Vinay Maruri

STAT135 HW6

1] (1) expand LHS:

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{..})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[(Y_{ij}-\bar{Y}_{i.})+(\bar{Y}_{i.}-\bar{Y}_{..})\right]^2$$

notice no j terms in summand

$$= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{i.})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{Y}_{i.}-\bar{Y}_{..})^2 +$$

$$2\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{i.})(\bar{Y}_{i.}-\bar{Y}_{..})$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{i.})^2 + \sum_{i=1}^{k} n_i(\bar{Y}_{i.}-\bar{Y}_{..})^2 +$$

$$2\left\{\sum_{i=1}^{k}(\bar{Y}_{i.}-\bar{Y}_{..})\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{i.})\right\}$$

by assumption, sum of deviations from mean is zero.

$$= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{i.})^2 + \sum_{i=1}^{k} n_i(\bar{Y}_{i.}-\bar{Y}_{..})^2$$

Hence,

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{..})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{i.})^2 + \sum_{i=1}^{k} n_i(\bar{Y}_{i.}-\bar{Y}_{..})^2$$

where $\bar{Y}_{i.} = \frac{1}{n_i}\sum_j Y_{ij}$, $\bar{Y}_{..} = \frac{\sum_i n_i \bar{Y}_i}{\sum_i n_i}$

(2) expand LHS:

$$\sum_{i=1}^{k} n_i(\bar{Y}_{i.}-\bar{Y}_{..})^2 = \sum_{i=1}^{k} n_i\left[(\bar{Y}_{i.}-\mu)-(-\mu+\bar{Y}_{..})\right]^2$$

$$= \sum_{i=1}^{k} n_i(\bar{Y}_{i.}-\mu)^2 - \sum_{i=1}^{k} n_i(-\mu+\bar{Y}_{..})^2$$

← notice no i terms except n_i, $\sum_i n_i = n$

$$= \sum_{i=1}^{k} n_i(\bar{Y}_{i.}-\mu)^2 - n(\bar{Y}_{..}-\mu)^2$$

Hence, for any $\mu \in \mathbb{R}$,

$$\sum_{i=1}^{k} n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2 = \sum_{i=1}^{k} n_i(\bar{Y}_{i.} - \mu)^2 - n(\bar{Y}_{..} - \mu)^2$$

2] __WTS__ $SS_{TOT} = SS_A + SS_B + SS_{AB} + SS_E$

Assume n_{ij} = constant across all cells, balanced design

From Rice 12.3, $SS_A = JK \sum_{i=1}^{I} (\bar{Y}_{i..} - \bar{Y}_{...})^2$

$$SS_B = IK \sum_{j=1}^{J} (\bar{Y}_{.j.} - \bar{Y}_{...})^2$$

$$SS_{AB} = K \sum_{i=1}^{I} \sum_{j=1}^{J} (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$$

$$SS_E = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (Y_{ijk} - \bar{Y}_{ij.})^2$$

$$SS_{TOT} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (Y_{ijk} - \bar{Y}_{...})^2$$

we prove identity as follows:

$$(Y_{ijk} - \bar{Y}_{...}) = (Y_{ijk} - \bar{Y}_{ij.}) + (\bar{Y}_{i..} - \bar{Y}_{...}) + (\bar{Y}_{.j.} - \bar{Y}_{...}) + (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})$$

$$(Y_{ijk} - \bar{Y}_{...})^2 = \left[(Y_{ijk} - \bar{Y}_{ij.}) + (\bar{Y}_{i..} - \bar{Y}_{...}) + (\bar{Y}_{.j.} - \bar{Y}_{...}) + (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})\right]^2 =$$

$$(Y_{ijk} - \bar{Y}_{ij.})^2 + (\bar{Y}_{i..} - \bar{Y}_{...})^2 + (\bar{Y}_{.j.} - \bar{Y}_{...})^2 + (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2 + 2(Y_{ijk} - \bar{Y}_{ij.})\left[(\bar{Y}_{i..} - \bar{Y}_{...}) + (\bar{Y}_{.j.} - \bar{Y}_{...}) + (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})\right] + 2(\bar{Y}_{i..} - \bar{Y}_{...})\left[(\bar{Y}_{.j.} - \bar{Y}_{...}) +\right.$$

$$(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})] + 2(\bar{Y}_{.j.} - \bar{Y}_{...})(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})$$

$$= (Y_{ijk} - \bar{Y}_{ij.})^2 + (\bar{Y}_{i..} - \bar{Y}_{...})^2 + (\bar{Y}_{.j.} - \bar{Y}_{...})^2 + (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$$

$$+ 2(Y_{ijk} - \bar{Y}_{ij.})[\bar{Y}_{ij.} - \bar{Y}_{...}) + 2(\bar{Y}_{i..} - \bar{Y}_{...})[\bar{Y}_{ij.} - \bar{Y}_{i..}]$$

$$+ 2(\bar{Y}_{.j.} - \bar{Y}_{...})(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})$$

Now, plug back into SS_{TOT} and expand:

$$SS_{TOT} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(Y_{ijk} - \bar{Y}_{...})^2 = SS_E + SS_A + SS_B + SS_{AB}$$

$$+ 2\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}[\overset{(1)}{(Y_{ijk} - \bar{Y}_{ij.})(\bar{Y}_{ij.} - \bar{Y}_{...})} + \overset{(2)}{(\bar{Y}_{i..} - \bar{Y}_{...})(\bar{Y}_{ij.} - \bar{Y}_{i..})}$$

$$+ \overset{(3)}{(\bar{Y}_{.j.} - \bar{Y}_{...})(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})}]$$

I will show each of labelled terms equals zero:

(1) $$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}[(Y_{ijk} - \bar{Y}_{ij.})(\bar{Y}_{ij.} - \bar{Y}_{...})] =$$

$$\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{Y}_{ij.} - \bar{Y}_{...})\underbrace{\sum_{k=1}^{K}(Y_{ijk} - \bar{Y}_{ij.})}_{\text{0 since this is sum of deviations from mean}} = 0$$

(2) $$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}((\bar{Y}_{i..} - \bar{Y}_{...})(\bar{Y}_{ij.} - \bar{Y}_{i..})]$$

0 since sum of deviations from mean

$$= \sum_{i=1}^{I}(\bar{Y}_{i..} - \bar{Y}_{...})\sum_{j=1}^{J}(\bar{Y}_{ij.} - \bar{Y}_{i..})\sum_{k=1}^{K}1\ )$$

$$= 0$$

well.

$$(3)\ \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left[(\bar{Y}_{.j.}-\bar{Y}_{...})(\bar{Y}_{ij.}-\bar{Y}_{i..}-\bar{Y}_{.j.}+\bar{Y}_{...})\right]$$

$$=\sum_{i=1}^{I}\sum_{j=1}^{J}\left[(\bar{Y}_{.j.}-\bar{Y}_{...})(\bar{Y}_{ij.}-\bar{Y}_{i..}-\bar{Y}_{.j.}+\bar{Y}_{...})\right]\sum_{k=1}^{K}1$$

$$=\left\{\sum_{j=1}^{J}(\bar{Y}_{.j.}-\bar{Y}_{...})\sum_{i=1}^{I}(\bar{Y}_{ij.}-\bar{Y}_{i..}-\bar{Y}_{.j.}+\bar{Y}_{...})\right\}K$$

$$=K\sum_{j=1}^{J}(\bar{Y}_{.j.}-\bar{Y}_{...})\underbrace{(I\bar{Y}_{.j.}-I\bar{Y}_{...}-I\bar{Y}_{.j.}+I\bar{Y}_{...})}_{0}$$

$$=0$$

Hence, $SS_{TOT}=SS_E+SS_A+SS_B+SS_{AB}+0$
$=SS_E+SS_A+SS_B+SS_{AB}$

Proving independence between SS_E and other SS's:

Model assumption – $(l=1,\dots,n_{ij},\ i=1,\dots,I,\ j=1,\dots,J)$

$$Y_{ijl}=\mu+\alpha_i+\beta_j+\delta_{ij}+\underbrace{\epsilon_{ijl}}_{\overset{iid}{\sim}N(0,\sigma^2)}$$

By lecture 17, corollary D:

$\frac{SS_E}{\sigma^2}\sim\chi^2_{n-IJ}$ and it's independent of each cell mean $\bar{Y}_{ij.}$

$\bar{Y}_{i..}$, $\bar{Y}_{.j.}$, $\bar{Y}_{...}$ can be expressed as a sum of $\bar{Y}_{ij.}$, hence ~~these are~~ SS_E are independent of these as well.

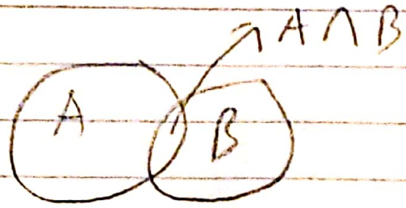
Finally, since SS_A, SS_B, SS_{AB} are functions of $\overline{Y}_{ij\cdot}$, $\overline{Y}_{i\cdot\cdot}$, $\overline{Y}_{\cdot j\cdot}$, and $\overline{Y}_{\cdot\cdot\cdot}$, these imply that ~~SSE~~ SS_E are independent of SS_A, SS_B, and SS_{AB}.

Hence, SS_E is independent from the other SS's.

spatial statistics

dependencies in extreme events

3] (1)



case 1: There is overlap between A & B

we can see that

since $P(A)+P(B)-P(A\cap B)=P(A\cup B)$

and since $P(A\cup B)\leq 1$

$$P(A)+P(B)-P(A\cap B)\leq 1$$
$$P(A)+P(B)-1\leq P(A\cap B)$$

case 2:

case 2: There is no overlap between A & B

Then, $P(A)+P(B)=1$, $P(A\cap B)=0$

and $P(A)+P(B)-P(A\cap B)=P(A\cup B)$

$$\Rightarrow P(A)+P(B)-0\leq 1$$
$$\Rightarrow P(A)+P(B)\leq 1+P(A\cap B)$$
$$\Rightarrow P(A)+P(B)-1\leq P(A\cap B)$$

true since $P(A\cap B)=0$.

Hence, completes proof.

(2) expand the LHS:

$$P\left(\bigcap_{i=1}^{m}A_i\right)=1-P\left(\bigcup_{i=1}^{m}A_i^c\right)$$

since $\left(\bigcap_{i=1}^{m}A_i\right)^c=\bigcup_{i=1}^{m}A_i^c$ using De Morgan's law.

By boole's inequality,

$$P\left(\bigcup_{i=1}^{m} A_i^c\right) \leq \sum_{i=1}^{m} P(A_i^c)$$

$$\Rightarrow 1 - P\left(\bigcup_{i=1}^{m} A_i^c\right) \geq 1 - \sum_{i=1}^{m} P(A_i^c)$$

$$\Rightarrow P\left(\bigcap_{i=1}^{m} A_i\right) \geq 1 - \sum_{i=1}^{m} P(A_i^c)$$

$$= P\left(\bigcap_{i=1}^{m} A_i\right) \geq 1 - \sum_{i=1}^{m} (1 - P(A_i))$$

$$\Rightarrow P\left(\bigcap_{i=1}^{m} A_i\right) \geq (1 - m) + \sum_{i=1}^{m} P(A_i)$$

$$\boxed{\text{Hence, } P\left(\bigcap_{i=1}^{m} A_i\right) \geq \sum_{i=1}^{m} P(A_i) - (m-1)}$$

(3) let m parameters be $A_1, \ldots, A_m$

DeMorgan's Law $\rightarrow$ $P(A_1 \cap A_2 \ldots \cap A_m) = 1 - P(A_1^c \cup A_2^c \cup \ldots \cup A_m^c)$

$$\Rightarrow P(A_1^c \cup A_2^c \cup \ldots \cup A_m^c) \leq \sum_{i=1}^{m} P(A_i^c)$$

$$\Rightarrow P(A_1 \cap A_2 \cap \ldots \cap A_m) \geq 1 - \sum_{i=1}^{m} P(A_i^c)$$

let A_i denote event that

$\Rightarrow$ $\theta_i \in I_i$ where θ_i represents parameter, I_i is interval

$P(\theta_i \in I_i) = 1 - \alpha_i$

$$\Rightarrow P(\theta_1 \in I_1, \theta_2 \in I_2, \ldots, \theta_m \in I_m) \geq 1 - \sum_{i=1}^{m} P(\theta_i \notin I_i)$$

$$\geq 1 - \sum_{i=1}^{m} \alpha_i$$

where $\sum_{i=1}^{m} \alpha_i = m\alpha$

if all α_i are equal, the simultaneous confidence level is $\geq 1 - m\alpha$. This is smaller than $1-\alpha$.

if each $\alpha_i = \frac{\alpha}{m}$, $\sum_{i=1}^{m} \alpha_i = \alpha$, and the confidence level is $\geq 1-\alpha$. This implies each confidence interval is set to level $1 - \frac{\alpha}{m}$.

(4) Compared to Tukey's method, Bonferroni's procedure produces more conservative CIs.

4]

Source	Sum Sq	Df	Mean Sq	F-stat	P-value
Treatment	64.42	3	21.47	8.98	0.0008647
Residual	40.63	17	2.39		
Total	105.05	20			

$df_{treatment} = 4-1 = 3$ $\quad$ $df_{residual} = 20-3 = 17$

$$MS_T = \frac{64.42}{3} \approx 21.47$$

$$MS_R = \frac{SS_R}{df_R} \qquad SS_R = MS_R \cdot df_R = 17 \cdot 2.39 = 40.63$$

P-value: $pf\left(\frac{MS_T}{MS_R}, 3, 17, lower.tail = False\right)$

≈ 0.0008647

5]

(1)

Source	df	SS	MS	F-stat
Treatment	2	1568.715	784.3577	14.44844
Residuals	15	814.3002	54.28668	
Total	17	2383.0152		

p-value

0.0003180036

Please see attached code file for computations.

(2) The F-stat in (1) is the F-stat for testing the null hypothesis that:

$$H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0$$

with a p-value of 0.0003180036, we reject the H_0 at a significance level of $\alpha = 0.05$.

(3) Bonferroni CI's 95%;

low: [2.768800, 4.154534]
med: [8.444959, 10.320755]
high: [13.121101, 40.130899]

← See attached code for computation.

None of the CI's overlap with each other.

The information we learn is about how the three diets differ, which pairs are significantly different, and we draw conclusions about each hypothesis test simultaneously.

*see attached code for computations

6] (1) No idea. [Honestly not sure where to start...]

Note to reader: these #'s should be flipped for I w/ larger negative on left, smaller negative on right.

(2) $\alpha_1 - \alpha_2$: $[-0.01816447, -19.64850770]$

$\alpha_1 - 2\alpha_2 + \alpha_4$, $[-15.93667, -43.39666]$

$3\alpha_3 - \alpha_1 - 2\alpha_2$: $[-9.904615, -41.928778]$

$5\alpha_2 - 4\alpha_3 - \alpha_4$: $[41.71705, 92.61628]$

$2\alpha_2 - \alpha_1 - \alpha_3$: $[9.217395, 34.15938]$

(3) Scheffe simultaneous CI's:

$\alpha_1 - \alpha_2$: $[-13.554476, -6.112191]$

$\alpha_1 - 2\alpha_2 + \alpha_4$: $[-34.77118, -24.56216]$

$3\alpha_3 - \alpha_1 - 2\alpha_2$: $[-31.60378, -19.73006]$

$5\alpha_2 - 4\alpha_3 - \alpha_4$: $[57.58134, 76.75199]$

$2\alpha_2 - \alpha_1 - \alpha_3$: $[16.97778, 26.35555]$

We can see across the board that Scheffe CI's are tighter than the analogous Bonferroni CI's.

*see code for computations
*note: I used the Lab 11 version of dataset
→ there are 2 or 6 causes

7] $\alpha_2 - \alpha_1$:

flip these 2 numbers in interval.

Bonferroni → [0.04523681, -0.12773681]

Scheffe → [-0.0243117, 0.0576888]

Tukey → [-0.104134026, 0.022134026]

$\alpha_3 - \alpha_1$:

Bonferroni: [-0.06616321, 0.06216321]

Scheffe: [-0.01441707, 0.01041707]

Tukey: [-0.065134026, 0.061134026]

$\alpha_3 - \alpha_2$:

Bonf.: [-0.03239284, 0.11039284]

Scheffe: [0.02518383, 0.05281617]

Tukey: [-0.024134026, 0.102134026]

Bonferroni produces the widest intervals, Scheffe produces the narrowest intervals, and Tukey is in between the two.

Bonferroni is least accurate because it collects for the worst case, independent comparisons, namely by reducing confidence level of each individual interval. This makes each individual CI in Bonferroni wider than Scheffe or Tukey, and hence the least precise.

8] (1) See attached code for plot.

Yes, there are interactions between the 2 factors.

(2) See attached code for computations:

Source	df	SS	MS	F	p-value
Species	2	3.52	1.76	0.09	0.913
Fertilizer	3	154.82	51.61	2.66	0.056
Interaction	6	211.89	35.32	1.82	0.110
Error	60	1164.31	19.41		
Total	71	1534.54			

Yes it is the same as output from Anova() in R.

(3) $H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0$, $\alpha = 0.05$

This is main effect of species —

with a F-stat of 0.09 and 2 df, yields a p-value of 0.913. We fail to reject H_0 at significance level of 0.05.

(4) $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

This is main effect of fertilizer —

with a F-stat of 2.66 and 3 df, yields a p-value of 0.056. We fail to reject H_0 at significance level of 0.05.

(5) $H_0: \delta_{ij} = 0$ for $i = 1, 2, 3$; $j = 1, 2, 3, 4$

This is main effect of interaction — with a F-stat of 1.82 and 6 df, yields a p-value of 0.110. We fail to reject H_0 at a significance level of 0.05.

9] 3-way layout:

Three factors A, B, C, with levels a, b, c respectively.

Y: response variable; Factor A with levels i=1 to a; Factor B with levels j=1 to b; Factor C with levels k=1 to c. Y_{ijkl} is lth observation in cell (i,j,k), l=1 to n_{ijk}

balance condition — $n_{ijk} = n$

Cell means: $Y_{ijkl} = \mu_{ijk} + \varepsilon_{ijkl}$

- μ: expected value of all obs. in cell (i,j,k)
- $\varepsilon_{ijkl} \overset{iid}{\sim} N(0, \sigma^2)$
- $Y_{ijkl} \sim N(\mu_{ijk}, \sigma^2)$ independent

Factor effects:

$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl}$

μ: overall mean; $\alpha_i, \beta_j, \gamma_k$: main effects of factors A, B, C.

$\alpha\gamma, \beta\gamma, \alpha\beta$: two-factor interactions

$\alpha\beta\gamma$: three-factor interaction

$\alpha_i = \mu_{i..} - \mu_{...}$ $\quad (\alpha\gamma)_{ik} = \mu_{i.k} - \mu_{i..} - \mu_{..k} + \mu_{...}$

$\beta_j = \mu_{.j.} - \mu_{...}$ $\quad (\beta\gamma)_{jk} = \mu_{.jk} - \mu_{.j.} - \mu_{..k} + \mu_{...}$

$\gamma_k = \mu_{..k} - \mu_{...}$ $\quad (\alpha\beta)_{ij} = \mu_{ij.} - \mu_{i..} - \mu_{.j.} + \mu_{...}$

$$(\alpha\beta\gamma)_{ijk} = \mu_{ijk} - \mu_{ij\cdot} - \mu_{i\cdot k} - \mu_{\cdot jk} + \mu_{i\cdot\cdot} + \mu_{\cdot j\cdot} + \mu_{\cdot\cdot k} - \mu_{\cdot\cdot\cdot}$$

constraint: sums of effects for any of indices are zero. $\mu_{\cdot\cdot\cdot}$ is overall mean hence.

Interpretations!

<u>main effect</u> – the effect of solely 1 factor on the response variable

<u>two-factor interaction</u> – combined effects of 2 factors on the response variable.

<u>three-factor interaction</u>: combined effect of 3 factors on response variable. This implies interaction among 2 factors (A & B) is different across levels of 3rd factor (C) if it's significant.

10] Please see attached code for MANOVA computations.

- Pillai: approx F: 4.2984, Pillai: 1.5539, p-value: 2.413e-05
- Wilks: approx F: 13.088, Wilks: 0.01201, p-value: 1.84e-12
- Hotelling-Lawley (HL): approx F: 39.76, HL: [illegible], p-value: 2.2e-16
- Roy: approx F: 130.64, Roy: 34.161, p-value: 9.444e-15

for $H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$, $\alpha = 0.05$

we reject the H_0 at a significance level of 0.05 under all 4 test statistics. All F-stats, p-values and test statistics show strong significance, providing robust evidence that the alternative hypothesis is likely true, and H_0 is not.