the x_i's about their mean. Consider linear regression models

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$
$$Y_i = \gamma_0 + \gamma_1 x_i^* + \epsilon_i$$

with independent identically distributed errors $\epsilon_i \sim (0, \sigma^2)$. Let $\widehat{\vec{\beta}}$ and $\widehat{\vec{\gamma}}$ be OLSE from the corresponding models.

a. How does $Var(\hat{\beta}_1)$ compare to $Var(\hat{\gamma}_1)$?

Ans: **Centering of the predictors does not affect the estimates of slopes in OLS, thus their distributions must be the same. Hence $Var(\hat{\beta}_1) = Var(\hat{\gamma}_1)$.** (see Homework #2 Key)

b. How does $Var(\hat{\beta}_0)$ compare to $Var(\hat{\gamma}_0)$?

Ans: **Let $\mathbf{X}$ and $\mathbf{X}^*$ be the design matrices for the uncentered and centered models, respectively, and let $S_{XX} = \sum X_i^2$ and $V_{XX} = S_{XX}/n - \overline{X}^2$. Then**

$$Var(\widehat{\vec{\beta}}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} = \sigma^2 \begin{pmatrix} \frac{S_{XX}}{n^2 V_{XX}} & -\frac{\overline{X}}{nV_{XX}} \\ -\frac{\overline{X}}{nV_{XX}} & \frac{1}{nV_{XX}} \end{pmatrix}$$
$$Var(\widehat{\vec{\gamma}}) = \sigma^2(\mathbf{X}^{*T}\mathbf{X}^*)^{-1} = \sigma^2 \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{nV_{XX}} \end{pmatrix}$$

Hence $Var(\hat{\gamma}_0) = \sigma^2/n < Var(\hat{\beta}_0) = \sigma^2 S_{XX}/(n^2 V_{XX})$. *(Note that in the centered model, the intercept is estimating the mean response for the average value of the predictor, and that is the case for which we have greatest precision for estimating the mean response.)*

c. Suppose we want to make inference about the average response when $x = x_0$. Specifically, we wish to test that $E[Y \mid x = x_0] = c$. Describe a hypothesis test based on $\widehat{\vec{\beta}}$ that is asymptotically valid for this setting. (I want formulas, but matrix notation is fine, providing you have defined your notation.)

Ans: **We want to test $H_0 : \beta_0 + x_0\beta_1 = c$. This can be expressed as $H_0 : \mathbf{A}\vec{\beta} = 0$ for $\mathbf{A} = (1 \quad x_0)$. The test then is based on $Q = (\mathbf{A}\widehat{\vec{\beta}} - c)\left(\sigma^2\mathbf{A}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A}^T\right)^{-1}(\mathbf{A}\widehat{\vec{\beta}} - c)$ and rejects H_0 when $Q > \chi^2_{1,1-\alpha}$, where $\chi^2_{1,1-\alpha}$ is the $1-\alpha$ quantile of the chi square distribution with 1 degree of freedom (note $rank(\mathbf{A}) = 1$).**

d. Suppose we were to also construct the hypothesis test of part (c) using $\widehat{\vec{\gamma}}$. Which test is more efficient to make this inference?

Ans: **The models based on the uncentered and centered predictors are merely different parameterizations of each other. Hence, inference about the estimable function $\beta_0 + x_0\beta_1$ is unique and independent of the particular parameterization. Neither model is more efficient than the other.**