Biostat 250B HW 4

Elvis Cui
Han Cui
Dept. of Biostat
UCLA

1 a In a standard 1-way ANOVA with k groups and n observations per group with the usual notation, show that the statistic

$$T_{k,n} = \max_{1\le i<j\le k} \; n^{1/2} \,|\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}| - (\mu_i - \mu_j)| / \hat{\sigma}$$

is distributed as the studentized range distribution and identify its degrees of freedom.

b Find the asymptotic distribution of $T_{k,n}$ as n tends to infinity.

2 The data you received contains results from a 1-way ANOVA setup with 5 patients' responses in each of the 5 treatment groups.

a Perform Hartley's test, Levene's test, Brown-Forsythe's test and Bartlett's test to ascertain whether variances from the 5 groups are equal at the 0.05 significance level. Report the 4 p-values for the 4 tests.

b Someone told you that there is also a O'Brien's test for testing homogeneity of variances in a 1-way ANOVA setup. Do some research on this test and write a description of O'Brien's test, up to ¾ page in length.

3 Refer to the Cook and Weisberg's paper on diagnosing heteroscedasticity.

a Describe, in not more than ¾ of a page length, the statistical setup and the basics of conducting a score test to draw statistical inference on model parameters.

b Derive, on your own, the test statistic for testing heteroscedasticity using the score test for the situation described in the above paper and see whether your results agree with equations (8-10).

4 Let $U = (U_1, U_2, .. U_p)'$ be a vector of principal components of X. Then $U_i = a_i' X$ for some vector a_i of length 1, i=1,2,.., p. Show that

a $\text{var } a'X \le \text{var } U_1$

b if $a'X$ is uncorrelated with $U_1, U_2, \ldots, U_{i-1}$, then $\text{var } a'X \le \text{var } U_i$.

[In the class, we discussed principal components in the context of a linear regression model $Ey=X\beta$. Here we have a very long vector X of random variables, and we want to work with a shorter vector formed by extracting a small number of linear combinations of the vector X (i.e. first few principal components of the vector X.]

5 Refer to the paper posted on the website where the authors used principal components to analyze a public health data. Read the paper to appreciate how principal components analysis is used in practice.

6 Let b be the least squares estimator and let $b_k = (X'X+kI)^{-1}X'y$ be the ridge estimator for estimating β in the standard linear model $EY=X\beta$ and k is a positive constant. Recall the methodology employed in class on how we expressed b_k in terms of b and then derived the bias and mean square error (MSE) of b_k. Show that the same methodology can be used to show that there are values of k for which the predictor of the mean response from the ridge regression perform better than that from the least squares regression, i.e. show that tr MSE of Xb_k is smaller than that of Xb for some values of k.

Q1:

1 a In a standard 1-way ANOVA with k groups and n observations per group with the usual notation, show that the statistic

$$T_{k,n} = \max_{1\le i<j\le k} n^{1/2}\, |\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}| - (\mu_i - \mu_j)|/\hat{\sigma}$$

is distributed as the studentized range distribution and identify its degrees of freedom.

b Find the asymptotic distribution of $T_{k,n}$ as n tends to infinity.

Sol. (a): $\sqrt{n}\,(\bar{Y}_{i\cdot} - \mu_i)/\sigma \sim N(0,1)$

By Cochran: $k(n-1)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{k(n-1)}(0)$

$$\Rightarrow T_{k,n} = \max_{1\le i<j\le k} \frac{\sqrt{n}\,((\bar{Y}_{i\cdot}-\bar{Y}_{j\cdot})-(\mu_i-\mu_j))}{\hat{\sigma}}$$

$$= \max_{1\le i<j\le k} \frac{\sqrt{n}\,((\bar{Y}_{i\cdot}-\mu_i)-(\bar{Y}_{j\cdot}-\mu_j))/\sigma}{\sqrt{\frac{(n-k)\hat{\sigma}^2}{(n-k)\sigma^2}}}$$

$$\sim q_{k,k(n-1)} \quad \text{(Tukey's } q\text{)}$$

Since by Cochran, numerator & denominator are independent.

(b): we know that $\hat{\sigma}^2 \xrightarrow{a.s.} \sigma^2$

So by continuous mapping: $\hat{\sigma} \xrightarrow{a.s.} \sigma$.

Also $\sqrt{n}\,(\bar{Y}_i - \mu_i)/\sigma \xrightarrow{\mathcal{L}} N(0,1)$.

So by Slutsky,

$$\sqrt{n}(\bar{Y}_i-\mu_i)/\hat{\sigma} \xrightarrow{\mathcal{L}} N(0,1)$$

Next, since $\max(\cdot)$ & $f(x_1,x_2)=x_1-x_2$ are both continuous mappings, we have

$$\max_{1\le i<j\le k} \frac{\sqrt{n}\,|(\bar{Y}_i-\mu_i)-(\bar{Y}_j-\mu_j)|}{\hat{\sigma}} \xrightarrow{\mathcal{L}}$$

$$\max_{1\le i<j\le k} |Z_i - Z_j|, \quad Z_i \overset{iid}{\sim} N(0,1).$$

Q2:

2 The data you received contains results from a 1-way ANOVA setup with 5 patients' responses in each of the 5 treatment groups.

a Perform Hartley's test, Levene's test, Brown-Forsythe's test and Bartlett's test to ascertain whether variances from the 5 groups are equal at the 0.05 significance level. Report the 4 p-values for the 4 tests.

b Someone told you that there is also a O'Brien's test for testing homogeneity of variances in a 1-way ANOVA setup. Do some research on this test and write a description of O'Brien's test, up to ¾ page in length.

Sol. (a)

Hartley's test: $\dfrac{\max_{1\le i\le k} S_i^2}{\min_{1\le i\le k} S_i^2}$

Levene's test: $\dfrac{\sum_i n_i (Z_{i\cdot}-Z_{\cdot\cdot})^2/(k-1)}{\sum_i\sum_j (Z_{ij}-Z_{i\cdot})^2/(n-k)}$, $Z_{ij}=|Y_{ij}-\bar{Y}_{i\cdot}|$

Brown-Forsythe's test: replace $\bar{Y}_{i\cdot}$ with median.

Bartlett's test: $\dfrac{(n-k)\ln S^2 - \sum_{i=1}^k (n_i-1)\ln S_i^2}{1+\frac{1}{3(k-1)}\left(\sum_{i=1}^k \left(\frac{1}{n_i-1}\right)-\frac{1}{n-k}\right)}$

where $S^2 = \frac{1}{n-k}\sum_i (n_i-1)S_i^2$.

Test	Statistic	P-value
Hartley	$F(4,5)=3.57$	0.7576
Levene	$F=0.674$	0.6179
Brown-Forsythe	$F=0.567$	0.6897
Bartlett	$\chi^2_4=2.569$	0.6323

At level $\alpha=0.05$, all tests suggest we do not reject the null: $\sigma_1^2=\cdots=\sigma_5^2$.

(b): See next page.

Q2 (b): O'Brien's Test

WLOG, assume k groups & n_i obs, $i=1,\cdots,k$, $n \equiv \sum_i n_i$

STEP 1: O'Brien's transform

$\forall w \in [0,1]$, Let

$$r_{ij} = \left[(w+n_j-2)n_j(x_{ij}-\bar{x}_{j})^2 - w s_j^2(n_j-1)\right] \Big/ \left[(n_j-1)(n_j-2)\right]$$

where n_j = size of group j

S_j^2 = Variance of the j^{th} group

x_{ij} = resp. of the i^{th} item in the j^{th} group

$\bar{x}_{j.}$ = mean of the j^{th} group

Transformed resp.	group
r_{11}	1
$\vdots$	2
$\vdots$	3
$\vdots$	4
$r_{n_5 5}$	5

STEP 2: perform a regular 1-way ANOVA

Q3:

3 Refer to the Cook and Weisberg's paper on diagnosing heteroscedasticity.

a Describe, in not more than ¾ of a page length, the statistical setup and the basics of conducting a score test to draw statistical inference on model parameters.

b Derive, on your own, the test statistic for testing heteroscedasticity using the score test for the situation described in the above paper and see whether your results agree with equations (8-10).

Sol. (a): Step 1: Calculate $\frac{\partial l(\beta;Y,X)}{\partial\beta} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial l(\beta;Y_i,X_i)}{\partial\beta}$

Step 2: Calculate $I(\beta)$: Fisher info. for β.

Step 3: Use the test statistic

$$\frac{1}{n}\left(\sum_{i=1}^{n}\frac{\partial l(\beta;Y_i,X_i)}{\partial\beta^T}\right) I(\beta)^{-1}\left(\sum_{i=1}^{n}\frac{\partial l(\beta;Y_i,X_i)}{\partial\beta}\right)$$

$$\xrightarrow{d} \chi^2_p,$$ p is the dim of β.

to test the hypothesis.

(b): $Y=\beta_0 1 + X\beta + \varepsilon$, $X_0=[1,X]$,

$\varepsilon \sim N(0,\sigma^2 W)$ where

$W = \mathrm{Diag}(w(\lambda, z_i))_{i=1}^{n} \Rightarrow$

$$l(\lambda;Y_i,X_i) = \log\frac{1}{\sqrt{2\pi\sigma_i^2 w_i}}\, e^{-(Y_i - X_i^T\beta)^2/\sigma_i^2 w_i}$$

$$= -\frac{1}{2}\left[\log 2\pi + 2\log\sigma_i + \log w_i\right] - \frac{(Y_i - X_i^T\beta)^2}{\sigma_i^2 w_i}$$

$$\frac{\partial l_i}{\partial\lambda} = -\frac{1}{2}\frac{w_i'}{w_i} - \frac{(Y_i - X_i^T\beta)^2}{\sigma_i^2 w_i^2}\, w_i' \quad (1)$$

substitute β with $\hat\beta$, σ_i^2 with $\hat\sigma^2$

(1) $\Rightarrow -\frac{1}{2}\frac{w_i'}{w_i} - \frac{e_i^2}{\hat\sigma^2 w_i^2} w_i'$, when $W=I$, we have:

$$\frac{\partial l}{\partial\lambda} = -\frac{1}{2}D^T 1 + \frac{1}{2}D^T u \quad \text{where}$$

$$D = \begin{bmatrix} (w_1')^T \\ \vdots \\ (w_n')^T \end{bmatrix}, \quad u = \begin{bmatrix} e_1^2 \\ \vdots \\ e_n^2 \end{bmatrix} / \hat\sigma^2$$

To compute Fisher's info.:

$$\hat{I}(\beta,\sigma^2,\lambda) = \begin{bmatrix} A & B^T \\ B & C \end{bmatrix} \quad \&$$

$$A = \begin{bmatrix} X_0^T X_0/\sigma^2 & 0 \\ 0^T & \frac{1}{2}n/\sigma^4 \end{bmatrix} \quad \text{and}$$

$$B = (0, \tfrac{1}{2}D^T 1/\sigma^2) \quad \text{and}$$

$C = \frac{1}{2}D^T D$ where C is due to the fact that

$$\mathbb{E}(u-1)(u-1)^T = 2I$$

Thus, the inverse of Fisher's info:

$$\hat{I}^{-1} = \begin{bmatrix} \tilde{A} & \tilde{B}^T \\ \tilde{B} & \tilde{C} \end{bmatrix} \quad \&$$

$\tilde{C} = 2(\bar{D}^T\bar{D})^{-1}$ where

$\bar{D} = (I - \frac{11^T}{n})D$. This follows from partitioned matrix formula.

Thus, score statistic for λ is:

$$\frac{1}{2}(u-1)^T D(\bar{D}^T\bar{D})^{-1}D^T(u-1)$$

$$= \frac{1}{2}u^T\bar{D}(\bar{D}^T\bar{D})^{-1}\bar{D}^T u \equiv S \quad \text{eq8}$$

The equality is due to:

$$1^T u = n.$$

When $g_\lambda = 1$,

$D = \begin{bmatrix} w_1' \\ \vdots \\ w_n' \end{bmatrix}$ is a $n\times 1$ vector,

Thus, $\bar{D} = (I - \frac{11^T}{n})D = \begin{bmatrix} \vdots \\ w_i' - \bar{w}' \\ \vdots \end{bmatrix}$ where

$\bar{w}' = \frac{1}{n}\sum_{i=1}^{n} w_i' = \frac{1}{n}\sum_{i=1}^{n} w'(z_i, \lambda)$. $\Rightarrow$

$$S = \frac{\left(\sum_i (w_i' - \bar{w}')(e_i^2/\hat{\sigma}^2 - 1)\right)^2}{2\sum_i (w_i' - \bar{w}')^2}$$ eq 9

Further, let $A = \mathrm{Diag}(w_i')_{i=1}^{n}$ and we know

$$e = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = (I - V)Y \text{ where}$$

$V = X_0(X_0^T X_0)^{-1} X_0^T$.

$\Rightarrow$ The numerator of S is:

$Y^T(I-V)A(I-V)Y$

$= \varepsilon^T(I-V)A(I-V)Y$ since $X \perp (I-V)$.

Thus, we can use

$$S' = \varepsilon^T(I-V)A(I-V)Y / \varepsilon^T(I-V)\varepsilon$$

$$= \sum_{i=1}^{n-p-1} y_i \chi_i^2 / \sum_{i=1}^{n-p-1} \chi_i^2$$ eq. 10. 11.

where χ_i^2 are $\chi^2(1)$ RVs & y_i are eig-vals of $(I-V)A(I-V)$.

Q4:

4 Let $U = (U_1, U_2, .. U_p)'$ be a vector of principal components of X. Then $U_i = a_i' X$ for some vector a_i of length 1, i=1,2,.., p. Show that

a var $a'X \le$ var U_1

b if $a'X$ is uncorrelated with $U_1, U_2, \ldots, U_{i-1}$, then var $a'X \le$ var U_i.

[In the class, we discussed principal components in the context of a linear regression model $Ey=X\beta$. Here we have a very long vector X of random variables, and we want to work with a shorter vector formed by extracting a small number of linear combinations of the vector X (i.e. first few principal components of the vector X.]

Sol.(a): Denote $\Sigma = \mathrm{Var}X$, then

$\Sigma a_1 = \lambda_1 a_1$. WLOG, $\|a\| = \|a_1\| = 1$

By SDT, $\Sigma = UDU^T$, $D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{bmatrix}$

$$
\begin{aligned}
\mathrm{Var}(a^T X) &= a^T \Sigma a \\
&= a^T U D U^T a \\
&= \sum_{i=1}^{p} \lambda_i (U^T a)_i^2, \quad \|U^T a\| = 1. \\
&\le \lambda_1 \quad \text{Since } (U^T a)_i^2 \le 1 \ \& \ \textstyle\sum_i (U^T a)_i^2 = 1. \\
&= a_1^T \Sigma a_1 \quad \text{Since } \|a_1\| = 1. \\
&= \mathrm{Var}\, U_1
\end{aligned}
$$

$\Rightarrow \mathrm{Var}(a^T X) \le \mathrm{Var}\, U_1$.

(b): Let $U = [a_1, \ldots, a_p]$ then $U_j = a_j^T X$.

Then
$$
\begin{aligned}
&\mathrm{Cov}(a^T X, U_j) \\
&= \mathrm{Cov}(a^T X, a_j^T X) \\
&= a^T \Sigma a_j = 0 \quad \forall j = 1, \ldots, i-1.
\end{aligned}
$$

Since $\Sigma = UDU^T$, we have

$$
\begin{aligned}
a^T \Sigma a_j &= a^T U D U^T a_j \\
&= (a^T a_j) \lambda_j \\
&= 0
\end{aligned}
$$

$\Rightarrow a^T a_j = 0 \ \forall j = 1, \ldots, i-1$

$$
\begin{aligned}
\mathrm{Var}\, a^T X &= a^T \Sigma a \\
&= a^T U D U^T a \\
&= \sum_{j=i}^{p} \lambda_j (U^T a)_j^2 \\
&\le \lambda_i \quad \text{since } \sum_{j=i}^{p} (U^T a)_j^2 = 1 \\
&= \mathrm{Var}\, a_i^T X
\end{aligned}
$$

$\Rightarrow \mathrm{Var}\, a^T X = \mathrm{Var}\, a_i^T X$.

Q5:

5 Refer to the paper posted on the website where the authors used principal components to analyze a public health data. Read the paper to appreciate how principal components analysis is used in practice.

Done.

Q6:

6 Let b be the least squares estimator and let $b_k = (X'X+kI)^{-1}X'y$ be the ridge estimator for estimating β in the standard linear model EY=X β and k is a positive constant. Recall the methodology employed in class on how we expressed b_k in terms of b and then derived the bias and mean square error (MSE) of b_k. Show that the same methodology can be used to show that there are values of k for which the predictor of the mean response from the ridge regression perform better than that from the least squares regression, i.e. show that tr MSE of Xb_k is smaller than that of Xb for some values of k.

Sol. $MSE(\hat\beta) = E(\hat\beta-\beta)(\hat\beta-\beta)^T$

$= Var\hat\beta = \sigma^2 (X^TX)^{-1}$

$MSE(b_k) = G_k(\sigma^2 X^TX + k^2\beta\beta^T)G_k$ (shown in class)

where $G_k = (X^TX+kI)^{-1}$.

Let $g(k) = Tr(MSE(b_k))$, then

$g(0) = Tr(MSE(\hat\beta))$.

By SDT, $X^TX = TDT^T$ & $TT^T = T^TT = I$.

then

$$g(k) = Tr\left[(TDT^T+kTT^T)^{-1}(\sigma^2X^TX+k^2\beta\beta^T)(TDT^T+kTT^T)^{-1}\right]$$

$$= Tr\left[T(D+kI)^{-1}T^T(\sigma^2X^TX+k^2\beta\beta^T)T(D+kI)^{-1}T^T\right]$$

$$= Tr\left[(D+kI)^{-2}(\sigma^2D+k^2T^T\beta\beta^TT)\right]$$

$$= \sum_{i=1}^{p}(\sigma^2\lambda_i+k^2\beta_i^2)/(\lambda_i+k)^2 \text{ where}$$

we assume $D = \begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_p\end{bmatrix}$.

$$g(0) = \sum_{i=1}^{p}\sigma^2\lambda_i \Rightarrow$$

$$g(k)-g(0) = \sum_{i=1}^{p}\left[\frac{\sigma^2\lambda_i+k^2\beta_i^2}{(\lambda_i+k)^2} - \sigma^2\lambda_i\right] \quad (\Delta)$$

Then

$$\frac{\partial(\Delta)}{\partial k} = \sum_i \frac{2k\beta_i^2(\lambda_i+k)-2(\sigma^2\lambda_i+k^2\beta_i^2)}{(\lambda_i+k)^3}$$

$$= \sum_{i=1}^{p}\frac{2\lambda_i[k\beta_i^2-\sigma^2]}{(\lambda_i+k)^3} \Rightarrow$$

Next Page

Clearly,

$\frac{\partial(*)}{\partial k}$ is a continuous fct of k.

When $k=0$:

$$\left.\frac{\partial(*)}{\partial k}\right|_{k=0} = -2\sum_{i=1}^{p}\frac{\sigma^2}{\lambda_i^2} < 0$$

when k large:

$$k > \max\left(\frac{\sigma^2}{\beta_i^2}\right)$$

$$\frac{\partial(*)}{\partial k} > 0.$$

So $\exists$ a solution to

$\frac{\partial(*)}{\partial k} = 0$. & this is a

minima to $g(k)-g(0)$.

Since $g(0)-g(0)=0$.

There $\exists$ a k s.t.

$$g(k)-g(0)<0.$$

Since $g(k)-g(0)$ is a continuous

mapping & decreases when k small.

Thus, $\exists$ k s.t.

$$Tr(MSE(\hat{\beta})) > Tr(MSE(b_k)) \square.$$