```
/* Montgomery 8.1 */
proc import datafile="/home/u63048916/STAT571B/Homework/Homework 6/Q8-1repl.xlsx"
    dbms=xlsx
    out=mont8_1
    replace;
    getnames=yes;
run;

data inter;
    set mont8_1;
        A=A;
        B=B;
        C=C;
        D=D;
        AB=A*B;
        AC=A*C;
        AD=A*D;
        y=Yield;
    run;

proc glm data=inter;                    /* GLM Proc to Obtain Effects */
 class A B C D AB AC AD;
 model y=A B C D AB AC AD;
 estimate 'A' A 1 -1;  estimate 'B' B 1 -1;
  estimate 'C' C 1 -1;  estimate 'D' B 1 -1;
estimate 'AB' AB 1 -1;
estimate 'AC' AC 1 -1;
estimate 'AD' AD 1 -1;
 run;

 proc reg outest=effects data=inter;    /* REG Proc to Obtain Effects */
 model y=A B C D AB AC AD;
 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;
heffect=abs(effect);
run;

/* draw normal probablity plot */
proc sort data=effect4; by effect;
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;

/*fidelity with half normal prob plot*/

/* draw half normal probablity plot */
proc sort data=effect4; by heffect;
run;

proc rank data=effect4 out=hnranks;
var heffect;
ranks hneffect;
run;

data hnormals;
set hnranks nobs=n;
hneff=probit(((hneffect-1/3)/(n+1/3))/2+.5); /* calculate half normal scores */
run;

proc print data=hnormals;
run;

title 'Half Normal Probability Plot';
proc sgplot data=hnormals;
scatter x=hneff y=heffect/datalabel=_NAME_;
xaxis label='Half Normal Score';
run;

/* run ANOVA model on the selected terms */
/* terms A,C,AB,AD of interest, include B and D in model for hierarchy principle */

title 'Selected factor analysis';
proc glm data=inter;                    /* GLM Proc to Obtain Effects */
 class A B C D AB AD;
 model y=A B C D AB AD;
run;

proc reg data=inter;
model y=A B C D AB AD;
output out=two r=res1 p=pred1;
run;
```

8.1. Suppose that in the chemical process development experiment described in Problem 6.7, it was only possible to run a one-half fraction of the 2^4 design. Construct the design and perform the statistical analysis, using the data from replicate I.

---

3. **Montgomery 8.1**
We are refining the data from 6.7 to create a half-fractional data set to analyze. The resulting data is as follows:

A	B	C	D	Yield
-1	-1	-1	-1	90
1	1	-1	-1	83
1	-1	1	-1	81
-1	1	1	-1	88
1	-1	-1	1	72
-1	1	-1	1	87
-1	-1	1	1	99
1	1	1	1	80

6.7. An experiment was performed to improve the yield of a chemical process. Four factors were selected, and two replicates of a completely randomized experiment were run. The results are shown in the following table:

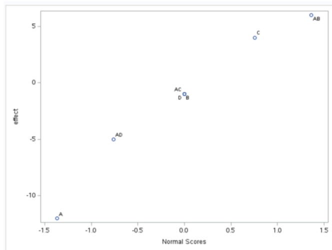
Treatment Combination	Replicate I	Replicate II	Treatment Combination	Replicate I	Replicate II
(1)	90	93	*d*	98	95
a	74	78	*ad*	72	76
b	81	85	*bd*	87	83
ab	83	80	*abd*	85	86
c	77	78	*cd*	99	90
ac	81	80	*acd*	79	75
bc	88	82	*bcd*	87	84
abc	73	70	*abcd*	80	80

```
proc univariate data=two normal;
var res1;
qqplot;
run;
/* check constant variance using graph*/
title 'residual plot: res vs predicted value ';
proc sgplot data=two;
scatter x=pred1 y=res1;
refline 0;
run;

proc reg data=inter;
model y=A;
output out=three r=res2 p=pred2;
run;

proc univariate data=three normal;
var res2;
qqplot;
run;
/* check constant variance using graph*/
title 'residual plot: res vs predicted value ';
proc sgplot data=three;
scatter x=pred2 y=res2;
refline 0;
run;
```

Full Model NPP



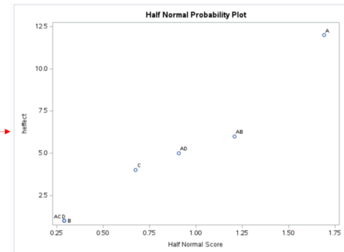


Figure 8.1.4

Our first step was to generate the Normal Probability Plot for the model with all factors (Figure 8.1.1) to aid in identifying the possible significant factors. In order to better identify the important effects we rerun the data with a Half Normal Probability plot. The potentially significant factors identified were A, C, AB, and AD. In order to account for hierarchy principle, we run an analysis of a refined model using the following factors: A, B, C, D, AB, and AD. The ANOVA table in Figure 8.1.2 indicates only factor A is close to significant (not actually significant at $\alpha = 0.05$ level), but our normality diagnostics (Figure 8.1.3) indicate there are problems with the normality assumptions and thus the model is not useful for inference.

Given that A appears closest to significance, we rerun the model using only factor A, which we find is significant at $\alpha = 0.05$ level with P value of 0.0167 (Figure 8.1.2), and normality and constant variance assumptions are acceptable by diagnostics in Figure 8.1.4.

The parameter estimate for factor $A$is -6.000 with an intercept of 85, so the final model should be $yield = 85 - 6A$.

ANOVA

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	446.00000	74.33333	37.17	0.1249
Error	1	2.00000	2.00000		
Corrected Total	7	448.00000			

Root MSE	1.41421	R-Square	0.9955
Dependent Mean	85.00000	Adj R-Sq	0.9688
Coeff Var	1.66378		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	85.00000	0.50000	170.00	0.0037
A	A	1	-6.00000	0.50000	-12.00	0.0529
B	B	1	-0.50000	0.50000	-1.00	0.5000
C	C	1	2.00000	0.50000	4.00	0.1560
D	D	1	-0.50000	0.50000	-1.00	0.5000
AB		1	3.00000	0.50000	6.00	0.1051
AD		1	-2.50000	0.50000	-5.00	0.1257

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	288.00000	288.00000	10.80	0.0167
Error	6	160.00000	26.66667		
Corrected Total	7	448.00000			

Root MSE	5.16398	R-Square	0.6429
Dependent Mean	85.00000	Adj R-Sq	0.5833
Coeff Var	6.07527		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	85.00000	1.82574	46.56	<.0001
A	A	1	-6.00000	1.82574	-3.29	0.0167