```
/* Montgomery 6.28 */
proc import datafile="/home/u63048916/STAT571B/Homework/Homework 5/Q6-28.xlsx"
    dbms=xlsx
    out=mont6_28
    replace;
    getnames=yes;
run;

data inter;                              /* Define Interaction Terms */
 set mont6_28;
 AB=A*B; AC=A*C; AD=A*D; BC=B*C; BD=B*D; CD=C*D; ABC=AB*C; ABD=AB*D;
 ACD=AC*D; BCD=BC*D; block=ABC*D; y=Yield;
run;

proc glm data=inter;                   /* GLM Proc to Obtain Effects */
 class A B C D AB AC AD BC BD CD ABC ABD ACD BCD block;
 model y=block A B C D AB AC AD BC BD CD ABC ABD ACD BCD;
 run;


proc reg outest=effects data=inter;    /* REG Proc to Obtain Effects */
 model y=A B C D AB AC AD BC BD CD ABC ABD ACD BCD block;
 run;

 proc print data=effects;
 run;

data effect2; set effects;
 drop y intercept _RMSE_;
 run;

proc transpose data=effect2 out=effect3;
run;

data effect4; set effect3; effect=col1*2;
run;

proc sort data=effect4; by effect;
run;

proc print data=effect4;
run;

proc rank data=effect4 out=effect5 normal=blom;
 var effect;
ranks neff;
 run;

proc sgplot data=effect5;
scatter x=neff y=effect/datalabel=_NAME_;
xaxis label='Normal Scores';
run;

/*%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%*/

proc glm data=inter;
class A C D AC AD;
model y= A C D AC AD;   /* use A*B if you want to generate an interaction plot */
estimate "A" A -1 1;
estimate "C" C -1 1;
estimate "D" D -1 1;
estimate "AC" AC -1 1;
estimate "AD" AD -1 1;
output out=diag r=res p=pred;
run;


proc reg data=inter;
model y= A C D AC AD;
output out=outres r=res p=pred;
run;

/* check normality */
proc univariate data=outres normal;
var res;
qqplot res / normal (mu=est sigma=est);
Run;

/* check constant variance using graph*/
title 'residual plot: res vs predicted value ';
proc sgplot data=outres;
scatter x=pred y=res;
refline 0;
run;

title 'residual plot: res vs A (h) ';
proc sgplot data=outres;
scatter x=A y=res;
refline 0;
run;

title 'residual plot: res vs C(psi) ';
proc sgplot data=outres;
scatter x=C y=res;
refline 0;
run;

title 'residual plot: res vs D(°C) ';
proc sgplot data=outres;
scatter x=D y=res;
refline 0;
run;
```

6.28. In a process development study on yield, four factors were studied, each at two levels: time (A), concentration (B), pressure (C), and temperature (D). A single replicate of a 2^4 design was run, and the resulting data are shown in Table P6.7.

(a) Construct a normal probability plot of the effect estimates. Which factors appear to have large effects?

(b) Conduct an analysis of variance using the normal probability plot in part (a) for guidance in forming an error term. What are your conclusions?

(c) Write down a regression model relating yield to the important process variables.

(d) Analyze the residuals from this experiment. Does your analysis indicate any potential problems?

(e) Can this design be collapsed into a 2^3 design with two replicates? If so, sketch the design with the average and range of yield shown at each point in the cube. Interpret the results.

TABLE P6.7
Process Development Experiment from Problem 6.28

Run Number	Actual Run Order	A	B	C	D	Yield (lbs)	Factor Levels	Low (−)	High (+)
1	5	−	−	−	−	12	A (h)	2.5	3
2	9	+	−	−	−	18	B (%)	14	18
3	8	−	+	−	−	13	C (psi)	60	80
4	13	+	+	−	−	16	D (°C)	225	250
5	3	−	−	+	−	17			
6	7	+	−	+	−	15			
7	14	−	+	+	−	20			
8	1	+	+	+	−	15			
9	6	−	−	−	+	10			
10	11	+	−	−	+	25			
11	2	−	+	−	+	13			
12	15	+	+	−	+	24			
13	4	−	−	+	+	19			
14	16	+	−	+	+	21			
15	10	−	+	+	+	17			
16	12	+	+	+	+	23			

```
title 'residual plot: res vs AC ';
proc sgplot data=outres;
scatter x=AC y=res;
refline 0;
run;

title 'residual plot: res vs AD ';
proc sgplot data=outres;
scatter x=AD y=res;
refline 0;
run;
```

a.) See figure 6.28.1, the normal probability plot indicates that the factors that are significant (deviate visibily from line through data near zero) are A, C, D, and interactions AC and AD.

b.) See figure 6.28.2 and 6.28.3, the ANOVA table indicates that the model as well as

each factor A, C, D, AC, AD are significant at an $\alpha = 0.05$ level.

c.) The regression equation is of the form:

$$y = intercept + \frac{A}{2}x_1 + \frac{C}{2}x_3 + \frac{D}{2}x_4 + \frac{AC}{2}x_1x_3 + \frac{AD}{2}x_1x_4$$

coefficients from figure 6.28.8 are half of the factor effects from a.) with $intercept = 17.375$. The estimated regression equation is

$$y = 17.375 - 2.25x_1 + x_3 + 1.625x_4 - 2.125x_1x_3 + 2x_1x_4$$

The factor estimates and subsequent model coefficient estimates are coded factors, so x_1, x_3, and x_4 can be subbed by the below factors to utilize actual Flow rate and Deoposition time data values:

$$x_1 = \frac{Time - (3 + 2.5)/2}{(3 - 2.5)/2} = \frac{Flow - 2.75}{0.25}$$

$$x_3 = \frac{Pressure - (80 + 60)/2}{(80 - 60)/2} = \frac{Pressure - 70}{10}$$

$$x_4 = \frac{Temperature - (250 + 225)/2}{(250 - 225)/2} = \frac{Pressure - 237.5}{12.5}$$

d.) See figures 6.28.4-6.28.7. The normality diagnostics indicate no issues with the normality of the resiudals, which is supported by the QQ plot. The residual plots in figures 6.28.6 and 6.28.7 show no patterns that would indicate deviation from constant variance assumption, so the model is appropriate.