1. (28pts) Consider the model  $\mathbf{y}_{8\times 1} = \mathbf{X}_{8\times 3}\boldsymbol{\beta}_{3\times 1} + \boldsymbol{\epsilon}_{8\times 1}$ , where  $\mathbf{y}$  is blood pressure of 8 individuals,  $\mathbf{X}$  includes intercept (1st column of  $\mathbf{X}$ ) and two covariates: age (2nd column of  $\mathbf{X}$ ) and body weight (lbs) (3rd column of  $\mathbf{X}$ ). More specifically,

$$\mathbf{y} = \begin{bmatrix} 137 \\ 126 \\ 114 \\ 95 \\ 111 \\ 112 \\ 107 \\ 121 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & 26 & 134 \\ 1 & 27 & 138 \\ 1 & 23 & 118 \\ 1 & 24 & 124 \\ 1 & 22 & 123 \\ 1 & 30 & 135 \\ 1 & 20 & 128 \\ 1 & 25 & 131 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \quad \text{and } \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_8 \end{bmatrix} \sim N(0, \sigma^2 \mathbf{I})$$

You should NOT run any software to answer the following questions. However, some computation by calculator maybe needed given the following potential helpful facts.

- The corrected total sum of squares of y is 1476.
- $\bullet \ (\mathbf{X}^T\mathbf{X})^{-1} =$

- $\hat{\sigma}^2 = 145.37$ .
- (a) (5pts) Is each of the following statement correct or not? If it is not correct, please explain why it is wrong and try to correct it.
  - i.  $\beta$  are statistics.

Invorvent,  $\beta$ 's one parameters that we can't observe. We use  $\hat{\beta}$  to estimate them.

ii.  $\epsilon$  are parameters.

Invorcet. E's are random errors.

iii. y is a random variable following multivariate normal distribution with mean value  $0_{8\times 1}$  and variance  $\sigma^2 I_{8\times 8}$ .

Interest, y is a random variable following multivariate around distribution but the mean value E(y)=  $X\beta$ , not E(E), covariance =  $\sigma^2 I_{228}$ 

iv.  $\hat{\sigma}^2$  is a random variable.

Correct. J2 is the estimator of T2 and is a vardom variable

v. 61 is independent with 62.

Correct. random errors are assumed to be independent of each other

(b) (3pts) Fill in the following t-table and please show your work on calculating the Standard Errors.

Standard Parameter Estimate Error t value Pr(>|t|) (Intercept) -22.0801 0.823 0.959 -0.1105 age weight 1.0877 0.299 0.8339

$$\begin{aligned} & ( \circ \vee ( \beta ) = 0^{2} ( \times \times )^{-1} & Var( \beta_{0} ) = 8345.11 & Se( \beta_{0} ) = \sqrt{8345.11} = 9135.16 \\ & = \hat{\sigma}^{2} ( \times \times )^{-1} & Var( \beta_{0} ) = 4.07036 & Se( \beta_{0} ) = \sqrt{4.07036} = 2.07556 \\ & = 145.37 & 57.406 & 0.435 & -0.528 & Var( \beta_{0} ) = 4.07036 & Se( \beta_{0} ) = \sqrt{4.07036} = 2.07556 \\ & = -0.435 & 0.028 & -0.009 & 0.006 & Var( \beta_{0} ) = 0.87222 & Se( \beta_{0} ) = \sqrt{4.07036} = 2.07339 \\ & = -0.528 & -0.009 & 0.006 & Var( \beta_{0} ) = 0.87222 & Se( \beta_{0} ) = \sqrt{4.07036} = 0.02417 \\ & = -0.1055 & -0.02417 & = -0.0548 \\ & = -0.1055 & -0.0548 & + 2 & = \frac{1.0877 - 0}{0.07329} = 1.1647 \end{aligned}$$
(5.765) Test  $\beta_{0} = -\beta_{0}$  using GLH approach. Write out the contrast

(c) (5pts) Test  $\beta_0 = \beta_1 = \beta_2$  using GLH approach. Write out the contrast matrix C, calculate test statistic and specify its null distribution and the corresponding degree of freedom. Though you do not need to calculate the p-value.

$$\beta_0 \ge \beta_1$$

$$\beta_1 = \beta_2$$

$$\beta_0 - \beta_1 = 0$$

$$\beta_1 - \beta_2 = 0$$

$$C = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$$

Hor 
$$\theta = \begin{bmatrix} \beta_0 - \beta_1 \\ \beta_1 - \beta_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
, corresponding contrast matrix  $C = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$ 

Because X is full rank, so D is estimable.

Mso since C is full rank, so & is testable.

$$M_{242} = C(X|X)^{7} C^{1} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 57.406 & 0.435 & -0.528 \\ 643.5 & 0.028 & -0.009 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 0.528 & -0.009 \end{bmatrix} \begin{bmatrix} -0.009 & 0.006 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 0 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 56.504 & 0.926 \\ 0.926 & 0.052 \end{bmatrix} \begin{bmatrix} 0.025 & -0.444 \\ -0.444 & 27.144 \end{bmatrix}$$

$$= \begin{bmatrix} -21.9696 \\ 1.0877 \end{bmatrix} = \begin{bmatrix} -1.1982 \\ -1.1982 \end{bmatrix}$$

 $F-obs = \frac{(\hat{\theta}-\theta_0)'M''(\hat{\theta}-\theta_0)/\alpha}{\hat{\sigma}^2} = \frac{[-21.969b + 1.982][-0.444][-21.969b]/2}{[445.37]}$ 

(d) (5pts) Test  $\beta_1 = \beta_2 = 0$  using GLH approach. Write out the contrast matrix C, calculate test statistic and specify its null distribution and the degree of freedom. Though you do not need to calculate the p-value.

value. Ho:  $Q = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ , corresponding contrast matrix  $C = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ 

Because X is full rank, C is full rank, so & is testable.

$$M_{242} = C(X|X)^{-1}C' = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 5790b & 0.925 & -0.528 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0.925 & 0.008 \\ -0.528 & -0.009 & 0.006 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.028 & -0.009 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.028 & -0.009 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.028 & -0.009 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.028 & -0.009 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.028 & -0.009 \\ 0 &$$

 $\mathcal{V}_{\mathbf{f}}$ : 2, 5
(e) (5pts) Calculate the correlation between  $\hat{\beta}_0$  and  $\hat{\beta}_1$ .

Correlation = 
$$\frac{\text{Cov}(\hat{\beta}_0, \hat{\beta}_1)}{\text{Vor}(\hat{\beta}_0)\text{Vai}(\hat{\beta}_1)} = \frac{63.23595}{8345.11 \times 4.07036} = 0.3431$$

(f) (5pts) What is the interpretation of  $\beta_0$ ,  $\beta_1$ , and  $\beta_2$ , respectively. Is the interpretation of  $\beta_0$  meaningful, if so, why? If not, how to fix this problem?

Bo — the expected blood pressure when age and body weight the valve zero.

& - the expected increase in blood pressure for one unit increase in age.

Bz — the experted increase in blood pressure for one unit increase ix body weight.

The interpretation of Bo is not meaningful because of no biological meaning for BP with age =0, body weight =0. To fix the problem, we can center the age veriable and weight variable by subtracting the average of age and body weight from each observation respectively. In doing so, the interest Bo will be the experted blood pressure toher age is at the observed average and body weight is at the observed average value.

- 2. (20pts) Still use the data presented in problem 1. Suppose we are interested in the event of whether blood pressure is larger than 120. Let  $\tilde{y}_i = 1$ , if  $y_i > 120$ , and  $\tilde{y}_i = 0$  otherwise. Here i = 1, 2, ..., 8 is the index of the 8 individuals. Let  $p_i = Pr(y_i > 120)$ .
  - (a) (5pts) Is  $p_i$  a parameter or a statistic? Given  $p_i$ , what the distribution of  $\tilde{y}_i$ ? Calculate  $\tilde{y}_i$ 's expectation and variance.

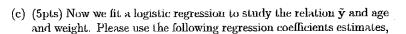
Pi is a parameter

$$\hat{y}_i = \begin{cases} 1 & Pr = Pi; \\ 0 & Pr = I - Pi \end{cases}$$
 Given Pi,  $\hat{y}_i \sim \text{Bernoulli}(p_i)$ 
 $E(\hat{y}_i) = P_i$ 
 $Var(\hat{y}_i) = P_i(I - P_i)$ 

(b) (5pts) Calculate the odds ratio of the event  $y_i > 120$  vs. the event weight > 132.

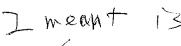
For 
$$y_i > 120$$
,  $\frac{p_i}{1-p_i} = \frac{3/8}{5/8} = \frac{3}{5} = 0.60$   
For weight >132,  $\frac{p_0}{1-p_0} = \frac{3/8}{5/8} = \frac{3}{5} = 0.60$ 

$$OR = \frac{p_{1}(1-p_{1})}{p_{2}/(1-p_{2})} = \frac{0.60}{0.60} = 1$$

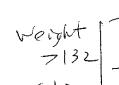


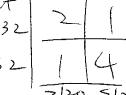
Estimate Std. Error (Intercept) -118.9085 135.9180 age -0.7111 0.9114 weight 1.0373 1.1950

I count this as correct,



 $\frac{\frac{2/3}{1-2/3}}{(1/5)/(4/5)}$ 





What

3

to estimate the probability that blood pressure is larger than 120 for an individual of age 30 and weight 133.

$$\beta = \frac{\exp(\beta_0 + \beta_1 \log_2 + \beta_2 \omega t)}{1 + \exp(\beta_0 + \beta_1 \log_2 + \beta_2 \omega t)} = \frac{\exp(-118.9085 + (-0.711) \times 30 + 1.0373 \times 133)}{1 + \exp(-118.9085 + (-0.711) \times 30 + 1.0373 \times 133)}$$

$$= 0.093$$

(d) (5pts) Please use the regression coefficient estimates in part (c) to calculate the odds ratio of the event  $y_i > 120$  for person B vs. person A. They are of the same age, but B is 10 pounds heavier than A.

$$\log (odds_B) = \hat{\beta}_0 + \hat{\beta}_1 \times age_B + \hat{\beta}_2 \times Wt_B$$

$$\log (odds_A) = \hat{\beta}_0 + \hat{\beta}_1 \times age_A + \hat{\beta}_2 \times Wt_A, \quad wt_B = 10 + Wt_A$$

$$\log (OR_{BVSA}) = \log (odds_B) - \log (odds_A) = \hat{\beta}_2 (VUt_B - Wt_A)$$

$$= \hat{\beta}_2 \times 10$$

$$OR_{BVSA} = e^{10\hat{\beta}_0} = e^{10(1.03\hat{\beta}_0)}$$

$$= 31984$$

- 3. (12pts) Now suppose we know the 8 individuals are from two family. The first four are from one family and the next four are from the other family. In order to accommodate the correlations between individuals within one family, we decide to use a random effect model to study the relation between blood pressure versus age and weight.
  - (a) (4pts) If we use "unstructured" covariance structure, how many parameters of the covariance matrix of the 8 individuals need to be estimated? Write out the covariance matrix using concise notations (you just need to present the form of the matrix, but do not need to calculate the actual values of the matrix elements).

Unstrumed covariance matrix in one family:

Yij = Xij $\beta$  + bi + 2ij two famalies i=2four from a family, j=4 (b) (4pts) If we used "compound symmetry" covariance structure, how many parameters of the covariance matrix of the 8 individuals need to be estimated? Write out the covariance matrix using concise notations.

Using compound symmetry covariance structure, we need to estimate 2 parameters:  $Tb^2$  and  $Tw^2$ . For one family:

(c) (2pts) Which covariance structure (unstructured or compound symmetric) should we use for this dataset and why?

Because of the small sample size in this dataset, we should use compound symmetric covariance matrix because it has fewer parameters than unstructured. If assumption for compound symmetry is not valid, we may need to force a compound symmetry structure with

(d) (2pts) Mixed model parameters can be estimated using either Maximum Likelihood (ML) method or Restricted maximum likelihood (REML) method. In order to compare a model with fixed effects of age and weight vs. the other model with only one fixed effect weight, should we use ML or REML method, and why? (Assume the same covariance structure is used both models.)

We should use ML to compare the two models because the likelihood obtained for models with different fixed effects are not comparable when REML is used to estimate the models. REML maximises the likelihood of the observed residuels, so different degrees of freedom by two models, thus they're not comparable.

4. (25pts) We want to compare two drugs (denoted by A and B) for their effects of reducing cholesterol levels (LDL, in the unit of mg/dL). The following table shows the sample size for each combination of drug and

Drug	Dose	Sample Size $(n_{ij})$	i (drug index)	j (dose index)
	1	100	1	1
Α	<b>2</b>	100	1	2
	3	100	1	3
	1	100	2	1
$\mathbf{B}$	2	100	2	2
	3	100	2	3

(a) (3pts) First consider the dose variable as a categorical variable with 3 levels, and employ an additive model:  $y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk},$ 

where  $i=1, j=1,2, k=1, 2, ..., n_{ij}$ . We use reference cell coding with drug B and dose 3 as reference. Therefore  $\alpha_1$  models the effect of drug A (drug B is reference),  $\beta_j$  models the effect for dose j (j=1 or 2) (dose 3 is reference); and  $e_{ijk}$  ( $k=1, 2, ..., n_{ij}$ ) indicates residual error. If we write this ANOVA model as a regression model: y = Xb + e, what is the dimension of y, X, b and e, and for an ANOVA model, what kind of distribution we usually assume e should follow?

Ybooxi , X100x4, b4x1, e 600x1.
e follows a Gaussian distribution within cell.

(b) (4pts) For the model specified in part (a), write the cell mean for

each combination of drug and dose in terms of  $\mu$ ,  $\alpha_i$  and  $\beta_j$ .

Drug	Dose	Mean
A	1	M+ 21+ B1
Α	2	Mt 21+B
$\mathbf{A}$	3	M+dr.
$\mathbf{B}$	1	u+ B.
В	2	M+ B2
В	3	u'

y = druga dose 1 dose 2

## Y=drug A dose 1 dose 2

(c) (3pts) For the model specified in part (a), fill the following ANOVA table.

8 672849 mass 57.186	
Hudel 22420.3 / (1662)	<.0001
Error 596 233751.2 392.2	
Corrected Total 599 30   036.	

(d) (3pts) If we model the interaction between dose and drug, the model can be written as

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$$

where  $\gamma_{ij}$  indicates interaction effects. Write the cell mean for each dozel dozeld dozeld combination of drug and dose in terms of  $\mu$ ,  $\alpha_i$ ,  $\beta_j$  and  $\gamma_{ij}$ . Explain the meaning of interaction effect  $\gamma_{11}$  by comparing the table in question (b) and the table in this question.

	Drug	Dose	
_	Α.	1	Ut dit Bit Tin
	$\mathbf{A}$	2	M+21+32+12
	${\bf A}$	3	
	В	1	Mt Bi
	В	2	U+β2
	В	3	M+β2 M
			- 1

Til — the dofference in doing effect for dose I versus dose 3.

(e) (2pts) Now if we model dose as a <u>interval variable</u>, with doses equals to 1, 2, 3 and fit a model of LDL with main effects of dose and drug, but <u>no interaction</u>, fill the following ANOVA table

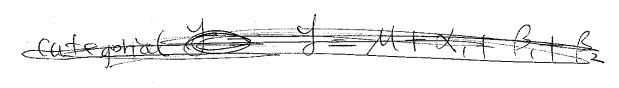
Source DF Squares Mean Square F Value Pr > F Model 2 67203.b 33601.8 85.785 <.0001

Error 597 233844.9 391.7

Corrected Total 599 301048.5

y=drug A dose

of the land of the



(a tegs vical

(f) (3pts) Compare the model using dose as a categorical variable (part (c)) and the model using dose as a interval variable (part (2)) by F-test. Please write down  $H_0$ , calculate F-Statistic, and give the degree of freedom of the corresponding F-distribution when  $H_0$  is true. Though you do not need to calculate the p-value.

We can view the model using close as an interval volvable as a model nested in the categorical dose parameterization model.

J=M+X, drag

+B, (dose=1)+B2(dose=2)

233751,2/596

humerial/interval

Y=N+X, drug

(g) (4pts) Let  $\mu_A$  and  $\mu_B$  be the overall mean values of LDL for drug A and B, respectively. Write  $\mu_A$  and  $\mu_B$  in terms of  $\alpha_i$ ,  $\beta_i$  and  $\gamma_{ii}$ . If we want to test  $H_0: \mu_A = \mu_B$ , write  $H_0$  in terms of  $\alpha_i, \beta_j$  and  $\gamma_{ij}$ , the contrast matrix, and the degrees of freedom.

dose categorical

+ B3 dose

 $M_A$ :  $(M+\partial_1+\beta_1+\Gamma_1)+(M+\partial_1+\beta_2+\Gamma_{12})+(M+\partial_1)$  $= \mu + \lambda_1 + \frac{\beta_1 + \beta_2 + \gamma_1 + \gamma_{12}}{2}$ 

UB: 
$$\frac{(\mathcal{U}+\beta_1)+(\mathcal{U}+\beta_2)+\mathcal{U}}{3}=\mathcal{U}+\frac{\beta_1+\beta_2}{3}$$

interact: 1,594

(h) (3pts) If the design is unbalanced, with sample size shown in the following table. Test  $H_0: \mu_A = \mu_B$ . Write  $H_0$  in terms of  $\alpha_i$ ,  $\beta_j$  and  $\gamma_{ij}$ , the contrast matrix, and the degrees of freedom.

dose categorical

Drug	Dose	Sample Size $(n_{ij})$	i (drug index)	j (dose index)
	1	100	1	1
A	2	100	1	2
	3	50	1	3
	1	100	2	1
B	2	100	<b>2</b>	2
	3	50	2	3

$$\mathcal{U}_{A} = \frac{100 \left( \mathcal{U} + \partial_{1} + \beta_{1} + \gamma_{1} \right) + 100 \left( \mathcal{U} + \partial_{1} + \beta_{2} + \gamma_{1} \right) + 50 \left( \mathcal{U} + \partial_{1} \right)}{250}$$

$$= \frac{250 \mathcal{U} + 250 \partial_{1} + 100 \gamma_{1} + 100 \gamma_{2} + 100 \beta_{1} + 100 \beta_{2}}{250}$$

$$= \mathcal{U} + \partial_{1} + \frac{2}{5} \left( \beta_{1} + \beta_{2} + \gamma_{1} + \gamma_{1} \right)$$

$$\mathcal{U}_{B} = \frac{100 \left( \mathcal{U} + \beta_{1} \right) + 100 \left( \mathcal{U} + \beta_{2} \right) + 50 \left( \mathcal{U} \right)}{250}$$

$$= \frac{250 \mathcal{U} + 100 \left( \beta_{1} + \beta_{2} \right)}{250}$$

$$= \mathcal{U} + \frac{2}{5} \left( \beta_{1} + \beta_{2} \right)$$

$$= \mathcal{U} + \frac{2}{5} \left( \beta_{1} + \beta_{2} \right)$$

$$= \mathcal{U} + \frac{2}{5} \left( \gamma_{1} + \gamma_{12} \right) = \mathcal{U}$$

$$= \mathcal{U} + \frac{2}{5} \left( \gamma_{1} + \gamma_{12} \right) = \mathcal{U}$$

$$= \mathcal{U} + \frac{2}{5} \left( \gamma_{1} + \gamma_{12} \right) = \mathcal{U}$$

$$= \mathcal{U} + \frac{2}{5} \left( \gamma_{1} + \gamma_{12} \right) = \mathcal{U}$$

$$= \mathcal{U} + \frac{2}{5} \left( \gamma_{1} + \gamma_{12} \right) = \mathcal{U}$$

$$= \mathcal{U} + \frac{2}{5} \left( \gamma_{1} + \gamma_{12} \right) = \mathcal{U}$$

$$= \mathcal{U} + \frac{2}{5} \left( \gamma_{1} + \gamma_{12} \right) = \mathcal{U}$$

$$= \mathcal{U} + \frac{2}{5} \left( \gamma_{1} + \gamma_{12} \right) = \mathcal{U}$$

$$= \mathcal{U} + \frac{2}{5} \left( \gamma_{1} + \gamma_{12} \right) = \mathcal{U}$$

$$= \mathcal{U} + \frac{2}{5} \left( \gamma_{1} + \gamma_{12} \right) = \mathcal{U}$$

$$= \mathcal{U} + \frac{2}{5} \left( \gamma_{1} + \gamma_{12} \right) = \mathcal{U}$$

$$= \mathcal{U} + \frac{2}{5} \left( \gamma_{1} + \gamma_{12} \right) = \mathcal{U}$$

$$= \mathcal{U} + \frac{2}{5} \left( \gamma_{1} + \gamma_{12} \right) = \mathcal{U}$$

$$= \mathcal{U} + \frac{2}{5} \left( \gamma_{1} + \gamma_{12} \right) = \mathcal{U}$$

5. (15pts) Still using the data of Problem 4 (with balanced design of 100 samples in each cell). Now we introduce another interval variable "age" and the interaction between drug and dose, fit a model using the following SAS code

proc glm;
class drug;
model LDL= age dose drug drug\*dose/ solution;
run;

and obtained the following output.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	86791.9439	21697.9860	60.26	<.0001
Error	595	214247.6617	360.0801		
Corrected Total	599	301039.6056			

	R-Square	Coeff Var	Root MSE	LDL Mean
0.288307 15.19667 18.97578 124.8	0.288307	15.19667	18.97578	124.8680

Source	DF	Type I SS	Mean Square	F Value	Pr > F
age	1	21309.98280	21309.98280	59.18	<.0001
dose	1	6664.73548	6664.73548	18.51	<.0001
drug	1	58218.74750	58218.74750	161.68	<.0001
dose*drug	1	598.47814	598.47814	1.66	0.1978

Source	DF	Type III SS	Mean Square	F Value	Pr>F
age	1	19205.20980	19205.20980	53.34	<.0001
dose	1	6683.65025	6683.65025	18.56	<.0001
đrug	1	4691.53105	4691.53105	13,03	0.0003
dose*drug	1	598.47814	598.47814	1.66	0.1978

Parameter	Estimate	-	Standard Error	t Value	Pr >  t
Intercept	104.9204188	В	3.94321568	26.61	<.0001
age	0.4855570		0.06648601	7.30	<.0001
dose	5.3125392	В	1.34185926	3.96	<.0001
drug ()	-14,8100813	В	4.10298291	-3.61	0.0003
drug 1	0.0000000	В			
dose*drug 0	-2.4479126	В	1.89876546	-1.29	0.1978
dose*drug 1	0.0000000	В			

plug in the values of

Bis?

(a) (3pts) Write down the fitted model based on the above output. Is dose treated as categories or interval variable?

LDL= Bo+ B, +age + B2 x dose + B3 x drug + B4 x drug + dose + E Dose was receted as continuous here since only drug was used in the class statement.

(b) (2pts) Why is the regression coefficient estimate for "drug 1" is 0 without estimate for standard error? Note the numerical value of drug is 0 for drug A and 1 for drug B.

Because doing ( was used on the reference group and embedded in the intercept. (drug B)

(c) (3pts) Briefly explain what is the difference between Type I SS and Type III SS. Why the Type I SS of age is larger than the Type III SS of age, but the Type I SS of dose\*drug is the same as the Type III SS of dose\*drug?

Type ISS are from added-in-order tests, and they are mutually exclusive and together exhaustive pieces of the model SS. The sizes of Type ISS for a covariate depends on the order the covariate is added to the model, except when Type IISS are from added-last tests, and they are SS for each variable if it was entered last in the model. The size of Type IISS tells how much variance being explained by this variable after accounting for all other variables. Here

all predictors are uncorrelated

Age was added first in the model, so its Type ISS is much larger than its Type III error, 13

The variable added last into the model in added in-order test is equivalent to the added last test of this variable since SS. from these two tests are SS explained by this variable beyond other variables. This is the reason why for doset drug, the

Type ISS is the same as the Type IIISS.

(d) (4pts) Write the contrast matrix to estimate the average LDL level when drug A is used for an individual of age 40. Similarly, Write the contrast matrix to estimate the average LDL level when drug B is used for an individual of age 40. LDL = Pot Brage + Bz dow + Bz dong + By dong-dose+ E

doug A por individual 40:

[ 1 40 dose 1 dose ]

doug A: doug o dright: drigt

drug B for individual 40: because drug B was the reference

[1 40 dose 0 0]

where dose = grand mean of the dose variable

(e) (3pts) Write the contrast matrix to test the hypothesis that the average LDL level for the individuals of age 40 taking drug A is different from the average LDL level for the individuals of age 40 taking drug B. Write the formula to calculate the test-statistic and what is the degree of freedom of this test?

$$\theta = \mathcal{U}_1 - \mathcal{U}_2 = 0$$

$$\theta = M_{1} - M_{2} = \beta_{0} + \beta_{1}(x_{0}) + \beta_{2}(x_{0}) + \beta_{3}(x_{1}) + \beta_{4}(x_{0}) - [\beta_{0} + \beta_{1}(4_{0}) + \beta_{2}(x_{0}) + \beta_{3}(0) + \beta_{4}(0)]$$

$$= \beta_{3} + \beta_{4}(x_{0}) + \beta_{2}(x_{0}) + \beta_{3}(x_{0}) + \beta_{3}(x_{0}) + \beta_{4}(0)$$

$$= \beta_{3} + \beta_{4}(x_{0}) + \beta_{3}(x_{0}) + \beta_{3}(x_{0}) + \beta_{3}(x_{0}) + \beta_{4}(0)$$

$$= \beta_{3} + \beta_{4}(x_{0}) + \beta_{3}(x_{0}) + \beta_{3}(x_{0}) + \beta_{3}(x_{0}) + \beta_{4}(0)$$

$$= \beta_{3} + \beta_{4}(x_{0}) + \beta_{3}(x_{0}) + \beta_{3}(x_{0}) + \beta_{3}(x_{0}) + \beta_{4}(0)$$

$$= \beta_{3} + \beta_{4}(x_{0}) + \beta_{3}(x_{0}) + \beta_{3}(x_{0}) + \beta_{4}(x_{0})$$

$$= \beta_{3} + \beta_{4}(x_{0}) + \beta_{4}(x_{0})$$

$$= \beta_{3} + \beta_{4}(x_{0})$$

$$= \beta_{4} + \beta_{4}(x$$

df: 1, 595