NFL 2013 Combine Data Multivariate Analysis

John Michael Croft, Brian Ginburg, Gary Keller and William Ward

Kennesaw State University

**Abstract**

Since 1982, the NFL Combine (an invitation only event) evaluates college football players’ physical abilities and mental awareness.  NFL teams use the results to make targeted evaluations of draft prospects. The purpose of this research is to examine the difference in multiple response variables between groups of player positions via multivariate methods. All charts, graphs, figures, &c… can be found in the appendices at the end of the analysis while some have been placed within the body to emphasize the importance of the topic being addressed. Table 1 contains the original dataset variables, a brief description, general and specific types, and measurement units.

Due to exploratory analyses and data cleansing seeking to reduce multicollinearity among response variables, the final analysis suggests multivariate normality reducing the probability of Type I errors when compared with a series of univariate analyses of variances. The analysis provides strong evidence of significant differences between groups across multiple response variables. Contrasts are utilized to highlight the most significant differences between Group1 (FS; SS; CB; WR) vs Group 3 (OT; OC; OG; DT) in response variables: Hands, Bench, Vertical (-.7inches, -11.87 reps, 8.7 inches, respectively, on average) and Group 3 (OLB, ILB, DE, TE) vs Group 4 (RB) in response variable: Height (5.97 inches on average).

**Introduction**

Player positions form the basis of this analysis. Kickers (K), Long snappers (LS), and Punters (P) are not found in the 2013 data subset, whereas Quarterbacks (QB) have been omitted due to lack of observations (n=14<20). Table A displays the initial groups (A - F) prior to the exploratory analysis and final groups (1 - 4) after the exploratory analysis.



The initial groups above are based on an assumption that players at similar positions have similar attributes. Tight Ends have been arbitrarily assigned to Group C primarily for group sample size consistency as well as expecting similar attributes (e.g. height, weight, &c...). The final groups above will be discussed later but reclassify certain positions to better align with adjusted expectations after the exploratory analysis. Significant differences in response variables due to perceived group attribute differences (e.g. big v. small; fast v. slow; short v. tall) were expected. Figure 1 shows approximately equal initial group sizes. The global hypothesis expects significant group differences in at least one response variable.

**Data Cleansing**

The following variables are considered redundant or inconsequential and have been omitted from this analysis: College, FirstName, HeightFeet, HeightInches, LastName, Name, Pick, PickRound, PickTotal, Round, and Year.

Missing values are assumed missing at random and have been set to missing to observe percent missing per variable and per observation (see Tables 2 & 3). Variables missing more than 20% were omitted from the analysis: Wonderlic, TwentyYD, ThreeCone, TwentySS. Observations missing more than 33.34% were omitted from the analysis: ID #’s 9225, 8984, 9107, 9140. All remaining missing values were imputed via linear regression (by position). The normality and constant variance of errors assumptions are reasonably met for each regression.

Whilemoderate response variable correlations are desirable, significant correlations (>.7) were examined to reduce multicollinearity and increase the power of the analysis. Table 4 shows all possible correlations with significant correlations highlighted. All response variables, other than Hands and Bench, are significantly correlated with at least one other response variable. In conjunction with evaluating standardized effect sizes (Figure 2), Broad and TenYd have been omitted from further analysis. Acknowledging FortyYD has marginally higher correlations than TenYD, assumed industry preference is to keep FortyYD in the analysis.



**Figure 2: Initial Group Variable Profile Plot**



**Assumptions**

The initial Mardia’s test (Table 5) suggests non-multivariate normality in the symmetry (*p* = .003) with marginal multivariate normality in the distributional spread (*p* = .133).  Attempting to refine the analysis, individual variables were examined for univariate normality (Figures 3 - 9). Weight (bimodal), FortyYD (skewed), and Arms (skewed) were omitted from further analysis due to apparent non-univariate normality. The final Mardia’s test (Table 6) suggests multivariate normality in both symmetry (*p* = .293) and distributional spread (*p* = .428).

At this time the reader is reminded of and encouraged to review Table A, delineating the initial groups (A - F) from the final groups (1 - 4). Figure 10 suggests concerns with variance homogeneity between the initial groups - the Vertical boxplot is provided as an example. Other variables’ boxplots suggest similar concerns but have been omitted as redundant. Table 7 supports nonhomogeneous variance between the initial groups (*p* < .001).

Players were reclassified into final groups (1-4) attemptimg to correct for non-homogeneous variance. Group 1 is a combination of Groups A plus E; Group 2 is a combination of Group C plus DE; Group 3 is a combination of Group D plus DT; Group 4 is the same as Group F. Group sample sizes remain similar (Figure 11). Table 8 supports variance homogeneity between final groups (*p* < .552).

Observations are assumed independent from each other as players are measured separately from one another (i.e. One player’s results do not influence another player’s results.)

Mahalanobis distances were calculated per observation. An upper limit of 13 was approximated using the mean and adding three standard deviations (3.9 + 3\*(2.9)) to determine outliers. Five outliers were detected but were not removed due to low marginal impact on the analysis.

**Results**

Tables 9 & 10 contain multivariate analysis of variance test criteria, F-stat approximations, and characteristic roots. A Wilk’s lambda of  .113 indicates at least one group is significantly different from another for at least one response variable (*p* < .0001), rejecting the null hypothesis. Consideration could be given to evaluating our model in one dimension with a single variable dominating the model (89.63% characteristic root) suggesting Roy’s greatest root should be the test criteria utilized. However, all test criteria are satisfied to support rejecting the null hypothesis (*p*<.0001).



Univariate analyses of variances were analyzed per response variables (Table 11). The univariate results indicate significant differences between groups per response variable, suggesting contrasts be analyzed per response variable.

Figure 12 shows a standardized profile plot of the final groups across all remaining response variables to aid in determining which contrasts to examine.

**Figure 12: Final Group Variable Profile Plot**



Table 12 summarizes all contrasts consider:

* Vertical: All contrasts significantly different (all *p* values ≤ .01) except Group 2 vs Group 4 (*p* = 0.6245) with Group 1 vs Group 3 being most significant (SS = 2970.98, Estimate = 8.70).
* Bench: All contrasts significantly different (all *p* values < .0001) except Group 2 vs Group 3 (*p* = .468) with Group 1 vs Group 3 being most significant (SS = 5474.55, Estimate = -11.81).
* Hands: All contrasts significantly different (all *p* values ≤ .0243) except Group 2 vs Group 3 (*p* = .6897) with Group 1 vs. Group 3 being most significant (SS = 19.18, Estimate = -0.70).
* Height: All contrasts significantly different (all *p* values ≤ .0012) with Group 3 vs. Group 4 being most significant (SS = 874.20, Estimate = 5.97).



**Conclusion**

The analysis supports the expected hypothesized significant differences between groups of 2013 NFL draft combine participants. The most significant differences are found between Group 1 vs Group 3 (Vertical; Bench; Hands); i.e. Defensive backs and wide receivers, on average, jump 8.7 inches higher, bench press 11.87 less reps, and have hands .7 inches less than offensive linemen and defensive tackles. On average, this is expected due to the nature of positions within each group – defensive backs and wide receivers are required to be more athletic overall, running faster longer, jumping higher to catch passes while offensive linemen and defensive tackles require stamina and stability to pass block and run block constantly coming in contact with the opposing team.

However, the most significant difference in height is between Group 3 vs Group 4; i.e. Running backs, on average, are 5.97 inches shorter than offensive linemen and defensive tackles. On average, this is expected due to the nature of positions within each group – running backs are required to be more mobile and agile to break tackles, hurdle defenders and outrun the opposing team while offensive linemen and defensive tackles were discuss above. Additionally defensive tackles are looking to disrupt passing attempts with maximum vertical extension utilizing the additional 5.97 inches in height.

Overall, the analysis provide strong evidence toward significant differences between groups primarily due to the inherent athleticism commonly found within each group allowing similar within group performances across response variables.

Recommend offensive linemen and defensive tackles focus primarily on stamina and stability while defensive backs, wide receivers and running backs focus more on mobility and agility. Linebackers, defensive ends, and tight ends should attempt to focus on some combination of stamina, stability, mobility and agility as versatility is required at those positions; recommend heavier players focus on stamina and stability while lighter players focus on mobility and agility.

While linear combinations were not compared, it is noted the groups somewhat achieve this organically by grouping positions of players with similar size, weight and athleticism.

**Future Research**

Comparing the results of the current analysis with same players’ production over the first 2-5 years of their career may be of interest (both drafted and undrafted participants) as well as predicting future combine participant responses. Recommend future studies focus on the differences among drafted and undrafted combine participants per same response variables. Additionally, focusing only on drafted combine participants would allow draft picks to be evaluated as an additional response variable.

**Appendix 1: Tables**











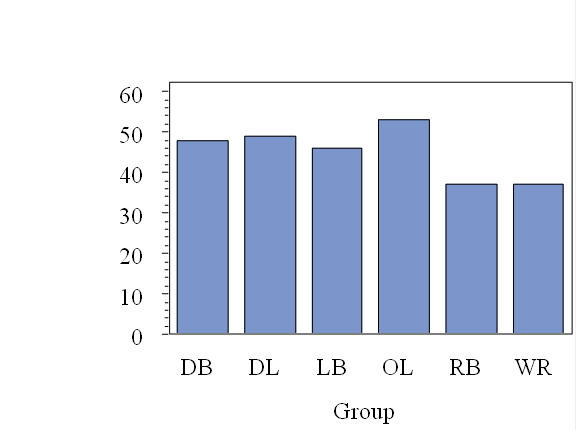






**Appendix 2: Figures**

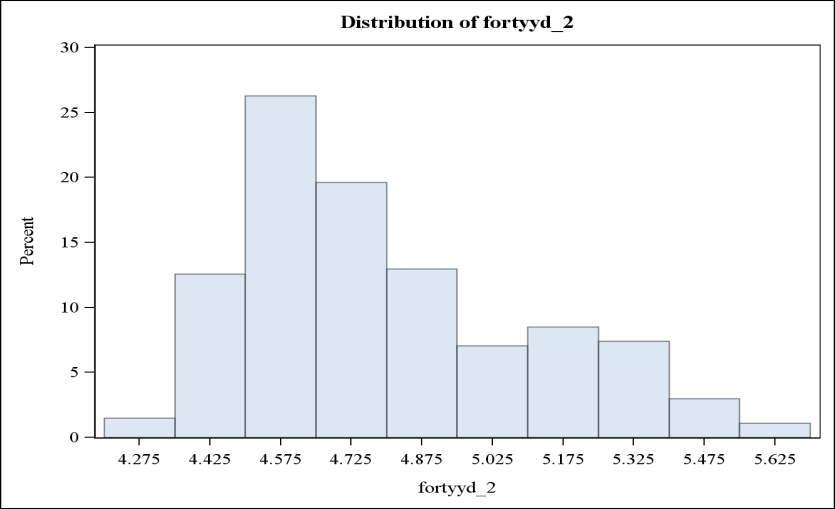
**Figure 1: Initial Group Frequency Distribution**

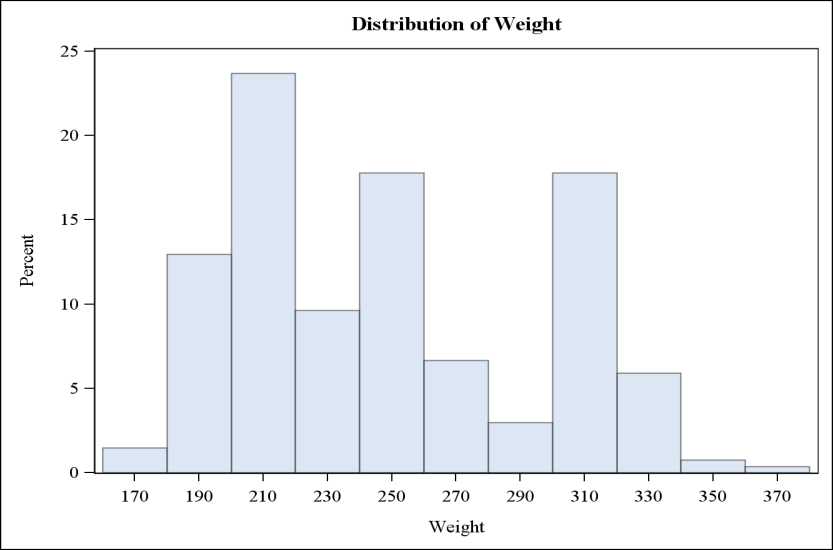


**Figure 2: Initial Group Variable Profile Plot**

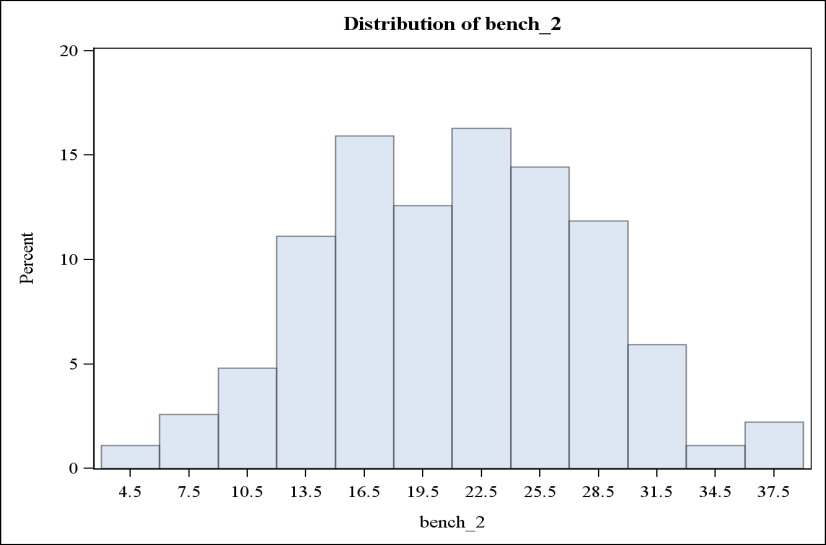


**Figure 3: Forty Yard Time Histogram (in seconds)**

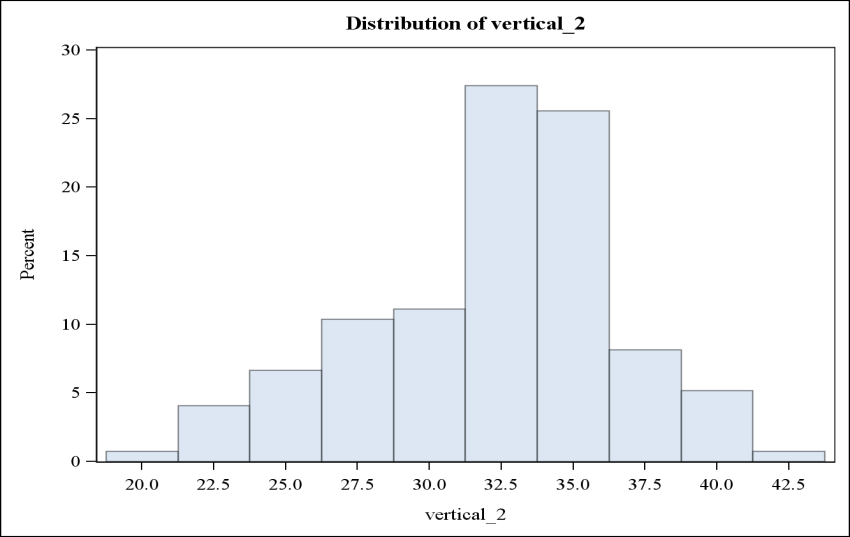


**Figure 4: Weight Histogram (in pounds)**

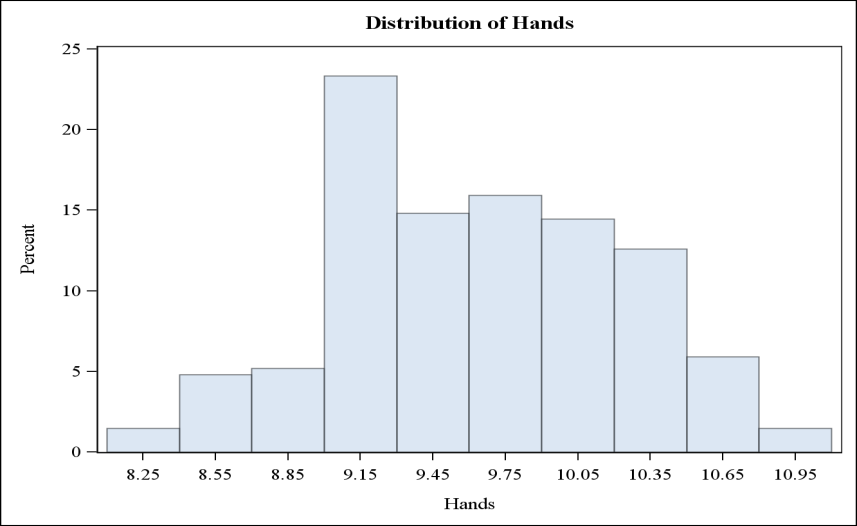
**Figure 5: Bench Press Histogram (# of reps)**



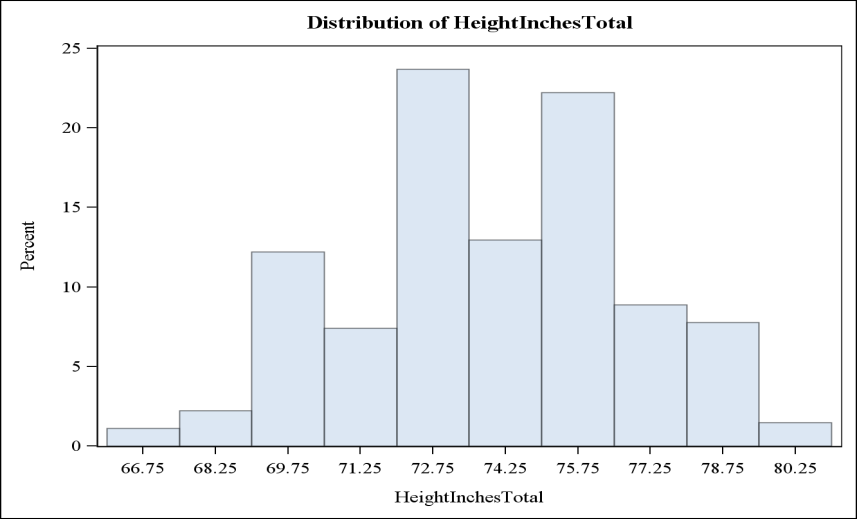
**Figure 6: Vertical Jump Histogram (in inches)**



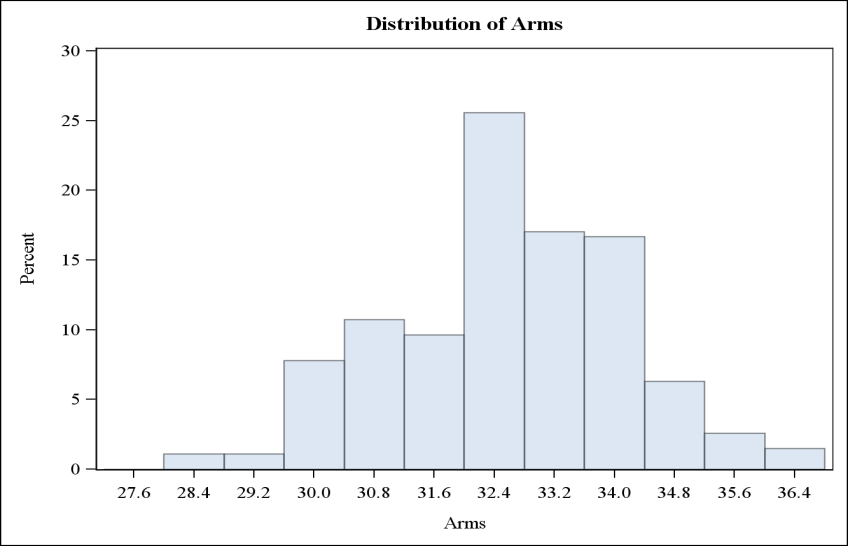
**Figure 7: Hand Length Histogram (in inches)**



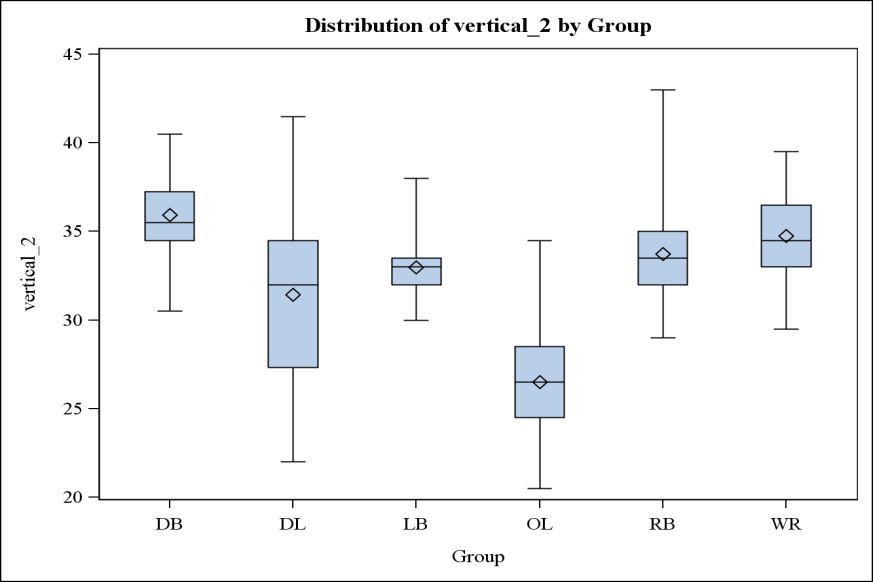
**Figure 8: Height Histogram (in inches)**



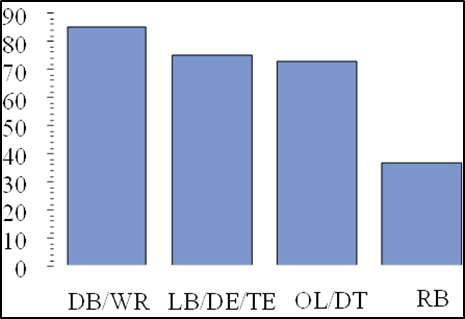
**Figure 9: Arms Histogram (in inches)**



**Figure 10: Vertical Jump Boxplot (in inches)**



**Figure 11: Final Group Frequency Distribution**



**Figure 12: Final Group Variable Profile Plot**



**Appendix 3: SAS Code**

\*========================================================================================================================\*

Create Library and Read Data to the Library

\*========================================================================================================================\*;

libname C13 "\\Client\F$\Stat Classes\Current\Multivariate Data Analysis\Project1";

**proc** **import** datafile="\\Client\F$\Stat Classes\Current\Multivariate Data Analysis\Project1\combine.csv"

out=combine

dbms=csv

replace;

getnames=yes;

**run**;

**data** C13.combine;

set combine;

**run**;

\*========================================================================================================================\*

Variable Audit

\*========================================================================================================================\*;

**proc** **means** data = C13.combine;

**run**;

\*========================================================================================================================\*

Set all other 0 Values to missing

\*========================================================================================================================\*;

**data** C13.combine\_2 (drop = i);

set C13.combine;

array var{\*} arms hands fortyyd twentyyd tenyd twentyss threecone vertical broad bench round pickround picktotal wonderlic;

do i = **1** to **14**;

if var{i} = **0** then var{i} = **.** ;

end;

**run**;

**proc** **means** data = C13.combine\_2 n nmiss min max mean std;

**run**;

**data** C13.combine\_2 (drop = wonderlic twentyyd threecone twentyss);

set C13.combine\_2;

**run**;

\*========================================================================================================================\*

Use a transpose to identify individuals

that have several missing values.

\*========================================================================================================================\*;

**data** temp (drop = college firstname lastname name pick pickround picktotal round year) ;

set C13.combine\_2;

**run**;

**proc** **transpose** data = temp out = transpose;

**run**;

**proc** **means** data = transpose n nmiss;

**run**;

\*========================================================================================================================\*

Remove Individuals with more than 33%

missing values.

\*========================================================================================================================\*;

**data** C13.combine\_3;

set C13.combine\_2;

if id = **9225** or id = **8984** or id = **9107** or id = **9140** then delete;

**run**;

**proc** **means** data = C13.combine\_3 n nmiss;

**run**;

\*========================================================================================================================\*

Need to impute the following variables:

fortyyd tenyd vertcal broad bench

Regression Imputation: use height in inches

weight, and position as predictors

Run Regression Imputation on all 5 to get in one dataset

\*========================================================================================================================\*;

**proc** **freq** data = C13.combine\_3;

tables position;

**run**;

\*\*\* Create Dummy Variables for Postion with QB the base \*\*\*;

**data** C13.combine\_3;

set C13.combine\_3;

if position = "CB" then CB = **1**;

else CB = **0**;

if position = "DE" then DE = **1**;

else DE = **0**;

if position = "DT" then DT = **1**;

else DT = **0**;

if position = "FS" then FS = **1**;

else FS = **0**;

if position = "IL" then IL = **1**;

else IL = **0**;

if position = "OC" then OC = **1**;

else OC = **0**;

if position = "OG" then OG = **1**;

else OG = **0**;

if position = "OL" then OL = **1**;

else OL = **0**;

if position = "OT" then OT = **1**;

else OT = **0**;

if position = "WR" then WR = **1**;

else WR = **0**;

if position = "RB" then RB = **1**;

else RB = **0**;

if position = "SS" then SS = **1**;

else SS = **0**;

if position = "TE" then TE = **1**;

else TE = **0**;

**run**;

\*\*\* Regression Imputation \*\*\*;

**proc** **reg** data = C13.combine\_3;

model fortyyd = CB DE DT FS IL OC OG OL OT WR RB SS TE / VIF;

output out=impute\_1 p=predicted\_fortyyd;

**run**;

**quit**;

**proc** **reg** data = impute\_1;

model tenyd = CB DE DT FS IL OC OG OL OT WR RB SS TE / VIF;

output out=impute\_2 p=predicted\_tenyd;

**run**;

**quit**;

**proc** **reg** data = impute\_2;

model vertical = CB DE DT FS IL OC OG OL OT WR RB SS TE / VIF;

output out=impute\_3 p=predicted\_vertical;

**run**;

**quit**;

**proc** **reg** data = impute\_3;

model Broad = CB DE DT FS IL OC OG OL OT WR RB SS TE / VIF;

output out=impute\_4 p=predicted\_broad;

**run**;

**quit**;

**proc** **reg** data = impute\_4;

model Bench = CB DE DT FS IL OC OG OL OT WR RB SS TE / VIF;

output out=impute\_5 p=predicted\_bench;

**run**;

**quit**;

**data** C13.combine\_Imputation\_GK;

set impute\_5;

/\*=====================================================

fortyy\_2, vertical\_2, etc. are the imputed values

\*=====================================================\*/

if fortyyd = **.** then fortyyd\_2 = predicted\_fortyyd;

else fortyyd\_2 = fortyyd;

if tenyd = **.** then tenyd\_2 = predicted\_tenyd;

else tenyd\_2 = tenyd;

if vertical = **.** then vertical\_2 = predicted\_vertical;

else vertical\_2 = vertical;

if broad = **.** then broad\_2 = predicted\_broad;

else broad\_2 = broad;

if bench = **.** then bench\_2 = predicted\_bench;

else bench\_2 = bench;

**run**;

\*===========================================================================================\*

Remove unnecessary variable and create the groups.

\*==========================================================================================\*;

**data** master;

set C13.combine\_imputation\_gk;

**run**;

**proc** **freq** data = master;

table position;

**run**;

**data** master\_2 (keep= id name position group weight arms hands fortyYd tenyd vertical broad bench

heightinchestotal fortyyd\_2 tenyd\_2 vertical\_2 broad\_2 bench\_2);

set master;

if position = "QB" then delete;

else if Position = "DE" then Group = "DL";

else if Position = "DT" then Group = "DL";

else if Position = "IL" then Group = "LB";

else if Position = "OL" then Group = "LB";

else if Position = "CB" then Group = "DB";

else if Position = "SS" then Group = "DB";

else if Position = "FS" then Group = "DB";

else if Position = "OT" then Group = "OL";

else if Position = "OC" then Group = "OL";

else if Position = "OG" then Group = "OL";

else if Position = "TE" then Group = "LB";

else if Position = "RB" then Group = "RB";

else if Position = "WR" then Group = "WR";

else group = "";

**run**;

**proc** **freq** data = master\_2;

tables position\*group;

**run**;

**data** C13.master;

set master\_2;

**run**;

\*===========================================================================================\*

Profile Analysis

\*==========================================================================================\*;

\*\*\* Standardize the values for each possible Y \*\*\*;

**proc** **means** data = C13.master;

var weight arms hands fortyyd\_2 tenyd\_2 vertical\_2 broad\_2 bench\_2 heightinchestotal;

output out = standard mean = avg\_weight avg\_arms avg\_hands avg\_forty avg\_ten avg\_vert avg\_broad avg\_bench avg\_height

std = std\_weight std\_arms std\_hands std\_forty std\_ten std\_vert std\_broad std\_bench std\_height;

**run**;

**proc** **sql**;

create table standard\_2 as

select \*

from C13.master, standard;

**quit**;

**data** standard\_3 (drop= weight arms hands fortyyd\_2 tenyd\_2 vertical\_2 broad\_2 bench\_2 heightinchestotal

avg\_weight avg\_arms avg\_hands avg\_forty avg\_ten avg\_vert avg\_broad avg\_bench avg\_height

std\_weight std\_arms ste\_hands std\_forty std\_ten std\_vert std\_broad std\_bench std\_height

\_type\_ \_freq\_ fortyyd tenyd vertical broad bench name position id);

set standard\_2;

s\_weight = (weight-avg\_weight)/std\_weight;

s\_arms = (arms-avg\_arms)/std\_arms;

s\_hands = (hands-avg\_hands)/std\_hands;

s\_forty = (fortyyd\_2-avg\_forty)/std\_forty;

s\_ten = (tenyd\_2-avg\_ten)/std\_ten;

s\_vert = (vertical\_2-avg\_vert)/std\_vert;

s\_broad = (broad\_2-avg\_broad)/std\_broad;

s\_bench = (bench\_2-avg\_bench)/std\_bench;

s\_height = (heightinchestotal-avg\_height)/std\_height;

**run**;

\*\*\* Obtain the average of the standardized values and plot per group \*\*\*;

**proc** **means** data = standard\_3;

class group;

var s\_weight s\_arms s\_hands s\_forty s\_ten s\_vert s\_broad s\_bench s\_height;

output out = temp mean = avg\_weight avg\_arms avg\_hands avg\_forty avg\_ten avg\_vert avg\_broad avg\_bench avg\_height;

**run**;

**data** temp2 (drop= \_freq\_ \_type\_);

set temp;

**run**;

**proc** **transpose** data = temp2 out=trans;

by group;

**run**;

**proc** **format**;

value varfmt

**1** = "Weight"

**2** = "Arms"

**3** = "Hands"

**4** = "Forty"

**5** = "Ten"

**6** = "Vert"

**7** = "Broad"

**8** = "Bench"

**9** = "Height";

**run**;

**data** temp3;

set trans;

if \_name\_ = "avg\_weight" then name = **1**;

else if \_name\_ = "avg\_arms" then name = **2**;

else if \_name\_ = "avg\_hands" then name = **3**;

else if \_name\_ = "avg\_forty" then name = **4**;

else if \_name\_ = "avg\_ten" then name = **5**;

else if \_name\_ = "avg\_vert" then name = **6**;

else if \_name\_ = "avg\_broad" then name = **7**;

else if \_name\_ = "avg\_bench" then name = **8**;

else if \_name\_ = "avg\_height" then name = **9**;

else name = **10**;

format name varfmt.;

**run**;

symbol1 interpol=join value=dot;

**proc** **gplot** data = temp3;

plot col1\*name=group;

**run**;

\*\*\* Check correlations for vert and broad and ten and forty \*\*\*;

**proc** **corr** data = C13.master;

var vertical\_2 broad\_2;

**run**;

**proc** **corr** data = C13.master;

var fortyyd\_2 tenyd\_2;

**run**;

\*\*\* Drop Broad\_2 and Ten\_2 \*\*\*;

**data** C13.master\_2 (drop= broad\_2 tenyd\_2 broad tenyd);

set C13.master;

**run**;

\*========================================================================================================================\*

Multivariate Normality Check: Mardia's Kurtosis / Skewness

\*========================================================================================================================\*;

%let newinpt= vertical\_2 bench\_2 hands heightinchestotal;

**proc** **iml**;

use C13.master\_2;

read all var {&newinpt} into y;

n = nrow(y) ;

p = ncol(y) ;

dfchi = p\*(p+**1**)\*(p+**2**)/**6** ;

q = i(n) - (**1**/n)\*j(n,n,**1**);

s = (**1**/(n))\*y`\*q\*y ; s\_inv = inv(s) ;

g\_matrix = q\*y\*s\_inv\*y`\*q;

beta1hat = ( sum(g\_matrix#g\_matrix#g\_matrix) )/(n\*n);

beta2hat =trace( g\_matrix#g\_matrix )/n ;

k=(p+**1**)\*(n+**1**)\*(n+**3**)/(n\*((n+**1**)\*(p+**1**)-**6**));

kappa1 = n\*beta1hat\*k/**6** ;

kappa2 = (beta2hat - p\*(p+**2**) ) /sqrt(**8**\*p\*(p+**2**)/n) ;

pvalskew = **1** - probchi(kappa1,dfchi) ;

pvalkurt = **2**\*( **1** - probnorm(abs(kappa2)) );

print s ;

print s\_inv ;

print 'TESTS:';

print 'Based on skewness: ' beta1hat kappa1 pvalskew ;

print 'Based on kurtosis: ' beta2hat kappa2 pvalkurt;

**quit**;

\*\*\* Macro to look at Univariate Normality \*\*\*;

**%Macro** Hist(var= );

proc univariate data = C13.master\_2;

var &var;

histogram;

run;

**%Mend**;

%***Hist*** (var=fortyyd\_2);

%***Hist*** (var=vertical\_2);

%***Hist*** (var=bench\_2);

%***Hist*** (var=heightinchestotal);

%***Hist*** (var=weight);

%***Hist*** (var=arms);

%***Hist*** (var=hands);

\*\*\* Ran several iterations of this test to get a set of variables that are multivariate normal \*\*\*;

**data** C13.master\_3 (drop= fortyyd vertical bench fortyyd\_2 weight arms);

set C13.master\_2;

**run**;

\*========================================================================================================================\*

Covariance Matrix Structure

\*========================================================================================================================\*;

**proc** **discrim** data = C13.master\_3 pool=test;

class group;

var vertical\_2 bench\_2 hands heightinchestotal;

**run**;

\*\*\* This assumption is highly violated. Try to group differently \*\*\*;

**data** regroup;

set C13.master\_3;

if position = "QB" then delete;

else if Position = "DE" then group\_2 = "LB/DE/TE";

else if Position = "DT" then group\_2 = "OL/DT";

else if Position = "IL" then group\_2 = "LB/DE/TE";

else if Position = "OL" then group\_2 = "LB/DE/TE";

else if Position = "CB" then group\_2 = "DB/WR";

else if Position = "SS" then group\_2 = "DB/WR";

else if Position = "FS" then group\_2 = "DB/WR";

else if Position = "OT" then group\_2 = "OL/DT";

else if Position = "OC" then group\_2 = "OL/DT";

else if Position = "OG" then group\_2 = "OL/DT";

else if Position = "TE" then group\_2 = "LB/DE/TE";

else if Position = "RB" then group\_2 = "RB";

else if Position = "WR" then group\_2 = "DB/WR";

else group\_2 = "";

**run**;

**proc** **discrim** data = regroup pool=test;

class group\_2;

var vertical\_2 bench\_2 hands heightinchestotal;

**run**;

**data** C13.master\_4;

set regroup;

**run**;

\*========================================================================================================================\*

Redo Profile Analysis Based on New Groups

\*========================================================================================================================\*;

**data** new\_standard;

set c13.master;

if position = "QB" then delete;

else if Position = "DE" then group\_2 = "LB/DE/TE";

else if Position = "DT" then group\_2 = "OL/DT";

else if Position = "IL" then group\_2 = "LB/DE/TE";

else if Position = "OL" then group\_2 = "LB/DE/TE";

else if Position = "CB" then group\_2 = "DB/WR";

else if Position = "SS" then group\_2 = "DB/WR";

else if Position = "FS" then group\_2 = "DB/WR";

else if Position = "OT" then group\_2 = "OL/DT";

else if Position = "OC" then group\_2 = "OL/DT";

else if Position = "OG" then group\_2 = "OL/DT";

else if Position = "TE" then group\_2 = "LB/DE/TE";

else if Position = "RB" then group\_2 = "RB";

else if Position = "WR" then group\_2 = "DB/WR";

else group\_2 = "";

**run**;

\*\*\* Standardize the values for each possible Y \*\*\*;

**proc** **means** data = new\_standard;

var weight arms hands fortyyd\_2 tenyd\_2 vertical\_2 broad\_2 bench\_2 heightinchestotal;

output out = standard mean = avg\_weight avg\_arms avg\_hands avg\_forty avg\_ten avg\_vert avg\_broad avg\_bench avg\_height

std = std\_weight std\_arms std\_hands std\_forty std\_ten std\_vert std\_broad std\_bench std\_height;

**run**;

**proc** **sql**;

create table standard\_2 as

select \*

from new\_standard, standard;

**quit**;

**data** standard\_3 (drop= weight arms hands fortyyd\_2 tenyd\_2 vertical\_2 broad\_2 bench\_2 heightinchestotal

avg\_weight avg\_arms avg\_hands avg\_forty avg\_ten avg\_vert avg\_broad avg\_bench avg\_height

std\_weight std\_arms ste\_hands std\_forty std\_ten std\_vert std\_broad std\_bench std\_height

\_type\_ \_freq\_ fortyyd tenyd vertical broad bench name position id);

set standard\_2;

s\_weight = (weight-avg\_weight)/std\_weight;

s\_arms = (arms-avg\_arms)/std\_arms;

s\_hands = (hands-avg\_hands)/std\_hands;

s\_forty = (fortyyd\_2-avg\_forty)/std\_forty;

s\_ten = (tenyd\_2-avg\_ten)/std\_ten;

s\_vert = (vertical\_2-avg\_vert)/std\_vert;

s\_broad = (broad\_2-avg\_broad)/std\_broad;

s\_bench = (bench\_2-avg\_bench)/std\_bench;

s\_height = (heightinchestotal-avg\_height)/std\_height;

**run**;

\*\*\* Obtain the average of the standardized values and plot per group \*\*\*;

**proc** **means** data = standard\_3;

class group\_2;

var s\_weight s\_arms s\_hands s\_forty s\_ten s\_vert s\_broad s\_bench s\_height;

output out = temp mean = avg\_weight avg\_arms avg\_hands avg\_forty avg\_ten avg\_vert avg\_broad avg\_bench avg\_height;

**run**;

**data** temp2 (drop= \_freq\_ \_type\_);

set temp;

**run**;

**proc** **transpose** data = temp2 out=trans;

by group\_2;

**run**;

**proc** **format**;

value varfmt

**1** = "Weight"

**2** = "Arms"

**3** = "Hands"

**4** = "Forty"

**5** = "Ten"

**6** = "Vert"

**7** = "Broad"

**8** = "Bench"

**9** = "Height";

**run**;

**data** temp3;

set trans;

if \_name\_ = "avg\_weight" then name = **1**;

else if \_name\_ = "avg\_arms" then name = **2**;

else if \_name\_ = "avg\_hands" then name = **3**;

else if \_name\_ = "avg\_forty" then name = **4**;

else if \_name\_ = "avg\_ten" then name = **5**;

else if \_name\_ = "avg\_vert" then name = **6**;

else if \_name\_ = "avg\_broad" then name = **7**;

else if \_name\_ = "avg\_bench" then name = **8**;

else if \_name\_ = "avg\_height" then name = **9**;

else name = **10**;

format name varfmt.;

**run**;

symbol1 interpol=join value=dot;

**proc** **gplot** data = temp3;

plot col1\*name=group\_2;

**run**;

\*\*\* Profile Analysis Leads to the Same Y's to remove

Move on to Outlier Detection and MANOVA \*\*\*;

\*========================================================================================================================\*

Check for Outliers

\*========================================================================================================================\*;

%INCLUDE "\\Client\F$\Stat Classes\Current\Multivariate Data Analysis\Project1\mnorm.sas";

\*EXAMPLE 1;

%***MNORM***(DATA=C13.master\_4,CLASS=Group\_2 ,RESPONSE=vertical\_2 bench\_2 hands heightinchestotal ,ID=id)

**proc** **means** data = C13.master\_4\_mnorm mean median std;

var MNORM\_SMD;

**run**;

\*\*\* Mean is about 3.94 and STD is about 3.07 \*\*\*;

**data** outlier;

set C13.master\_4\_mnorm;

if MNORM\_SMD > **3.94** + (**3**\***3.07**) then Outlier = **1**;

else outlier = **0**;

**run**;

**proc** **sort** data = outlier;

by descending MNORM\_SMD;

**run**;

**proc** **print** data = outlier (obs=**20**);

var ID name MNORM\_SMD outlier;

**run**;

\*\*\* Limited Outliers (only 5) Assumption met \*\*\*;

\*========================================================================================================================\*

Profile Analysis Pre-MANOVA

\*========================================================================================================================\*;

\*\*\* Standardize the values for each possible Y \*\*\*;

**proc** **means** data = C13.master\_4;

var hands vertical\_2 bench\_2 heightinchestotal;

output out = standard mean = avg\_hands avg\_vert avg\_bench avg\_height

std = std\_hands std\_vert std\_bench std\_height;

**run**;

**proc** **sql**;

create table standard\_2 as

select \*

from C13.master\_4, standard;

**quit**;

**data** standard\_3;

set standard\_2;

s\_hands = (hands-avg\_hands)/std\_hands;

s\_vert = (vertical\_2-avg\_vert)/std\_vert;

s\_bench = (bench\_2-avg\_bench)/std\_bench;

s\_height = (heightinchestotal-avg\_height)/std\_height;

**run**;

\*\*\* Obtain the average of the standardized values and plot per group \*\*\*;

**proc** **means** data = standard\_3;

class group\_2;

var s\_hands s\_vert s\_bench s\_height;

output out = temp mean = avg\_hands avg\_vert avg\_bench avg\_height;

**run**;

**data** temp2 (drop= \_freq\_ \_type\_);

set temp;

**run**;

**proc** **transpose** data = temp2 out=trans;

by group\_2;

**run**;

**proc** **format**;

value re\_varfmt

**1** = "Hands"

**2** = "Vert"

**3** = "Bench"

**4** = "Height";

**run**;

**data** temp3;

set trans;

if \_name\_ = "avg\_hands" then name = **1**;

else if \_name\_ = "avg\_vert" then name = **2**;

else if \_name\_ = "avg\_bench" then name = **3**;

else if \_name\_ = "avg\_height" then name = **4**;

format name re\_varfmt.;

**run**;

symbol1 interpol=join value=dot;

**proc** **gplot** data = temp3;

plot col1\*name=group\_2;

**run**;

\*========================================================================================================================\*

MANOVA

\*========================================================================================================================\*;

**proc** **sort** data = C13.master\_4 out=test;

by group\_2;

**run**;

/\*==================\*

Order of Groups

"DB/WR"

"LB/DE/TE"

"OL/DT"

"RB"

\*==================\*/

**proc** **glm** data = C13.master\_4;

class group\_2;

model vertical\_2 bench\_2 hands heightinchestotal = group\_2;

manova h = group\_2;

contrast "DB/WR vs LB/DE/TE" group\_2 **1** -**1** **0** **0**;

contrast "DB/WR vs OL/DT" group\_2 **1** **0** -**1** **0**;

contrast "DB/WR vs RB" group\_2 **1** **0** **0** -**1**;

contrast "LB/DE/TE vs OL/DT" group\_2 **0** **1** -**1** **0**;

contrast "LB/DE/TE vs RB" group\_2 **0** **1** **0** -**1**;

contrast "OL/DT vs RB" group\_2 **0** **0** **1** -**1**; MANOVA H = \_ALL\_;

estimate "DB/WR vs LB/DE/TE" group\_2 **1** -**1** **0** **0**;

estimate "DB/WR vs OL/DT" group\_2 **1** **0** -**1** **0**;

estimate "DB/WR vs RB" group\_2 **1** **0** **0** -**1**;

estimate "LB/DE/TE vs OL/DT" group\_2 **0** **1** -**1** **0**;

estimate "LB/DE/TE vs RB" group\_2 **0** **1** **0** -**1**;

estimate "OL/DT vs RB" group\_2 **0** **0** **1** -**1**;

**run**;