Source of Variation	df	SS	MS	F_{obs}
Intercept	1	SSI	SSI	SSI/MSE
Model(uncorrected)	q	USS(model)	USSM/q	(USSM/q)/MSE
Model(corrected)	q-1	CSS(model)	CSSM/q-1	(CSSM/q-1)/MSE
Error(residual)	N-q	SSE	SSE/(N-q)=MSE	-
Total(uncorrected)	N	USS(total)	-	-
Total(corrected)	N-1	CSS(total)	-	-

T=total	M=Model	E=Error
Source	**Formula**	**Alternate**
CSS(M)	$\sum\hat{y}_i^2 - N\bar{y}^2 = \sum(\hat{y}_i - \bar{y}^2)$	USS(M) - SSI
CSS(T)	$\sum y_i^2 - N\bar{y}^2 = \sum(y_i - \bar{y}^2)$	USS(T) - SSI & CSS(M) + SSE
CSS(E)/USS(E)/SSE	CSS(T) - CSS(M)	USS(T) - USS(M) = USS(E)
SSE	$\sum(\hat{y} - y_i)^2$	$\sum\hat{\epsilon}^2$
USS(M)	$\sum\hat{y}_i^2$	USS(T) - SSE
USS(T)	$\sum y_i^2$	USS(M) + SSE

Source	DF	Type I SS	MS	F-value
x_i	1	Type I SS	T1SS/df = T1SS	MS/MSE(model)
$\Sigma_{i=1}^{q-1} T1SS_{x_i} = CSS(M)$				

(8 points) Compute the 95% prediction interval for a subject with $x_1 = 1$ and $x_2 = 2$.

Solution: $\hat{y} = (1,1,2)\hat{\beta} = 3.27$ and $var(\hat{y}) = (1,1,2)\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}(1,1,2)' + \hat{\sigma}^2 = 3.9$ a 95% prediction interval is $3.27 \pm 4.3\sqrt{3.9} = (-5.3, 11.7)$.

(8 points) Calculate the corrected R^2_c, interpret its value, and test the hypothesis that its corresponding population value is zero, that is, $H_0 : \rho_c^2 = 0$.

Solution: Since $\bar{y} = 3.4$, we have $CSS(regression) = \sum_i \hat{y}_i^2 - 5\bar{y}^2 = 11.2$ $CSS(total) = \sum_i y_i^2 - 5\bar{y}^2 = 17.2$ and corrected $R^2_c = 11.2/17.2 = 0.65$.

$H_0 : \rho_c^2 = 0$ is equivalent to $H_0 : \beta_1 = \beta_2 = 0$. Thus we have $F - test = \frac{11.2/2}{\hat{\sigma}^2} = 2.25 \sim F_{2,2}$.

Compare full model $post = \beta_0 + \beta_1 pre + \beta_2 age + \beta_3 gender + error$ with $post = \beta_0 + \beta_2 age + \beta_3 gender + error$ or testing $H_0 : \beta_1 = 0$.

t-test $= \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = 1.15/\sqrt{\hat{\sigma}^2 * 0.011} = 13.1 \sim t(96) \approx N(0,1)$ thus significant at $\alpha = 0.05$. Thus F-test $= (t-test)^2 = 13.1^2 = 171.6 \sim F_{1,96}$.

Two-way ANOVA coding schemes

Reference Cell: The total number of columns of **X**: 1 col of 1's for reference cell mean, (a-1) for factor A, (b-1) for factor B and (a-1)(b-1) for interaction.

$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

Cell Means:

$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Classic ANOVA: The total number of columns of **X** is 1 + a + b + a*b

$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$

Effect: same as reference cell but you replace the zeros for the reference row and range with -1s

$\begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

Cell Means Per Coding 2 drugs A and B (α) and 3 dosages 1, 2, and 3 (β). The model is $y_{ijk} = \mu + \alpha_1 I(drugA) + \beta_1 I(dose1) + \beta_2 I(dose2) + e_{ijk}$

Reference Cell Drug B and Dose 3 are reference

Drug	*Dose*	*Mean*	*With interactions*
A	1	$\mu + \alpha_1 + \beta_1$	$+\gamma_{A1}$
A	2	$\mu + \alpha_1 + \beta_2$	$+\gamma_{A2}$
A	3	$\mu + \alpha_1$	$+\gamma_{A3}$
B	1	$\mu + \beta_1$	
B	2	$\mu + \beta_2$	
B	3	μ	

Cell Means For 2-way ANOVA can only have max a*b terms, which are the interaction terms.

Drug	*Dose*	*Mean*
A	1	γ_{A1}
A	2	γ_{A2}
A	3	γ_{A2}
B	1	γ_{B1}
B	2	γ_{B2}
B	3	γ_{B3}

$(X'X)^{-1} = \begin{vmatrix} .308 & -.06 & -.017 \\ -.06 & -.025 & -.004 \\ -.017 & -.004 & .006 \end{vmatrix}$ $X'y = \begin{vmatrix} 405 \\ 1402 \\ 2350 \end{vmatrix}$

Source	DF	SS	MS	Fval	$P > F$
Model	2	79	239.5	349.6	$< .001$
Error	97	11	.113		
Ctotal	99	90			

param	est	se	tval	$p > \|t\|$
x0	.67	.187	3.58	.009
x1	1.35	.053	25.47	$< .001$
x2	1.61	.026	61.81	$< .001$

2) c) Test $H_0 : \beta_1 = 1$

t-test$= \frac{\hat{\beta}_1 - 1}{se(\hat{\beta}_1)} = 6.6 \sim t_{97}$ $6.6 > 1.96$ reject H_0

d) Test $H_0 : \beta_1 = \beta_2 = 1$

$B_1 = 1 \quad B_2 = 1$

$C = \begin{vmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}$ $\theta_0 = \begin{vmatrix} 1 \\ 1 \end{vmatrix}$ $H_0 : C\theta = \theta_0$ $\hat{\theta} = \begin{vmatrix} 1.35 \\ 1.61 \end{vmatrix}$

calc M calc M^{-1}

F-test$= \frac{43.83}{.113} = 387.8 \sim F_{2,97}$

e) 95% CI of $\beta_1 + \beta_2$ (θ)

$\hat{\theta} = \hat{\beta}_1 + \hat{\beta}_2 = 2.96$ $C = [0,1,1]$

$se(\hat{\theta}) = \hat{\sigma}^2_{\hat{\theta}} = \sqrt{.0026}$

$2.96 \pm 1.96\sqrt{.0026} = [2.86, 3.06]$

f) Transform x_1, x_2 to $z_1 = x_1 - 2$, $z_2 = x_2 - 4$

Refit with $y^* = \beta_0^* + \beta_1^* z_1 + \beta_2^* z_2$

$= (\beta_0^* - 2\beta_1^* - 4\beta_2^*) + \beta_1^* x_1 + \beta_2^* x_2$

$\beta_0^* = \beta_0^* + 2\beta_1^* + 4\beta_2^* = 9.8$ $C = [1,2,4]$

$\sigma^2 = \hat{\sigma}^2 C(X'X)^{-1}C^{-1} = .085$ $t = 115.3$

3) f) is $H_0 : \beta_0 - \beta_2 = 0$ and $\beta_1 + 2\beta_2 = 2$ and $2\beta_0 + \beta_1 = 2$ testable? Reduce to ETH?

$y = \begin{vmatrix} 0 \\ 2 \\ 3 \\ 6 \\ 10 \end{vmatrix}$ $X = \begin{vmatrix} 1 & 6 & 11 \\ 1 & 7 & 13 \\ 1 & 8 & 15 \\ 1 & 9 & 17 \\ 1 & 11 & 21 \end{vmatrix}$ $\beta = \begin{vmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{vmatrix}$

$C = \begin{vmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 2 & 1 & 0 \end{vmatrix}$ reduces $\begin{vmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{vmatrix}$ 2pivots $\theta_0 = \begin{vmatrix} 0 \\ 2 \\ 2 \end{vmatrix}$

not FR, not testable, reduces to C^* testable

$C^* = \begin{vmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{vmatrix}$ $\theta_0^* = \begin{vmatrix} 0 \\ 2 \end{vmatrix}$ $H_0 : \begin{vmatrix} \beta_0 - \beta_2 = 0 \\ \beta_1 + 2\beta_2 = 2 \end{vmatrix}$

1) a) $x = \begin{vmatrix} x_1 \\ x_2 \\ x_3 \end{vmatrix} \sim N(0, \Sigma)$ $\Sigma = \begin{vmatrix} 1 & 0 & .6 \\ 0 & 1 & .5 \\ .6 & .5 & 1 \end{vmatrix}$

$c = \begin{vmatrix} 2 & 1 & -1 \end{vmatrix}$

$2x_1 + x_2 - x_3 \sim N(c\mu, c\Sigma c') = N(0, 2.6)$

b) $Cov(x_1 - x_2, 2x_2 + x_3) = c_1\Sigma c_2' = -1.9$

$c_1 = \begin{vmatrix} 1 & -1 & 0 \end{vmatrix}$ $c_2 = \begin{vmatrix} 0 & 2 & 1 \end{vmatrix}$

2011 1) $\hat{\beta} = \begin{vmatrix} 24.877 \\ -.231 \\ -.185 \end{vmatrix}$ $\hat{\sigma^2} = 5.59$ $dfE = 197$

$diag((X'X)^{-1}) = [.03752, .0017, .00056]$

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$

$se\beta_1 = .032 \quad se\beta_2 = .05605$

95% CI $x_1 = 2$ $x_2 = 6$ $\quad C = [1\ 2\ 6]$

$\hat{\beta}^* \pm 1.96 * se(\hat{\beta}^*) \quad \hat{\beta}^* = \hat{\beta}_0 + 2\hat{\beta}_1 + 6\hat{\beta}_2 = 23.3$

$se = \sqrt{Var(\hat{\beta}^*)} = \sqrt{\hat{\sigma^2} M} = .1946$

Test $H_0 : \beta_1 = 3\beta_2$

$\theta = \beta_1 - 3\beta_2$ $H_0 : \theta = 0$ $C = [0\ 1\ -3]$ $t = .327/.1868 = 1.75 \sim t_{97} \approx Z = 1.96$ FTR H_0

Center x_1 x_2 at means 1,5 $z_1 = x_1 - 1$ $z_2 = x_2 - 5$ $y^* = \beta_0^* + \beta_1^* z_1 + \beta_2^* z_2$

$\beta_0^* + \beta_1^*(x_1 - 1) + \beta_2^*(x_2 - 5)$

$y^* = (\beta_0^* - \beta_1^* - 5\beta_2^*) + \beta_1^* x_1 + B_2^* x_2$

$\beta_0^* = \beta_0 + \beta_1 + 5\beta_2 = -.9116$ $\beta_1^* = \beta_1$ $\beta_2^* = \beta_2$

$C = [1\ 1\ 5]$ $se(\beta_0^*) = \sqrt{M * \hat{\sigma^2}} = .254$

2) 4 dose levels (1,2,3,4) 2 methods (oj, aa)

n=800 ref cell $\Rightarrow$ ref group: aa dose 1

$y_i = \mu + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_3 x_{3i} + \beta x_{4i} + \epsilon_i$

dummy variables:

$x_{1i} = \begin{cases} 1 \text{ if dose=2} \\ 0 \text{ otherwise} \end{cases}$ all xs defined similarly

$x_{4i} = I(method = oj)$

cell mean of each group

Method	Dose	Mean
oj	1	$\mu + \beta$
oj	2	$\mu + \alpha_1 + \beta$
oj	3	$\mu + \alpha_2 + \beta$
oj	4	$\mu + \alpha_3 + \beta$
aa	1	μ
aa	2	$\mu + \alpha_1$
aa	3	$\mu + \alpha_2$
aa	4	$\mu + \alpha_3$

add interaction btwn method, dose $\Rightarrow \theta : 8 \times 1$

let μ_o and μ_a be overall means for the two methods

write them for models w/ and w/o interaction

Derive 2 C matrices for testing $H_0 : \mu_o = 2\mu_a$ under the 2 models

w/o: $\mu_o = \mu + \frac{\alpha_1 + \alpha_2 + \alpha_3}{4} + \beta$

$\mu_a = \mu + \frac{\alpha_1 + \alpha_2 + \alpha_3}{4}$

$H_0 : \mu_o = 2\mu_a \Leftrightarrow \mu + \frac{\alpha_1 + \alpha_2 + \alpha_3}{4} - \beta = 0$

$C = (1\ 1/4\ 1/4\ 1/41)$ for $\theta = (\mu, \alpha_1, \alpha_2, \alpha_3, \beta)$

w/:$H_0 : \mu_o = 2\mu_a \Leftrightarrow$

$\mu + \frac{\alpha_1 + \alpha_2 + \alpha_3}{4} - \frac{\gamma_[11] + \gamma_{12} + \gamma_13}{4} - \beta = 0$

$C = (1\ 1/4\ 1/4\ 1/411/4\ -1/4\ -1/4)$ for $\theta = (\mu, \alpha_1, \alpha_2, \alpha_3, \beta, ga_{11}, \gamma_{12}, \gamma_{13})$

treat dose as a continuous variable and fit model with additive effects of method and dose level with no interaction

nested w/in

$y_i = \mu + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_3 x_{3i} + \beta x_{4i} + \epsilon_i$? yes

write down H_0 for comparing the 2 models and derive C matrix. df of F test under H_0? Let $\alpha_2 = 2\alpha_1$, $\alpha_3 = 3\alpha_1$ (treating dose as continuous)

$C = \begin{vmatrix} 0 & -2 & 1 & 0 & 0 \\ 0 & -3 & 0 & 1 & 0 \end{vmatrix}$

assuming $\theta = (\mu, \alpha_1, \alpha_2, \alpha_3, \beta)'$

df=2 , 7955

2010 1) $n = 1000$ drugs = A,B doses=1,2,3,4,5

ref cell $y = \mu + \alpha_i + \beta_j + \epsilon$ (w/out interaction)

Drug B, Dose 5 ref

α_1 models effect of drug A

β_j models effect for dose j

dims: $y = 1000 \times 1$ $X = 1000 \times 6$ $\beta = 6 \times 1$

$\epsilon = 1000x1$ $\epsilon \sim N(0, \sigma^2 I)$

Cell mean for each combo of drug and dose:

Drug	Dose	Mean
A	1	$\mu + \alpha_1 + \beta_1$
A	2	$\mu + \alpha_1 + \beta_2$
A	3	$\mu + \alpha_1 + \beta_3$
A	4	$\mu + \alpha_1 + \beta_4$
A	5	$\mu + \alpha_1$
B	1	$\mu + \beta_1$
B	2	$\mu + \beta_2$
B	3	$\mu + \beta_3$
B	4	$\mu + \beta_4$
B	5	μ

for ref cell model p=6 df model $= 6 - 1 = 5$

with interaction $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon$

$\mu + \alpha_i + \beta_4 + \beta_3 + \beta_2 + \beta_1 + \gamma 14 + \gamma 13 + \gamma 12 + \gamma 11 + \epsilon$

dim ref model: $\beta = 10 \times 1$ everything else same

Drug	Dose	Mean
A	1	$\mu + \alpha_1 + \beta_1 + \gamma 11$
A	2	$\mu + \alpha_1 + \beta_2 + \gamma 12$
A	3	$\mu + \alpha_1 + \beta_3 + \gamma 13$
A	4	$\mu + \alpha_1 + \beta_4 + \gamma 14$
A	5	$\mu + \alpha_1$
B	1	$\mu + \beta_1$
B	2	$\mu + \beta_2$
B	3	$\mu + \beta_3$
B	4	$\mu + \beta_4$
B	5	μ

γ_{11} is the diff btwn the diff of dose1 and dose5 given drug A and b respectively

$= [E(y|A,1) - E(y|A,5)] - [E(y|B,1) - E(y|B,5)]$

Let μ_A and μ_B be the overall mean values of cholesterol level for drug A and B

$\mu_A = [(\mu + \alpha_1 + \beta_1 + \gamma_{11}) + (\mu + \alpha_1 + \beta_2 + \gamma_{12}) + (\mu + \alpha_1 + \beta_3 + \gamma_{13}) + (\mu + \alpha_1 + \beta_4 + \gamma_{14}) + (\mu + \alpha_1)]/5$

$\mu_B = [(\mu + \beta_1) + (\mu + \beta_2) + (\mu + \beta_3) + (\mu + \beta_4) + \mu]/5$

$H_0 : \mu_A = \mu_B \Leftrightarrow H_0 \frac{\alpha_1 + \gamma_{11} + \gamma_{12} + \gamma_{13} + \gamma_{14}}{5} = 0$

2) full model in each cell, dose is treated as interval variable

$\hat{y} = 9.82 + 3.55 * I(drug = A) + 2.14 dose + .296 age$ $+ 2.21 I(drug = A) * dose + .3 I(drug = A) * age$

fitted model when drug B is used (ref level)

$\hat{y} = 98.82 + 2.145 dose + 2.96 age$

Write down fitted model when drug A is used

$\hat{y} = 98.82 + 3.55 * 1 + 2.14 dose + .296 age + 2.21 * 1 * dose + .3 * 1 * age$

$\Rightarrow \hat{y} = 102.37 + 4.35 dose + .596 age$