1) An overparameterized two way Anova model is $Y_{ijk} = \mu + \alpha_i + \beta_j + \tau_{ij} + \epsilon_{ijk}$ for i = 1, ..., a and j = 1, ..., b and k = 1, ..., m. Suppose a = 2, b = 2, and m = 2. Then

 $egin{array}{c} Y_{112} \ Y_{121} \ Y_{122} \ Y_{211} \ Y_{212} \end{array} = m{X} \ egin{array}{c} X \ Y_{211} \ Y_{212} \end{array}$

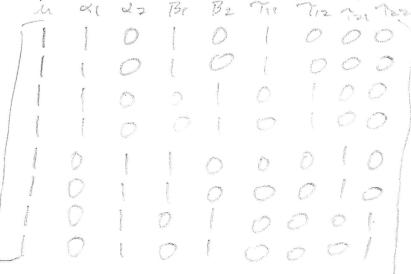
 $\left[egin{array}{c} Y_{111} \ Y_{112} \ Y_{121} \ Y_{122} \ Y_{211} \ Y_{212} \ Y_{221} \ Y_{222} \ \end{array}
ight] = oldsymbol{X} \left[egin{array}{c} \mu \ lpha_1 \ lpha_2 \ eta_1 \ lpha_2 \ eta_1 \ lpha_2 \ eta_1 \ lpha_1 \ lpha_2 \ eta_1 \ lpha_1 \ lpha_2 \ eta_1 \ lpha_1 \ lpha_2 \ eta_2 \ e$

NB = u + cx + 1/2 + 7/1

u + cx + 1/2 + 7/2

u

a) Give the matrix X.



b) Write the above model as $Y = X\beta + \epsilon$. This model is **not full rank**. What is the projection matrix P (onto the column space of X)? Hint: X^TX is singular, so use the generalized inverse.



 \mathcal{C} 2) In numerical linear algebra, the least squares solution to "Ax = b" is of interest where the problem is actually the multiple linear regression model $b = Ax + \epsilon$ where A has full rank p, and we will assume that $E(\epsilon) = 0$, and $Cov(\epsilon) = \sigma^2 I_n$.

a) What is the (formula for the) projection matrix P onto the column space of A?

A (A/A) / A/



b) What is the OLS estimator \hat{x} ?



c) What is the vector of fitted values $\hat{\boldsymbol{b}}$?



d) What is the residual vector e?



- 3) Consider the multivariate linear regression model Z = XB + E.
- a) What is $E(\mathbf{Z})$?



b) What is $E(\widehat{\boldsymbol{B}})$?



c) What is $E(\mathbf{E})$?



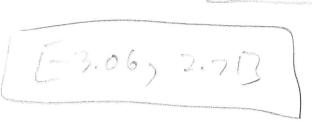
d) Consider testing $H_0: \mathbf{LB} = \mathbf{0}$. What is the limiting distribution of $(n-p)U(\mathbf{L})$?

$$(n-p)U(L) \stackrel{D}{\rightarrow}$$

4) The data below are a sorted residuals from a lasso regression where n=1000 and p=17. Find shorth(997) of the residuals.

number 1 2 3 4 ... 997 998 999 1000 residual -3.28 -3.06 -3.04 -2.96 ... 2.66 2.71 2.81 3.62

5.94 = 2.66 - (-7.28) 5.77 = 2.71 - (-3.06) 5.85 = 2.81 - (-3.04) 6.58 = 3.62 - (-2.96)



= Shorth 997



- 5) Suppose that $\mathbf{Y} = (Y_1, Y_2)'$, $\operatorname{Var}(\mathbf{Y}) = \sigma^2 \mathbf{I}$, $E(Y_1) = E(Y_2) = \beta_1 2\beta_2$. Show whether or not the following functions are estimable. Hint $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$, so find \mathbf{X} .
 - a) β_1 = $\langle O/E \rangle$ $\langle O/E \rangle$

07821



- b) β2 = 0 1/B, (?/¢ C 8 T/
 - (10)

3.8

c) $-\beta_1 + 2\beta_2 = (-1 \ 7) \ \beta$) $\left(-\frac{1}{2}\right) \in \mathbb{C}[\mathbb{Z}^7]$

(Yes)

d) $4\beta_1 - 8\beta_2 = (4 - 8) \vec{E}$, $(4 - 8) \vec{E}$, $(4 - 8) \vec{E}$





6) Suppose $Y^* = X\hat{\beta} + r^W$ where where $E(r^W) = 0$ and $Cov(r^W) = Cov(Y^*) = 0$ $MSE I_n$. Then $\hat{\boldsymbol{\beta}}^* = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{Y}^*$. Recall that \boldsymbol{X} is an $n \times p$ constant matrix. Simplify quantities when possible.

a) What is $E(\beta^*)$?

$$= (x^{7}x)^{7}x^{7}Ey^{8} = (x^{2}x)^{7}x^{7}x^{2}$$

$$= |\hat{x}|$$

$$= |\hat{x}|$$

b) What is
$$Cov(\hat{\boldsymbol{\beta}}^*)$$
? = $A could* AT = (\chi^T \chi)^T \chi^T MSE_{A} \chi (\chi^T \chi)^T$

$$= MSE (\chi^T \chi)^T \chi^T \chi (\chi^T \chi)^T = MSE (\chi^T \chi)^T$$

c) Recall that $X\hat{\boldsymbol{\beta}} = \boldsymbol{P}\boldsymbol{Y}$. What is $E(\hat{\boldsymbol{\beta}}_I^*) = E[(\boldsymbol{X}_I^T\boldsymbol{X}_I)^{-1}\boldsymbol{X}_I^T\boldsymbol{Y}^*]$?

$$= (8\overline{1} \times \overline{1}' \times \overline{1} \times \widehat{1} \times \widehat{1}) = (8\overline{1} \times \overline{1})' \times \overline{1} \times \underline{1}' = (8\overline{1} \times \overline{1})' \times \underline{1}' \times \underline{1}' \times \underline{1}' = (8\overline{1} \times \overline{1})' \times \underline{1}' \times \underline{$$

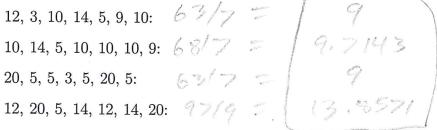
d) What is $Cov(\hat{\boldsymbol{\beta}}_I^*)$? = (x, x, / x, (o, y, y) x, (x, x,) = (8=1x=1), X=1 MSE. I X= X=1x=1)

- MGE(8] 81) 8 FEL (2) 81)

COU(AYX) = A COU(YX) AT

7) Suppose you are estimating the mean μ of losses with $T = \overline{X}$. actual losses 14, 3, 5, 12, 20, 10, 9: $\overline{X} = 10.4286$,

a) Compute $T_1^*,...,T_4^*$, where T_i^* is the sample mean of the i bootstrap sample. bootstrap samples:



b) Now compute the bagging estimator which is the sample mean of the T_i^* : the bagging estimator $\overline{T}^* = \frac{1}{B} \sum_{i=1}^{B} T_i^*$ where B = 4 is the number of bootstrap samples.

robstat old Q8

8) The output below is for forward selection and I_{min} is the minimum C_p model. Here Y = height, the constant $x_{i,1} \equiv 1$, $x_{i,2} = height$ when sitting, $x_{i,3} = height$ when kneeling, $x_{i,4} = head length, x_{i,5} = nasal breadth, and x_{i,6} = span.$

Intercept -42.4846 51.2863 [-192.281, 52.492] 0.000,

0.268]

ХЗ 1.1707 0.0598 [0.992, 1.289]

X4 0.000, 0.840]**X5**

0.000, 1.916]

X6 0.0368 [0.0747, 0.215]

(Intercept) b е 1

TRUE FALSE TRUE FALSE FALSE 2 TRUE FALSE TRUE FALSE FALSE

3 TRUE FALSE TRUE TRUE FALSE TRUE

4 TRUE FALSE TRUE TRUE TRUE TRUE

5 TRUE TRUE TRUE TRUE TRUE TRUE

> tem2\$cp

[1] 14.389492 0.792566 2.189839 4.024738 6.000000

What is the value of $C_p(I_{min})$ and what is $\hat{\boldsymbol{\beta}}_{I_{min},0}$?

CO(IMIN) = 0,792566



