Problem 1

a)State an appropriate model and its assumptions.

We can have a two-factor effect model .

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

Where :

$\mu_{..}$ is a constant.

ρ_i are constants for the block(row) effects , subject to the restriction $\sum \rho_i$ =0 .

τ_j are constants for the treatment effects , subject to the restriction $\sum \tau_j$ =0 .

ε_{ij} are independent N(0, σ^2)

i=1...5, j=1...4.

b)Test whether or not the main effect of variety is present.

H_0: All τ_j are equeal 0

H_a: Not all τ_j are equeal 0

$$F^* = \frac{MSTR}{MSBL.TR} = 127.60$$

F(0.95; 3, 12) = 3.49 < F^*

Then, we can conclude Ha, the main effect is present. The p_value is less than 0.0001.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
variety	3	154.4920000	51.4973333	127.60	<.0001
block	4	12.3930000	3.0982500	7.68	0.0026

```
proc glm data= Plants alpha=0.05;

 class variety block;

 model height = variety block;

 lsmeans variety block ;
```

```
run;

quit;
```

C) Obtain confidence interval for all pairwise comparison between the variety means; use the most efficient multiple comparison procedure with a 90%family confidence coefficient.Interpret your results.

Since, Tukey procedure is the most efficient in this problem.
Then we can get

Least Squares Means for Effect variety				
i	j	Difference Between Means	Simultaneous 90% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-3.400000	-4.428660	-2.371340
1	3	2.220000	1.191340	3.248660
1	4	4.060000	3.031340	5.088660
2	3	5.620000	4.591340	6.648660
2	4	7.460000	6.431340	8.488660
3	4	1.840000	0.811340	2.868660

Since all the pairwise comparison doesn't include 0 , it indicates that they are all significant.

```
proc glm data= Plants;

 class variety block;

 model height = variety block;

 lsmeans variety block /pdiff cl adjust=tukey alpha=0.1;

 estimate 'mu.1 vs. mu.2'   variety 1 -1  0 0;

 estimate 'mu.1 vs. mu.3'   variety 1  0 -1 0;

 estimate 'mu.1 vs. mu.4'   variety 1  0  0 -1;

 estimate 'mu.2 vs. mu.3'   variety 0  1 -1 0;

 estimate 'mu.2 vs. mu.4'   variety 0  1  0 -1;

 estimate 'mu.3 vs. mu.4'   variety 0  0  1 -1 ;
```

```
 output out=out p=p stdp = stdp uclm=uclm lclm=lclm;
run;
quit;
```

d) estimate the difference in the mean plant height for the first two groups of variety with 95% C.I. Interpret your results.

Parameter	Estimate	Standard Error	t Value	Pr > |t|
mu.1 vs. mu.2	-3.40000000	0.40178767	-8.46	<.0001

Parameter	95% Confidence Limits	
mu_.1 - mu_.2	-4.27542013	-2.52457987

It is significant , since zero is not in the C.I .

```
proc glm data= Plants;
 class variety block;
 model height = variety block;
 lsmeans variety block /pdiff cl adjust=bon alpha=0.05;
 estimate 'mu.1 vs. mu.2'   variety 1 -1  0 0;
 run;
```

e)Test for Ho: $\mu_{.1} = \mu_{.2}$ VS Ha: $\mu_{.1} \neq \mu_{.2}$, with α =0.05. State the test statistics ,p-value and your conclusion.Doses your conclusion agree with the result of (d)? Explain.

Since $t^* = \frac{\mu_{.1} - \mu_{.2}}{Sd} \sim t(0.975, 12)$ and t(0.975,8)=-2.178

By t test if $|t^*| \leq t(0.975, 12)$ conclude H0

If $|t^*| \geq t(0.975, 12)$ conclude Ha

Since t^* = -8.46

Thus , we conclude Ha, the p-value is less than 0.0001.
It is agree with the conclusion in d.

Parameter	Estimate	Standard Error	t Value	Pr > |t|
mu.1 vs. mu.2	-3.40000000	0.40178767	-8.46	<.0001

f)Test for H0: $\mu_{.1} \geq \mu_{.2}$ VS Ha: $\mu_{.1} \leq \mu_{.2}$ with α =0.025.

It is a one-side test.

Since $t^* = \dfrac{\mu_{.1} - \mu_{.2}}{Sd} \sim t(0.975,12)$ and t(0.975,8)=-2.178

By t test if $\left|t^*\right| \leq t(0.975,12)$ reject H0

If $\left|t^*\right| \geq t(0.975,12)$ reject Ha

Since t^* =-8.46

Thus , we conclude Ha, the p-value is less than 0.0001.

g) Comment on the efficiency of the blocking variable.

$E = \dfrac{\sigma_r^2}{\sigma_b^2}$ then we can get

$$E = \frac{(n_b - 1)MSBL + n_b(r-1)MSBL.TR}{(n_b r - 1)MSBL.TR} = \frac{4*3.09825 + 5*3*0.4035}{(5*4-1)0.4035}$$

Thus E=2.40598 .It means the MSBL>MSBL.TR , since it is greater than 1.
And it indicates that about 2.3 times the replications need be with a completely randomized design as compared to a randomized complete block design in order that the variance of any estimated treatment contrast be the same for the two designs.

Problem 2

a)State an appropriate model and its assumptions.

$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + \varepsilon_{ijk}$

Where:

$\mu_{..}$ is a constant.

α_i are independent N(0,σ^2) , it is treatment .

β_j are constants for the block effects , subject to the restriction $\sum \beta_j$ =0 .

ε_{ij} are independent N(0,σ^2)

i=1...4, j=1...5.

b) ANOVA table

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	166.8850000	23.8407143	59.07	<.0001
Error	12	4.8430000	0.4035833		
Corrected Total	19	171.7280000			

R-Square	Coeff Var	Root MSE	height Mean
0.971798	9.397664	0.635282	6.760000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
variety	3	154.4920000	51.4973333	127.60	<.0001
block	4	12.3930000	3.0982500	7.68	0.0026

```
proc glm data= Plants;

class variety block;

model height = variety block/ss3;

random variety ;

lsmeans variety block;
```

```
run;
```

C) Test whether there are difference in plant heights between varieties.State the hypotheses ,test statistics ,p-value and conclusion.

H_0: All α_i are equeal 0

H_a: Not all α_i are equeal 0

$$F^* = \frac{MSTR}{MSBL.TR} = 127.60$$

$F(0.95; 3, 12) = 3.49 < F^*$

Then, we can conclude Ha, the main effect is present. The p_value is less than 0.0001.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
variety	3	154.4920000	51.4973333	127.60	<.0001
block	4	12.3930000	3.0982500	7.68	0.0026

```
proc glm data= Plants;
 class variety block;
 model height = variety block/ss3;
 random variety ;
 lsmeans variety block;
 run;
```

Problem 3

a)Write down an appropriate model and assumptions.Check the appropriateness of the model.

$$Y_{ij} = \mu_{.} + \tau_i + \gamma(X_{ij} - \bar{X}_{..}) + \varepsilon_{ij}$$

$\mu_{.}$ is an overall mean

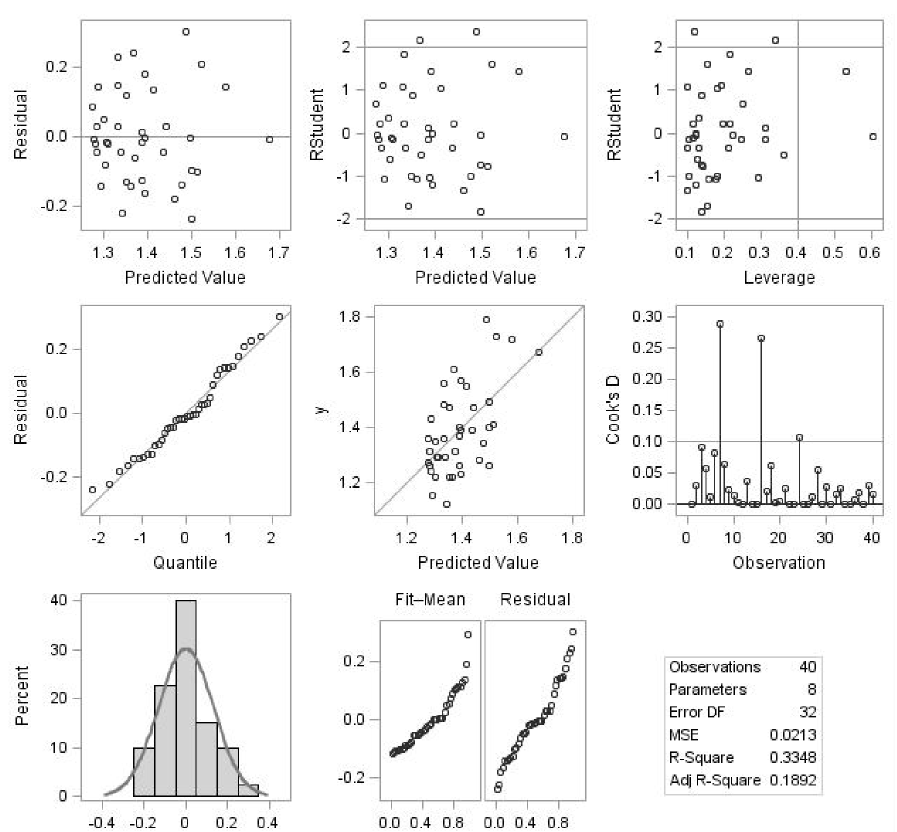
τ_i are the fixed treatment effects subject to the restriction $\sum \tau_j$ =0

γ is a regression coefficient for the relation between Y and X

X_{ij} are constants

$\bar{X}_{..}$ is overall mean of concomitant variable X_{ij}

ε_{ij} are independent N(0, σ^2)



From the above plots , we can say that the residual distribution is normally . And the variance is constant . Thus , the model is basically appropriate for this problem.

b)Create an ANOVA table and discuss , in as much detail as possible, the results that you observe.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	0.31577099	0.07894275	3.91	0.0100
Error	35	0.70700651	0.02020019		
Corrected Total	39	1.02277750			

R-Square	Coeff Var	Root MSE	y Mean
0.308739	10.23788	0.142127	1.388250

Source	DF	Type III SS	Mean Square	F Value	Pr > F
trts	3	0.16231353	0.05410451	2.68	0.0620
x2	1	0.13822349	0.13822349	6.84	0.0130

From the above anova table, we can know the F value , and the p-value and type III SS. We can use it to do the treatment effect.

And from type III SS , treatment 's p value is greater than 0.05, it means the presence of X, the treatment becomes not important . P value for x2 is less than 0.05 which means even the presence of treatment variable, X still important .

```
ods graphics on;

proc glm data = Diet plots=DIAGNOSTICS residuals;;

class trts;

model y = trts x2/ ss3 solution clparm;

run;

run;

ods graphics off;
```

C)obtain a point estimate and a 95% C.I for the slope of the regression line.

x2	0.005384841		0.00205854	2.62	0.0130	0.001205780	0.009563903

```
proc glm data = Diet;

class trts;

model y = trts x2 / solution clparm alpha = .05;

run;
```

d) obtain a 95% C.I to estimate the mean weight gain for pigs in the second treatment group with an initial weight of 58 pounds.

Parameter	Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
mean response for tx 2 at X=58	1.33515452	0.04536691	29.43	<.0001	1.24305479	1.42725426

```
proc glm data = Diet;

class trts;

model y = trts x2 / clparm alpha = .05;

 estimate "mean response for tx 2 at X=58" intercept 1 trts 0 1
0 0 x2 58;

run;

quit;
```

e) Obtain C.I for all pairwise comparisons between the treatment means .use the most efficient multiple comparison procedure with a 90% family confidence coefficient . State your finding.

Bonferroni procedure B = 2.515

Scheffe procedure S = 2.597

Since S>B, Scheffe procedure will lead to a wider confidence interval, Bonferroni procedure is more efficient. Let recall that Tukey procedure can not be use for pair wise comparisons for the covariance model.

Least Squares Means for Effect trts				
i	j	Difference Between Means	Simultaneous 90% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	0.148231	-0.011627	0.308089
1	3	0.011461	-0.148531	0.171453
1	4	0.112614	-0.049213	0.274442
2	3	-0.136770	-0.296932	0.023393
2	4	-0.035617	-0.197960	0.126727
3	4	0.101153	-0.059698	0.262004

Since all the pairwise comparisons' confidence interval doesn't include zero. We can say that all the difference are not significant.

```
***** Bon ;

data _null_;

df = 35;

alpha = 1- .1/12;

file print;

t =tinv(alpha, df);

put t 6.3;

run;



***** Scheffe ;

data _null_;

df2 = 35;

df1 = 3;

r = 4;

alpha = .1;
```

```
file print;

S2 =(r-1)*(finv(1 - alpha,df1,df2));

S=sqrt(S2);

put s 6.3;

run;

proc glm data = Diet;

class trts;

model y = trts x2;

 lsmeans trts / cl adjust=bon alpha= .1;

run;
```

Problem 4

a)state an appropriate model . Test whether or not the main effect of diet are present by assuming that all level combinations of diet and initial age are equally important.

$$Y_{ij} = \mu_{..} + \tau_i + \rho_j + (\tau\rho)_{ij} + \gamma(X_{ij} - \bar{X}_{..}) + \varepsilon_{ij}$$

$\mu_{..}$ is an overall mean

τ_i are the fixed treatment effects subject to the restriction $\sum \tau_j$ =0

ρ_j are the fixed block effects subject to the restriction $\sum \rho_j$ =0

$(\tau\rho)_{ij}$ is the interaction effect when factor τ is at the ith level and factor ρ at jth level.

γ is a regression coefficient for the relation between Y and X

X_{ij} are constants

$\bar{X}_{..}$ is overall mean of concomitant variable X_{ij}

ε_{ij} are independent N(0, σ^2)

The GLM Procedure

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	0.38277642	0.04784705	2.32	0.0446
Error	31	0.64000108	0.02064520		
Corrected Total	39	1.02277750			

R-Square	Coeff Var	Root MSE	y Mean
0.374252	10.35004	0.143684	1.388250

Source	DF	Type III SS	Mean Square	F Value	Pr > F
diets	3	0.17800565	0.05933522	2.87	0.0520
block	1	0.03518804	0.03518804	1.70	0.2013
x	1	0.16585630	0.16585630	8.03	0.0080
diets*block	3	0.03512550	0.01170850	0.57	0.6408

Ho: $\tau_i = 0$ vs Ha: $\tau_i \neq 0$

p-value =0.052< 0.1. conclude Ha . Therefore , we say the main effects of diet is present .

```
Proc GLM Data = diet;

Class  diets Block;

Model Y =  diets Block x Block*diets / ss3 solution clparm;

 run;
```

b)

$$H_0: \frac{7\mu_{11}+3\mu_{12}}{10} = \frac{7\mu_{21}+3\mu_{22}}{10} = \frac{7\mu_{31}+3\mu_{32}}{10} = \frac{7\mu_{41}+3\mu_{42}}{10}$$

H_a : not all equalities hold

The SAS System

The GLM Procedure

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
L	3	0.14403102	0.04801034	2.33	0.0941

Thus F-value equal to 2.33. F(1-α, 3, 31)= 2.9113 . $F^* \le F$.

Thus , conclude H0 we say the main effects of diet are not present .

```
proc glm data = diet order=data alpha=.1;

 class diets Block ;

 model gain =diets Block weight Block*diets/ solution clparm;

 lsmeans block diets /adjust=tukey cl  ;

 contrast  'L' diets 10 -10 0 0 diets*block 7 3 -7 -3 0 0 0 0 ,

           diets 0 10 -10 0 diets*block 0 0 7 3 -7 -3 0 0 ,

           diets 0 0 10 -10  diets*block 0 0 0 0 7 3  -7 -3 ;


run;
```

c) As in (b) ,Obtain a 95% C.I for the mean weight gain of pigs with an initial weight of 58 pounds , and the second type of diet.

The SAS System

The GLM Procedure

Parameter	Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
mean response for tx 2 at x=58	1.37268184	0.05222376	26.28	<.0001	1.26617079	1.47919290

```
proc glm data = diet order=data;

class diets Block ;

model Y =diets Block x Block*diets/solution clparm;

lsmeans block diets /alpha=.05 ;

estimate 'mean response for tx 2 at x=58' intercept 10 diets 0
10 0 0 block 7 3 diets*block 0 0 7 3 0 0 0 0 X 580 /divisor=10 ;

run;
```

d)Test whether or not the mean weight gain obtained considered in(c) is positive. State the hypotheses , test statistics , p_value and conclusion.

Since $H_0 : Y_k < 0$ VS $H_a : Y_k > 0$

$t^* = \frac{\hat{Y}}{sd} = 26.28$ and $t^* \sim t(0.975, 31) = 2.042$

By t test if $\left|t^*\right| \le t(0.975, 31)$ conclude H0

If $\left|t^*\right| \ge t(0.975, 31)$ conclude Ha

and the p-value is less than 0.0001, the t value is really great over the t value table.
Thus we can conclude Ha , the mean weight gain in (c) is positive.