/\*\*\*\*\*\*\*\*\*\* FINAL PROJECT STAT 7010 \*\*\*\*\*\*\*\*\*\*/

/\*Importng data from Excel sheet \*/

options center nodate ps=**5000** ls=**240** formdlim="-" symbolgen;

%let path= H:\Stat7010\hitters;

libname hitters "&path";

**PROC** **IMPORT** OUT=hit

DATAFILE= "&path"

DBMS=EXCEL2000 REPLACE;

Sheet=hitters;

GETNAMES=YES;

**RUN**;

**proc** **print**; **run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* CLEANING THE DATA AND RENAMING VARIABLES \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**data** hit2;set hit;

rename name\_of\_player=player;

if sal= **.** then delete;**run**;

**PROC** **PRINT**;**RUN**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* CREATING RESPONSE VARIABLE (GROSS PRODUCTION AVERAGE )\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**data** hit3; set hit2;

OBP= (x2+x6+Err)/(x1+x6+Err+Put); /\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*On Base Percentage \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

SLG= (x4+**3**\*x5+**4**\*x3)/x1; /\*\*\*\*\*\*\*\*\*\*\*\*\*\*Slugging Percentage \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

GPA1= ((**1.8**)\*OBP+SLG)/**4** ; /\*\*\*\*\*\*\*\*\*\*\*\*\*\*Gross Production Average \*\*\*\*\*\*\*\*\*\*\*\*\*/

GPA= ROUND (GPA1,**.01**); /\*\*\*\*\*\*\*\*\*\*\*\*\*\*Rounding Off GPA to two decimal place \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**RUN**;

**PROC** **PRINT**;**RUN**;

**data** hit4;

set hit3;

if team1='Pit.' then delete;

if team1='Mon.' then delete;

if team1='Mil.' then delete;

if team1='Bal.' then delete;

if team2='Pit.' then delete;

if team2='Mon.' then delete;

if team2='Mil.' then delete;

if team2='Bal.' then delete;

**run**;**quit**;

**proc** **print**;**run**;**quit**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* SORTING THE DATA \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**proc** **sort** data=hit4;

by league1 division1 pos;

**proc** **print**;**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* CHECKING MODEL ASSUMPTIONS \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

ods html;

**proc** **glm** data = hit4;

class league1 division1 pos team1 ;

model GPA =league1 division1 pos team1 ;

output out = diag r = res p = pred;

**run**;

**proc** **print** data=diag ;**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Test Of Normality using QQ Plot and Histogram \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**proc** **univariate** normal data = diag;

var res;

qqplot res / normal (L = **1** mu = est sigma = est);

histogram res / normal;

**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Testing Homogeneity of Variance \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**proc** **gplot** data = diag;

plot res \* (League1 Division1 pos team1 pos pred) ;

**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Testing the Independence of Covariate (Salary) and other Factors\*\*\*\*\*\*\*\*\*\*\*/

**proc** **glm** data= hit4;

class League1 Division1 pos team1 ;

model sal = League1 Division1 pos team1 ;

**run**;

**proc** **glm** data=hit4;

class League1 Division1 pos team1;

model GPA= League1 Division1 pos team1 sal;

output out = diag1 r=res1 p=pred1;

**run**;

**proc** **univariate** data = diag1;

qqplot res1 / normal (L = **1** mu = est sigma = est);

**run**;

**proc** **gplot** data = diag1;

plot res1 \* (pred1 League1 Division1 pos team1 sal);

**run**;

ods html close;

/\*\*\*\*\*\*\*\*\*\* ANOVA FOR 1986 DATA ON GPA WITH COVARIATE "SALARY" \*\*\*\*\*\*\*\*\*\*/

**proc** **mixed** data=hittrans cl covtest method=type1 ;

class league1 division1 team1 pos logsal ;

model logGPA =league1|division1|pos|logsal;

random team1 ;

**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* No significant results with Sal so excluding it \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**proc** **mixed** data=hit3 cl covtest method=type1 ;

class league1 division1 team1 pos sal ;

model GPA =league1|division1|pos ;

random team1 ;

**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Modifying Sal to be significant \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**proc** **insight** data=hittrans;

**run**;**quit**;

**data** hittrans;

set hit3;

logsal=log(sal);

salsq=sal\*\***2**;

sqrtsal=sqrt(sal);

recsal=**1**/sal;

logGPA=log(GPA);

GPAsq=GPA\*\***2**;

recGPA=**1**/GPA;

sqrtGPA=sqrt(GPA);

**run**;**quit**;

**proc** **print** data=hittrans;**run**;**quit**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Tukey's Multiple Comparison Test for Pos and Team1 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*ods html;

proc glimmix data=hit3;

class league1 division1 team1 pos sal ;

model GPA/sal =league1|division1|pos ;

random team1 ;

lsmeans pos /adjust=tukey;

run;

ods html close;run;

proc glimmix data=hit3 ;

class league1 division1 team1 player pos sal;

model GPA/sal = league1|division1|pos ;

random team1;

lsmeans pos\*division1 /adjust=tukey lines;

run;

proc glimmix data=hit3 ;

class league1 division1 team1 player pos sal;

model GPA/sal= league1|division1|pos ;

random team1;

lsmeans league1\*division1\*pos /adjust=tukey lines;

run; \*/

/\*\*\*\*\*\*\* ANOVA FOR CAREER DATA ON GPA WITH AND WITHOUT COVARIATE "SALARY" \*\*\*\*\*\*\*\*\*\*/

ods html;

**data** hit5; set hit2;

OBP1= (x9+x13)/(x8+x13); /\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*On Base Percentage \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

SLG1= (x11+**3**\*x12+**4**\*x10)/x8; /\*\*\*\*\*\*\*\*\*\*\*\*\*\*Slugging Percentage \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

GPA2= ((**1.8**)\*OBP1+SLG1)/**4** ; /\*\*\*\*\*\*\*\*\*\*\*\*\*\*Gross Production Average \*\*\*\*\*\*\*\*\*\*\*\*\*/

GPA3= ROUND (GPA2,**.01**); /\*\*\*\*\*\*\*\*\*\*\*\*\*\*Rounding Off GPA2 to two decimal place \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**RUN**;

**PROC** **PRINT**;**RUN**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* SORTING THE DATA \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**proc** **sort** data=hit5;

by league2 division1 pos;

**proc** **print**;**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* CHECKING MODEL ASSUMPTIONS \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**proc** **glm** data = hit5;

class league2 division1 pos team2 ;

model GPA3 =league2 division1 pos team2 ;

output out = diag2 r = res2 p = pred2;

**run**;

**proc** **print** data=diag2 ;**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Test Of Normality using QQ Plot and Histogram \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**proc** **univariate** normal data = diag2;

var res2;

qqplot res2 / normal (L = **1** mu = est sigma = est);

histogram res2 / normal;

**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Testing Homogeneity of Variance \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**proc** **gplot** data = diag2;

plot res2 \* (League2 Division1 pos team2 pos pred2) ;

**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Testing the Independence of Covariate (Salary) and other Factors\*\*\*\*\*\*\*\*\*\*\*/

**proc** **glm** data= hit5;

class League2 Division2 pos team2 ;

model sal = League2 Division1 pos team2 ;

**run**;

**proc** **glm** data=hit5;

class League2 Division1 pos team2;

model GPA3= League2 Division1 pos team2 sal;

output out = diag3 r=res3 p=pred3;

**run**;

**proc** **univariate** data = diag3;

qqplot res3 / normal (L = **1** mu = est sigma = est);

**run**;

**proc** **gplot** data = diag3;

plot res3 \* (pred3 League2 Division1 pos team2 sal);

**run**;

ods html close;

/\*\*\*\*\*\*\*\*\*\* ANOVA FOR CAREER DATA ON GPA WITH & WITHOUT COVARIATE "SALARY" \*\*\*\*\*\*\*\*\*\*/

**proc** **mixed** data=hit4 cl covtest method=type1 ;

class league2 division1 team2 pos sal ;

model GPA3 =league2|division1|pos ;

random team2 ;

**run**;

**proc** **mixed** data=hit4 cl covtest method=type1 ;

class league2 division1 team2 pos sal ;

model GPA3 =league2|division1|pos|sal ;

random team2 ;

**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* TUKEY'S TEST FOR MULTIPLE COMPARISON \*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**proc** **glimmix** data=hit4;

class league2 division1 team2 player pos sal;

model GPA3/sal =league2|division1|pos ;

random team2;

lsmeans league2|division1|pos /adjust=tukey lines;

**run**;