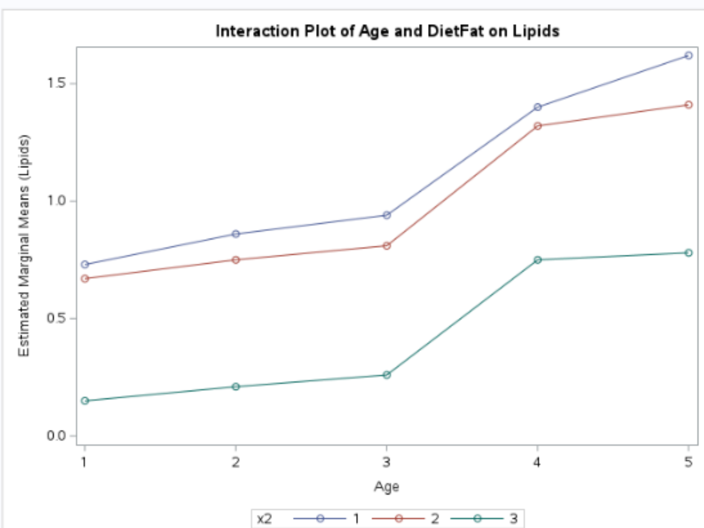
Priyasri Sankaran

Project2

2. RCB design, Fat in diets: Consider the data of problem21.7 (Applied Linear
Statistical Models
Fifth Edition
Michael H. Kutner, EmOlY University
Christopher J. Nachtsheim, University of Minnesota
John Neter, University of Georgia
William Li,Universlty of Minnesotap. 913).
Note that the outcome Y_{ij} is the reduction in lipids after being on the diet.

(1) Obtain the interaction plot for these data. In SAS you can type plots=all in your proc glm statement, and including the interaction, e.g. model lipids = age|dietfat. Considering the sampling variability of estimating each μ_{ij} by one observation Y_{ij}, are the curves reasonably parallel? If so, that would indicate the additive model IV

iid 2 $Y_{ij} = \mu + \rho_i + \tau_j + \varepsilon_{ij}$, $\varepsilon_{ij} \sim N(0, \sigma)$ is appropriate.

	Yes, the curves are reasonably parallel. We observe the plot levels are different on some spots, but reasonably parallel. This suggests that the interaction effects are minimal and the additive model IV is likely appropriate.

1) Still just using the plot in part (1), what generally happens to the ability to reduce lipids as age increases? Why might this be? Which diet is best at reducing lipids, i.e. has the highest lipid reduction?
   Comments:
   By looking at the plot, we see that as the age increases the lipid reduction decreases.

From the least-square means output(please refer the plot, green line), diet (x2)=3 had the highest mean lipid reductions. This diet appears to be the best at reducing lipids across all age groups.

2) As a double check, perform Tukey's one-degree-of-freedom test for additivity for these data and report the p-value and conclusions of your test. That is, test H_0:$\delta = 0$ in the model

$$Y_{ij} = \mu + \rho_i + \tau_j + \delta\rho_i\tau_j + \varepsilon_{ij}.$$

H_0: The interaction term is zero
H_a: The interaction term$\neq 0$

SAS

Tukey's Additivity

The GLM Procedure

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	2.74850145	0.39264306	273.25	<.0001
Error	7	0.01005855	0.00143694		
Corrected Total	14	2.75856000			

R-Square	Coeff Var	Root MSE	y Mean
0.996354	4.491343	0.037907	0.844000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
x1	4	1.41896000	0.35474000	246.87	<.0001
x2	2	1.32028000	0.66014000	459.41	<.0001
fitted*fitted	1	0.00926145	0.00926145	6.45	0.0387

Source	DF	Type III SS	Mean Square	F Value	Pr > F
x1	4	0.08460150	0.02115038	14.72	0.0016
x2	2	0.12430033	0.06215016	43.25	0.0001
fitted*fitted	1	0.00926145	0.00926145	6.45	0.0387

R

```
> library(asbio)
Loading required package: tcltk
> with(diets,tukey.add.test(y,x1,x2))

Tukey's one df test for additivity
F = 6.4452749   Denom df = 7    p-value = 0.0387344
>
```

We plotted the same in SAS and R, we got F=6.45 and the P-value=0.0387. Decision Rule: P-value>0.05 we accept the null, P-value<0.05 we reject the null hypothesis. Conclusion: P-value=0.0387<0.05, we reject the null hypothesis. Interaction exist in the model. The P-value is very close to the alpha value 0.05.

Observation: Form the plot we observe that there is an interaction, the line is still parallel, the test might not be detecting the interaction in the model correctly.

3) Fit the additive model IV to the data and write down the fitted model. Test whether there's differences in diet, e.g. H_0:$\tau_j = 0$ at the 5% level using the Type III p-value. Was blocking on age effective?

H_0:$\tau_j = 0$; tests whether there are significant difference in lipid reduction among the levels of Dietfat(x2)

SAS

The GLM Procedure

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	2.73924000	0.45654000	189.04	<.0001
Error	8	0.01932000	0.00241500		
Corrected Total	14	2.75856000			

R-Square	Coeff Var	Root MSE	y Mean
0.992996	5.822589	0.049143	0.844000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
x1	4	1.41896000	0.35474000	146.89	<.0001
x2	2	1.32028000	0.66014000	273.35	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
x1	4	1.41896000	0.35474000	146.89	<.0001
x2	2	1.32028000	0.66014000	273.35	<.0001

R

```
> fit <- aov(y~ x1+x2, diets)
> summary(fit)
            Df Sum Sq Mean Sq F value   Pr(>F)    
x1           4 1.4190  0.3547   146.9 1.61e-07 ***
x2           2 1.3203  0.6601   273.4 4.33e-08 ***
Residuals    8 0.0193  0.0024                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Results:
Sum of Squares for x2=1.32
F value=273.35, P-value=0.0001<0.05 significant level
Decision Rule:P-value<0.05, reject the null hypothesis.P-value>0.05,accept null hypothesis.
Conclusion:
P<0.05, we reject the null hypothesis and conclude the there is a significant differences in lipid reduction amoung the levels of Dietfat(x2)

Was blocking on Age effective?
Results for x1:
F value: 146.89, P-value:0.0001.
Decision Rule:P-value<0.05, reject the null hypothesis.P-value>0.05,accept null hypothesis.
Conclusion:
P<0.05, we reject the null hypothesis and conclude the there is a significant differences in lipid reduction amoung the levels of Dietfat(x1). Efficiency of clocking on age was effective.

4) Look at all three pairwise differences for diet using Tukey’s procedure and make a “lines plot” for the three diet types with an overall FER of 5%.

```
> # Get confidence intervals for pairwise differences
> confint(Tukey)
 contrast  estimate     SE df lower.CL upper.CL
 x21 - x22    0.118 0.0311  8   0.0292    0.207
 x21 - x23    0.680 0.0311  8   0.5912    0.769
 x22 - x23    0.562 0.0311  8   0.4732    0.651
```

R result Interpretation: non of the intervals contain zero, proving that there is a significant difference.

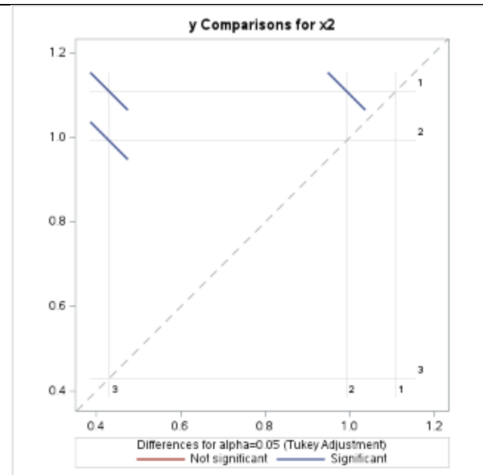
The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey

x2	y LSMEAN	LSMEAN Number
1	1.11000000	1
2	0.99200000	2
3	0.43000000	3

Least Squares Means for effect x2
Pr > |t| for H0: LSMean(i)=LSMean(j)

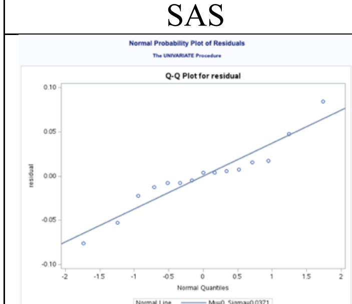
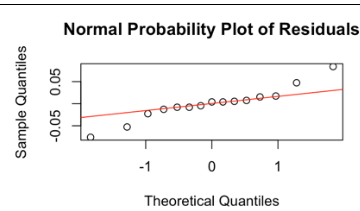
Dependent Variable: y

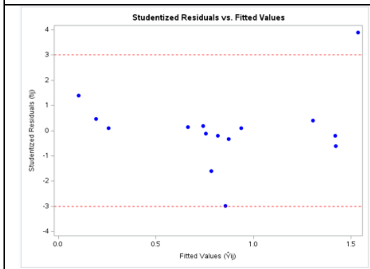
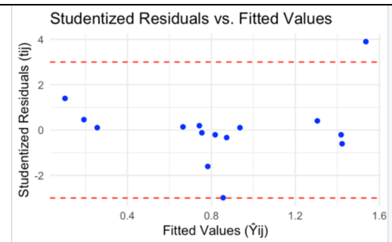
i/j	1	2	3
1		0.0130	<.0001
2	0.0130		<.0001
3	<.0001	<.0001	



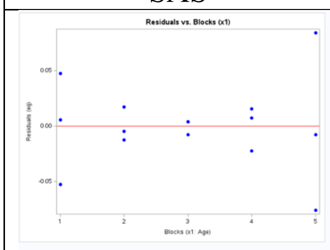
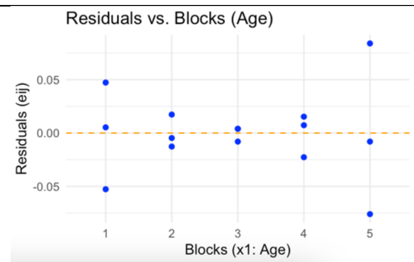
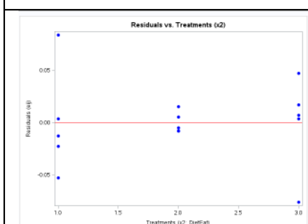
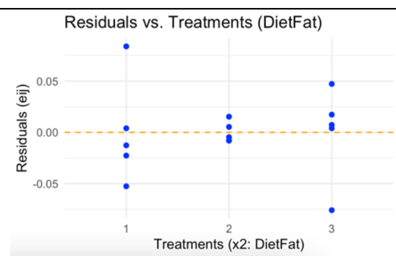
SAS Result interpretation:
The least Squares Means shows the average response(y) for each diet type(x2)
Diet1:1.11, Diet2: 0.992, Diet3: 0.43.
Pair value comparison table :
P-values for Diet1 vs 2=0.013< 0.05→ Significant
P-values for Diet1 vs 3=0.0001< 0.05→ Significant
P-values for Diet2 vs 3=0.0.0001< 0.05→ Significant
Conclusion: All the P-value<0.05, all diet groups are significantly different from each other. Diet 1 has the highest mean response, followed by Diet2, and Diet3 has the lowest. The lineplot confirms all three diet groups differ significantly. The over all FER is conrolled at 5%. The difference remain valid with all comparisons. This suggests diet has a strong effect on y(response), and the impact of different diet levels is significant

5) (6) Examine the standard diagnostic plot and comment on the e_{ij}'s vs. the $\hat{Y}_{ij}$'s, the normal probability plot of the $\{e_{ij}\}$, and the t_{ij}'s vs. the $\hat{Y}_{ij}$'s. Is normality reasonable? Does variance seem roughly constant with the mean? Are there any outliers (e.g. $|t_{ij}| > 3$) ?

SAS	R	Interpretation
		Q-Q plot: We notice that the data points are little off form the 45^0 line. We also notice outliers. We see that the data points are reasonab normally distributed. The Shapiro-wilk test: P-value=0.3135>0.05,

		indicated that the residuals are approximately normally distributed
		We notice a data point $\lvert t_{ij} \rvert > 3$ threshold, indicating outlier. We also see a spread of studentized residuals seems consistent across the range of fitted values, supporting the constant variance assumption.

(7) Finally, prepare plots of the e_{ij} vs. i, and e_{ij} vs. j. Does constant variance seem reasonable across blocks and treatments?

We see constant variance across blocks and treatments, they are reasonable.		
SAS	R	Interpretation
		We don't see any pattern or trend in the residuals across the x1 (age). The spread of residuals appears consistent for all levels of x1, which supports the assumption of constant variance is reasonable across blocks(x1).
		We see no pattern or trend in the residuals across the treatment. The spread of residuals appears consistent for all levels of x2. We see evidence of constant variance is reasonable across treatment(x2).