Biost/Stat 533
Emerson, Spring 09
Midterm Examination Key
May 13, 2009

This examination is closed book, closed notes. You have 80 minutes to complete the exam. Concise answers are greatly to be preferred, however, in all cases you should provide enough detail to make clear the reasoning behind your answer. Each problem is worth 25 points.

If there are any problems that you believe are not solvable without making additional assumptions, state clearly the (reasonable) assumptions you made in order to solve the problem.

For all problems you may adopt the following notation.

- $\vec{1}_n$ is the n-vector having every element equal to 1.
- $\mathbf{I}_n$ is the n dimensional identity matrix.
- For random n-vector $\vec{X} = (X_1, \ldots, X_n)^T$:

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$S_{XX} = \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 \qquad v_X^2 = \frac{1}{n} S_{XX} \qquad s_X^2 = \frac{n}{n-1} v_X^2$$

- Given two random n-vectors $\vec{X}$ and $\vec{Y}$:

$$S_{XY} = \sum_{i=1}^{n} (X_i - \overline{X}_n)(Y_i - \overline{Y}_n) \qquad r_{XY} = \frac{\frac{1}{n} S_{XY}}{\sqrt{v_X^2 \, v_Y^2}}$$

- Analogous notation can be used for random vectors $\vec{W}$, $\vec{Z}$, etc.

1. Consider a regression model in which response variables $\vec{Y} = (Y_1, \ldots, Y_n)^T$ satisfy

$$\vec{Y} \,|\, \mathbf{X} \;\sim\; (\mathbf{X}\vec{\beta}, \Sigma),$$

with $\Sigma = \sigma^2 \mathbf{I}_n$.

a. Express ordinary least squares estimators (OLSE) $\widehat{\vec{\beta}} = (\hat{\beta}_0,\ \hat{\beta}_1)^T$ in terms of n and the descriptive statistics $\overline{X}_n$, $\overline{Y}_n$, v_X^2, v_Y^2, and r_{YX}.

Ans: **In this simple regression model, I will presume $\mathbf{X}$ is of full rank. $\hat{\beta}_1 = S_{XY}/S_{XX}$ and $\hat{\beta}_0 = \overline{Y}_n - \overline{X}_n\hat{\beta}_1$, so**

$$\widehat{\vec{\beta}} \;=\; \begin{pmatrix} \overline{Y}_n - r_{YX}\frac{v_Y}{v_X}\overline{X}_n \\ r_{YX}\frac{v_Y}{v_X} \end{pmatrix}.$$

OLSE have $\widehat{\vec{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\vec{Y}$. Because I asked for $\hat{\beta}_0$, you may not assume that $\overline{X}_n = 0$ without loss of generality. Straightforward matrix multiplication and matrix inversion yields

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} n & n\overline{X}_n \\ n\overline{X}_n & n(v_X^2 + \overline{X}_n^2) \end{pmatrix} \Rightarrow (\mathbf{X}^T\mathbf{X})^{-1} = \frac{1}{nv_X^2}\begin{pmatrix} v_X^2 + \overline{X}_n^2 & -\overline{X}_n^2 \\ -\overline{X}_n^2 & 1 \end{pmatrix}$$

$$\Rightarrow \widehat{\vec{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\vec{Y} = \begin{pmatrix} \overline{Y}_n - \frac{S_{XY}}{S_{XX}}\overline{X}_n \\ \frac{S_{XY}}{S_{XX}} \end{pmatrix} = \begin{pmatrix} \overline{Y}_n - r_{YX}\frac{v_Y}{v_X}\overline{X}_n \\ r_{YX}\frac{v_Y}{v_X} \end{pmatrix}.$$

Note that in simple linear regression, the assumption that $\mathbf{X}$ is of full rank is just assuming that we do not have a one-sample problem. If rank$(\mathbf{X}) = 1$, then $\widehat{\vec{\beta}}$ is not unique. One obvious choice would be $\widehat{\vec{\beta}} = (\overline{Y}_n, 0)^T$.

b. What are the conditional moments $E[\widehat{\vec{\beta}}\,|\mathbf{X}]$ and $Var(\widehat{\vec{\beta}}\,|\,\mathbf{X})$in terms of n and the descriptive statistics $\overline{X}_n$, $\overline{Y}_n$, v_X^2, v_Y^2, and r_{YX}?

Ans: **Because the errors have mean 0, OLSE are unbiased, and $E[\widehat{\vec{\beta}}] = \vec{\beta}$. Because the errors are uncorrelated, $Var(\widehat{\vec{\beta}}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$, which is found by straightforward matrix multiplication and matrix inversion to be**

$$Var(\widehat{\vec{\beta}}\,|\,\mathbf{X}) = \begin{pmatrix} \frac{\sigma^2(v_X^2 + \overline{X}_n^2)}{nv_x^2} & \frac{-\sigma^2\overline{X}_n^2}{nv_x^2} \\ \frac{-\sigma^2\overline{X}_n^2}{nv_x^2} & \frac{\sigma^2}{nv_x^2} \end{pmatrix}.$$

By linearity of integration,

$$E[\widehat{\vec{\beta}}\,|\mathbf{X}] = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T E[\vec{Y}|\mathbf{X}] = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\vec{\beta} = \vec{\beta},$$

and

$$Var(\widehat{\vec{\beta}} \,|\, \mathbf{X}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \, Var(\vec{Y} \,|\, \mathbf{X}) \, \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}$$
$$= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{I}_n\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}.$$

c. Provide a consistent estimator $\hat{\sigma}^2$ of σ^2 in the setting of part b. Briefly justify your answer.

Ans: **We use**

$$\hat{\sigma}^2 = \frac{1}{n-2}(\vec{Y} - \mathbf{X}\widehat{\vec{\beta}})^T(\vec{Y} - \mathbf{X}\widehat{\vec{\beta}}) = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2.$$

The consistency of this estimate follows from being unbiased and its variance decreasing to 0 as $n \to \infty$. *The latter places some restriction on the distribution of the errors and the sampling of the X_i's, but those restrictions are basically the same as required for the asymptotic normality of the OLSE.*

d. Describe a hypothesis test of $H_0 : \beta_1 = 0$. Briefly describe any necessary assumptions for your test to be statistically valid and the sense in which it might be optimal.

Ans: **From the CLT for simple linear regression as proved in class, we know that as $n \to \infty$ with the eigenvalues of $\mathbf{X}^T\mathbf{X}$ all going to infinity,**

$$\widehat{\vec{\beta}} \,\dot{\sim}\, \mathcal{N}_2\left(0, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\right).$$

The marginal distribution of $\hat{\beta}_1$ is therefore approximately

$$\hat{\beta}_1 \,\dot{\sim}\, \mathcal{N}\left(\beta_1, \frac{\sigma^2}{nv_X^2}\right),$$

and we can use a test statistic

$$Z = \sqrt{n}v_X\frac{\hat{\beta}_1}{\hat{\sigma}} \,\dot{\sim}\, \mathcal{N}(0,1) \quad \textbf{under } H_0.$$

We would therefore reject H_0 in a two-sided test if $|Z| > z_{1-\alpha/2}$. The test is consistent in the sense that as $n \to \infty$, the power approaches 1 under every alternative. Even in the absence of a straight line relationship among the means, this test is interpretable as a test of a first order trend in the mean of Y. *While you could have constructed a test based on the quadratic form, such a test is exactly equivalent to the above formulation. Because the OLSE are BLUE, this test might be expected to tend to be most powerful among tests based on linear unbiased estimators. The overall efficiency of the test will depend upon the true distribution of the ϵ_i. I note that we typically use a t distribution instead* *of a standard normal distribution for Z, because when the errors are normally distributed, $(n-2)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-2}$ independently of $\widehat{\vec{\beta}}$, and the t distribution is exact. The only real rationale for this approach in the more general distribution-free case is that the "small sample adjustment" of the t distribution for the critical value is necessary when the data are normal, so we might as well use that as a standard approach. As proven in HW #1, the t distribution converges in distribution to a standard normal as $n \rightarrow \infty$.*

e. Describe a hypothesis test of $H_0 : \beta_0 = 0$. Briefly describe any necessary assumptions for your test to be statistically valid and the sense in which it might be optimal.

Ans: **As in part d, we can construct a test based on the asymptotic normal distribution of $\hat{\beta}_0$:**

$$Z = \frac{\hat{\beta}_0}{\widehat{se}(\hat{\beta}_0)},$$

which is distributed according to the standard normal under the null hypothesis. All of the comments about optimality and validity in part d apply here plus the need for linearity to hold across the group means.

2. Consider again the regression model of problem 1, but now assume $\Sigma = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$.

a. What are the conditional (on $\mathbf{X}$) moments of the OLSE $\widehat{\vec{\beta}}$ in this setting?

Ans: **The OLSE are conditionally unbiased, so $E[\widehat{\vec{\beta}}\,|\,\mathbf{X}] = \vec{\beta}$. The variance of $\widehat{\vec{\beta}}$ is easily found from the results for linear transformations**

$$Var(\widehat{\vec{\beta}}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\Sigma\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}.$$

b. What does the estimator $\hat{\sigma}^2$ from problem 1c consistently estimate in this setting?

Ans: **The estimator $\hat{\sigma}^2$ will tend toward $\sum_{i=1}^n \sigma^2/n$ (presuming that tends toward some constant).**

c. What would be the impact in terms of optimality and statistical validity of using the hypothesis test described in problem 1d in this setting?

Ans: **The OLSE estimator used in problem 1d was BLUE for the case of independent, homoscedastic observations. In the presence of heteroscedasticity, that test will not be based on a BLUE. Also that test presumes equal variability across groups when estimating the standard error of $\widehat{\vec{\beta}}$. When this does not hold, the statistical test may be conservative, anticonservative, or approximately correct depending upon the trends in σ_i^2**

relative to the X_i's. *Drawing on the results from a homework problem, if there is a linear trend toward higher variability of measurements in the same direction as any skewness in X, the test will be anti-conservative. If the first order trend is toward higher variability of measurements in the opposite direction as the skewness of X, the test will tend to be conservative. If there is no linear trend in the variability, or if the distribution of the X_i's is not skewed, the inference will be approximately valid, but it will still be inefficient.*

d. Describe more optimal, statistically valid inference for testing $H_0 : \beta_1 = 0$ in this setting. Be sure to describe the sense in which your inference is optimal.

Ans: **If Σ is known or estimable from the data, we would rather base our inference on the GLSE**

$$\widehat{\vec{\beta}}_G = (\mathbf{X}^T\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}^T\Sigma^{-1}\vec{Y},$$

which will be approximately normally distributed with mean $\vec{\beta}$ and variance $(\mathbf{X}^T\Sigma^{-1}\mathbf{X})^{-1}$.

3. We are interested in designing an experiment to estimate $\vec{\beta}$ in the simple regression settings of problems 1 and 2. We are able to sample subjects at J distinct values of X, $a_1 < a_2 < \cdots < a_J$. We are interested in determining an optimal number of subjects $n_1, n_2, \ldots, n_J$ to sample at each of the J levels in order to have the most precise inference about β_1.

a. What is the optimal design under the setting of problem 1 (so $\Sigma = \sigma^2\mathbf{I}_n$) using an optimal testing strategy ? Justify your answer.

Ans: **We know the squared standard error of $\hat{\beta}_1$ is**

$$Var(\hat{\beta}_1 \,|\, \mathbf{X}) = \frac{\sigma^2}{nv_X^2},$$

so we merely need to sample in order to maximize the variance of the X_i's. This is effected by sampling $n/2$ subjects with $X_i = a_1$ and $n/2$ subjects with $X_i = a_J$.*Hence, we do best when we use a t test in which the two groups have the largest possible difference in means: $\beta_1(a_J - a_1)$.*

b. What is the optimal design under the setting of problem 2 (so $\Sigma = diag(\sigma_1^2, \ldots, \sigma_n^2)$) using an optimal testing strategy ? Justify your answer.

Ans: **When using GLSE, we know the squared standard error of GLSE $\widehat{\vec{\beta}}$ is $(\mathbf{X}^T\Sigma^{-1}\mathbf{X})^{-1}$, which leads to $Var(\hat{\beta}_1) = 1/W_X^2$, where W_X^2 is related to the weighted variance:**

$$W_X^2 = \sum_{j=1}^{J} \frac{n_j a_j^2}{\sigma_j^2} - \frac{\left(\sum_{j=1}^{J} \frac{n_j a_j}{\sigma_j^2}\right)^2}{\sum_{j=1}^{J} \frac{n_j}{\sigma_j^2}}.$$

In maximizing that weighted variance, we might intuitively deduce that a t test will be the best choice, and recall from HW #1 that sample sizes should be chosen in proportion to standard deviations. Then we want to choose j and k to maximize

$$\frac{\beta_1^2(a_k - a_j)^2}{\sigma_k^2/n_k + \sigma_j^2/n_j},$$

where $n_k = n\sigma_k/(\sigma_k + \sigma_j)$ and $n_j = n\sigma_j/(\sigma_k + \sigma_j)$. Hence we choose j and k to maximize

$$\frac{(a_k - a_j)^2}{(\sigma_k + \sigma_j)^2}.$$

I did not really expect you to work out all of the above. I was just hoping that there would be a recognition that GLSE was the way to go and that the results from the t test were applicable in some way.

c. What scientific issues might you have when using either of the above designs?

Ans: **The major scientific issue is that we are probably not really willing to presume that a straight line relationship holds exactly. Of course, we also probably do not know the within group variances.**

4. Consider again the setting of problem 2 (with an arbitrary covariance matrix).

   a. Provide a definition for an estimable function.

   b. State and prove the Gauss-Markov Theorem in the setting of an arbitrary covariance matrix for $\vec{Y}$.

5. Consider again the setting of problem 1, but suppose that instead of conditioning on $\mathbf{X}$, we desire inference in the setting that $\vec{X} \sim (\mu\vec{1}_n, \tau^2\mathbf{I}_n)$. Derive an expression for the unconditional moments $E(\hat{\beta}_1)$ and $Var(\hat{\beta}_1)$. How does the sampling distribution of $\vec{X}$ affect the difference between the conditional and unconditional moments?

Ans: **By the double expectation formula, we know**

$$E[\hat{\beta}_1] = E_{\mathbf{X}}[\ E[\hat{\beta}_1 \,|\, \mathbf{X}]\ ] = E_{\mathbf{X}}[\beta_1] = \beta_1.$$

Using the formula for unconditional variance based on conditional moments, we have

$$\begin{aligned} Var(\hat{\beta}_1) &= E_{\mathbf{X}}[\ Var(\hat{\beta}_1 \,|\, \mathbf{X})\ ] + Var_{\mathbf{X}}(\ E[\hat{\beta}_1 \,|\, \mathbf{X}]\ ) \\ &= E_{\mathbf{X}}\left[\ \frac{\sigma^2}{nVar(X)}\ \right] + Var_{\mathbf{X}}(\ \beta_1\ ) \\ &= E_{\mathbf{X}}\left[\ \frac{\sigma^2}{nVar(X)}\ \right] \end{aligned}$$

So the sampling distribution of $\vec{X}$ does not affect the unconditional expectation of $\hat{\beta}_1$ at all, and it only affects the unconditional standard error through the sampling distribution of the observed $Var(X)$ in each sample.

6. Consider an interventional experiment in which baseline measurements Y_{Bi} ('B' for baseline) are obtained prior to the start of the experiment on each subject, and the subjects are then randomized to either receive an experimental treatment (in which case covariate $X_i = 1$) or some control experiment (in which case covariate $X_i = 0$). Final measurements Y_{Fi} ('F' for final) are obtained at the completion of the therapy. We are ultimately interested in the effect of treatment on the mean value of Y_{Fi} for each treatment group.

   Let $Var(Y_{Bi} \,|\, X_i) = Var(Y_{Fi} \,|\, X_i) = \sigma^2$, and suppose $Corr(Y_{Bi}, Y_{Fi} \,|\, X_i) = \rho$, with all other pairs of observations being uncorrelated. Assume blocked randomization in which $corr(Y_{Bi}, X_i) = 0$. (Hint: What does randomization say about the conditional distribution of $Y_{Bi}|X_i$?)

   We consider three different regression analysis models:

A. $Y_{Fi} \,|\, X_i = \alpha_0 + \alpha_1 X_i + \epsilon_{Ai}$, with $\epsilon_{Ai} \sim (0, \sigma_A^2)$,

B. $D_i \,|\, X_i = \beta_0 + \beta_1 X_i + \epsilon_{Bi}$, with $D_i \equiv Y_{Fi} - Y_{Bi}$ and $\epsilon_{Bi} \sim (0, \sigma_B^2)$,

C. $W_i \,|\, X_i = \gamma_0 + \gamma_1 X_i + \epsilon_{Ci}$, with $W_i \equiv Y_{Fi} - wY_{Bi}$ for some w and $\epsilon_{Ci} \sim (0, \sigma_C^2)$,

   a. What is the conditional sampling distribution for OLSE $\hat{\alpha}_1$ when using Model A?

   Ans: **Linear regression on a binary predictor is equivalent to a t test, with α_1 representing the difference in means $E[Y_{Fi} \,|\, X_i = 1] - E[Y_{Fi} \,|\, X_i = 0])$, $\hat{\alpha}_1$ being the difference in sample means, and $se^2(\alpha_1) = \sigma_A^2(1/n_0 + 1/n_1)$. The OLSE parameter estimate $\hat{\alpha}_1$ is unbiased for α_1, $\sigma_A^2 = \sigma^2$ is just the within group variance of $Y_{Fi}|X_i$, and the sampling distribution is asymptotically normal.**

   b. What is the conditional sampling distribution for OLSE $\hat{\beta}_1$ when using Model B?

   Ans: **Linear regression on a binary predictor is equivalent to a t test, with β_1 representing the difference in mean change $E[D_i \,|\, X_i = 1] - E[D_i \,|\, X_i = 0])$, $\hat{\beta}_1$ being the difference in sample mean change, and $se^2(\beta_1) = \sigma_B^2(1/n_0 + 1/n_1)$. The OLSE parameter estimate $\hat{\beta}_1$ is unbiased for β_1, $\sigma_B^2 = 2\sigma^2(1-\rho)$ is just the within group variance of $D_i|X_i$, and the sampling distribution is asymptotically normal. It should be noted that $E[D_i \,|\, X_i] = E[Y_{Fi} \,|\, X_i] - E[Y_{Bi} \,|\, X_i]$ and by randomization $E[Y_{Bi} \,|\, X_i = 1] = E[Y_{Bi} \,|\, X_i = 0]$. Hence $\beta_1 = \alpha_1$.**

   c. What is the conditional sampling distribution for OLSE $\hat{\gamma}_1$ when using Model C? What would be the optimal choice of w?

Ans: **Linear regression on a binary predictor is equivalent to a t test, with γ_1 representing the difference in mean adjusted change $E[W_i \mid X_i = 1] - E[W_i \mid X_i = 0]$), $\hat{\gamma}_1$ being the difference in sample mean adjusted change, and $se^2(\gamma_1) = \sigma_C^2(1/n_0 + 1/n_1)$. The OLSE parameter estimate $\hat{\gamma}_1$ is unbiased for γ_1, $\sigma_C^2 = \sigma^2 + w^2\sigma^2 - 2w\rho\sigma^2$ is just the within group variance of $W_i|X_i$, and the sampling distribution is asymptotically normal. The minimal value for σ_C^2 is found by differentiating with respect to w, to yield $w = \rho$. For that choice of w, $\sigma_C^2 = \sigma^2(1-\rho^2)$. It should again be noted that by randomization $E[Y_{Bi} \mid X_i = 1] = E[Y_{Bi} \mid X_i = 0]$. Hence $\gamma_1 = \beta_1 = \alpha_1$.**

e. How do the conditional expectations of the above models relate to the scientifically relevant parameter?

Ans: **As noted above, $\gamma_1 = \beta_1 = \alpha_1$, and they all estimate the scientifically relevant parameter.**

d. Suppose that the value of w has to be estimated. Describe a linear regression model that might be expected to provide similar precision.

Ans: **From problem 1, we note that the regression parameter is related to the correlation between the outcome and the predictor. Hence, we could merely perform a linear regression of Y_{Fi} or D_i on both X_i and Y_{Bi}. The parameter estimate for the baseline term will estimate ρ in this setting where we presumed a common variance for the baseline and final measurements.**

e. What is the optimal analysis approach for this data?

Ans: **Model C is the best, because $\sigma_C^2 \leq \sigma_B^2$ and $\sigma_C^2 \leq \sigma_A^2$, no matter what the value of ρ. The ANCOVA model regressing on the baseline would be optimal when the value of ρ is unknown.**

7. Consider the setting of problem 1: $\vec{Y} \mid \mathbf{X} \sim (\mathbf{X}\vec{\beta}, \sigma^2\mathbf{I}_n)$.

a. Derive a relationship between the unconditional $Var(Y_i)$, the conditional $Var(Y_i \mid X_i)$, and the correlation between Y_i and X_i.

Ans: **Using the formula for unconditional variance related to the conditional moments:**

$$
\begin{aligned}
Var(Y_i) &= E_X[\ Var(Y_i \mid X_i)\ ] + Var_X(\ E[Y_i \mid X_i]\ ) \\
&= E_X[\ \sigma^2\ ] + Var_X(\ \beta_0 + \beta_1 X_i\ ) \\
&= \sigma^2 + \beta_1^2 v_X^2 = \sigma^2 + \rho^2 \frac{v_Y^2}{v_X^2} v_X^2
\end{aligned}
$$

,

which yields

$$\frac{Var(Y \mid X)}{Var(Y)} = 1 - \rho^2.$$

b. Provide the OLSE for a model of $Y_i = \gamma_1 X_i + \eta_i$ (that is, fitting no intercept).

Ans: **Using standard formulas for OLSE**

$$\hat{\gamma}_1 = (\vec{X}^T \vec{X})^{-1} \vec{X}^T \vec{Y} = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2} = \frac{S_{XY} + n\overline{X}\ \overline{Y}}{S_{XX} + n(\overline{X})^2}.$$

c. Contrast the scientific interpretation of a test of $H_0 : \beta_1 = 0$ from problem 1 and a test of $H_0 : \gamma_1 = 0$ from the regression fit in part b of this problem.

Ans: **In problem 1, when $\beta_1 = 0$, we presume that every group would have some common mean (which we would estimate by $\overline{Y}$). In part b of this problem, we presume that when $\gamma_1 = 0$, every group would have mean 0.**

d. Contrast the relationship between the slope estimates from problem 1 and part b or this problem and the correlation between Y_i and X_i.

Ans: **In problem 1, the slope is directly related to the correlation between Y and X. When we leave out an intercept, the slope is more driven by assumption that the line goes through the origin.** *Note that when regressing without an intercept, the R^2 term will be considering how the model does better than predicting that all groups have mean 0, rather than comparing the slope to a model with a flat line through an arbitrary mean.*