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.

a. in one-way ANOVA, the n observations in the ith group follows: $Y_{i1} \dots Y_{in} \overset{iid}{\sim} N(\mu_i, \sigma^2)$

then: $\bar{Y}_{i\cdot} \sim N(\mu_i, \frac{\sigma^2}{n})$. $i = 1, 2, \dots k$

$\Rightarrow \bar{Y}_{i\cdot} - \mu_i \overset{iid}{\sim} N(0, \frac{\sigma^2}{n})$, $i = 1, 2 \dots k$

$$\Rightarrow q = \max_{1\le i<j\le k} \frac{|(\bar{Y}_{i\cdot} - \mu_i) - (\bar{Y}_{j\cdot} - \mu_j)|}{\sqrt{s^2/n}}$$

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

$\sim$ studentized range distribution with $df = (a, b)$

where $b = (nk-k)$. as $s^2/\sigma^2 \sim \chi^2(nk-k)$; $a = k$. as $(\bar{Y}_{i\cdot} - \mu_i) \overset{iid}{\sim} N(0, \frac{\sigma^2}{n})$, $i = 1, \dots k$

$\Rightarrow T_{k,n}$ follows studentized range distribution with $df = (k, nk-k)$ (notated in the setting)

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

b. under setting of standard 1-way ANOVA:

denote $e_i = \bar{Y}_{i\cdot} - \mu_i$. then $e_i \overset{iid}{\sim} N(0, \frac{\sigma^2}{n})$, $i = 1, 2, \dots k$

$\frac{e_i}{\sigma/\sqrt{n}} \overset{iid}{\sim} N(0, 1)$, $i = 1, 2, \dots k$

$$T_{k,n} = \max_{1\le i<j\le n} \frac{|e_i - e_j| / (\sigma/\sqrt{n})}{\hat{\sigma}/\sigma}$$

$$= \left(\frac{\hat{\sigma}}{\sigma}\right) \cdot \max_{1\le i<j\le n} |e_i - e_j| / \sigma/\sqrt{n}$$

$$\triangleq \left(\frac{\hat{\sigma}}{\sigma}\right) \cdot Q_{k,n}$$

as $\lim_{n\to\infty} \left(\frac{\hat{\sigma}}{\sigma}\right) = 1$, and by definition, $Q_{k,n} \sim$ range distribution $q_{k,n}$

by slutsky theorem: $T_{k,n} \xrightarrow{d} q_{k,n}$

Also, the result remains valid when normality assumption is violated as $e_i / \sigma/\sqrt{n} \xrightarrow{d} N(0,1)$ by CLT.

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

a. $U = \begin{pmatrix} U_1 \\ \vdots \\ U_p \end{pmatrix} = A'X = \begin{pmatrix} a_1' \\ \vdots \\ a_p' \end{pmatrix}$, $cov(X) = \Sigma > 0$

be definition of PCA, the A is an orthogonal matrix s.t. $A \Sigma A' = \Lambda = diag(\lambda_1 \dots \lambda_n)$

$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$

therefore, $cov(A'X) = cov(U) = A' cov(X) A$

$= A' \Sigma A$

$= \Lambda = diag(\lambda_1 \dots \lambda_n)$

then: $var(U_1) = \lambda_1$

let a be any $n \times 1$ vector s.t. $a'a = 1$

$$\begin{aligned} var(a'X) &= a' \Sigma a \\ &= a'(A \Lambda A') a \\ &= (A'a)' \Lambda (A'a) \\ &= \begin{pmatrix} a_1' a \\ \vdots \\ a_n' a \end{pmatrix}^T diag(\lambda_1 \dots \lambda_n) \begin{pmatrix} a_1' a \\ \vdots \\ a_n' a \end{pmatrix} \\ &= \sum_{i=1}^n a' a_i \lambda_i a_i' a \\ &= a' \left(\sum_{i=1}^n a_i \lambda_i a_i' \right) a \\ &\le a' \sum_{i=1}^n \lambda_1 \cdot a_i a_i' a \end{aligned}$$

since $AA' = I$, $\sum_{i=1}^n a_i a_i' = AA' = I$

$var(a'X) \le \lambda_1 \cdot a'a = \lambda_1 = var(U_1)$, for any a s.t $||a|| = 1$ $\square$

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

b. $a'X \perp\!\!\!\perp U_1 \dots U_{i-1} \iff cov(a'X, a_k' X) = 0$, $k = 1, 2, \dots i-1$

$cov(a'X, a_k' X) = a' \Sigma a_k = a' A \Lambda A' a_k = a' \left(\sum_{i=1}^n a_i \lambda_i a_i' \right) a_k$

for $A'A = I$: $= a' a_k \lambda_k a_k' a_k = \lambda_k \cdot a' a_k = 0$ $\Rightarrow a' a_k = 0$, $k = 1, \dots i-1$

then $var(a'X) = a' \Sigma a = a' A \Lambda A' a$

$$= \sum_{i=1}^{n} a' a_i \lambda_i a_i' a$$

for $a_k' a = 0,\ k < i$: $$= \sum_{k=i}^{n} a' a_k \lambda_k a_k' a$$

$$\le \lambda_i \cdot a' \sum_{k=i}^{n} a_k a_k' \cdot a$$

$$= \lambda_i \cdot a' \cdot \begin{bmatrix} 0 & \\ & I_{n-i+1} \end{bmatrix} a$$

for $a' I a = 1$: $$\le \lambda_i \cdot 1 = \lambda_i$$

therefore, for any $a'X$ s.t. $a'X \perp\!\!\!\perp u_1 \ldots u_{i-1}$, $var(a'X) \le \lambda_i$. $\square$