STAT6302_HW3

Heesu Choi

2025-03-03

Part A

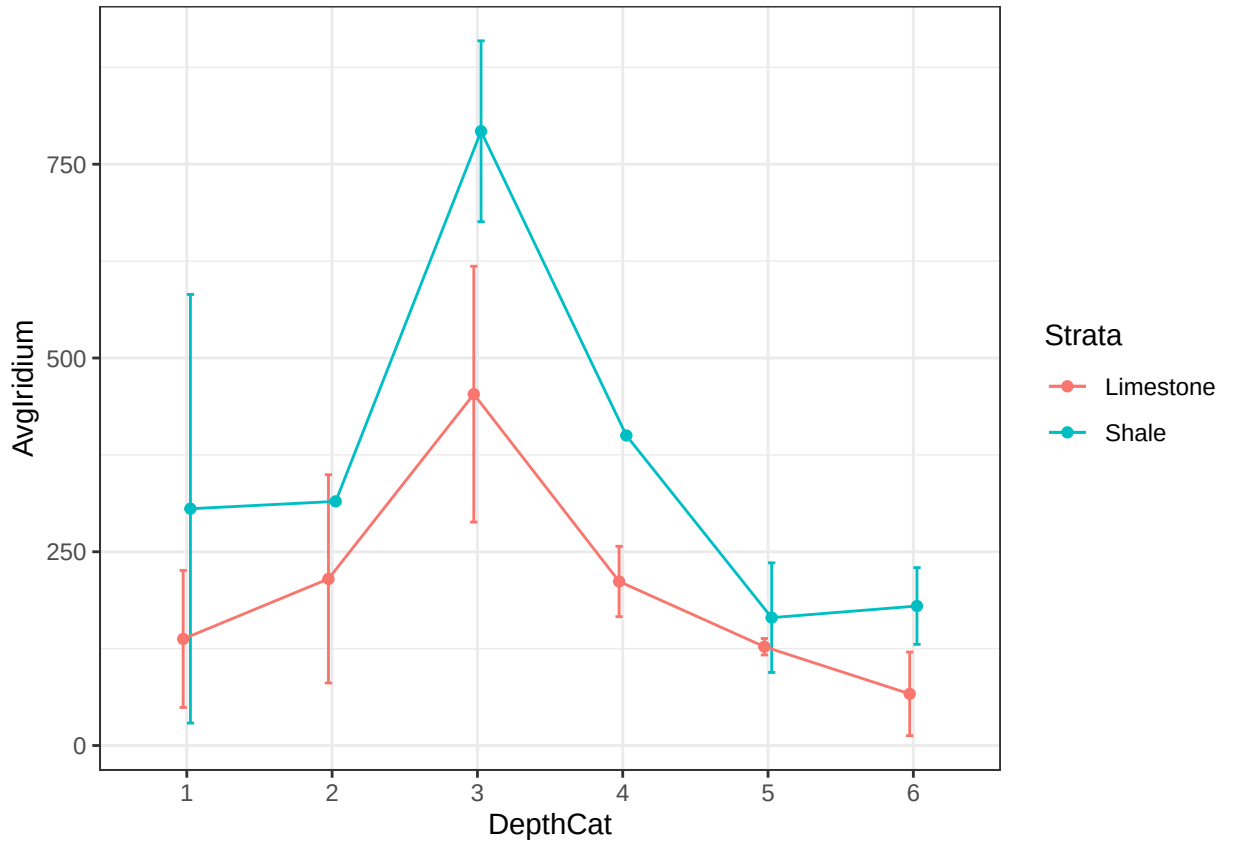
```
head(ex1317)
```

```
##   Iridium    Strata DepthCat
## 1      75 Limestone        1
## 2     200 Limestone        1
## 3     120 Limestone        2
## 4     310 Limestone        2
## 5     290 Limestone        3
## 6     450 Limestone        3
```

```
#reproduce the table and plot
strata.act.summary <- ex1317 %>%
                      group_by(Strata, DepthCat) %>%
                      summarize(N = n(),
                      AvgIridium = mean(Iridium, na.rm = TRUE),
                      SDIridium = sd(Iridium, na.rm = TRUE),
                      SEIridium = SDIridium/sqrt(N),.groups = "drop")

print(strata.act.summary)
```

```
## # A tibble: 12 x 6
##    Strata    DepthCat     N AvgIridium SDIridium SEIridium
##    <fct>     <fct>    <int>      <dbl>     <dbl>     <dbl>
##  1 Limestone 1            2      138.       88.4      62.5
##  2 Limestone 2            2      215       134.       95
##  3 Limestone 3            3      453.      165.       95.3
##  4 Limestone 4            3      212.       45.4      26.2
##  5 Limestone 5            2      128.       10.6       7.5
##  6 Limestone 6            3       66.7      53.9      31.1
##  7 Shale     1            2      306.      276.      196.
##  8 Shale     2            1      315        NA        NA
##  9 Shale     3            2      792.      117.       82.5
## 10 Shale     4            1      400        NA        NA
## 11 Shale     5            5      165        70.7      31.6
## 12 Shale     6            2      180        49.5      35
```

```
ggplot(strata.act.summary, aes(x = DepthCat, y = AvgIridium, group = Strata, color = Strata)) +
      geom_errorbar(aes(ymin = AvgIridium - SDIridium, ymax = AvgIridium + SDIridium),
                    width = 0.1, position = position_dodge(0.1)) +
      geom_line(position = position_dodge(0.1)) +
      geom_point(position = position_dodge(0.1)) +
      theme_bw()
```



Part B

: There is no interaction effect between Limestone and Shale. The lines representing these two groups do not touch or cross, indicating no visible interaction. Additionally, if the graph alone is inconclusive, we should refer to the statistical results. The sample size is too small, and the p-value ($p = 0.4314$) is too large to suggest any significant interaction.

Part C

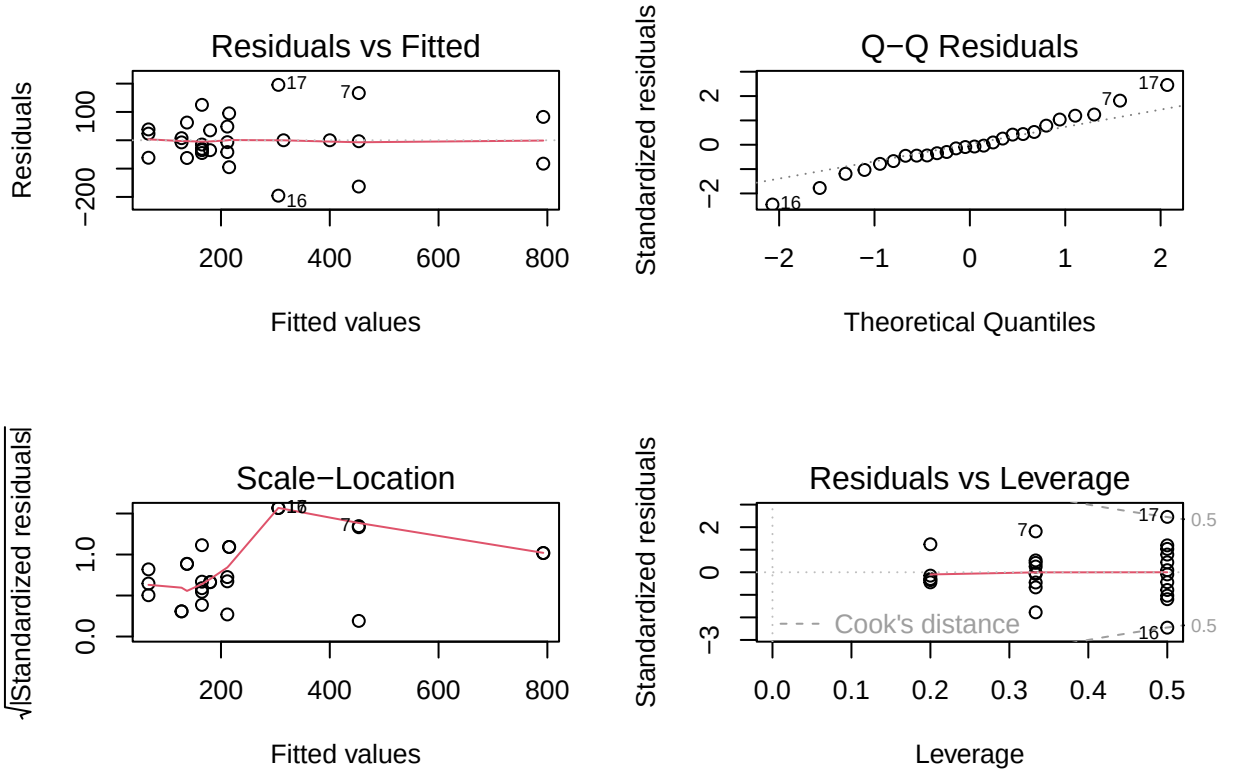
```
strata <- ex1317
strata$Strata <- as.factor(strata$Strata)
strata$DepthCat <- as.factor(strata$DepthCat)
```

```r
strata.anova <- aov(Iridium ~ Strata + DepthCat + Strata*DepthCat, data=strata)
summary(strata.anova)
```

```
##                 Df Sum Sq Mean Sq F value   Pr(>F)
## Strata           1  76407   76407   6.026   0.0259 *
## DepthCat         5 810293  162059  12.781 4.07e-05 ***
## Strata:DepthCat  5  65523   13105   1.033   0.4314
## Residuals       16 202878   12680
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
##To get diagnostic plots, we need to use a regression model
stratalm <- lm(Iridium ~ Strata + DepthCat + Strata*DepthCat, data=strata)
par(mfrow=c(2,2))
plot(stratalm)
```

```
## Warning: not plotting observations with leverage one:
##   18, 21
```




```r
#comparisons
##No need to test for this example since it wasn't significant
##BUT, had the interaction been significant
emmeans(strata.anova, pairwise ~ DepthCat)
```

```
## NOTE: Results may be misleading due to involvement in interactions

## $emmeans
##  DepthCat emmean   SE df lower.CL upper.CL
##  1           222 56.3 16    102.1      341
##  2           265 69.0 16    118.8      411
##  3           623 51.4 16    514.0      732
##  4           306 65.0 16    168.0      444
##  5           146 47.1 16     46.4      246
##  6           123 51.4 16     14.4      232
##
## Results are averaged over the levels of: Strata
## Confidence level used: 0.95
##
## $contrasts
##  contrast              estimate   SE df t.ratio p.value
##  DepthCat1 - DepthCat2    -43.5 89.0 16  -0.489  0.9959
##  DepthCat1 - DepthCat3   -401.4 76.2 16  -5.266  0.0009
##  DepthCat1 - DepthCat4    -84.3 86.0 16  -0.981  0.9176
##  DepthCat1 - DepthCat5     75.2 73.4 16   1.025  0.9027
##  DepthCat1 - DepthCat6     98.2 76.2 16   1.288  0.7873
##  DepthCat2 - DepthCat3   -357.9 86.0 16  -4.162  0.0080
##  DepthCat2 - DepthCat4    -40.8 94.8 16  -0.431  0.9977
##  DepthCat2 - DepthCat5    118.8 83.5 16   1.422  0.7142
##  DepthCat2 - DepthCat6    141.7 86.0 16   1.647  0.5819
##  DepthCat3 - DepthCat4    317.1 82.9 16   3.826  0.0155
##  DepthCat3 - DepthCat5    476.7 69.7 16   6.837  <.0001
##  DepthCat3 - DepthCat6    499.6 72.7 16   6.873  <.0001
##  DepthCat4 - DepthCat5    159.6 80.3 16   1.988  0.3902
##  DepthCat4 - DepthCat6    182.5 82.9 16   2.202  0.2892
##  DepthCat5 - DepthCat6     22.9 69.7 16   0.329  0.9994
##
## Results are averaged over the levels of: Strata
## P value adjustment: tukey method for comparing a family of 6 estimates
```

```
emmeans(strata.anova, pairwise ~ Strata)
```

```
## NOTE: Results may be misleading due to involvement in interactions

## $emmeans
##  Strata    emmean   SE df lower.CL upper.CL
##  Limestone    202 29.7 16      139      265
##  Shale        360 36.1 16      283      436
##
## Results are averaged over the levels of: DepthCat
## Confidence level used: 0.95
##
## $contrasts
##  contrast          estimate   SE df t.ratio p.value
##  Limestone - Shale     -158 46.7 16  -3.375  0.0039
##
## Results are averaged over the levels of: DepthCat
```

```
knitr::include_graphics("C:/Users/choih/OneDrive/Desktop/STAT6302/hw4_1.jpg")
```

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Strata	1	16	11.39	0.0039
DepthCat	5	16	12.63	<.0001
Strata*DepthCat	5	16	1.03	0.4314

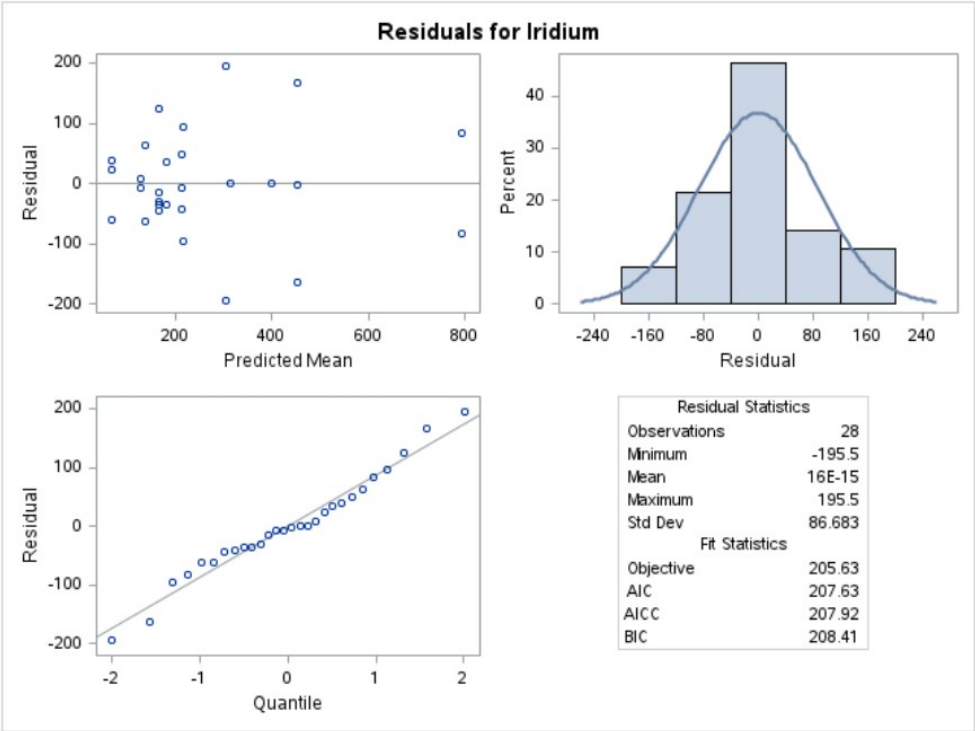
Least Squares Means

Effect	Strata	DepthCat	Estimate	Standard Error	DF	t Value	Pr > |t|
DepthCat		1	221.50	56.3025	16	3.93	0.0012
DepthCat		2	265.00	68.9562	16	3.84	0.0014
DepthCat		3	622.92	51.3969	16	12.12	<.0001
DepthCat		4	305.83	65.0125	16	4.70	0.0002
DepthCat		5	146.25	47.1060	16	3.10	0.0068
DepthCat		6	123.33	51.3969	16	2.40	0.0289
Strata	Limestone		201.94	29.6740	16	6.81	<.0001
Strata	Shale		359.67	36.1000	16	9.96	<.0001

Differences of Least Squares Means

Effect	Strata	DepthCat	_Strata	_DepthCat	Estimate	Standard Error	DF	t Value	Pr > |t|	Adjustment	Adj P
DepthCat		1		2	-43.5000	89.0220	16	-0.49	0.6317	Tukey-Kramer	0.9959
DepthCat		1		3	-401.42	76.2339	16	-5.27	<.0001	Tukey-Kramer	0.0009
DepthCat		1		4	-84.3333	86.0035	16	-0.98	0.3414	Tukey-Kramer	0.9176
DepthCat		1		5	75.2500	73.4094	16	1.03	0.3206	Tukey-Kramer	0.9027
DepthCat		1		6	98.1667	76.2339	16	1.29	0.2162	Tukey-Kramer	0.7873
DepthCat		2		3	-357.92	86.0035	16	-4.16	0.0007	Tukey-Kramer	0.0080
DepthCat		2		4	-40.8333	94.7712	16	-0.43	0.6723	Tukey-Kramer	0.9977
DepthCat		2		5	118.75	83.5101	16	1.42	0.1742	Tukey-Kramer	0.7142
DepthCat		2		6	141.67	86.0035	16	1.65	0.1190	Tukey-Kramer	0.5819
DepthCat		3		4	317.08	82.8750	16	3.83	0.0015	Tukey-Kramer	0.0155
DepthCat		3		5	476.67	69.7181	16	6.84	<.0001	Tukey-Kramer	<.0001
DepthCat		3		6	499.58	72.6862	16	6.87	<.0001	Tukey-Kramer	<.0001
DepthCat		4		5	159.58	80.2845	16	1.99	0.0642	Tukey-Kramer	0.3902
DepthCat		4		6	182.50	82.8750	16	2.20	0.0427	Tukey-Kramer	0.2892
DepthCat		5		6	22.9167	69.7181	16	0.33	0.7466	Tukey-Kramer	0.9994
Strata	Limestone		Shale		-157.72	46.7307	16	-3.38	0.0039	Bonferroni	0.0039

```
knitr::include_graphics("C:/Users/choih/OneDrive/Desktop/STAT6302/hw4_2.jpg")
```




```
knitr::include_graphics("C:/Users/choih/OneDrive/Desktop/STAT6302/hw4_3.jpg")
```

```
PROC IMPORT OUT=newdata
    DATAFILE="/home/u64010957/strata_data.csv"
    DBMS=CSV
    REPLACE;
    GETNAMES=NO;
RUN;

proc mixed data = IMPORT2 PLOTS = (RESIDUALPANEL);
class Strata DepthCat;
model Iridium = Strata DepthCat Strata*DepthCat;

lsmeans DepthCat / pdiff tdiff adjust=tukey;
lsmeans Strata / pdiff tdiff adjust=bon;
run;
```

Part D

: The residual diagnostics indicate that the model assumptions are not fully met for three key reasons. First, the residuals exhibit heteroscedasticity, meaning their spread is not constant across the fitted values. Second,

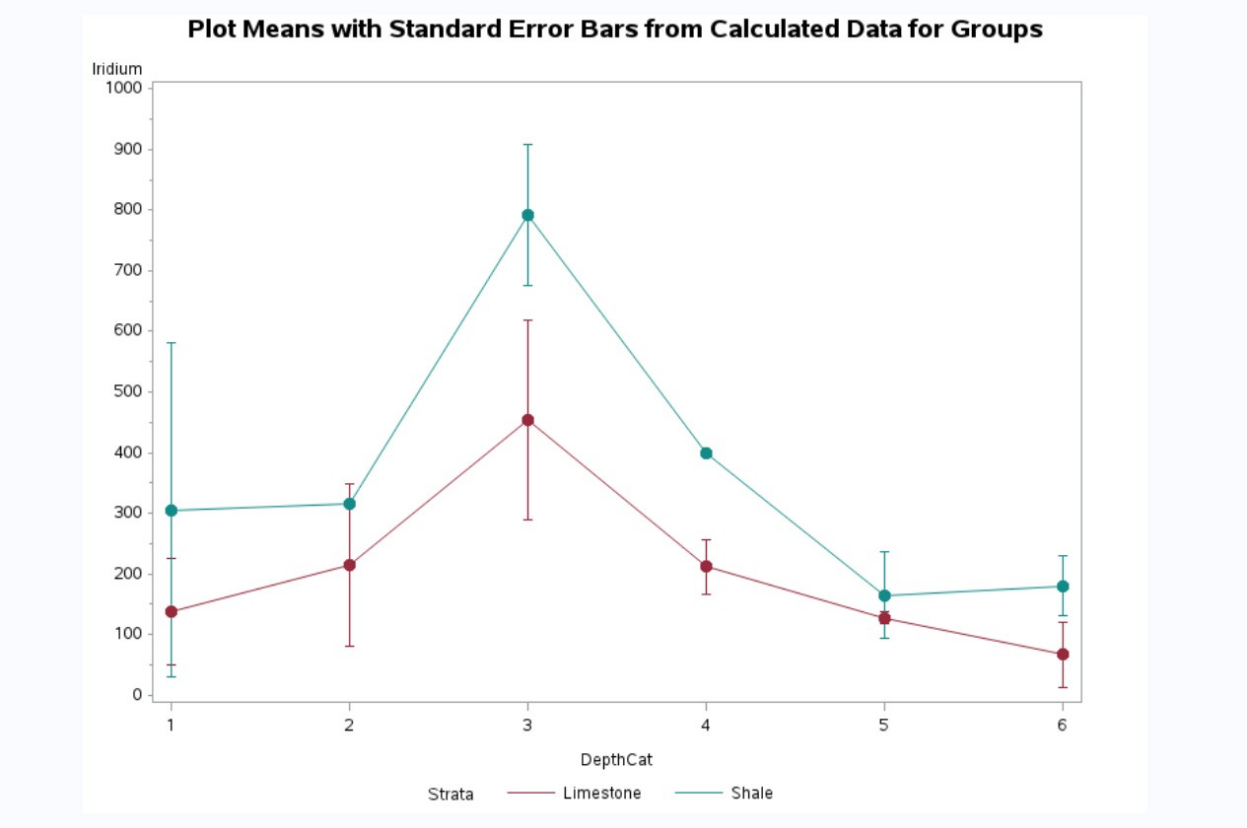
the QQ plot shows deviations from the expected straight line, suggesting that the normality assumption is violated. Lastly, there is a potentially influential outlier, observation "17," which may be impacting the model's performance.

Part E

: A two-way ANOVA was conducted to examine the effects of "DepthCat" and "Strata" on "AvgIridium". The analysis revealed a significant main effect of "DepthCat" ($F(5, 16) = 12.781$, $p < 0.0001$), indicating that the average Iridium levels differed significantly across the six DepthCat categories. There was also a significant main effect of "Strata" ($F(1, 16) = 6.026$, $p = 0.0259$), suggesting that the average Iridium levels differed between the Limestone and Shale strata. However, the interaction between "DepthCat" and "Strata" was not statistically significant ($F(5, 16) = 1.033$, $p = 0.4314$), meaning that the effect of DepthCat on average Iridium levels was consistent across both Limestone and Shale strata. The comparison revealed several statistically significant differences between specific DepthCat categories. DepthCat 3 exhibited significantly higher average Iridium levels compared to DepthCat 1 ($p = 0.0009$), DepthCat 2 ($p = 0.0080$), DepthCat 5 ($p < 0.0001$), and DepthCat 6 ($p < 0.0001$). These results suggest that DepthCat 3 consistently demonstrates a substantially higher average Iridium level compared to several other categories

Part F

```
knitr::include_graphics("C:/Users/choih/OneDrive/Desktop/STAT6302/hw4_4.jpg")
```



```
knitr::include_graphics("C:/Users/choih/OneDrive/Desktop/STAT6302/hw4_5.jpg")
```

```
proc means data=IMPORT2 n mean max min range std fw=8;
class DepthCat Strata; var Iridium;
output out=meansout mean=mean std=std;
title 'Summary of Strata DepthCat Iridium';
run;


data summarystrata;
set meansout;
if _TYPE_=0 then delete;
if _TYPE_=1 then delete;
if _TYPE_=2 then delete;
run;

data plottingdata(keep= DepthCat Strata mean std newvar);
set summarystrata;
by DepthCat Strata;
newvar=mean;
output;
newvar=mean - std; output;
newvar=mean + std; output;
run;

title1 'Plot Means with Standard Error Bars from Calculated Data for Groups';
symbol1 interpol=hiloctj color=depk line=1;
symbol2 interpol=hiloctj color=vibg line=1;
symbol3 interpol=none color=depk value=dot height=1.5;
symbol4 interpol=none color=vibg value=dot height=1.5;

axis1 offset=(2,2) ;
axis2 label=("Iridium") order=(0 to 1000 by 100) minor=(n=1);


proc sort data=plottingdata;
by DepthCat;
run;

proc gplot data=plottingdata;
plot NewVar*DepthCat = Strata / vaxis=axis2 haxis=axis1;
plot2 Mean*DepthCat = Strata / vaxis=axis2 noaxis nolegend;

run;
quit;
```

ome/u64010957/HW4Part_a_.sas~

Part G

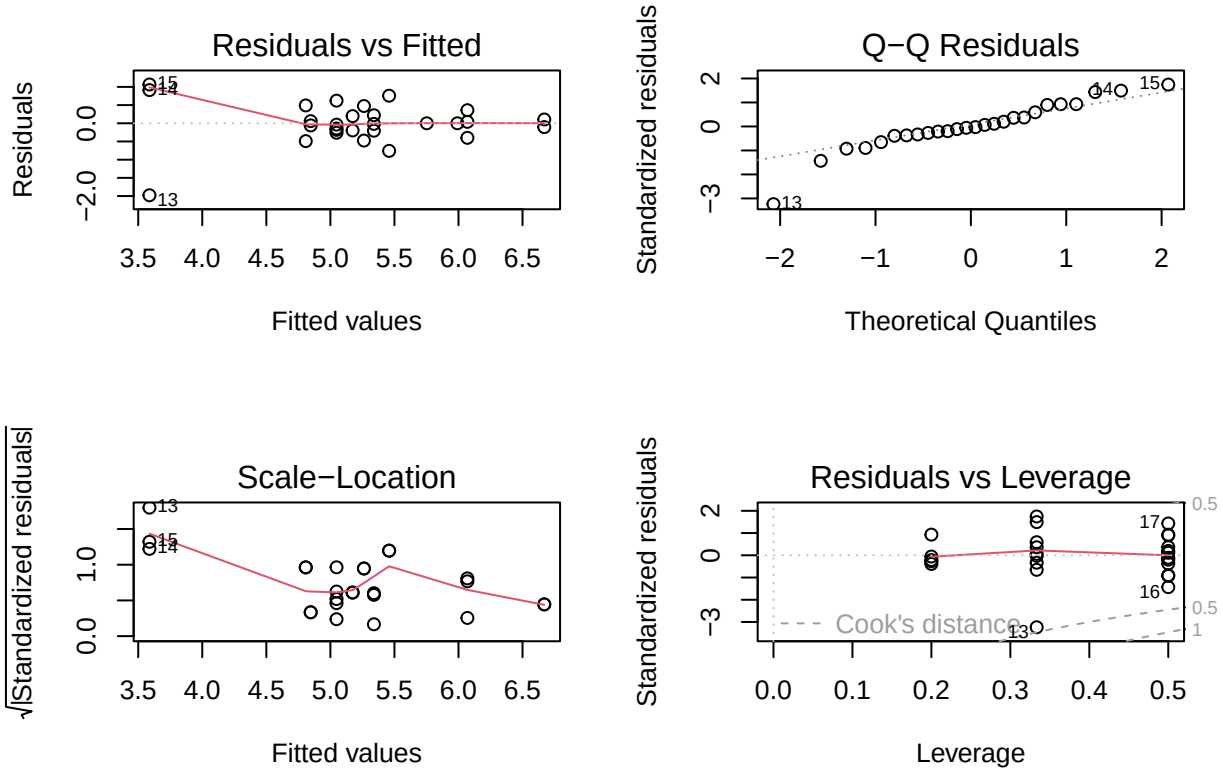
```
# Assuming 'strata' is your dataset
strata$Strata <- as.factor(strata$Strata)
strata$DepthCat <- as.factor(strata$DepthCat)

# Apply log transformation to Iridium
strata$log_Iridium <- log(strata$Iridium)
```

```r
# Run the ANOVA on log-transformed outcome
strata.anova_log <- aov(log_Iridium ~ Strata + DepthCat + Strata*DepthCat, data=strata)
summary(strata.anova_log)
```

```
##                 Df Sum Sq Mean Sq F value Pr(>F)
## Strata           1  1.873  1.8727   3.344 0.0862 .
## DepthCat         5 12.955  2.5910   4.627 0.0084 **
## Strata:DepthCat  5  1.342  0.2684   0.479 0.7866
## Residuals       16  8.960  0.5600
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# To get diagnostic plots, use a regression model
stratalm_log <- lm(log_Iridium ~ Strata + DepthCat + Strata*DepthCat, data=strata)
par(mfrow=c(2,2))  # Set up the plot area for 4 plots
plot(stratalm_log)  # Create diagnostic plots for log-transformed data
```

```
## Warning: not plotting observations with leverage one:
##   18, 21
```




```r
# Perform pairwise comparisons if interaction is significant
# No need to test for this example if the interaction wasn't significant
emmeans(strata.anova_log, pairwise ~ DepthCat)
```

```
## NOTE: Results may be misleading due to involvement in interactions

## $emmeans
##  DepthCat emmean    SE df lower.CL upper.CL
##  1          5.13 0.374 16     4.34     5.93
##  2          5.51 0.458 16     4.54     6.48
##  3          6.37 0.342 16     5.65     7.09
##  4          5.67 0.432 16     4.75     6.58
##  5          4.95 0.313 16     4.28     5.61
##  6          4.38 0.342 16     3.66     5.10
##
## Results are averaged over the levels of: Strata
## Confidence level used: 0.95
##
## $contrasts
##  contrast              estimate    SE df t.ratio p.value
##  DepthCat1 - DepthCat2   -0.374 0.592 16  -0.632  0.9867
##  DepthCat1 - DepthCat3   -1.236 0.507 16  -2.441  0.2001
##  DepthCat1 - DepthCat4   -0.532 0.572 16  -0.932  0.9324
##  DepthCat1 - DepthCat5    0.186 0.488 16   0.381  0.9987
##  DepthCat1 - DepthCat6    0.753 0.507 16   1.485  0.6777
##  DepthCat2 - DepthCat3   -0.862 0.572 16  -1.509  0.6639
##  DepthCat2 - DepthCat4   -0.158 0.630 16  -0.251  0.9998
##  DepthCat2 - DepthCat5    0.560 0.555 16   1.009  0.9082
##  DepthCat2 - DepthCat6    1.127 0.572 16   1.971  0.3988
##  DepthCat3 - DepthCat4    0.704 0.551 16   1.278  0.7922
##  DepthCat3 - DepthCat5    1.422 0.463 16   3.070  0.0665
##  DepthCat3 - DepthCat6    1.989 0.483 16   4.118  0.0087
##  DepthCat4 - DepthCat5    0.718 0.534 16   1.346  0.7562
##  DepthCat4 - DepthCat6    1.285 0.551 16   2.333  0.2373
##  DepthCat5 - DepthCat6    0.567 0.463 16   1.223  0.8197
##
## Results are averaged over the levels of: Strata
## P value adjustment: tukey method for comparing a family of 6 estimates
```

```
emmeans(strata.anova_log, pairwise ~ Strata)
```

```
## NOTE: Results may be misleading due to involvement in interactions

## $emmeans
##  Strata    emmean    SE df lower.CL upper.CL
##  Limestone   4.99 0.197 16     4.57     5.40
##  Shale       5.68 0.240 16     5.17     6.19
##
## Results are averaged over the levels of: DepthCat
## Confidence level used: 0.95
##
## $contrasts
##  contrast          estimate    SE df t.ratio p.value
##  Limestone - Shale   -0.697 0.311 16  -2.244  0.0394
##
## Results are averaged over the levels of: DepthCat
```

: After transformation data, the residuals and leverage are a bit better after the transformation but the Q-Q plot still shows deviations from normality