Biost/Stat 533
Emerson, Spr 99
Homework #2 Key
April 27, 1999

1. Suppose $Y_{0i} \sim (\mu_0, \sigma_0^2)$ for $i = 1, \ldots, n_0$ and $Y_{1i} \sim (\mu_1, \sigma_1^2)$ for $i = 1, \ldots, n_1$, with $Cov(Y_i, Y_j) = 0$ for $i \neq j$ and $0 < \sigma_k^2 < \infty$ for $k = 0, 1$. Let $\overline{Y}_0 = \sum_{i=1}^{n_0} Y_{0i}/n_0$ and $\overline{Y}_1 = \sum_{i=1}^{n_1} Y_{1i}/n_1$ denote the sample means for each group, and let $s_0^2 = \sum_{i=1}^{n_0}(Y_{0i} - \overline{Y}_0)^2/(n_0 - 1)$ and $s_1^2 = \sum_{i=1}^{n_1}(Y_{1i} - \overline{Y}_1)^2/(n_1 - 1)$ denote the sample variances for each group, with pooled variance estimate $s_p^2 = ((n_0 - 1)s_0^2 + (n_1 - 1)s_1^2)/(n_0 + n_1 - 2)$. Suppose further that as $n_0 \to \infty$ and $n_1 \to \infty$, $n_0/(n_0 + n_1) \to \lambda$ for some known constant $0 < \lambda < 1$. We are interested in making inference about $\delta = \mu_1 - \mu_0$.

   Under the assumption of equal variances (so $\sigma_0^2 = \sigma_1^2$), the usual statistical inference for this problem is based on the two sample t test for independent samples assuming equal variances using test statistic $T_e = (\overline{Y}_1 - \overline{Y}_0)/(s_p\sqrt{1/n_0 + 1/n_1})$ and assuming that T_e is distributed according to the t distribution with $n_0 + n_1 - 2$ degrees of freedom under the null hypothesis $H_0 : \delta = 0$. A confidence interval is constructed by inverting that test statistic.

   Under the assumption of unequal variances (so $\sigma_0^2 \neq \sigma_1^2$), the usual statistical inference for this problem is based on the two sample t test for independent samples assuming unequal variances using test statistic $T_u = (\overline{Y}_1 - \overline{Y}_0)/\sqrt{s_0^2/n_0 + s_1^2/n_1}$ and assuming that T_u is distributed according to the t distribution with k degrees of freedom under the null hypothesis $H_0 : \delta = 0$, where k might be determined by the Satterthwaite or Aspin-Welch approximations. A confidence interval is constructed by inverting that test statistic.

   a. Show that under the assumption of equal variances, the statistical inference based on T_e is asymptotically correct in that the size of the hypothesis test is asymptotically at the correct level and the confidence interval has the correct coverage probability asymptotically.

   Ans: It is easiest in the long run if I solve for some general results assuming unequal variances first, and then at a later stage consider equal variances. So, based on the Levy CLT, I know

$$\sqrt{n_0}(\overline{Y}_0 - \mu_0) \to_d \mathcal{N}(0, \sigma_0^2)$$
$$\sqrt{n_1}(\overline{Y}_1 - \mu_1) \to_d \mathcal{N}(0, \sigma_1^2)$$

   and since $n_0/n \to \lambda$ and $n_1/n \to 1 - \lambda$, by Slutsky's theorem we have

$$\sqrt{n}(\overline{Y}_0 - \mu_0) \to_d \mathcal{N}(0, \frac{\sigma_0^2}{\lambda})$$
$$\sqrt{n}(\overline{Y}_1 - \mu_1) \to_d \mathcal{N}(0, \frac{\sigma_1^2}{1 - \lambda})$$

   Thus by independence we get

$$\sqrt{n}D_n \equiv \sqrt{n}(\overline{Y}_1 - \overline{Y}_0 - (\mu_1 - \mu_0)) \to_d \mathcal{N}(0, \frac{\sigma_0^2}{\lambda} + \frac{\sigma_1^2}{1 - \lambda}) \quad (1.1)$$

   Now by WLLN we know for $j = 0, 1$

$$\frac{1}{n_j}\sum_{i=1}^{n_j}(Y_{ji} - \mu_j)^2 \to_p \sigma_j^2$$

   We also have

$$\frac{1}{n_j}\sum_{i=1}^{n_j}(Y_{ji} - \mu_j)^2 = \frac{1}{n_j}\sum_{i=1}^{n_j}(Y_{ji} - \overline{Y}_j)^2 + (\overline{Y}_j - \mu_j)^2$$

and by WLLN $\overline{Y}_j \to_p \mu_j$, so $(\overline{Y}_j - \mu_j)^2 \to_p 0$. Thus using Slutsky's along with the fact that $n/(n-1) \to_p 1$, we have

$$s_j^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (Y_{ji} - \mu_j)^2 \to_p \sigma_j^2$$

Using this, we find

$$\sqrt{n}\sqrt{\frac{s_0^2}{n_0} + \frac{s_1^2}{n_1}} \to_p \sqrt{\frac{\sigma_0^2}{\lambda} + \frac{\sigma_1^2}{1-\lambda}} \tag{1.2}$$

and

$$s_p^2 = \frac{(n_0 - 1)s_0^2 + (n_1 - 1)s_1^2}{n - 2} \to_p \lambda\sigma_0^2 + (1-\lambda)\sigma_1^2$$

so

$$\begin{aligned} \sqrt{n}s_p\sqrt{\frac{1}{n_0} + \frac{1}{n_1}} &= \sqrt{s_p^2(\frac{n}{n_0} + \frac{n}{n_1})} \\ &\to_p \sqrt{\frac{\sigma_0^2}{1-\lambda} + \frac{\sigma_1^2}{\lambda}} \end{aligned} \tag{1.3}$$

Now combining results (1.1) and (1.3), we find by Slutsky's that under the null hypothesis $H_0 : \mu_0 = \mu_1$

$$T_e = \frac{\sqrt{n}D_n}{\sqrt{n}s_p\sqrt{\frac{1}{n_0} + \frac{1}{n_1}}} \to_d \mathcal{N}\left(0, \frac{\frac{\sigma_0^2}{\lambda} + \frac{\sigma_1^2}{1-\lambda}}{\frac{\sigma_0^2}{1-\lambda} + \frac{\sigma_1^2}{\lambda}}\right) \tag{1.4}$$

and combining results (1.1) and (1.2), we find by Slutsky's that under the null hypothesis $H_0 : \mu_0 = \mu_1$

$$T_u = \frac{\sqrt{n}D_n}{\sqrt{n}\sqrt{\frac{s_0^2}{n_0} + \frac{s_1^2}{n_1}}} \to_d \mathcal{N}\left(0, \frac{\frac{\sigma_0^2}{\lambda} + \frac{\sigma_1^2}{1-\lambda}}{\frac{\sigma_0^2}{\lambda} + \frac{\sigma_1^2}{1-\lambda}}\right) = \mathcal{N}(0, 1) \tag{1.5}$$

Now to answer part a, we find from (1.4) that when $\sigma_0^2 = \sigma_1^2 = \sigma^2$, $T_e \to_d \mathcal{N}(0, 1)$. The usual inference for T_e is based on a t distribution with $n-2$ degrees of freedom, which asymptotically is equivalent to the standard normal, so the usual inference in this case is valid.

b. Show that under the assumption of equal variances, the statistical inference based on T_u is asymptotically correct.

Ans: From (1.5) we find that T_u is asymptotically standard normal no matter whether the variances are equal or not. Hence, as the t with k degrees of freedom is asymptotically standard normal, and as the degrees of freedom for the Aspin-Welch test is bounded below by the smaller of $n_0 - 1$ and $n_1 - 1$, we know that the usual inference based on T_u is asymptotically valid.

c. Show that under the assumption of unequal variances, the statistical inference based on T_u is asymptotically correct.

Ans: Same answer as part b.

d. Show that under the assumption of unequal variances, the statistical inference based on T_e is not necessarily asymptotically correct. Under what conditions will inference based on T_e be asymptotically valid in this setting? Under what conditions will it be conservative? anti-conservative?

Ans: When the group variances are not equal, we see from (1.4) that the asymptotic variance of T_e is 1 only if $\lambda = 1/2$. Otherwise, the asymptotic variance is

$$\frac{(1-\lambda)\sigma_0^2 + \lambda\sigma_1^2}{\lambda\sigma_0^2 + (1-\lambda)\sigma_1^2}$$

Now the inference will be conservative if the true variance is less than 1, and the inference will be anti-conservative if the true variance is greater than 1. We find that the true variance is greater than 1 precisely when

$$(1-2\lambda)\sigma_0^2 > (1-2\lambda)\sigma_1^2$$

This occurs when $\lambda < 0.5$ and $\sigma_0^2 > \sigma_1^2$ or when $\lambda > 0.5$ and $\sigma_0^2 < \sigma_1^2$. That is, the t test based on equal variances is asymptotically anti-conservative when the group with the largest variance has a smaller sample size than the group with the smallest variance. If we sample the group with the larger variance in greater numbers than the group with the smaller variance, the use of T_e leads to conservative inference (the true type I error is smaller than the nominal level, the coverage probability of confidence intervals is greater than the stated level, and the statistical power to detect an alternative hypothesis is diminished).

e. What do the above results suggest about the validity of regression based on linear regression models in the presence of heteroscedasticity?

Ans: The t test with equal variances is the special case of unweighted linear regression in which the single predictor is a binary variable. Because the unweighted linear regression model produces variance estimates based on the assumption of equal error variance across all observations, the above results tell us that that inference will be incorrect in the absence of a balanced design.

2. Consider the simple linear regression model $Y_i = \beta_0 + x_i\beta_1 + \epsilon_i$ for $i = 1, \ldots, n$, with x_i known predictors, $\vec{\beta} = (\beta_0, \beta_1)^T$ an unknown parameter vector to be estimated and/or tested, and $Cov(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$. Without loss of generality, we will assume that $\sum_{i=1}^n x_i = 0$. Let $\sigma_i^2 = \alpha_i + x_i\gamma > 0$ with γ and $\vec{\alpha}$ unknown nuisance parameters subject to $\vec{\alpha}^T\vec{x} = 0$. Let $\widehat{\vec{\beta}}$ be the ordinary least squares estimate of $\vec{\beta}$.

a. What is the mean and variance of $\hat{\beta}_1$?

Ans: In this problem with $\overline{Y} = \sum_{i=1}^n Y_i/n$, $S_{xx} = \sum_{i=1}^n x_i^2$, and $S_{xY} = \sum_{i=1}^n x_iY_i$

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

so

$$\widehat{\vec{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\vec{Y} = \begin{pmatrix} \overline{Y} \\ \frac{S_{xY}}{S_{xx}} \end{pmatrix}$$

$E[\widehat{\vec{\beta}}] = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T E[\vec{Y}] = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\vec{\beta} = \vec{\beta}$ and the variance of $\widehat{\vec{\beta}}$ is found by

$$\mathbf{X}^T\mathbf{V}\mathbf{X} = \begin{pmatrix} \sum\alpha_i + \gamma\sum x_i & \sum\alpha_i x_i + \gamma\sum x_i^2 \\ \sum\alpha_i x_i + \gamma\sum x_i^2 & \sum\alpha_i x_i^2 + \gamma\sum x_i^3 \end{pmatrix} = \begin{pmatrix} \sum\alpha_i & \gamma\sum x_i^2 \\ \gamma\sum x_i^2 & \sum\alpha_i x_i^2 + \gamma\sum x_i^3 \end{pmatrix}$$

$$\begin{aligned} Var(\widehat{\vec{\beta}}) &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\ &= \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{S_{xx}} \end{pmatrix} \begin{pmatrix} \sum\alpha_i & \gamma\sum x_i^2 \\ \gamma\sum x_i^2 & \sum\alpha_i x_i^2 + \gamma\sum x_i^3 \end{pmatrix} \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{S_{xx}} \end{pmatrix} \\ &= \begin{pmatrix} \sum\alpha_i/n^2 & \gamma/n \\ \gamma/n & \sum\alpha_i x_i^2/S_{xx}^2 + \gamma\sum x_i^3/S_{xx}^2 \end{pmatrix} \end{aligned}$$

From this we find that $E[\hat{\beta}_1] = \beta_1$ and $Var(\hat{\beta}_1) = (\sum\alpha_i x_i^2 + \gamma\sum x_i^3)/S_{xx}^2$.

b. Under what conditions will the estimated variance of $\hat{\beta}_1$ based on the ordinary least squares regression analysis be consistent for the true variance of $\hat{\beta}_1$.

Ans: Under OLS, we assume that the errors have constant variance, and our estimate

$$\hat{\sigma}^2 = \frac{1}{n-p}(\vec{Y} - \mathbf{X}\widehat{\vec{\beta}})^T(\vec{Y} - \mathbf{X}\widehat{\vec{\beta}})$$

consistently estimates the limit of $\sum_{i=1}^{n}(\alpha_i + \gamma x_i)/n = \sum_{i=1}^{n} \alpha_i/n = \overline{\alpha}$, assuming such a limit exists. Let α be that limit. Then, the OLSE variance estimate of $\widehat{\vec{\beta}}$ is

$$\widehat{Var}(\widehat{\vec{\beta}}) = \hat{\sigma}^2(\mathbf{X}^T\mathbf{X})^{-1} = \overline{\alpha} \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{S_{xx}} \end{pmatrix}$$

and the variance estimate for $\hat{\beta}_1$ is therefore just

$$\widehat{Var}(\hat{\beta}_1) = \frac{\overline{\alpha}}{\sum_{i=1}^{n} x_i^2} = \frac{\overline{\alpha}}{nV_x}$$

where $V_x = S_{xx}/n$ is the variance of the x_i's.

From part (a), we find

$$Var(\hat{\beta}_1) = \frac{\sum \alpha_i x_i^2}{n^2 V_x^2} + \frac{\gamma \sum x_i^3}{n^2 V_x^2}$$

From this we see that $\widehat{Var}(\hat{\beta}_1)$ will tend toward $Var(\hat{\beta}_1)$ certainly when $\alpha_i \equiv \alpha$ for all i and either $\gamma = 0$ or the distribution of the x_i's is unskewed.

Now it is stipulated that $\sum \alpha_i x_i = 0$. If this were to be strengthened to the case that the α_i's are also sampled independently of the x_i's (and note that $\sum \alpha_i x_i = 0$ merely suggests that they are uncorrelated, not necessarily independent), then $\sum \alpha_i x_i^2/n = (\sum \alpha_i/n)(\sum x_i^2/n)$. Thus

$$Var(\hat{\beta}_1) = \frac{\overline{\alpha}}{nV_x} + \frac{\gamma \sum x_i^3}{n^2 V_x^2}$$

and $\widehat{Var}(\hat{\beta}_1)$ will tend toward $Var(\hat{\beta}_1$ when either $\gamma = 0$ or the distribution of the x_i's is unskewed. Notice that the requirement for the x_i's to be unskewed is analogous to the equal sample sizes required for the answer to problem 1(d) above.

c. What restrictions on the problem would be necessary for $\widehat{\vec{\beta}}$ to be asymptotically normally distributed? (You need not rigorously derive an asymptotic distribution, instead just briefly discuss the ways that this setting differs from the assumptions under which we derived the asymptotic distribution in class, and what general requirements might address those problems.)

Ans: When the errors are uncorrelated and identically distributed, in order to derive the asymptotic normal results for the OLSE we had to place restrictions on the sampling of the x_i's to ensure that the contribution of any particular x_i to the total variance of the x_i's was negligible as $n \rightarrow \infty$ in simple linear regression. This restriction translates into a requirement that the smallest eigenvalue of $\mathbf{X}^T\mathbf{X}$ tend to infinity as $n \rightarrow \infty$, along with some requirements that none of the cases sampled be too influential (see class notes). In this more general case we must be concerned in the way that we sample the x_i's <u>and</u> the ϵ_i's. This will translate into a requirement that the smallest eigenvalue of $\mathbf{X}^T\mathbf{V}^{-1}\mathbf{V}$ approach infinity as $n \rightarrow \infty$, along with some requirements that the most influential cases not also tend to have the errors with the largest variance.

d. What would be the effect of using the asymptotic results for ordinary least squares regression analysis on tests of $H_0 : \hat{\beta}_1 = 0$? Consider the effect that the variance of the α_is and the value of γ has on your answer.

Ans: Assuming that asymptotic normality holds, we need only worry about when the variance estimate under OLSE over- or underestimates the true variance. When the variance estimate underestimates the true variance, tests of H_0 will be anti-conservative in that the true type I error will be larger than desired. When the variance estimate overestimates the true variance, tests of H_0 will tend to have a smaller type I error than desired.

When the α_i's are independent of the x_i's (which would include the case when $Var(\alpha_i) = 0$), then the variance estimate $\widehat{Var}(\hat{\beta}_1)$ will underestimate the true variance when γ and the skewness of the x_i's are of the same sign, and it will overestimate the true variance when γ and the skewness of the x_i's are of opposite sign. For instance, if the distribution of the x_i's is positively skewed, then a tendency for larger variance with larger values of x_i will lead to anti-conservative tests, while a tendency for smaller variance with larger values of x_i will lead to conservative tests (and loss of statistical power).

When $\gamma = 0$ or the distribution of the x_i's is unskewed, then the variance estimate $\widehat{Var}(\hat{\beta}_1)$ will underestimate the true variance when the weighted average of the α_i's based on weights x_i^2 is greater than $\overline{\alpha}$ and will overestimate the true variance when the weighted average is less than $\overline{\alpha}$. The weighted average will tend to be greater than $\overline{\alpha}$ when the more extreme values of x_i are associated with larger α_i. Consider for example the simple example where $n = 5$ and $\vec{x} = (-2, -1, 0, 1, 2)^T$ (so $\sum x_i = 0$). Now if $\vec{\alpha} = (3, 1, 2, 1, 3)^T$ (so $\sum \alpha_i x_i = 0$, but the α_i's are not independent of the x_i's), then $\overline{\alpha} = 2$, but the weighted average $(\sum \alpha_i x_i^2)/(\sum x_i^2) = 2.6$ and inference based on the OLSE estimate of the variance is anti-conservative. On the other hand, if $\vec{\alpha} = (1, 3, 2, 3, 1)^T$ (so $\sum \alpha_i x_i = 0$, but the α_i's are not independent of the x_i's), then $\overline{\alpha} = 2$, but the weighted average $(\sum \alpha_i x_i^2)/(\sum x_i^2) = 1.4$ and inference based on the OLSE estimate of the variance is conservative.

Clearly, as both the distribution of the α_i's relative to the x_i's and the value of γ and/or the skewness of the x_i's are allowed to vary, the tendency for the OLS variance estimate to over- or underestimate the true variance will reflect the combination of those effects.

e. What would be the effect of using the asymptotic results for ordinary least squares regression analysis on confidence intervals for β_1? Consider the effect that the variance of the α_is and the value of γ has on your answer.

Ans: Assuming that asymptotic normality holds, we need only worry about when the variance estimate under OLSE over- or underestimates the true variance. When the variance estimate underestimates the true variance, confidence intervals for β_1 will tend to be too narrow, and thus will have a coverage probability that is less than the desired level. When the variance estimate overestimates the true variance, confidence intervals for β_1 will tend to be too wide, and thus will have a coverage probability that is greater than the desired level. Discussion of the cases that such over- or underestimation occurs is exactly the same as for part (d).

3. Let $Y_i \sim Bernoulli(p_i), i = 1, \ldots, n$ be independent random variables with $p_i = \vec{x}_i^T \vec{\beta}$ for known predictor vector $\vec{x}_i$.

a. Is inference about $\vec{\beta}$ using linear regression analysis asymptotically valid for this problem? If so, provide justification. If not, are there any situations in which it might be approximately valid?

Ans: In the Bernoulli model, we have regression model $Y_i = \vec{x}_i^T \vec{\beta} + \epsilon_i$ with ϵ_i's being independently distributed with mean 0 and variance $p_i(1 - p_i)$. Hence, unless $\vec{x}_i^T \vec{\beta}$ is constant for all $i = 1, \ldots, n$ (as it would be under the null hypothesis $H_0 : \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$), there is heteroscedasticity, with a relationship between the predictors and the magnitude of the error variance. Hence, inference based on OLS will in general not be correct. Note, however, that if $.4 \leq p_i \leq .6$, then $.24 \leq p_i(1 - p_i) \leq .25$, and the heteroscedasticity is not too severe. In fact, a commonly quoted rule of thumb is that if $.3 \leq p_i \leq .7$, then $.21 \leq p_i(1 - p_i) \leq .25$, and OLS based inference in this setting will generally be okay. Some others will further relax this to $.2 \leq p_i \leq .8$. I note that if the x_i's are sampled such that the distribution of p_i's is fairly symmetric about .5, then we can use the results of problem 2 about the α_i's (because there will be no linear trend in the error variances with the x_i's) to infer that the OLS inference will tend to be conservative, because the more extreme values of x_i will tend to be associated with the smaller error variance. On the other hand, if the x_i's are sampled such that the distribution of the p_i's is skewed about .5, then the results of problem 2 about γ can be used, because there will tend to be a trend in the error variance with the values of the x_i's.

b. Describe an iterative approach in which weighted least squares might be used to address this problem. What undesirable small sample behavior with respect to the range of estimates $\hat{p}_i$ might persist under this analysis scheme?

Ans: If we knew the error variances, we could use weighted least squares (generalized least squares with a diagonal covariance matrix for $\vec{\epsilon}$) to obtain valid inference. One approach would be to first use OLS to estimate $\widehat{\vec{\beta}}^{(0)}$ and $\hat{p}_i^{(0)} = \vec{x}_i^T \widehat{\vec{\beta}}^{(0)}$. Then use $\mathbf{V}^{(0)}$ with $V_{ii}^{(0)} = \hat{p}_i^{(0)}(1 - \hat{p}_i^{(0)})$ and $V_{ij}^{(0)} = 0$ for $i \neq j$ to find GLSE $\widehat{\vec{\beta}}_G^{(1)}$. These estimates are then used to find $\hat{p}_i^{(1)}$ and $\mathbf{V}^{(1)}$. The process is then repeated with GLSE $\widehat{\vec{\beta}}_G^{(k)}$ estimated using $\mathbf{V}^{(k-1)}$ until $(\widehat{\vec{\beta}}_G^{(k)} - \widehat{\vec{\beta}}_G(k-1))^T(\widehat{\vec{\beta}}_G^{(k)} - \widehat{\vec{\beta}}_G(k-1))$ is sufficiently small. Inference is then based on estimates of the variance of the regression parameter vector derived under weighted (generalized) least squares theory.

This is an asymptotically valid procedure under the correct model. However, in the setting of small samples, it may well happen that estimates $\hat{p}_i$ would be less than 0 or greater than 1. This is often regarded as undesirable. This problem does not occur in logistic regression, where instead of the mean p_i, the log odds $\log(p_i/(1-p_i))$ (which can range from $-\infty$ to ∞) is modeled by $\vec{x}_i^T \vec{\beta}$. Finding estimates for logistic regression uses an iteratively reweighted least squares approach very similar to that described above, except transformations of the observations are used. (See Biost 570/Stat 570)