Question 1/

2013, MS-2

a) Write a statistical model to determine if there is a linear trend in glucose values over time. Explain the meaning of the parameters in your model

glucose measurements = intercept month

Intercept = β_0 = expected glucose measurement when month = 0

slope = β_1 = expected change in glucose measurement per month

$$E(g) = \beta_0 + \beta_1(\text{month})$$

b) Estimate the parameters in your model.

$$Y = \begin{bmatrix} g_{11} \\ \vdots \\ g_{j1} \\ \vdots \\ g_{18} \\ \vdots \\ g_{j8} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ & \vdots \\ 1 & \vdots \\ 1 & \vdots \\ \end{bmatrix} \begin{matrix} \}\, n_1 = 100 \\ \\ \}\, n_8 = 300 \text{ etc.} \end{matrix}$$

$$\hat{\beta} = (X'X)^{-1}(X'Y)$$

$$\text{or } \hat{\beta} = \frac{\Sigma X_i Y_i - n\bar{X}\bar{Y}}{\Sigma X_i^2 - n\bar{X}}$$

$$X = \begin{bmatrix} 1 & 1_{100} \\ \vdots & 2_{200} \\ 1 & 3_{400} \\ 1 & 4_{400} \\ 1 & 5_{800} \\ 1 & 6_{500} \\ & 7_{300} \\ 1 & 8_{300} \end{bmatrix}_{3500 \times 2}$$

$$(X'X) = \begin{pmatrix} 1 & \cdots & 1 \\ 1 & 2 \cdots & 8 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ \vdots & 2 \\ \vdots & 3 \\ & 4 \\ & 5 \\ & 6 \\ & 7 \\ & 8 \end{pmatrix} = \begin{pmatrix} 3500 & 16300 \\ 16300 & 87300 \end{pmatrix}^{-1}$$

$$(X'Y) = \begin{pmatrix} 1 & \cdots & 1 \\ 1 & 2 \cdots & 8 \end{pmatrix} \begin{pmatrix} g. \\ \vdots \\ g. \end{pmatrix} = \begin{pmatrix} \Sigma g \\ \Sigma i \cdot g_i \end{pmatrix} = \begin{pmatrix} 394232.8 \\ 1839488.4 \end{pmatrix}$$

$$= \begin{pmatrix} 111.21 \\ 0.31 \end{pmatrix} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}$$

Question 1 | 2013, MS-2

c) Test whether there is a linear trend over time in the last 8 months.

⇒ ~~does~~ is the slope significant

⇒ $H_0: \beta_1 = 0$

$$Cov(\hat{\beta}) = \sigma^2 (X'X)^{-1} = \begin{pmatrix} 1.75 & -0.33 \\ -0.33 & 0.07 \end{pmatrix}$$

use t-test : $\dfrac{\hat{\beta}_1}{\sqrt{Var\hat{\beta}_1}}$

$$= \frac{0.31}{\sqrt{0.07}} = 1.17 \sim t_{df = n-2 = 3500-2}$$

linear trend ⇒ slope coefficient

$$t_{0.975,\, 3498} = 1.960642$$

our test statistic is not greater than the C.V. of t.
Therefore, we fail to reject the null hypothesis.
There is no evidence of a linear trend in time over the first 8 months.

---

d) Based on the results so far, predict the mean fasting glucose in months 9-12.

⇒ Because there is no linear trend, the mean fasting glucose in months 9-12 should be the same as months 1-8, so $\bar{X}_{1-8} = \dfrac{394232.8}{3500} = 112.638$ units

↳ Q's can be easy, like this. not always doing a test or something

Question 1 | 2013, MS-2

(now given SAS output using data from all 12 months)

e) Does mean fasting glucose vary significantly over time?

⇒ Yes. The slope coefficient is significant ($p < 0.0001$)

⇒ For every additional month of measurement (when x increases by 1) the expected mean glucose measurement ($E(Y)$) decreases by -0.60913 units.

---

f) Does the mean fasting glucose change btwn machines.

⇒ $\bar{g}_{1-8} = \bar{g}_{9-12}$

* list all assumptions

$$\bar{g}_{1-8} = \frac{294232.8}{3500} = 112.638$$

$$S^2_{1-8} = \frac{47204540}{3500} - (112.638)^2 = 799.6924$$

$$\bar{g}_{9-12} = \frac{119718.6}{1120} = 106.8916$$

$$S^2_{9-12} = \frac{13514127}{1120} - (106.8916)^2 = 640.3707$$

to test the diff btwn 2 means, I will assume that n is large enough to justify the use of large sample approximations and Slutsky's Thm

i.e. $$\frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{(112.638 - 106.8916)}{\sqrt{\frac{799.6924}{3500} + \frac{640.3707}{1120}}} \sim N(0,1)$$ so c.v. = 1.96 for $\alpha = 0.05$

$$= 6.42$$

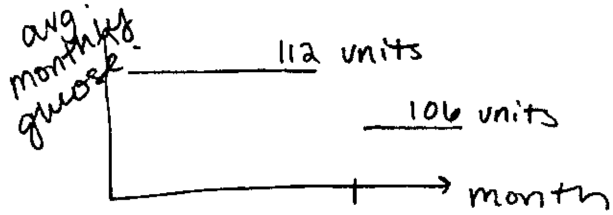
⇒ Reject the null. The two averages are not the same.

Question 1

2013, MS-2

g) Assuming no Δ over months 9-12, what is the interpretation of the model in e)?

It means that while there is a diff in months (1-8) and (9-12) there is no trend the graph would look like



---

h) assume month 13 is the same as 9-12

so avg = 106.8916

$$0.95 = \left(-1.96 < \frac{106.8916 - \mu_{13}}{\sqrt{640.3707/1120}} < 1.96\right)$$ then solve for μ_{13}

---

i) They should adjust measurements made in months 1-8 by subtracting the difference betwn averages per machine.

$(112.638 - 106.8916) = \delta_1$

each value in months 1-8 should be

$g^*_{ij} = g_{ij} - \delta_1$ for $i = 1$ to 8.

2013 Applied

pgl ≥ 126 mg/dl $\Rightarrow$ T II diabetes

STATE ASSUMP. for any tests

8 months (old) + 4 months (new) = 12 months of data total

n_i = # people who were assayed in month i

g_{ij} = pgl for person j in month i

Σ_j = summation over values for each indiv. per month i

a) write a model to determine whether there is a trend in glucose values over time. explain the meaning of the parameters in your model.

② PA levels pre- and post- trail construction
- adjust for # rain, # cold days, dewpoint (↑ = humid), # hot days

$$PAPOST = \beta_0 + PAPRE + (RAIN-5) + (COLD-1) + (RAIN-5)*(COLD-1)$$
$$+ (DEW-60) + (HOT-3) + (DEW-60)*(HOT-3)$$

a) Fill in missing parts on ANOVA table

$8 = p$ (7 + int)

Source	df	SS	MS	F
model	7	6467394.63	SS/df	MSM/MSE = 14.6083 (good, p-val is small)
error	293	18530998.39	SS/df	
total	300	24998393		

b) Report a test that HOT is unrelated postpa
(main effect + interaction should not be 0)

$H_0 : \beta_H = \beta_{H*D} = 0$

$$C = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \theta_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$F = \frac{(\hat{\theta}-\theta_0)' M^{-1} (\hat{\theta}-\theta_0) / a}{\hat{\sigma}^2}$$

where $\hat{\theta} = C\hat{\beta} = \begin{pmatrix} -275.93\ldots \\ 25.735\ldots \end{pmatrix}$

$M^{-1} = (C(X'X)^{-1}C')^{-1}$

$\theta_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

a = rank of $C = 2$

$\hat{\sigma}^2$ = MSE = from part a)

* need $(X'X)^{-1}$

c) $H_0: \beta_{PAPRE} = 0$

in-order: $F = 88.62$, $p < 0.0001$

added-last: $F = 80.83$, $p < 0.0001$

There is evidence that the level of a students PA pre-trail is significantly related to their PA post-trail const.

Both in / last are significant.

In a simpler reg. (in order), they are related in the absence of other vars. When added after all other covariates / adjusting for, there is still a sign. relationship.

d) same as b w/ diff C matrix

1) do you test the interaction yes

2) need $(X'X)^{-1}$ $\longrightarrow$ add T I SS's $\Rightarrow$ SSH

e) Discussion.

(int = ~~effect~~ if wed + raing etc ...) OK

$(X'X)^{-1}$

③ predicting sudden death using wt / BP / BMI / smoke / age
{w/o prior CHD}

a) assess sign. of each risk factor + explain

$\beta / SE(\beta) \sim N(0,1)$ in $n >>>$

	Wald	compared to 1.96
BP	0.2	NS
BMI	3	S
smoke	0.35	NS
age	4	S

C

BP + smoke do not significantly predict ... add interpretations etc.

BMI / age do ...

b) OR btwn women 10 yrs apart

$\exp(0.08 * 10) = 2.23$

For every 10 yr ↑ in age, the odds of sudden death is 2.23x greater

c) 95% CI for the OR. ~~craig has 0.2, which is wrong.~~

$\exp(0.08*10 \pm 1.96(0.02)) = (2.14, 2.31)$ → notwrong! $\boxed{0.2(10)}$

d) $p = \dfrac{\exp(X\beta)}{1+\exp(X\beta)} = \dfrac{\exp(-15.3 + 110*0.002 + 36*0.06 + 0.007*20 + 0.08*60)}{1+\exp(\text{''})}$

$= 0.0004775447$

probability is very small

$p = \dfrac{\exp(X\hat{\beta})}{1+\exp(X\hat{\beta})}$ bounds for this, then transform b/c invariant

e) We would need the SE of the probability

comes from $(X'X)^{-1}$ matrix, we don't have that

Craig: est = −8.34, SE = 3.3

$\dfrac{\exp(-8.34 \pm 1.96(3.3))}{1+\exp(-8.34 \pm 1.96(3.3))}$

4. Two treatments $\Rightarrow$ wt reduction in kg after 1 month
↳ diet
exercise

$n = 100$ total
$n_1 = 25 \Rightarrow$ diet
$n_2 = 25 \Rightarrow$ exercise
$n_3 = 25 \Rightarrow$ both
$n_4 = 25 \Rightarrow$ neither

x_1 = ind. diet
x_2 = ind. exercise
x_3 = diet * exercise

Several regression models were fitted

a) Fill in ?'s in table

$\hat{\beta}$'s are $(X'X)^{-1}(X'y)$

model 1: $(X'X)^{-1} = (100)^{-1}$

$X = (1)_{100\times1}$ $y = \begin{pmatrix} kg \\ kg \end{pmatrix}_{n\times1}$ $\Rightarrow (X'X)^{-1}X'y = \frac{1}{100}(\Sigma y_i) = 9.88$

$X'y = \Sigma y_i$ $\Leftarrow$

$\Sigma y_i = 988$

model 2

$X = \begin{pmatrix} 1 & 1_{25} \\ \vdots & \\ 1 & 0_{75} \end{pmatrix}$ ← 25 received diet

$X = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}$

$(X'X) = \begin{pmatrix} 1 & \cdots & 1 \\ 1 & \cdots 1\,0 \cdots & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & 0 \end{pmatrix} = \left\{ \begin{pmatrix} 100 & 25 \\ 25 & 25 \end{pmatrix} \right\}^{-1} =$ $(X'X) = \begin{pmatrix} 100 & 50 \\ 50 & 50 \end{pmatrix}$
$(2\times100)(100\times2) =$

$(X'y) = \begin{pmatrix} 1 & \cdots & 1 \\ 1 & & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_{100} \end{pmatrix} = \begin{pmatrix} \Sigma y_i \\ \Sigma y_i \text{ in diet} \end{pmatrix}$ Still OK.

→ (?) need this.
model 4 is a good shot.

model 4

$$X = \begin{pmatrix} 1_{100} & 1_{25} & 0_{25} \\ & 0_{25} & 1_{25} \\ & & 0_{50} \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 100 & 50 & 50 \\ 50 & 50 & 25 \\ 50 & 25 & 50 \end{pmatrix}$$

wrong essence

$$(X'X) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}^{*} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} = \left\{\begin{pmatrix} 100 & 25 & 25 \\ 25 & 25 & 0 \\ 25 & 0 & 25 \end{pmatrix}\right\}^{-1} \overset{?}{=} \text{(don't need yet...)}$$

$$(X'Y) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}^{*} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \Sigma y_i \\ \Sigma y_i \text{ group 1} \\ \Sigma y_i \text{ group 2} \end{pmatrix}$$

$$\underset{3\times 3}{(X'X)^{-1}} \underset{3\times 1}{(X'Y)} = \begin{pmatrix} 4.63 \\ 5.77 \\ 4.72 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} \Sigma y_i \\ \Sigma y_i\, g1 \\ \Sigma y_i\, g2 \end{pmatrix} = (X'X)(\hat\beta) = \begin{pmatrix} 987.5 \\ 638.00 \\ 611.75 \end{pmatrix}$$

Craigs way: weighted avg. of the two $E(Y)$ should equal the int. only overall $E(Y)$

$$\Rightarrow 9.88 = \frac{6.99(50) + (6.99 + \beta_1)50}{100}$$

$$\Rightarrow \beta_1 = 5.78$$

$\Rightarrow$

Model 2: $(X'X)^{-1} = \left\{\begin{pmatrix} 100 & 50 \\ 50 & 50 \end{pmatrix}\right\}^{-1}$

$(X'y) = \begin{pmatrix} 987.5 \\ 638.00 \end{pmatrix}$ $\quad \hat\beta = \begin{pmatrix} 6.99 \\ 5.77 \end{pmatrix}$

Model 3: $(X'X)^{-1} = \left\{\begin{pmatrix} 100 & 50 \\ 50 & 50 \end{pmatrix}\right\}^{-1}$

$X'y = \begin{pmatrix} 987.5 \\ 611.75 \end{pmatrix}$ $\quad \hat\beta = \begin{pmatrix} 7.515 \\ 4.720 \end{pmatrix}$

same logic

b) Test the hypothesis that diet and exercise neither enhance nor antagonize each other WRT weight L

↳ interaction of the two = 0

$H_0: \beta_3 = 0$

$C = (0 \quad 0 \quad 0 \quad 1) \qquad \theta_0 = 0$

$$F = \frac{(\hat\theta - \theta_0)' M^{-1} (\hat\theta - \theta_0)/a}{MSE}$$

$M^{-1} = C(X'X)^{-1}C'$

$= (0\,0\,0\,1)(XTXINV)\begin{pmatrix}0\\0\\0\\1\end{pmatrix}$

$\hat\theta = 0.23$

$MSE = 2263/96$

$= 0.00168$

$F \sim 1, 96$

⇒ Fail to reject the interaction estimate is not significant

(no antagonism / enhancement

~~craig's was~~ ~~$(SSE_S - SSE_L)/$~~ ~~$SSE_L/df_{E,L}$~~ 0.042

c) 95% CI for diet under Model 4

$5.77 \pm t_{0.75, df=97}(SE(\hat\beta_1))$

$5.77 \pm 1.9847(\sqrt{0.9336})$

$(3.86, 7.68)$

$SE\hat\beta_1 = \sqrt{\{\sigma^2 (X'X)^{-1}\}_{11}}$

$= \sqrt{0.9336}$

$(X'X)^{-1}$ given

d) compute a 95% interval for the expected wt red of diet = 1 and ex = 1

point = $4.63 + 5.77 + 4.72 = 15.12$

$\hat y \sim SN(X\beta, \sigma^2 H) \quad H = (X(X'X)^{-1}X') \quad (663)$

$\hat\beta = \begin{pmatrix}4.63\\5.77\\4.72\end{pmatrix} \quad x = (1 \quad 1 \quad 1)$

$\underset{1\times3}{X}\{\underset{(3\times1)(1\times3)}{X'X}\}^{-1}\underset{3\times1}{X'}$

craig got 6.7002 w/ same $X_f = (1\,1\,1)$

$= (X_f \cdot (X'X)^{-1} X_f') \cdot MSE$

more d)

e) H_0 : p_comp_diet = p_comp_ex

(15/25) (23/25)

then test w/ proportion testing

f) Write a short summary of the results... OK.