STAT 2132
orly olbum

HOMEWORK 8

1. $Y_{ij} = x_{ij}^T\beta + \delta_i + \epsilon_{ij}$

$j = 1, \dots, n_i \quad , \quad i = 1, \dots, r$

$x_{ij}, \beta \in \mathbb{R}^p$ are non-random

$\delta_i \overset{iid}{\sim} N(0, \sigma_\delta^2)$, $\epsilon_{ij} \sim N(0, \sigma^2)$, $\delta_i \perp \epsilon_{ij}$

* not necessarily balanced *

$n = \sum_{i=1}^{r} n_i$ $\qquad Y \in \mathbb{R}^n$, arranged by i then j

ⓐ Show that for $\delta \sim N(0, \sigma_\delta^2 I_r), \epsilon \sim N(0, \sigma^2 I_n)$ and $X \in \mathbb{R}^{n \times p}$, $Z \in \mathbb{R}^{n \times p}$,

$$Y = X\beta + Z\delta + \epsilon$$

Find an expression for Z.

$$Z = \begin{bmatrix} 1 & 0 & \cdots \\ 1 & 0 & \cdots \\ 0 & 1 & \cdots \\ \vdots & 1 & \cdots \\ \vdots & 0 & \vdots \\ 0 & 0 & 1 \end{bmatrix} \begin{matrix} \rightarrow n_i \\ 1\text{'s} \end{matrix}$$

$n = \sum_{i=1}^{r} n_i$

we have

$Y_i : n_i \times 1 \qquad Z_i : n_i \times r \rightarrow Z = \text{diag}(Z_1, \dots, Z_n)$

$X_i : n_i \times p \qquad \delta_i : r \times 1 \rightarrow \delta \sim N(0, G), G = \sigma_\delta^2 I_r$

$\beta : p \times 1 \qquad \epsilon : n_i \times 1 \rightarrow \epsilon \sim N(0, R), R = \sigma^2 I_n$

$$\underbrace{N \times 1}_{Y \,(n_i \times 1)} = \underbrace{N \times 1}_{X\beta \,(n_i \times p,\ p \times 1)} + \underbrace{N \times 1}_{Z\delta \,(n_i \times r,\ r \times 1)} + \underbrace{N \times 1}_{\epsilon \,(n_i \times 1)}$$

so $Y \sim (X\beta, V)$ where $V = \text{var}(Y) = \text{var}(X\beta + Z\delta + \epsilon)$

$= 0 + \text{var}(Z\delta) + \text{var}(\epsilon) = ZGZ^T + R$

ⓑ X has full column rank; Show that $Q \in \mathbb{R}^{n \times (p-q)}$ exists and for all $\beta \in \mathbb{R}^p$,

$\tilde{Y} = Q^T Y = (Q^T Z)\delta + \tilde{\epsilon} \quad , \quad \tilde{\epsilon} \sim N(0, \sigma^2 I_{n-p})$

Is Q unique? why or why not?

$Q : n \times (p-q) \qquad Q^T : (p-q) \times n \qquad Z : n \times r \qquad \delta : r \times 1$

$\tilde{Y} = Q^T Y = Q^T X\beta + Q^T Z\delta + Q^T\epsilon$

since Q is orthonormal and Q^T's rows are orthogonal to X, $Q^T X = 0$

$= Q^T Z\delta + \tilde{\epsilon}$ where $\tilde{\epsilon} = Q^T\epsilon$, $\tilde{\epsilon} \sim N(0, \sigma^2 I_{n-p})$

So, $E(\tilde{Y}) = 0$ and if $\text{var}(\delta) = \sigma_\delta^2 I_r = \Sigma_\delta$ & $\text{var}(\epsilon) = \sigma^2$,

$\text{var}(\tilde{Y}) = Q^T Z \Sigma_\delta Z^T Q + \sigma^2 Q^T Q$

which does not depend on the choice of Q, because Q is not unique → columns of Q are orthogonal to X and we can have many different Q's

(1, continued)

$\tilde{Z} = Q^T Z$,

col rank$(\tilde{Z}) = d \le r$

$\tilde{Y} = Q^T Y$

$= \tilde{Z}\delta + \tilde{\varepsilon}$

$= Q^T Z + \tilde{\varepsilon}$

ⓒ $\tilde{H}$ is orth proj matrix of $\tilde{Z}$

(i) For SSE $\perp\!\!\!\perp$ SSTR, need to satisfy

$(SSE)(SSTR) = 0$ and $SSE + SSR = I$

(Fisher & Cochran Theorem?)

→ assistance from 2131 HW6, q.1

$SSE = \tilde{Y}^T(I_{n-p} - \tilde{H})\tilde{Y}$

$SSTR \ \tilde{Y}^T \tilde{H} \tilde{Y}$

$A = I_{n-p} - \tilde{H}$

$B = \tilde{H}$

If we show $AB^T = 0$, $AY \perp\!\!\!\perp BY$ and we can show that $Y^TAY \perp\!\!\!\perp Y^TBY$

$AB^T = (I_{n-p} - \tilde{H})\tilde{H}^T = I_{n-p}\tilde{H}^T - \tilde{H}\tilde{H}^T = \tilde{H} - \tilde{H} = 0$

by earlier homeworks

$Y \sim N(X\beta, Z\Sigma_\delta Z^T + \sigma^2 I_n)$

$\tilde{Y} \sim N(0, Q^T Z \Sigma_\delta Z^T Q + \sigma^2 Q^T Q)$

From 2132, If $AB^T = 0$, $AY \perp\!\!\!\perp BY$

$AY \sim N(AX\beta, A[Z\Sigma_\delta Z^T + \sigma^2 I_n]A^T)$

$BY \sim N(BX\beta, B[Z\Sigma_\delta Z^T + \sigma^2 I_n]B^T)$

then $(AY, BY) \sim BVN$ and $AY \perp\!\!\!\perp BY$ if

$cov(AY, BY) = 0$

$cov(AY, BY) = E[A(Y - X\beta)(Y - X\beta)^T B^T]$

$= A\,E[(Y - X\beta)(Y - X\beta)^T]\,B^T$

$= A[Q^T Z \Sigma_\delta Z^T Q + \sigma^2 I_n]B^T$

$= AB^T[Q^T Z \Sigma_\delta Z^T Q + \sigma^2 I_n] = 0$

so, $AY \perp\!\!\!\perp BY$ and $A\tilde{Y} \perp\!\!\!\perp B\tilde{Y}$

we can show that $Y^TAY \perp\!\!\!\perp Y^TBY$ (and by extension, $\tilde{Y}^TA\tilde{Y} \perp\!\!\!\perp \tilde{Y}^TB\tilde{Y}$)

Suppose $AB = 0$ (and we know $AB^T = 0$),

say $q_1 = Y^TAY$ and $q_2 = Y^TBY$

also, $AB = ATT^TB = 0$

$\rightarrow T^TATT^TBT = 0$, say $C = T^TAT$, $D = T^TB^T$

Making $CK = (T^TAT)(T^TBT) = T^TABT = T^TT = 0$

so, $CK = KC$

→

(1, continued)

(c)(i) ...

Let $Q^T C Q = \begin{bmatrix} E_1 & 0 \\ 0 & 0 \end{bmatrix}$ and $Q^T D Q = \begin{bmatrix} 0 & 0 \\ 0 & E_2 \end{bmatrix}$

For $W = Q^T T^{-1} Y$, $E(w) = Q^T T^{-1} Y$

where W is a vector of standard normals

$var(w) = Q^T T^{-1} \mu = Q^T T^{-1} T T^T T^{T^{-1}} Q = I$

so, $Y = TQV$ and $Y^T = V^T Q^T T^T$

$q_1 = Y^T A Y = V^T Q^T T^T A T Q V = V^T Q^T T^T (T^{T^{-1}} C T^{-1}) T Q V$

$= V^T Q^T C Q V = V_1^T E_1 V_1$

and (by similar math) $q_2 = V_2^T E_2 V_2$

where V_1 is the first half, V_2 is the second half, they are separate, making

$q_1 \perp\!\!\!\perp q_2 \Rightarrow Y^T A Y \perp\!\!\!\perp Y^T B Y \Rightarrow \tilde{Y}^T A \tilde{Y} \perp\!\!\!\perp \tilde{Y}^T B \tilde{Y}$

$\longrightarrow$

(1, continued)

(ii) Show that $MSE = (n-p-d)^{-1} SSE$ and $MSTR = d^{-1} SSTR$ are unbiased estimators for σ^2 and $\sigma^2 + d^{-1} Tr(\tilde{Z}^T \tilde{Z}) \sigma_\delta^2$.

In other words, show that

$E(MSE) = \sigma^2$ and $E(MSTR) = \sigma^2 + d^{-1} Tr(\tilde{Z}^T \tilde{Z}) \sigma_\delta^2$

where $SSE = \tilde{Y}^T (I_{n-p} - \tilde{H}) \tilde{Y}$ and $SSTR = \tilde{Y}^T \tilde{H} \tilde{Y}$

$E(MSE) = E[(n-p-d)^{-1} SSE] = E[(n-p-d)^{-1} \tilde{Y}^T (I_{n-p} - \tilde{H}) \tilde{Y}]$

SSE is quadratic form, say $A = \sigma^{-2} (I_{n-p} - \tilde{H})$,

then $\tilde{Y}^T A \hat{Y} = \sigma^{-2} SSE \sim \chi^2_{rank(A)} \longrightarrow rank(I_{n-p} - \tilde{H}) = n-p-d$

so $E(MSE) = E[(n-p-d)^{-1} SSE]$ for $\chi^2 = \sigma^2$

$E(MSTR) = E[d^{-1} SSTR] = E[d^{-1} \tilde{Y}^T \tilde{H} \tilde{Y}]$

SSTR is also quadratic form, $A = d^{-1} Tr(\tilde{Z}^T \tilde{Z}) \sigma_\delta^2 \tilde{H}$

then $\tilde{Y}^T A \tilde{Y} = \sigma^{-2} SSTR \sim \chi^2_{rank(A)} \longrightarrow rank(A) = d^{-1} Tr(\tilde{Z}^T \tilde{Z}) \sigma_\delta^2$

so $E(MSTR) = E[d^{-1} SSTR]$ for $\chi^2 = \sigma^2 + d^{-1} Tr(\tilde{Z}^T \tilde{Z}) \sigma_\delta^2$

(iii) If the null $H_0: \sigma_\delta^2 = 0$ is true, show that

$F = \frac{MSTR}{MSE} \sim F_{d, n-p-d}$

Since the ratio of χ^2 variables is F-distribution, and we have shown that SSE is independent to SSTR, so their ratio

$\frac{SSTR/d}{SSE/(n-p-d)}$ will have

F-distribution with degrees of freedom $(d, n-p-d)$ for a true null

2. (a) Suppose $Y_1, \dots, Y_K$ all have finite second moments with $corr(Y_r, Y_s) = \rho$ for $\forall\, r \neq s \in \{1, \dots, K\}$ ($Y_1, \dots, Y_K$ do not necessarily have the same variance) Show that $\rho \geq \frac{-1}{K-1}$

If $Y_1, \dots, Y_K$ are stacked into a vector Y where $var(Y) = \sigma_P^2 B + \sigma^2 I_K$ and $B \in \mathbb{R}^{n \times n}$ is a <u>partition matrix</u>, by individuals

$$B_{rs} = \begin{cases} 1 & , r,s \text{ from same individual} \\ 0 & , \text{ow} \end{cases}$$

and we have the max eigenvalue $\to \lambda_{max}$ is the number of groups K,

$$\text{so } corr(Y_r, Y_s) = \frac{\sigma_P^2 \theta}{\sigma_P^2 + \sigma^2} \geq \frac{-1}{\lambda_{max} - 1} = \frac{-1}{K-1}$$

(b) Give an intuitive argument as to why $\rho > -1$ for $K \geq 3$.

If K is the number of samples we take from each individual and we take more and more ($K \to \infty$), we would expect the variance to stay the same, and small, which means the correlation between two observations from the same individual will grow towards 0

ie, as $K \to \infty$, $\rho \to 0$

and mathematically,

if $\rho = \frac{-1}{K-1}$, $K \geq 3$, $\rho \to 0$

3. $Y = X\beta + \epsilon$, $X \in \mathbb{R}^{n \times p}$ is full rank, $\beta \in \mathbb{R}^p$, $E(\epsilon) = 0$, $var(Y) = \sigma^2 I_n$

$var(Y) = \sigma^2 I_n$

(a) Let

$$\{\hat{\sigma}^2_{ML}, \hat{\beta}\} = \underset{\sigma^2 > 0, \beta \in \mathbb{R}^p}{\arg\max} \left[-\frac{1}{2} \log\{\det(\underbrace{\sigma^2 I_n}_{var(Y) = \sigma^2 I_n})\} - \frac{1}{2\sigma^2}(Y - X\beta)^T (Y - X\beta) \right]$$

be the max quasi-likelihood estimates for σ^2, β. Show that $E(\hat{\sigma}^2_{ML}) = \frac{n-p}{n}\sigma^2 < \sigma^2$

say $l = -\frac{1}{2}\log\{\det(\underbrace{V}_{V = var(Y)})\} - \frac{1}{2\sigma^2}(Y - X\beta)^T(Y - X\beta)$

need to take derivative of l wrt σ^2 to find $\hat{\sigma}^2_{ML}$

$$\nabla_{\sigma^2} l = -\frac{1}{2} Tr[(\sigma^2 I_n)^{-1} I_n] + \frac{1}{2\sigma^{4}}[(Y - X\beta)^T(Y - X\beta)]$$

$= \frac{-n}{\sigma^2} + \frac{1}{\sigma^4}[(Y - X\hat{\beta})^T(Y - X\hat{\beta})]$ → replace β with $\hat{\beta}$

$= 0$ (set to zero

$$\frac{n}{\sigma^2} = \frac{1}{\sigma^4}[(Y - X\hat{\beta})^T(Y - X\hat{\beta})]$$

$$\hat{\sigma}^2_{ML} = n^{-1}[(Y - X\hat{\beta})^T(Y - X\hat{\beta})]$$

$$E(\hat{\sigma}^2_{ML}) = E[n^{-1}(Y - X\hat{\beta})^T(Y - X\hat{\beta})]$$

$= n^{-1} E[(Y - X\hat{\beta})^T(Y - X\hat{\beta})]$

$= n^{-1} E(\hat{\epsilon}^T\hat{\epsilon})$

From 2131 HW5 proof

$= n^{-1} E[\hat{\epsilon}^T(I_n - H)\hat{\epsilon}]$

$= n^{-1} Tr[(I_n - H) var(\epsilon)] + \underbrace{E(\epsilon^T(I_n - H)\epsilon)}_{0}$

$= n^{-1}\sigma^2 Tr(I_n - H)$

$= n^{-1}\sigma^2[Tr(I_n) - Tr(H)]$

$= n^{-1}\sigma^2(n - p) < \sigma^2$

↳ $\frac{n-p}{n}\sigma^2$

Problem 4

The data in the file "sleep.txt" (with 40 rows of data) give the average hours (times 10) of sleep (in the third column) of 10 insomniacs (indexed with the numbers from 1 to 10, in the first column of the file) without treatment (A) and with three different drugs (B,C and D), of which C and D are of the same general type (but are not identical) and B is a different type of drug. The averages are over a varying number of nights (from 3 through 9), but the specific number of nights for each entry is unavailable.

(a)

Fitting an additive fixed effects model (treating treatment and individual as fixed effects), estimate sigma, the standard deviation of the errors. Next, using only the data for treatments C and D (i.e., leave out B and A), fit an additive fixed effects model and again estimate sigma, the standard deviation of the errors. Compare the two estimates of sigma and give an explanation for any difference you find.

```
model1 = aov(Hours ~ Individual + Treatment, data = sleep)
summary(model1)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Individual   9   9070  1007.8   7.504 2.02e-05 ***
## Treatment    3   4108  1369.4  10.197 0.000116 ***
## Residuals   27   3626   134.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
sleep_c = sleep[sleep$Treatment == 'C',]
sleep_d = sleep[sleep$Treatment == 'D',]
sleep_cd = rbind(sleep_c, sleep_d)

model2 = aov(Hours ~ Individual + Treatment, data = sleep_cd)
summary(model2)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Individual   9   5595   621.7  34.024 6.85e-06 ***
## Treatment    1      0     0.1   0.003    0.959
## Residuals    9    164    18.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The square root of MSE will be the estimate for sigma for each model. For the model with all four treatments, sigma-hat = 11.59 and for the model with just treatments C and D, sigma-hat = 4.28. The subsetted model has a much smaller sigma-hat, which indicates much less variation between treatments C and D, or conversely, much more variation between all of the treatments.

#(b) *Is there any evidence of a person x treatment interaction? If so, what is the nature of this interaction? In particular, does your answer to the previous point provide evidence of an interaction?*

```
model3 = aov(Hours ~ Individual + Treatment + Individual*Treatment, data = sleep)
summary(model3)
```

```
##                        Df Sum Sq Mean Sq
## Individual             9   9070  1007.8
## Treatment              3   4108  1369.4
## Individual:Treatment  27   3626   134.3
```

Once we add an interaction to the model between Individual and Treatment, we get almost the same exact Mean Sq calculations, which tells us that the interaction accounts for all errors from the first model in (a). If it accounts for all outside influences, we may not need to include the interaction in the model. This result may also indicate that one or both of these terms should be treated as random effects rather than as fixed effects.

#(c) *Suppose the 10 patients can be thought of as a random sample from the population of insomniacs. Assume for the rest of the problem that the average hours of sleep is normally distributed.*

(i) Write down a model for the average hours of sleep for insomniac patients, assuming treatment is a fixed effect. Make sure to allow for the possibility of a patient x treatment interaction if you found evidence for it in part (b).

If we assume there is an interaction, the model is as follows:

Hours = mean + Individual (random) + Treatment (fixed) Individual*Treatment + error

$$y_{ij} = \mu_{..} + \rho_i + \tau_j + (\rho\tau)_{ij} + \epsilon_{ij}$$

Where Yij are hours indexed by i treatments (i = A, B, C, D) and j individuals (j = 1, . . . , 10). We assume mu_i's are iid, e_ij's are iid, and mu_i's are independent of e_ij's (as well as the interaction term). We assume hours are normally distributed. We also consider the restriction on the sum of all tau's to be zero, and the sum of all interaction terms to be 0.

(ii) Assuming a patient x treatment interaction, use part (i) to derive a model for the average number of hours a patient will sleep after being treated with drug B, C or D, given that they slept y0 hours before treatment.

Hours|y0 = Individual (random) + Treatment (fixed) Individual*Treatment + error

$$y_{ij}|y0 = y0 + \rho_i + \tau_j + (\rho\tau)_{ij} + \epsilon_{ij}$$

Now, instead of a mean as the intercept, we have the y0 hours slept before treatment as a condition on the model. All other assumptions from (i) hold.

(iii) Given your results, do you think the mean effect of each drug over this population is a quantity of clinical importance? When answering this question, think about whether or not the drug you would recommend an insomniac take depends on how many hours they currently sleep..

What we've seen is that the effect of each drug is a quantity of clinical importance, based on the models above and their interpretations. We would want to treat individuals with different hours of sleep with different treatments - i.e., yes it does matter how many hours an insomniac sleeps when determining their treatment.