10

Identification Code 24

1. Let X be a random variable and assume that its moment generating function, $M_X(t) = E(e^{tX})$ exists for any $t \geq 0$.

   (a) *(2 points)* Markov's inequality states that if a random variable Z is nonnegative with finite mean $E(Z)$, then for any fixed $z > 0$, $P(Z \geq z) \leq \frac{E(Z)}{z}$. Use this inequality to show that for any fixed $x > 0$,

   $$P(X \geq x) \leq \frac{E(e^{tX})}{e^{tx}},$$

   for any fixed $t \geq 0$.

   (b) *(2 points)* Use Jensen's inequality to show that if $0 < x \leq E(X)$ and $t \geq 0$, then $\frac{E(e^{tX})}{e^{tx}} \geq 1$, i.e., the inequality in part (a) is trivial. Recall that Jensen's inequality states that if Z is a random variable and ϕ is a convex function (e.g., $\phi(x) = x^2$), then $\phi(E(Z)) \leq E(\phi(Z))$, assuming that the latter term exists.

   (c) *(2 points)* Let X be a discrete random variable following a Poisson distribution with parameter $\lambda > 0$, i.e.,

   $$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \text{ for } k = 0, 1, 2, ...$$

   Prove that $M_X(t) = e^{\lambda(e^t - 1)}$ for any $t \geq 0$.

   (d) *(4 points)* Assume the setting as in part (c). Let $h(x, t) = E(e^{tX})/e^{tx}$, which is a function of both x and t. Then part (a) gives,

   $$P(X \geq x) \leq h(x, t).$$

   For a fixed $x > 0$, we would like to make the above bound $h(x, t)$ as tight as possible by minimizing $h(x, t)$ as a function of t. Let $h_{\min}(x) = \min_{t \geq 0} h(x, t)$. Find an explicit expression of $h_{\min}(x)$. Note that we shall then have

   $$P(X \geq x) \leq h_{\min}(x).$$

Problem 1

(a). $P(X \geq x) = P(tX \geq tx)$ ← since $t \geq 0$

$= P(e^{tX} \geq e^{tx})$ ← since $e^{(\cdot)}$ is a monotonic transformation

$\leq \frac{E(e^{tX})}{e^{tx}}$ by Markov's inequality 2

(b). $\exp(\cdot)$ is convex, consider $g(x) = e^{tx}$, it's convex given $t \geq 0, x > 0$.

Thus $E(g(X)) \geq g(E(X))$ by Jensen's inequality

$= e^{tE(X)}$

$\geq e^{tx}$ since $x \leq E(X)$

Thus, $E(e^{tX}) \geq e^{tx}$

$\frac{E(e^{tX})}{e^{tx}} \geq 1$ 2

(c). $M_X(t) = E(e^{tK})$

$= \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k e^{-\lambda}}{k!}$

$= \sum_{k=0}^{\infty} \frac{(e^t\lambda)^k e^{-\lambda}}{k!}$

$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^t\lambda)^k}{k!}$ Recall $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda} = \lim(1+\frac{\lambda}{n})^n$

$= e^{-\lambda} e^{e^t\lambda}$

$= e^{\lambda(e^t-1)}$ $\forall t \geq 0$

(d). Aim to find minimum of $\frac{E(e^{tX})}{e^{tx}}$ w.r.t. t 4

$h(x,t) = \frac{E(e^{tX})}{e^{tx}} = e^{\lambda(e^t-1)-tx} = e^{\lambda(e^t-1)} e^{-tx}$

$\frac{\partial h(x,t)}{\partial t} = u'v + v'u$ $\quad u = e^{\lambda(e^t-1)}$ $\quad v = e^{-tx}$

$= \lambda e^{t+\lambda(e^t-1)} e^{-tx}$ $\quad u' = \frac{\partial h}{\partial g}\frac{\partial g}{\partial t}$ $\quad v' = -xe^{-tx}$

$- xe^{-tx} e^{\lambda(e^t-1)}$ $\quad = \lambda e^{\lambda(e^t-1)} \cdot e^t$

$= \lambda e^{t+\lambda(e^t-1)}$

$\lambda e^{t(1-x)+\lambda(e^t-1)} - xe^{\lambda(e^t-1)-tx} \overset{set}{=} 0$

$e^{\lambda(e^t-1)}(\lambda e^{t(1-x)} - xe^{-tx}) = 0$

$\lambda e^{t-tx} - xe^{-tx} = 0$ since $e^{\lambda(e^t-1)} \neq 0$

$e^{-tx}(\lambda e^t - x) = 0$

$\lambda e^t - x = 0$ since $e^{-tx} \neq 0$

$e^t = \frac{x}{\lambda}$

$\hat{t} = \log(\frac{x}{\lambda})$

$h_{min}(x) = \frac{E(e^{\log(\frac{x}{\lambda})X})}{e^{\log(\frac{x}{\lambda})x}} = \frac{E((\frac{x}{\lambda})^X)}{(\frac{x}{\lambda})^x}$

$\frac{\partial^2 h(x,t)}{\partial t^2} = \frac{\partial}{\partial t} e^{\lambda(e^t-1)}(\lambda e^{t(1-x)} - xe^{-tx})$

$= \frac{\partial}{\partial t}(e^{\lambda(e^t-1)})(\lambda e^{t(1-x)}) - (e^{\lambda(e^t-1)} xe^{-tx})$

$= \lambda e^t e^{\lambda(e^t-1)} \lambda e^{t(1-x)} + e^{\lambda(e^t-1)} \lambda(1-x) e^{t(1-x)} - \lambda e^t e^{\lambda(e^t-1)} xe^{-tx} + x^2 e^{-tx} e^{\lambda(e^t-1)}$

$= e^{\lambda(e^t-1)}(\lambda^2 e^{t(2-x)} + \lambda(1-x)e^{t(1-x)} - \lambda x e^{t(1-x)} + x^2 e^{-tx})$

$= e^{\lambda(e^t-1)}(\lambda^2 e^{t(2-x)} + \lambda e^{t(1-x)} + x^2 e^{-tx}) > 0$ thus convex, thus $\hat{t}$ is the minimum.

$\frac{9}{10}$

Identification Code 24

2. Suppose that $X_i \overset{iid}{\sim} f(x, \boldsymbol{p})$ for $i = 1, \ldots, n$ where $\boldsymbol{p} = (p_1, p_2, p_3)$ is unknown and $f(x, \boldsymbol{p})$ denotes the probability mass function of a discrete random variable taking values $x = 1, 2$ and 3 with probability p_1, p_2 and p_3, respectively, satisfying $p_j \geq 0$ for $j = 1, 2, 3$ and $p_1 + p_2 + p_3 = 1$. Consider testing the null hypothesis

$$H_0 : p_1 = \theta^2,\ p_2 = (1-\theta)^2,\ p_3 = 2\theta(1-\theta) \text{ for some } \theta \in (0, 1).$$

(a) *(3 points)* Obtain the MLE of θ under the above null hypothesis H_0.

(b) *(3 points)* Let $\Delta = \Delta(\boldsymbol{p}) = \sqrt{p_1} + \sqrt{p_2} - 1$. Show that the above-mentioned H_0 is equivalent to null hypothesis $\Delta = 0$.

(c) *(4 points)* Obtain a large sample test for testing the null hypothesis $H_0 : \Delta = 0$. Explicitly state the test statistic and rejection region.

Problem 2

(a) Recognize $X \sim$ multinomial (n, p_1, p_2, p_3) denote n_1, n_2, n_3 be the # of X_i taking 1, 2, 3 values.

$p(X) = \frac{n!}{n_1!n_2!n_3!} p_1^{n_1} p_2^{n_2} p_3^{n_3}$ $\qquad n_i = \sum_{j=1}^{n} I(X_j = i)$

$= \frac{n!}{n_1!n_2!n_3!} (\theta^2)^{n_1} ((1-\theta)^2)^{n_2} (2\theta(1-\theta))^{n_3}$

$= C \cdot (\theta^2)^{n_1} (1-\theta^2)^{n_2} (2\theta(1-\theta))^{n-n_1-n_2}$

$= C \cdot \left(\frac{\theta}{2(1-\theta)}\right)^{n_1} \left(\frac{1+\theta}{2\theta}\right)^{n_2} (2\theta(1-\theta))^{n}$

$\ell(X) = C + n_1 \log\left(\frac{\theta}{2(1-\theta)}\right) + n_2 \log\left(\frac{1+\theta}{2\theta}\right) + n \log(2\theta(1-\theta))$

$= n_1 \log(\theta) - n_1 \log(2(1-\theta)) + n_2 \log(1+\theta) - n_2 \log(2\theta) + n \log(2\theta) + n \log(1-\theta)$

$\frac{\partial \ell}{\partial \theta} = \frac{n_1}{\theta} + \frac{2n_1}{2(1-\theta)} + \frac{n_2}{1+\theta} - \frac{2n_2}{2\theta} + \frac{2n}{2\theta} - \frac{n}{1-\theta} \overset{set}{=} 0$

$\Rightarrow \frac{n_1}{\theta} + \frac{n_1}{1-\theta} + \frac{n_2}{1+\theta} - \frac{n_2}{\theta} + \frac{n}{\theta} - \frac{n}{1-\theta} = 0$

$\Rightarrow \frac{n_1}{(1-\theta)\theta} - \frac{n_2}{(1+\theta)\theta} + \frac{n(1-2\theta)}{\theta(1-\theta)} = 0$ $\qquad (1-2\theta)(1+\theta)$

$\Rightarrow \frac{n_1(1+\theta) - n_2(1-\theta) + n(1-2\theta)(1+\theta)}{(1-\theta)\theta(1+\theta)} = 0$ $\qquad = 1 - 2\theta + \theta - 2\theta^2$

$\Rightarrow n_1 + n_1\theta - n_2 + n_2\theta + n - n\theta - 2n\theta^2 = 0$ $\qquad = 1 - \theta - 2\theta^2$

$\Rightarrow (n_1 - n_2 + n) - n_3\theta - 2n\theta^2 = 0$

$\Rightarrow (2n_1 + n_3) - n_3\theta - 2n\theta^2 = 0$ $\qquad \Delta = \frac{b \pm \sqrt{b^2 - 4ac}}{2a}$

$= \frac{n_3 \pm \sqrt{n_3^2 -}}{2(2n_1 + n_3)}$

→ calculation please see next page.

→ 2nd order derivative check also see next page.

$\hat{\theta} = \frac{2n_1 + n_3}{2(n_1 + n_2 + n_3)}$

(b). $H_0: p_1 = \theta^2 \quad p_2 = (1-\theta)^2, \quad p_3 = 2\theta(1-\theta) \quad$ with $\theta \in (0,1)$

Then without more specification on the θ value, it is the relationship that we are testing on that is, $\sqrt{p_1} + \sqrt{p_2} = \theta + (1-\theta) = 1$

equivalent to $H_0: \sqrt{p_1} + \sqrt{p_2} = 1$ $\qquad$ * since $p_3 = 1 - p_1 - p_2$, no need to specify.

equivalent to $H_0: \sqrt{p_1} + \sqrt{p_2} - 1 = 0$

$\Leftrightarrow \Delta(p) = 0$

Thus original $H_0 \Leftrightarrow \Delta(p) = 0$.

You have only shown $\Delta = 0$ if H_0 holds

part (c) see last page.

You have to show if $\Delta = 0$ then H_0 holds for some $\theta \in (0,1)$.

−1

Problem 2 continuous.

(a). $P(X) = c \cdot (\theta^2)^{n_1} ((1-\theta)^2)^{n_2} (2\theta(1-\theta))^{n_3}$ c is constant, invariant of θ.

$$\ell(x) = 2n_1 \log\theta + 2n_2 \log(1-\theta) + n_3 \log(2\theta(1-\theta))$$

$$\frac{\partial \ell}{\partial \theta} = \frac{2n_1}{\theta} - \frac{2n_2}{1-\theta} + \frac{n_3(2-4\theta)}{2\theta(1-\theta)} \overset{set}{=} 0$$

$$\Rightarrow \frac{2n_1}{\theta} - \frac{2n_2}{1-\theta} + \frac{n_3(1-2\theta)}{\theta(1-\theta)} = 0$$

$$\Rightarrow \frac{2n_1-2n_1\theta}{\theta(1-\theta)} - \frac{2n_2\theta}{(1-\theta)\theta} + \frac{n_3(1-2\theta)}{\theta(1-\theta)} = 0$$

$$\Rightarrow 2n_1 - 2n_1\theta - 2n_2\theta + n_3 - 2n_3\theta = 0$$

$$\Rightarrow 2(n_1+n_2+n_3)\theta = 2n_1+n_3$$

$$\hat{\theta} = \frac{2n_1+n_3}{2(n_1+n_2+n_3)} = \frac{2n_1+n_3}{2n}$$

$$\frac{\partial^2 \ell}{\partial \theta^2} = \frac{-2n_1}{\theta^2} - \frac{2n_2}{(1-\theta)^2} + \frac{\partial}{\partial\theta}\left(\frac{n_3(1-2\theta)}{\theta-\theta^2}\right)$$

$$= -\frac{2n_1}{\theta^2} - \frac{2n_2}{(1-\theta)^2} + \frac{n_3(-2)(\theta-\theta^2) - n_3(1-2\theta)(1-2\theta)}{(\theta-\theta^2)^2}$$

$$= -\frac{2n_1}{\theta^2} - \frac{2n_2}{(1-\theta)^2} - \frac{2n_3(\theta-\theta^2) + n_3(1-2\theta)(1-2\theta)}{(\theta-\theta^2)^2}$$

$$\leq 0$$

Thus the log-likelihood is concave

Thus $\hat{\theta}$ is the MLE.

Problem 2 continuous

(c). Here we can do a likelihood ratio test, as we already have $\hat{\theta}$ under H_0.

Let $\hat{\theta}$ under H_0 be $\hat{\theta}_0$, we can have $\hat{P}_{10}, \hat{P}_{20}, \hat{P}_{30}$ by invariance of MLE under regularity condition.

$\hat{P}_{10} = \hat{\theta}_0^2 \quad \hat{P}_{20} = (1-\hat{\theta}_0)^2$

$\hat{P}_{30} = 2\hat{\theta}_0(1-\hat{\theta}_0)$

Then recall under H_1, $\hat{P}_1 = \frac{n_1}{n}$, $\hat{P}_2 = \frac{n_2}{n}$, $\hat{P}_3 = \frac{n_3}{n}$ ~~in general~~.

that is, $\hat{P}_1 = \frac{\sum_k I(X_k=1)}{n}$ $\hat{P}_2 = \frac{\sum_k I(X_k=2)}{n}$, $\hat{P}_3 = \frac{\sum_k I(X_k=3)}{n}$

Then $R = -2\log \frac{L(\hat{P}_{10}, \hat{P}_{20}, \hat{P}_{30})}{L(\hat{P}_1, \hat{P}_2, \hat{P}_3)} \overset{H_0}{\sim} \chi^2_1$ ~~under wilk's theorem~~

$$= 2\left(\ell(\hat{P}_1, \hat{P}_2, \hat{P}_3) - \ell(\hat{P}_{10}, \hat{P}_{20}, \hat{P}_{30})\right)$$

Thus we reject when $R > \chi^2_{1, 0.05} = 3.841$

R is the test statistics

Rejection region is $RR = \{ X : R(X) > 3.841 \}$

9

12 "13"

3. Suppose an agricultural researcher is investigating twelve new corn varieties as well as the currently recommended commercial variety. She is interested in learning whether any of the test (new) varieties have greater yield than the current variety. In the design of the experiment, four complete blocks were used (each treatment appeared once within each block). RCBD

   (a) *(2 points)* Explain the likely reason for the choice of the block design. Also, the agronomist assumed "random block effects." What does this imply in the context of the experiment?

   (b) *(2 points)* Carefully write the model for the experiment, defining any symbols and notation used. Include the usual distributional assumptions for an experiment of this type.

   (c) *(2 points)* Write the ANOVA table, including sources of variation and exact degrees of freedom. Give the formulas for the test statistics for testing about treatment effects and block effects.

   (d) *(2 points)* Under the assumptions of the model, derive the standard deviation of the sample mean for the j-th variety. Also derive the standard deviation of the sample mean for the i-th block.

   (e) *(2 points)* Suppose that after these data are gathered, four more independent yield measurements will be taken using block 1 and variety 1. Under the assumptions of the model, find the probability that the sample mean of these four new yields (for block 1, variety 1) will be at least 1.3 times the true expected yield value for block 1 and variety 3. Your answer should be given in terms of the model parameters.

Problem 3

(a). We are using the block design because the blocks "e.g. Soil" varies. For example, the water abstraction rate, the soil tension varies, which could be a confounder for the corn variety if NOT included in the model.

They are using the random block effect, because they are not interested in the effect of soil in this study, and if they repeat the study, block may vary.

OK

(b) $Y_{ij} = \mu + \alpha_i + B_j + \varepsilon_{ij}$ $\quad \varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$, $B_j \overset{iid}{\sim} N(0, \sigma_B^2)$ mutually independent

$i = 1, \cdots, 13$

$j = 1, 2, 3, 4$

μ: overall effect on output (yield)

α_i: main effect of i-th corn variety (fixed)

B_j: random effect of j-th block (random)

ε_{ij}: random error to i-th varity in j-th block (random)

* we don't include a interaction term because this is a RCBD, and we won't have enough degree of freedom to it. *

(c).

Source	df	sum of square	MS	EMS	F
$\alpha_{(trt)}$	$a-1=12$	$\Sigma_{ij}(\bar{Y}_{..}-\bar{Y}_{i.})^2$	SS/df	$\sigma^2 + bQ(\alpha) = \sigma^2 + \frac{1}{12}\Sigma_i\Sigma_j(\alpha_i - \bar{\alpha})^2$	MSA/MSE
Block	$b-1=3$	$\Sigma_{ij}(\bar{Y}_{..}-\bar{Y}_{.j})^2$		$\sigma^2 + a\sigma_B^2 = \sigma^2 + 13\sigma_B^2$	MSB/MSE
Error	36	$\Sigma_{ij}(Y_{ij}-\bar{Y}_{i.}-\bar{Y}_{.j}+\bar{Y}_{..})^2$		σ^2	
Total	$ab-1=51$	$\Sigma_{ij}(Y_{ij}-\bar{Y}_{..})^2$			

notize here, if we have an interaction that's random, then devide by the interaction term.

Thus, the test statis for treatment effect is

$$\frac{\Sigma_{ij}(\bar{Y}_{..}-\bar{Y}_{i.})^2/12}{\Sigma_{ij}(Y_{ij}-\bar{Y}_{i.}-\bar{Y}_{.j}+\bar{Y}_{..})^2/36} \overset{H_0}{\sim} F_{12,36}$$

the test statistics for block effect is

$$\frac{\Sigma(\bar{Y}_{..}-\bar{Y}_{.j})^2/3}{\Sigma_{ij}(Y_{ij}-\bar{Y}_{i.}-\bar{Y}_{.j}+\bar{Y}_{..})^2/36} \overset{H_0}{\sim} F_{3,36}$$

(d). okay, so up to this point I know the indices we are using are different, but may not matter.

$$SD(\bar{Y}_{j.}) = \sqrt{\hat{Var}(\bar{Y}_{j.})} = \sqrt{\hat{var}(\mu + \alpha_j + \bar{B}_. + \bar{\varepsilon}_{j.})}$$
$$= \sqrt{\tfrac{1}{4}\hat{Var}(B_k + \varepsilon_{jk})}$$
$$= \sqrt{\tfrac{1}{4}(\hat{\sigma}_B^2 + \hat{\sigma}^2)}$$
$$= \sqrt{\tfrac{1}{4}}(MSE + \hat{\sigma}_B^2)$$

$$\hat{\sigma}_B^2 = \frac{MSB - MSE}{13}$$

OK

$$SD(\bar{Y}_{.i}) = \sqrt{\hat{var}(\mu + \bar{\alpha}_. + B_i + \bar{\varepsilon}_{.i})}$$
$$= \sqrt{\hat{\sigma}_B^2 + \tfrac{1}{13}\hat{\sigma}^2}$$
$$= \sqrt{\tfrac{1}{13}}\sqrt{13\hat{\sigma}_B^2 + \hat{\sigma}^2} = \sqrt{\tfrac{1}{13}MSB}$$

Problem 3 continuous

(e) now let $Y_{111}, Y_{112}, Y_{113}, Y_{114}$ be the 4 new yieldings.

Then the sample mean is $\bar{Y}_{11.}$

The true expected yield value for block 1 and yield 3 is

$$E(Y_{13}) = E(\mu + \alpha_1 + B_3 + \varepsilon_{13})$$
$$= \mu + \alpha_1 \quad \text{since by assumption, } E(B_3) = E(\varepsilon_{13}) = 0.$$

$$P(\bar{Y}_{11.} \geq 1.3(\mu + \alpha_1))$$
$$= P(\mu + \alpha_1 + B_1 + \bar{\varepsilon}_{11.} \geq 1.3(\mu + \alpha_1))$$
$$= P(B_1 + \bar{\varepsilon}_{11.} \geq 0.3(\mu + \alpha_1))$$
$$= P\left(\frac{B_1 + \bar{\varepsilon}_{11.}}{\sqrt{\sigma_{B\varepsilon}^2}} \geq \frac{0.3(\mu + \alpha_1)}{\sigma_{B\varepsilon}^2}\right)$$

Recall $B_1 \sim N(0, \sigma_B^2)$ $\bar{\varepsilon}_{11.} \sim N(0, \frac{1}{4}\sigma^2)$

$B_1 + \bar{\varepsilon}_{11.} \sim N(0, \underbrace{\sigma_B^2 + \frac{1}{4}\sigma^2}_{\sigma_{B\varepsilon}^2})$

$$= 1 - \Phi\left(\frac{0.3(\mu + \alpha_1)}{\sigma_{B\varepsilon}^2}\right)$$

if σ_ε^2 known

7/10

Identification Code 24

4. Consider a regression problem with response vector $y_{n\times 1}$ and known and fixed model matrix $X_{n\times p} = [V_{n\times p_1} \vdots W_{n\times p_2}]$ with $p = p_1 + p_2$. Suppose X has full column rank. Consider the following two regression models:

$$\begin{aligned} \text{Model A:}\quad & y = V\beta_1 + W\beta_2 + e \\ \text{Model B:}\quad & y = V\eta_1 + (I - P_V)W\eta_2 + e \end{aligned}$$

Here P_V is the orthogonal projection matrix onto the column space of V. Assume that $e \sim N(0, \sigma^2 I_n)$. Answer the following questions.

(a) *(2 points)* Show that model B is a reparameterization of model A.

(b) *(2 points)* Find the relationship between (η_1, η_2) and (β_1, β_2).

(c) *(3 points)* Show that the ordinary least squares (OLS) estimators of η_1 and η_2 are

$$\widehat{\eta}_1 = (V^T V)^{-1} V^T y, \quad \widehat{\eta}_2 = [W^T(I - P_V)W]^{-1} W^T (I - P_V) y.$$

Note that you need to show inverses of the appropriate matrices exist.

(d) *(3 points)* Show that the OLS estimators of η_1 and η_2 are independent.

Problem 4

(a). A: $y = V\beta_1 + W\beta_2 + e$

B: $y = V\eta_1 + (I-P_V)W\eta_2 + e$

In definition, we can show model involving B is a reparametrization of model involving A if

$col(A) = col(B)$

Want to show $col(V \;\; W) = col(V \;\; (I-P_V)W)$

This is direct. ($\Rightarrow$)

Suppose $\exists\, x$ s.t. $x \in col(V\;W)$

Then $\exists\, a, b$ s.t. $x = Va + Wb$ ← here we

Then there must $\exists$ some c s.t. $(I-P_V)Wc = Wc - P_VWc = Wb$

($\Leftarrow$) direction follows the similar reasoning.

Thus, model B is a reparametrization of A

This can be seen as

$\underbrace{(V \;\; (I-P_V)W)}_{Z} = \underbrace{(V \;\; W)}_{X}\begin{pmatrix} I & 0 \\ 0 & I-P_V \end{pmatrix}$

Then indeed $col(X) \subseteq col(Z)$

Then as

$rank(Z) \le rank(X) \wedge rank\begin{pmatrix} I & 0 \\ 0 & I-P_V\end{pmatrix}$

$\le rank(X)$

Then $col(Z) = col(X)$

(C). $\beta = \begin{pmatrix}\beta_1 \\ \beta_2\end{pmatrix} = ((V\;W)'(V\;W))^{-1}(V\;W)'y$

$= \left(\begin{pmatrix}V' \\ W'\end{pmatrix}(V\;W)\right)^{-1}\begin{pmatrix}V' \\ W'\end{pmatrix}y$

$= \begin{pmatrix}V'V & V'W \\ W'V & W'W\end{pmatrix}^{-1}\begin{pmatrix}V'y \\ W'y\end{pmatrix}$

this is part (c), part (b) please see below (next page).

$\eta = \begin{pmatrix}\eta_1 \\ \eta_2\end{pmatrix} = \left(\begin{pmatrix}V' \\ W'(I-P_V)\end{pmatrix}(V \;\; (I-P_V)W)\right)^{-1}\begin{pmatrix}V' \\ W'(I-P_V)\end{pmatrix}y$

$= \begin{pmatrix}V'V & V'(I-P_V)W \\ W'(I-P_V)V & W'(I-P_V)W\end{pmatrix}^{-1}\begin{pmatrix}V'y \\ W'(I-P_V)y\end{pmatrix}$

$= \begin{pmatrix}V'V & 0 \\ 0 & W'(I-P_V)W\end{pmatrix}^{-1}\begin{pmatrix}V'y \\ W'(I-P_V)y\end{pmatrix}$ because $(I-P_V)V = V'(I-P_V)' = 0$

$= \begin{pmatrix}(V'V)^{-1}V'y \\ (W'(I-P_V)W)^{-1}W'(I-P_V)y\end{pmatrix}$

$\hat{\eta}_1 = (V'V)^{-1}V'y \qquad \hat{\eta}_2 = (W'(I-P_V)W)^{-1}W'(I-P_V)y$

exactly what's asked by the question, and $\hat{\eta}_1$, $\hat{\eta}_2$ are the OLS estimators because under Gauss-Markov settings, this is defined as the OLS.

Problem 4 continues

(b) recongnize

$$\underbrace{(V \quad (I-P_V)W)}_{Z} = \underbrace{(V \quad W)}_{X} \underbrace{\begin{pmatrix} I & 0 \\ 0 & (I-P_V) \end{pmatrix}_{2\times 2}}_{C}$$

Thus the design matrix in B is a linear transform of design matrix in A, we gave some names to these matrices, that is

$$Z = XC$$

Then we have

model A: $y = X\beta + e$ $\qquad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$

model B: $y = XC\eta + e$ $\qquad \eta = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}$

$$\Rightarrow \hat{\eta} = ((XC)'(XC))^{-1}(XC)'y$$
$$= (C'X'XC)^{-1}C'X'y$$
$$= (C^{-1})(X'X)^{-1}(C')^{-1}C'X'y \quad \leftarrow \text{C is invertible} \quad \text{Why??}$$
$$= C^{-1}(X'X)^{-1}X'y$$
$$= C^{-1}\hat{\beta}$$

the relation is $\beta = C\eta$

$\hat{\beta} = C\hat{\eta}$

(d). Recall from part (c),

$\hat{\eta}_1 = (V'V)^{-1}V'y \qquad \hat{\eta}_2 = (W'(I-P_V)W)^{-1}W'(I-P_V)y$

both follows normal distribution as y follows normal distribution, so they are indep. as long as their covariance is 0. (by normal dist property)

$$Cov(\hat{\eta}_2, \hat{\eta}_1) = Cov((W'(I-P_V)W)^{-1}W'(I-P_V)y, (V'V)^{-1}V'y)$$
$$= \sigma^2 (W'(I-P_V)W)^{-1}W'(I-P_V)V(X'X)^{-1}$$
$$= 0 \quad \text{since } (I-P_V)V = 0$$

Thus $\hat{\eta}_2$ and $\hat{\eta}_1$ are independent!

9.5/10 great

5. We consider the simple linear regression model:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \text{ for } i = 1, 2, \ldots$$

with independent ε_i's and we have $E(\varepsilon_i) = 0$, $\text{var}(\varepsilon_i) = \sigma^2$, $\forall\, i$.

Suppose we do not observe the responses directly, but only the average response $\bar{y}_0$ for n_0 observations with $x_i = 0$, and average $\bar{y}_1$ for n_1 observations with $x_i = 1$. (We use lower case y to denote observed/realized values). In other words, we observe only two data points: $\{\bar{x}_0 = 0, \bar{y}_0\}$, $\{\bar{x}_1 = 1, \bar{y}_1\}$, and n_0 and n_1 are known.

(a) *(3 points)* Using the two observations described above, state the ordinary least squares estimator of $\boldsymbol{\beta} = [\beta_0, \beta_1]^T$, simplified explicitly in terms of the known/observed quantities, and show that it is unbiased.

(b) *(5 points)* Using the two observations described above, state the weighted least squares estimator of $\boldsymbol{\beta}$ utilizing the variance-covariance matrix of $\{\bar{Y}_0, \bar{Y}_1\}$. Simplify the result to show that the estimator depends only on $\{\bar{y}_0, \bar{y}_1\}$, and not n_0 or n_1.

(c) *(2 points)* Whether or not you were able to simplify your answer to b), show that the weighted least squares estimator is unbiased for $\boldsymbol{\beta}$.

Problem 5

(a). we now have

$$\begin{pmatrix} \bar{y}_1 \\ \bar{y}_0 \end{pmatrix} = \beta_0 + \beta_1 \begin{pmatrix} \bar{X}_1 \\ \bar{X}_0 \end{pmatrix} + \varepsilon$$

β_0, β_1 here are scalars.

by definition, $\hat{\beta} = (X'X)^{-1}X'y$

$$= \left(\begin{pmatrix} 1 & \bar{x}_1 \\ 1 & \bar{x}_0 \end{pmatrix}' \begin{pmatrix} 1 & \bar{x}_1 \\ 1 & \bar{x}_0 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & \bar{x}_1 \\ 1 & \bar{x}_0 \end{pmatrix}' \begin{pmatrix} \bar{y}_1 \\ \bar{y}_0 \end{pmatrix}$$

$$= \left(\begin{pmatrix} 1 & 1 \\ \bar{x}_1 & \bar{x}_0 \end{pmatrix} \begin{pmatrix} 1 & \bar{x}_1 \\ 1 & \bar{x}_0 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & 1 \\ \bar{x}_1 & \bar{x}_0 \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_0 \end{pmatrix}$$

$$= \begin{pmatrix} 2 & \bar{x}_1 + \bar{x}_0 \\ \bar{x}_1 + \bar{x}_0 & \bar{x}_1^2 + \bar{x}_0^2 \end{pmatrix}^{-1} \begin{pmatrix} \bar{y}_1 + \bar{y}_0 \\ \bar{x}_1 \bar{y}_1 + \bar{y}_0 \bar{x}_0 \end{pmatrix}$$

$$= \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \bar{y}_1 + \bar{y}_0 \\ \bar{y}_1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} \bar{y}_1 + \bar{y}_0 \\ \bar{y}_1 \end{pmatrix}$$

clean

$$= \begin{pmatrix} \bar{y}_0 \\ \bar{y}_1 - \bar{y}_0 \end{pmatrix} \checkmark$$

$$E(\hat{\beta}) = E\begin{pmatrix} \bar{y}_0 \\ \bar{y}_1 - \bar{y}_0 \end{pmatrix}$$

because X_i are binary

$$= E\begin{pmatrix} \beta_0 + \beta_1 X_{0i} + \varepsilon_i \\ \beta_0 + \beta_1 X_{1i} + \varepsilon_i - (\beta_0 + \beta_1 X_{0i} + \varepsilon_i) \end{pmatrix}$$

$$= \begin{pmatrix} \beta_0 + 0 \\ \beta_0 + \beta_1 - (\beta_0 + 0) \end{pmatrix}$$

$$= \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$ that it unbiased

technically σ^2 still appears, although it can be shown to cancel a few steps later

$-\frac{1}{2}$

(b). This is we need to use the Aitken model setting.

$$Cov(\bar{Y}) = \begin{pmatrix} Var(\bar{Y}_1) & Cov(\bar{Y}_1, \bar{Y}_0) \\ Cov(\bar{Y}_1, \bar{Y}_0) & Var(\bar{Y}_0) \end{pmatrix} = \sigma^2 \begin{pmatrix} \frac{1}{n_1} & 0 \\ 0 & \frac{1}{n_0} \end{pmatrix} \Rightarrow V = \begin{pmatrix} \frac{1}{n_1} & 0 \\ 0 & \frac{1}{n_0} \end{pmatrix}$$

Then $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$

$$= \left(\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} n_1 & 0 \\ 0 & n_0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} n_1 & 0 \\ 0 & n_0 \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_0 \end{pmatrix}$$

$$= \left(\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} n_1 & n_1 \\ n_0 & 0 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} n_1 \bar{y}_1 \\ n_0 \bar{y}_0 \end{pmatrix}$$

$$= \begin{pmatrix} n & n_1 \\ n_1 & n_1 \end{pmatrix}^{-1} \begin{pmatrix} n_1 \bar{y}_1 + n_0 \bar{y}_0 \\ n_1 \bar{y}_1 \end{pmatrix} \rightarrow = \Sigma_i^n y_i$$

$$= \frac{1}{nn_1 - n_1 n_1} \begin{pmatrix} n_1 & -n_1 \\ -n_1 & n \end{pmatrix} \begin{pmatrix} n\bar{y} \\ n_1 \bar{y}_1 \end{pmatrix}$$

$$= \frac{1}{n_1 n_0} \begin{pmatrix} n_1 n \bar{y} - n_1 n_1 \bar{y}_1 \\ n n_1 \bar{y}_1 - n n_1 \bar{y} \end{pmatrix}$$ $\longrightarrow$ next page continuous.

(b).

$$\hat{\beta} = \frac{1}{n_1 n_0}\begin{pmatrix} n_1 n \bar{y} - n_1 n_1 \bar{y}_1 \\ n n_1 \bar{y}_1 - n n_1 \bar{y} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{n_0}\sum_{i=1} y_i - \frac{1}{n_0}\sum_{i=1} y_{i1} \\ \frac{(n_1+n_0)}{n_0}\bar{y}_1 - \frac{(n_1+n_0)}{n_0}\bar{y} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{n_0}(\sum_i y_{i1} + \sum_i y_{i0}) - \frac{1}{n_0}\sum y_{i1} \\ \frac{1}{n_0}\sum_i y_{i1} + \underbrace{\frac{1}{n_1}\sum y_{i1}}_{\bar{y}_1} - \frac{1}{n_0}(\sum y_{i1} + \sum y_{i0}) \end{pmatrix}$$

$$= \begin{pmatrix} \bar{y}_0 \\ \bar{y}_1 - \bar{y}_0 \end{pmatrix}$$

(c). This is exactly the same as output in (a), thus by the same reasoning,

$$E(\hat{\beta}) = E\begin{pmatrix} \bar{y}_0 \\ \bar{y}_1 - \bar{y}_0 \end{pmatrix}$$

$$= \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$ indeed unbiased.