HW I

1. $H_0: \mu_1 - \mu_2 = \Delta_0$ vs $H_a: \mu_1 - \mu_2 \neq \Delta_0$

β = Type II error = $P(\text{not reject } H_0 \mid H_1)$

$$= P\left(\left| \frac{\bar{y}_1 - \bar{y}_2 - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n}}} \right| < z_{\frac{\alpha}{2}} \,\Big|\, H_1 \right)$$

$$= P\left(-z_{\frac{\alpha}{2}} - \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} < \frac{\bar{y}_1 - \bar{y}_2 - \Delta}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} < z_{\frac{\alpha}{2}} - \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \,\Big|\, H_1 \right)$$

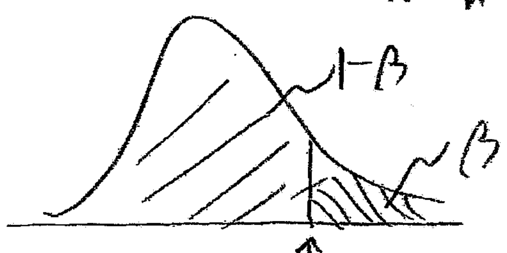
$\Delta = \mu_1 - \mu_2$

$$= \Phi\left(z_\alpha - \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \right) - \underbrace{\Phi\left(-z_{\frac{\alpha}{2}} - \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \right)}_{\text{if } \Delta - \Delta_0 > 0,}$$

$$\approx \Phi\left(z_\alpha + \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \right) \qquad \approx 0 .$$

$$= 1 - \Phi\left(-z_\alpha + \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \right)$$

$$\Rightarrow 1 - \beta \approx \Phi\left(-z_{\frac{\alpha}{2}} + \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \right)$$



$$-z_{\frac{\alpha}{2}} + \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}}$$

$$\parallel$$

$$z_\beta$$

$$\Rightarrow \sqrt{n} \approx \frac{\left(z_{\frac{\alpha}{2}} + z_\beta\right)\left(\sqrt{\sigma_1^2 + \sigma_2^2}\right)}{\Delta - \Delta_0}$$

$$\Rightarrow n \approx \frac{\left(z_{\alpha/2} + z_\beta\right)\left(\sigma_1^2 + \sigma_2^2\right)}{(\Delta - \Delta_0)^2}$$

#2 Compairing fasting serum cholesterol levels between 2 ethnic groups.

$n_1 = 20$, $\bar{Y}_1 = 180$ mg/dL, $S_1 = 41$ mg/dL

$n_2 = 30$, $\bar{Y}_2 = 165$ mg/dL, $S_2 = 36$ mg/dL

(a) Assumption for using indep. sample t-test for comparing mean in 2 groups

; population fasting serum cholesterol levels in subjects over 21 years of age belonging to two ethnic groups have normal distribution with mean μ_1, μ_2 and s.d. $\underbrace{\sigma_1 = \sigma_2 = \sigma}_{\text{(common var.)}}$ but unknown.

; Two samples are independent

(Samples are randomly and independently selected from the two popns.)

(b) Hypothesis test to see if μ_1 is at least 5 mg/dL higher than μ_2.

; $H_0 : (\mu_1 - \mu_2) < 5$

$H_1 : (\mu_1 - \mu_2) \geq 5$

$$\hat{\theta} = \bar{Y}_1 - \bar{Y}_2 = 180 - 165 = 15$$

$$\theta_0 = 5$$

$$\hat{\sigma}_{\hat{\theta}} = S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1)+(n_2-1)}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$= \sqrt{\frac{(20-1)(41)^2 + (30-1)(36)^2}{(20-1)+(30-1)}}\sqrt{\frac{1}{20}+\frac{1}{30}}$$

$$\approx 10.9863$$

T.S. $t = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}_{\hat{\theta}}} = \frac{15-5}{10.9863} \approx 0.91$

d.f $= (n_1 + n_2 - 2) = 48$

so p-value ; $0.15 < p < 0.2$ (See P3 of output for exact p-value = 0.18368)

If p-value < α, then reject H_0.

There is no sufficient evidence that the mean cholesterol level in group 1 is at least 5 mg/dL highter than that in group 2. (p-value = 0.18368)

(c) A subject presents 210 mg/dL

Reasonable to assume the subject does not belong to group 1?

By using Normal limits,

95% prediction interval ;

$$\bar{Y}_1 \pm t_{(n_1-1,\alpha/2)} \sqrt{(1+\frac{1}{n_1})S_1^2}$$

$$= 180 \pm t_{(19,0.025)} \sqrt{(1+\frac{1}{20})41^2}$$

$$= 180 \pm (2.093)(42.0125)$$

$$\Rightarrow (92.0678, 267.9322)$$

; 210 lies within the prediction interval

so we would say that the subject does belong to group 1.

(d) 95% C.I. for the diff. assume (i) with equal var. (ii) without equal var.

(i) with equal var ;

95% C.I. ; $(\bar{Y}_1-\bar{Y}_2) \pm t_{(n_1+n_2-2,\alpha/2)} S_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

$$= (180-165) \pm t_{(48,0.025)} (10.9863)$$

$$\approx 15 \pm (2.01063)(10.9863)$$, see P7 of output for $t_{(48,0.025)}$

$$\Rightarrow (-7.09, 37.09)$$

(ii) w/o equal var ;

95% C.I. ; $(\bar{Y}_1-\bar{Y}_2) \pm t_{(k,\alpha/2)} \sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$

, where $\frac{1}{k} = \frac{1}{n_1-1}\left(\frac{\frac{S_1^2}{n_1}}{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right)^2 + \frac{1}{n_2-1}\left(\frac{\frac{S_2^2}{n_2}}{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right)^2$

$$= \frac{1}{19}\underbrace{\left(\frac{41^2/20}{(41^2/20)+(36^2/30)}\right)^2}_{0.4363} + \frac{1}{29}\underbrace{\left(\frac{36^2/30}{(41^2/20)+(36^2/30)}\right)^2}_{0.1153}$$

$$\approx 0.0269 \iff k \approx 37.175$$

use $k = 37$

$$= (180-165) \pm t_{(37,0.025)} \sqrt{\frac{41^2}{20}+\frac{36^2}{30}}$$

$$\approx 15 \pm (2.02619)(11.281)$$, see P8 of output for $t_{(37,0.025)}$

$$\Rightarrow (-7.857, 37.857)$$

(e) Perform a test if the vars of two popns. are unequal.

$H_0: \sigma_1^2 = \sigma_2^2$

$H_1: \sigma_1^2 \neq \sigma_2^2$

T.S. $F_c = \frac{S_{max}^2}{S_{min}^2} = \frac{41^2}{36^2} \approx 1.30$

$F_{\alpha/2}[(n_{max}-1),(n_{min}-1)] = F_{\alpha/2}(19, 29)$

If we use $\alpha = 0.05$, $F_{0.025}(19,29) \approx 2.251$

With 95 % confidence, we can not reject H_0 since $F_\alpha < F$.

Hence with 95 % confidence, there is no sufficient reason to conclude that the variances of the two popns. are equal.

(f) d.f. = (19, 29) and $F_c = 1.3$

p-value = 0.25594 ; see p4 of output

⇒ I would say that there is no sufficient reason (p-value = 0.256) to conclude that the variances of the two popns. are equal.

$p = 2(.2563)$.

3 interested in comparing 4 employee training programs:

computer assisted training on site (CON),
computer assisted training off site (COF),
video tape training on site (VON),
video tape training off site (VOF)

Data are scores (weekly average increase in productivity) for 6 employees trained under each of the 4 training programs.

	program	scores						$t=4,\ N=24$
(Y_1)	CON	16	18	19	21	24	20	$\bar{Y}_1 = 19.67$
(Y_2)	COF	10	13	10	8	12	13	$\bar{Y}_2 = 11$
(Y_3)	VON	21	23	19	26	22	24	$\bar{Y}_3 = 22.5$
(Y_4)	VOF	12	16	13	14	16	13	$\bar{Y}_4 = 14$

(a) ANOVA Model ; $\mu + \tau_i$ where $\Sigma \tau_i = 0$

$Y_{ij} = \mu_i + E_{ij}$, $i = 1, 2, 3, 4$, $j = 1, 2, 3, 4, 5, 6$

, Y_{ij} = true jth response (score) in ith treatment (method)

μ overall mean score

μ_i = expected response (score) for ith treatment (method)

E_{ij} = random error in Y_{ij}

$\overset{iid}{\sim} N(0, \sigma^2)$

(b) ANOVA Table ; Also see the output on p. 2.

source	d.f	SS	MS	F_c	P
Between Groups (Between programs)	3	493.125	164.375	32.60	<0.0001
Error (Within Groups)	20	100.835	5.04175		
Total	23	593.96			

$$SS[Tot] = \sum_{i=1}^{4}\sum_{j=1}^{6} Y_{ij}^2 - \frac{(\sum_i\sum_j Y_{ij})^2}{N} = 7361 - \frac{(403)^2}{24} \approx 593.96$$

$$SS[T] = \sum_{i=1}^{4} \frac{Y_{i+}^2}{n_i} - \frac{Y_{++}^2}{N} = \left(\frac{(118)^2}{6} + \frac{(66)^2}{6} + \frac{(135)^2}{6} + \frac{(84)^2}{6}\right) - \frac{(403)^2}{24} = 493.125$$

$$SS[E] = SS[Tot] - SS[Tr] \approx 593.96 - 493.125 = 100.835$$

Hypothesis test to see if the mean productivity score differ for 4 methods

$$\begin{cases} H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 \\ H_1: \mu_i \neq \mu_j \text{ for at least one pair }, i \neq j \end{cases}$$

Reject H_0 (If p-value $< \alpha$, then we reject H_0)

There is sufficient evidence to conclude that

there is significant (p-value < 0.0001) difference in the mean productivity scores ~~Among 4 methods~~ of at least two methods.

At $\alpha = 0.01$, $F_{(3, 20, 0.01)} = 4.938 < F_c = 32.6$,

we reject H_0 and have the same conclusion

(C) Construct 95% C.I. for the diff. mean productivity scores for CON and COF.
On the basis of the interval, which method would you recommend? Why?

$\theta = \mu_1 - \mu_2$

$\hat{\theta} = \bar{Y}_1 - \bar{Y}_2 \approx 8.6667$, $C_1 = 1$, $C_2 = -1$

$\hat{\sigma}_{\hat{\theta}} = \sqrt{\left(\frac{C_1^2}{n_1} + \frac{C_2^2}{n_2}\right) MSE} = \sqrt{\left(\frac{1}{6} + \frac{1}{6}\right)(5.04125)} \approx 1.2964$

$100(1-\alpha)\%$ C.I for θ ; $\hat{\theta} \pm t_{(N-t, \alpha/2)} \hat{\sigma}_{\hat{\theta}}$

so 95% C.I. for θ

$= 8.6667 \pm \overset{2.086}{t_{(24-4, 0.025)}} (1.2964)$

$= 8.6667 \pm 2.7042$

$\Rightarrow (5.96, 11.37)$ ✓

; with 95% confidence, the diff. mean productivity scores for CON and COF is between 5.96 and 11.37

In other words, the mean productivity score for CON is at least 5.96 greater than that for COF and at most 11.37 greater than that for COF with α ~~0.025~~ 0.05 level.

Hence, ~~I~~ would recommend CON method.

(d) Construct 3 mutually orthogonal contrasts $\theta_1, \theta_2, \theta_3$ such that
θ_1 can be used to compare the computer and video methods
θ_2 can be used to compare the on site and off site methods
θ_3 can be used to compare the difference between computer and video methods when they are used on site with the corresponding difference when they are used off site.

$\hat{\theta}_1 = \frac{1}{2}(\bar{Y}_1 + \bar{Y}_2) - \frac{1}{2}(\bar{Y}_3 + \bar{Y}_4)$ ✓ ; $\hat{\theta}_1$ is contrast since $(\frac{1}{2})+(\frac{1}{2})+(-\frac{1}{2})+(-\frac{1}{2}) = 0$

$\hat{\theta}_2 = \frac{1}{2}(\bar{Y}_1 + \bar{Y}_3) - \frac{1}{2}(\bar{Y}_2 + \bar{Y}_4)$ ✓ ; $\hat{\theta}_2$ 〃 〃

$\hat{\theta}_3 = \frac{1}{2}(\bar{Y}_1 - \bar{Y}_3) - \frac{1}{2}(\bar{Y}_2 - \bar{Y}_4)$ ✓ ; $\hat{\theta}_3$ 〃 $(\frac{1}{2})+(-\frac{1}{2})+(-\frac{1}{2})+(\frac{1}{2}) = 0$

Checking for orthogonal ; (since $n_1 = \cdots = n_4$, check if $\sum c_i d_i = 0$)

$\langle \hat{\theta}_1$ and $\hat{\theta}_2 \rangle$

$(\frac{1}{2})(\frac{1}{2}) + (\frac{1}{2})(-\frac{1}{2}) + (-\frac{1}{2})(\frac{1}{2}) + (-\frac{1}{2})(-\frac{1}{2}) = 0$ so $\hat{\theta}_1, \hat{\theta}_2$ orthogonal

$\langle \hat{\theta}_1$ and $\hat{\theta}_3 \rangle$

$(\frac{1}{2})(\frac{1}{2}) + (\frac{1}{2})(-\frac{1}{2}) + (-\frac{1}{2})(-\frac{1}{2}) + (-\frac{1}{2})(\frac{1}{2}) = 0$ so $\hat{\theta}_1, \hat{\theta}_3$ orthogonal

$\langle \hat{\theta}_2$ and $\hat{\theta}_3 \rangle$

$(\frac{1}{2})(\frac{1}{2}) + (-\frac{1}{2})(-\frac{1}{2}) + (\frac{1}{2})(-\frac{1}{2}) + (-\frac{1}{2})(\frac{1}{2}) = 0$ so $\hat{\theta}_2, \hat{\theta}_3$ orthogonal

(e) Test significance of 3 contrasts in (d) using experimentwise error of 5%.
Interpret the result of the tests

$\cancel{\hat{\sigma}_{\hat{\theta}} = \sqrt{\left(\frac{c_1^2}{n_1} + \cdots + \frac{c_t^2}{n_t}\right) MSE}}$

$\hat{\theta}_1 = \frac{1}{2}(19.67 + 11) - \frac{1}{2}(22.5 + 14) = -2.915$

$\hat{\theta}_2 = \frac{1}{2}(19.67 + 22.5) - \frac{1}{2}(11 + 14) = 8.585$

$\hat{\theta}_3 = \frac{1}{2}(19.67 - 22.5) - \frac{1}{2}(11 - 14) = 0.085$

$\hat{\sigma}_{\hat{\theta}_1} = \sqrt{\frac{1}{6}\left((\frac{1}{2})^2 + (\frac{1}{2})^2 + (-\frac{1}{2})^2 + (-\frac{1}{2})^2\right)(5.04175)} \approx 0.9167$

$\cancel{\hat{\sigma}_{\hat{\theta}_2} = \sqrt{\frac{1}{6}\left((\frac{1}{2})^2 + (-\frac{1}{2})^2 + (\frac{1}{2})^2 + (-\frac{1}{2})^2\right)(5.04175)} \approx 0.9167}$

$\hat{\sigma}_{\hat{\theta}_3} = \sqrt{\frac{1}{6}\left((\frac{1}{2})^2 + (-\frac{1}{2})^2 + (-\frac{1}{2})^2 + (\frac{1}{2})^2\right)(5.04175)} \approx 0.9167$

Scheffe CCV at an experimentwise error rate $\alpha = 0.05$ is

$$\sqrt{(t-1)\,F(t-1, \nu, 0.05)}\ \hat{\sigma}_{\hat{\theta}_i} = \sqrt{3\,F(3, 20, 0.05)}\ \hat{\sigma}_{\hat{\theta}_i} \approx 3.0486\ \hat{\sigma}_{\hat{\theta}_i}$$

Bonferroni CCV at an experimentwise error rate $\alpha = 0.05$ is

$$t(\nu, 0.05, k)\ \hat{\sigma}_{\hat{\theta}_i} = t(20, 0.05, 3)\ \hat{\sigma}_{\hat{\theta}_i} = 2.613\ \hat{\sigma}_{\hat{\theta}_i}$$

$\hat{\theta}_i$	$\lvert\hat{\theta}_i\rvert$	$\hat{\sigma}_{\hat{\theta}_i}$	Scheffe CCV	Bonferroni CCV
$\hat{\theta}_1$	2.915	0.9167	2.79	2.40
$\hat{\theta}_2$	8.585	0.9167	2.79	2.40
$\hat{\theta}_3$	0.085	0.9167	2.79	2.40

test stat. 에서는 $\alpha/3$ 사용

If $\lvert\hat{\theta}_i\rvert >$ Sheffe CCV, then $\hat{\theta}_i$ is significant.

$\lvert\hat{\theta}_1\rvert = 2.915 > 2.79$ so $\hat{\theta}_1$ is signif.

$\lvert\hat{\theta}_2\rvert = 8.585 > 2.79$ so $\hat{\theta}_2$ is signif.

$\lvert\hat{\theta}_3\rvert = 0.085 < 2.79$ so $\hat{\theta}_3$ is not signif.

If $\lvert\hat{\theta}_i\rvert >$ Bonferroni CCV, then $\hat{\theta}_i$ is significant

$\lvert\hat{\theta}_1\rvert = 2.915 > 2.40$ so $\hat{\theta}_1$ is signif.

$\lvert\hat{\theta}_2\rvert = 8.585 > 2.40$ so $\hat{\theta}_2$ is signif.

$\lvert\hat{\theta}_3\rvert = 0.085 < 2.40$ so $\hat{\theta}_3$ is not signif.

Both tests lead the same conclusion
but Sheffee method is more conservative than Bonferroni method
($\because$ Scheffe CCV has larger cutoff value)

In this case, I prefer to use Bonferroni method (more liveral method).

$\hat{\theta}_1$ is significant

; With 95% confidence, there is significant difference in the mean productivity scores between computer method and video method.

$\hat{\theta}_2$ is significant

; With 95% confidence, there is significant difference in the mean productivity scores between on site method and off site method.

$\hat{\theta}_3$ is not significant

; With 95% confidence, there is no sufficient evidence to conclude that there is significant difference in the mean productivity scores between computer and video method when they are used on site with the corresponding diff. when they are used off site.

(f) Perform a pairwise multiple comparison of 4 treatment means by using Tukey method and Duncan method.

$\bar{Y}_1 - \bar{Y}_2 = 19.67 - 11 = 8.67$

$\bar{Y}_3 - \bar{Y}_1 = 22.5 - 19.67 = 2.83$

$\bar{Y}_1 - \bar{Y}_4 = 19.67 - 14 = 5.67$

$\bar{Y}_3 - \bar{Y}_2 = 22.5 - 11 = 11.5$

$\bar{Y}_4 - \bar{Y}_2 = 14 - 11 = 3$

$\bar{Y}_3 - \bar{Y}_4 = 22.5 - 14 = 8.5$

$$\begin{pmatrix} \bar{Y}_{(1)} = 11 \; ; \; \bar{Y}_2 \\ \bar{Y}_{(2)} = 14 \; ; \; \bar{Y}_4 \\ \bar{Y}_{(3)} = 19.67 \; ; \; \bar{Y}_1 \\ \bar{Y}_{(4)} = 22.5 \; ; \; \bar{Y}_3 \end{pmatrix}$$

$LSD_{ij}(T) = q(t, \nu, \alpha)\sqrt{\frac{MSE}{n}} = \overset{3.96}{q(4, 20, 0.05)}\sqrt{\frac{5.04175}{6}}$, at $\alpha = 0.05$

$= 3.63$

$q(t, \nu; \alpha)\sqrt{\frac{MSE}{n}}$

Y_1, Y_2 sig. diff.
Y_1, Y_4 sig. diff.
Y_2, Y_3 sig. diff.
Y_3, Y_4 sig. diff.

Y_2, Y_4 same group

Y_1, Y_3 same group

$n = \frac{t}{\frac{1}{n_1} + \cdots + \frac{1}{n_t}}$

$LSD_{ij}(D) = R(p, \nu, \alpha)\sqrt{\frac{MSE}{n}} = R(i-j+1, 20, 0.05)\sqrt{\frac{5.04175}{6}}$, at $\alpha = 0.05$

$R(q, \nu, \alpha)$

$p = i - j + 1$

LSD for $Y_1, Y_2 = R(3-1+1, \overset{3.1}{20}, 0.05)\,(0.9167) = 2.84$; sig. diff.

$Y_1, Y_3 = R(4-3+1, \overset{2.95}{20}, 0.05)\,(0.9167) = 2.70$; sig. diff.

$Y_1, Y_4 = R(3-2+1, \overset{2.95}{20}, 0.05)\,(0.9167) = 2.70$; sig. diff.

$Y_2, Y_3 = R(4-1+1, \overset{3.18}{20}, 0.05)\,(0.9167) = 2.92$; sig. diff.

$Y_2, Y_4 = R(2-1+1, \overset{2.95}{20}, 0.05)\,(0.9167) = 2.70$; sig. diff.

$Y_3, Y_4 = R(4-2+1, \overset{3.1}{20}, 0.05)\,(0.9167) = 2.84$; sig. diff.

; $Y_1, Y_2; Y_3, Y_4$ are all diff. group.

Also see the output on p. 3 (Duncan)

p. 4 (Tukey)

```
/* HW#4 : Bong-Rae Kim */

options ls=78 nodate;
data one;
input program score @@;
cards;
1 16 1 18 1 19 1 21 1 24 1 20
2 10 2 13 2 10 2  8 2 12 2 13
3 21 3 23 3 19 3 26 3 22 3 24
4 12 4 16 4 13 4 14 4 16 4 13
;
run;

proc ANOVA data=one;
  class program;
  model score=program;
  means program/bon scheffe tukey duncan;
run;
```

4.

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$$

$$\bar{y}_{i\cdot} = \mu + \tau_i + \bar{\varepsilon}_{i\cdot}, \quad \bar{\varepsilon}_{i\cdot} \overset{iid}{\sim} N(0, \tfrac{\sigma^2}{n})$$

$$\bar{y}_{\cdot\cdot} = \mu + \bar{\varepsilon}_{\cdot\cdot}, \quad \bar{\varepsilon}_{\cdot\cdot} \sim N(0, \tfrac{\sigma^2}{an})$$

(a)

$$SS_{Trt} = n\sum_{i=1}^{a} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 = n \sum_{i=1}^{a} (\tau_i + \bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})^2$$

$$= n\left\{ \sum_{i=1}^{a} \tau_i^2 + 2\sum_{i=1}^{a} \tau_i (\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot}) + \sum_{i=1}^{a} (\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})^2 \right\}$$

$$E\{SS_{Trt}\} = n\left\{ \sum_{i=1}^{a} \tau_i^2 + 2\sum_{i=1}^{a} \tau_i \underbrace{E(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})}_{=0} + E\left(\sum_{i=1}^{a} (\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})^2\right) \right\}$$

$$= n\sum_{i=1}^{a} \tau_i^2 + n\,\frac{\sigma^2}{n}\, \underbrace{E\left(\sum_{i=1}^{a} \frac{(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})^2}{\sigma^2/n} \right)}_{= a-1}$$

$$\frac{\sum_{i=1}^{a} (\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})^2}{\sigma^2/an} \sim \chi^2_{a-1}$$

$$= n\sum_{i=1}^{a} \tau_i^2 + (a-1)\sigma^2$$

Thus,

$$E(MS_{Trt}) = E\left(\frac{1}{a-1} SS_{Trt}\right) = \frac{1}{a-1}\left\{ n\sum_{i=1}^{a} \tau_i^2 + (a-1)\sigma^2 \right\}$$

$$= \sigma^2 + \frac{n}{a-1}\sum_{i=1}^{a} \tau_i^2$$

(b)

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{n} (y_{ij} - \bar{y}_{i\cdot})^2 = \sum_{i=1}^{a}\sum_{j=1}^{n} (\varepsilon_{ij} - \bar{\varepsilon}_{i\cdot})^2$$

Since $\sum_{j=1}^{n} (\varepsilon_{ij} - \bar{\varepsilon}_{i\cdot})^2/\sigma^2 \sim \chi^2_{n-1}$, $E\left(\sum_{j=1}^{n} (\varepsilon_{ij} - \bar{\varepsilon}_{i\cdot})^2/\sigma^2\right) = n-1$

$$E\{SS_E\} = \sum_{i=1}^{a} \sigma^2 \underbrace{E\left(\sum_{j=1}^{n} (\varepsilon_{ij} - \bar{\varepsilon}_{i\cdot})^2/\sigma^2\right)}_{n-1} = a(n-1)\sigma^2$$

$$E\{MS_E\} = E\left\{\frac{1}{a(n-1)} SS_E\right\} = \frac{1}{a(n-1)}\, a(n-1)\sigma^2 = \sigma^2$$

(c) Under $H_0: \tau_1 = \cdots = \tau_a = 0$,

$$\frac{SS_{Trt}}{\sigma^2} = \frac{n\sum_{i=1}^{a}(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot})^2}{\sigma^2} = \sum_{i=1}^{a}\frac{(\tau_i+\bar{\varepsilon}_{i\cdot}-\bar{\varepsilon}_{\cdot\cdot})^2}{\sigma^2/n} \underset{\text{under } H_0}{\uparrow =} \sum_{i=1}^{a}\frac{(\bar{\varepsilon}_{i\cdot}-\bar{\varepsilon}_{\cdot\cdot})^2}{\sigma^2/n} \sim \chi^2_{(a-1)} \quad (1)$$

$$\frac{SSE}{\sigma^2} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\cdot})^2}{\sigma^2} = \sum_{i=1}^{a}\underbrace{\sum_{j=1}^{n}(\varepsilon_{ij}-\bar{\varepsilon}_{i\cdot})^2/\sigma^2}_{\chi^2_{n-1}} \sim \chi^2_{a(n-1)} \quad (2)$$

(1) and (2) are independent.

$$F = \frac{\frac{SS_{Trt}}{\sigma^2}/(a-1)}{\frac{SSE}{\sigma^2}/a(n-1)} = \frac{MS_{Trt}}{MS_E} \sim F_{a-1,\,a(n-1)} \quad \text{under } H_0.$$