5

Multivariate Analysis of Covariance

CHAPTER CONTENTS

Assumptions
Multivariate Analysis of Covariance
MANCOVA Example
Dependent Variable: Adjusted Means
Reporting and Interpreting
Propensity Score Matching
Summary
Web Resources
References

Source: http://www.york.ac.uk/depts/maths/histstat/people/cochran.gif

William Gemmell Cochran (July 15, 1909, to March 29, 1980) was born in Scotland and spent much of his career in the United States. He attended Glasgow University, receiving an MA degree in 1931, and attended Cambridge next, but never received a doctorate, choosing instead to follow Frank Yates to the Rothamsted Experimental Station. Later, during the end of his career, he did receive honorary doctoral degrees from the University of Glasgow in 1970 and Johns Hopkins University in 1975. He was influenced by John Wishart (Wishart distribution), as well as R. A. Fisher (experimental design) and Frank Yates (Yates correction factor in chi-square), with whom he worked at the Rothamsted Experimental Station, the United Kingdom. W. G. Cochran also worked with George Snedecor

(Continued)

(Continued)

and Gertrude Cox at the University of Iowa, and taught courses in experimental design and sample survey. His books *Experimental Design* (1950), *Sampling Techniques* (1953), and *Statistical Methods* (1967) with these colleagues were the prominent textbooks of the time period. He eventually ended up in the Department of Statistics at Harvard University, in 1957 and retired as professor emeritus in 1976. He received many awards during his career, including two from the American Statistical Association. He was editor of the *Journal of the American Statistical Association* from 1945 to 1950. His many contributions to the field of statistics also included the use of data transformations, analysis of variance with percents (dependent variable), analyses of matched sample data, goodness of fit tests, and issues related to the chi-square test developed by Karl Pearson (Anderson, 1980; Dempster & Mosteller, 1981; Watson, 1982).

William Gemmell Cochran (1934) was recognized for his distribution of quadratic forms in a random normal system with applications to analysis of covariance (ANCOVA). His Cochran theorem was expanded to show that ANOVA can be extended to situations requiring adjustment for covariate variables. He therefore postulated analyzing adjusted means in ANCOVA. His applied work in this area was from his agriculture experimental design work at Rothamsted, where he addressed the practical concerns of farmers and breeders. He further addressed problems in biomedical research with the development and use of clinical trials and development of research protocols.

The ANCOVA technique adjusts group means for the influence by other variables not controlled in the study, which are called extraneous variables. The extraneous variables are assumed to influence variation in the dependent variable and therefore controlled by statistical adjustment, since not controlled by random assignment. Random assignment in experimental research designs control for bias in subject selection and other threats to the internal validity of the research design, which is not present in quasiexperimental and other types of nonexperimental research designs (Campbell & Stanley, 1966). The ANCOVA assumptions are more stringent than the ANOVA assumptions.

△ Assumptions

The ANOVA assumptions are listed below, and when not met, alternative approaches have been suggested (Lomax & Hahs-Vaughn, 2012, pp. 309–331).

1. Observations are independent of each other
2. Homogeneity of variance (population variances of groups are equal)
3. Normal distribution of dependent variable(s)

ANCOVA requires the following additional assumptions:

4. Dependent variable continuous measure and fixed factor independent group variable
5. Relation between dependent and independent variables are linear
6. Covariate variables and independent variables are not related
7. The regression line for the groups are parallel
8. Homoscedasticity of regression slopes

The continuous dependent variable is required to calculate means. The fixed factor indicates exclusive group membership categories. The linearity assumption can be assessed by visual inspection of scatter plots and the Pearson correlation of *X* and *Y*. There are nonlinear ANCOVA methods, but these are not covered in this book (Huitema, 1980). The covariate variables should be related to the dependent variable and not to the independent variable (group). If the regression lines are not parallel for each group, then separate regression lines should be used for each group for prediction. Generally, this assumption is not checked, and a common regression line is fit for all the data with the common slope (beta weight) used for computing the adjusted means. To check whether lines are parallel for each group, introduce an interaction term in the model statement: *Posttest* = *Group* + *Pretest* + *Group* * *Pretest*. The *Group* term would test if groups had different intercepts, *Pretest* would yield a common slope value, and the interaction term (*Group* * *Pretest*) would test if the group regression lines were parallel. To check whether the variance around the regression line is the same for groups (homoscedasticity), we would compare the mean square error (*MSE*) from the separate group regression analyses. The basic ANCOVA procedures for computing separate regression equations and a common regression equation when assumptions are met have been presented in numerous multiple regression textbooks, for example, Pedhazur (1997).

△ Multivariate Analysis of Covariance

The use of covariate variables to adjust means is linked to two basic research design objectives: (1) eliminate systematic bias and (2) reduce the within-group error *SS*. The best way to address systematic bias is to use random sampling techniques; however, intact designs by definition are not formed using random sampling. For example, students who qualify for the Head Start program would be considered an intact group. When random assignment is not possible, then covariate adjustment of the means helps reduce systematic bias (intact groups that differ systematically on several variables). The within-group *SS* is due to individual differences among the subjects in a group. This can be addressed by selecting more homogeneous groups of subjects, using a factorial design with blocking on key variables, using repeated measures ANOVA, or using covariate variables to adjust group means. The purpose of MANCOVA is to adjust post means for initial differences in groups (generally based on pretest measures of intact groups, where random selection and random assignment to groups was not possible).

ANCOVA techniques combine ANOVA and multiple regression. ANOVA would test for mean differences (intercepts), while the multiple regression technique would provide a common slope to compute adjusted group means. MANCOVA is an extension of ANCOVA, where extraneous variables that affect the dependent variables are statistically controlled, that is, the dependent variable means are adjusted. The adjustment of dependent variable means in different groups, given a single covariate, is computed as follows:

$$\bar{Y}_{j(\text{adj})} = \bar{Y}_j - b_w(\bar{X}_j - \bar{X}),$$

where $\bar{Y}_{j(\text{adj})}$ = adjusted dependent variable mean in group j, $\bar{Y}_j$ = dependent variable mean before adjustment, b_w = common regression coefficient in entire sample, $\bar{X}_j$ = mean of covariate variable for group j, and $\bar{X}$ = grand mean of covariate variable (covariate variable mean for entire sample). Obviously, if the covariate means of each group are the same, then no adjustment to the dependent variable would occur, that is, groups are initially equal prior to any treatment or intervention in the research design.

MANCOVA Example

MANCOVA extends the univariate ANCOVA to include more than one dependent variable and one or more covariate variables. The null hypothesis in MANCOVA is that the adjusted population means of the dependent variables are equal. This is tested with Wilks's Λ. A basic example with two dependent variables, two groups, and one covariate variable is presented using data from Stevens (2009, p. 302). The two dependent variables are posttest scores (*Postcomp* and *Posthior*), groups (male = 1, female = 2), and covariate variable (*Precomp*).

We would first install and load the necessary packages to conduct the various analyses. Next, we input the data for the two groups into matrices, which are then combined into a data frame with variable labels. The data set, *mancova*, is attached so that the variable names can be used in the **manova()** function. The R commands are specified as follows:

```
# MANCOVA example (Stevens, 2009, p. 302)
# Install Packages

> install.packages("MASS")     # MANOVA
> install.packages("car")      # Type III SS
> install.packages("psych")    # Descriptive statistics

# Load packages

> library(MASS); library(car); library(psych)

# Input data

> group1 =
+ matrix(c(1,15,17,3,1,10,6,3,1,13,13,1,1,14,14,8,1,12,12,3,1,10,9,9,1,
+ 12,12,3,1,8,9,12,1,12,15,3,1,8,10,8,1,12,13,1,1,7,11,10,1,12,16,1,1,9,
+ 12,2,1,12,14,8),nrow=15,ncol=4,byrow=TRUE)


> group2 =
+ matrix(c(2,9,9,3,2,13,19,5,2,13,16,11,2,6,7,18,2,10,11,15,2,6,9,9,2,16,20,
+ 8,2,9,15,6,2,10,8,9,2,8,10,3,2,13,16,12,2,12,17,20,2,11,18,12,2,14,18,16),
+ nrow=14,ncol=4,byrow=TRUE)
```

```
> mancova = data.frame(rbind(group1,group2))
> names(mancova) = c("GPID","Precomp","Postcomp","Posthior")
> attach(mancova)
> mancova      # data set for MANCOVA
```

	GPID	Precomp	Postcomp	Posthior
1	1	15	17	3
2	1	10	6	3
3	1	13	13	1
4	1	14	14	8
5	1	12	12	3
6	1	10	9	9
7	1	12	12	3
8	1	8	9	12
9	1	12	15	3
10	1	8	10	8
11	1	12	13	1
12	1	7	11	10
13	1	12	16	1
14	1	9	12	2
15	1	12	14	8
16	2	9	9	3
17	2	13	19	5
18	2	13	16	11
19	2	6	7	18
20	2	10	11	15
21	2	6	9	9
22	2	16	20	8
23	2	9	15	6
24	2	10	8	9
25	2	8	10	3
26	2	13	16	12
27	2	12	17	20
28	2	11	18	12
29	2	14	18	16

The MANCOVA with the ANOVA summary table for Wilks's Λ and Type III *SS* is run on the data set. The R commands are as follows:

```
# MANCOVA

> options(scipen=999)   # print p-values in decimal rather than scientific
                                                              notation
```

```
> outcome = cbind(mancova$Postcomp,mancova$Posthior)
> model = manova(outcome ~ GPID + Precomp + GPID * Precomp, data = mancova)
> summary(model,test = "Wilks",type ="III")

> summary(model,test = "Wilks",type ="III")

               Df     Wilks  approx F num  Df  den Df        Pr(>F)
GPID            1   0.63952         6.7639   2       24  0.00468 **
Precomp         1   0.34489        22.7938   2       24  0.000002832 ***
GPID:Precomp    1   0.86301         1.9048   2       24  0.17068
Residuals      25
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note: Stevens (2009) ran separate models, thus degrees of freedom differed. I ran a single model with the results.

You can run the other summary commands to obtain the Pillai, Hotelling–Lawley, and Roy values. These statistics will have the same values as Wilks's Λ because of specifying Type III *SS*. Also, the order of entry for the variables will not affect the partitioning of the *SS*. Recall that Type I *SS* would yield different results due to variable entry order.

```
> summary(model,test = "Pillai",type = "III")
> summary(model,test = "Hotelling",type = "III")
> summary(model,test = "Roy",type = "III")
```

The findings indicated that the interaction effect was nonsignificant. Therefore, the assumption of parallel slopes holds, that is, the two groups have the same linear relation between the dependent variables and the pretest variable. The group means on the joint dependent variables were statistically significantly different ($F = 6.76$, $df = 2, 24$, $p = .005$). However, the covariate variable was also statistically significant. This indicated that the two groups had significantly different pretest means on *Precomp*, thus the two groups did not start out the same. The fact that the two groups were initially different forms the basis for us wanting to adjust the posttest means of the dependent variables by including the pretest variable in the model.

Dependent Variable: Adjusted Means

The **manova()** function with the pretest variable tests the adjusted means of the dependent variable. We can run the **lm()** function to obtain the regression slope values for an equation to compute the adjusted means,

but it is easier to use the **aov()** function. To see the original dependent variable means, use the **describeBy()** function in the *psych* package. The R command for the original dependent variable means is given as follows:

```
# Original Dependent variable means
> library(psych)
> describeBy(mancova, mancova$GPID)
group: 1
      vars  n  mean   sd  median  trimmed  mad  min  max  range  skew  kurtosis      se
GPID     1 15  1.00 0.00       1     1.00 0.00    1    1      0   NaN       NaN    0.00
Precomp  2 15 11.07 2.31      12    11.08 2.97    7   15      8 -0.20     -1.13    0.60
Postcomp 3 15 12.20 2.91      12    12.31 2.97    6   17     11 -0.31     -0.64    0.75
Posthior 4 15  5.00 3.72       3     4.77 2.97    1   12     11  0.46     -1.45    0.96
------------------------------------------------------------
group: 2
      vars  n  mean   sd  median  trimmed  mad  min  max  range  skew  kurtosis      se
GPID     1 14  2.00 0.00     2.0     2.00 0.00    2    2      0   NaN       NaN    0.00
Precomp  2 14 10.71 2.97    10.5    10.67 3.71    6   16     10 -0.05     -1.13    0.79
Postcomp 3 14 13.79 4.56    15.5    13.83 5.93    7   20     13 -0.16     -1.75    1.22
Posthior 4 14 10.50 5.37    10.0    10.33 6.67    3   20     17  0.19     -1.26    1.44
```

The ANCOVA summary table, **aov()** function, for the two dependent variables using just the pretest variable and group membership variable are listed. The **effect()** function for adjusted means of each dependent variable is run after each ANCOVA. The R commands for each are listed below with their corresponding output.

```
# Postcomp ANOVA

> factor(GPID)
> modelA = aov(Postcomp ~ Precomp + GPID, data = mancova)
> summary(modelA,type="III")


            Df  Sum Sq  Mean Sq  F value       Pr(>F)
Precomp      1   237.7   237.69   43.899    0.0000005 ***
GPID         1    28.5    28.50    5.263    0.0301 *
Residuals   26   140.8     5.41
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> library(effects)
> adjmeanA = effect("GPID",modelA,se=TRUE,xlevels=2)
```

```
> summary(adjmeanA)
> adjmeanA$se

> summary(adjmeanA)

 GPID effect
GPID
  1  2
12.00555 13.99406    # Adjusted Postcomp Means

 Lower 95 Percent Confidence Limits
GPID
  1  2
10.76916 12.71417

 Upper 95 Percent Confidence Limits
GPID
  1  2
13.24193 15.27394
> adjmeanA$se
[1] 0.6014924 0.6226546


# Posthior ANOVA and adjusted dependent variable means

> modelB = aov(Posthior ~ Precomp + GPID, data = mancova)
> summary(modelB, type = "III")

            Df   Sum Sq   Mean Sq   F value     Pr(>F)
Precomp      1     17.7     17.66     0.821    0.37319
GPID         1    211.6    211.59     9.836    0.00422 **
Residuals 26     559.3     21.51
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> adjmeanB = effect("GPID",modelB,se=TRUE,xlevels=2)
> summary(adjmeanB)
> adjmeanB$se

> summary(adjmeanB)

 GPID effect
```

```
GPID
  1  2
 5.039439 10.457744       # Adjusted Posthior means

 Lower 95 Percent Confidence Limits
GPID
  1  2
2.575048 7.906650

 Upper 95 Percent Confidence Limits
GPID
  1  2
 7.50383 13.00884
> adjmeanB$se
[1] 1.198908 1.241089
```

The two separate ANOVA tables indicate that both dependent variables are contributing to the overall multivariate significance. It also helps our understanding of how the two dependent variables interact with the pretest variable. *Postcomp* group mean differences were statistically significant with a statistically significant pretest, *Precomp*. *Posthior* group mean differences were statistically significant, but there was no significant pretest difference. In MANCOVA, these two different ANOVA findings are taken together to yield significant group posttest adjusted mean differences.

MANCOVA tests the differences in the adjusted posttest means. It helps compute the original dependent variable means and compare them with the adjusted dependent variable means.

The R commands to compute the posttest means, standard deviations, and pretest means for each group and the entire sample are shown below.

```
# Postcomp Descriptive Statistics Males (n = 15) Females (n = 14)

> mean(Postcomp[1:15]); sd(Postcomp[1:15]);mean(Precomp[1:15]) #GPID1
> mean(Postcomp[16:29]); sd(Postcomp[16:29]);mean(Precomp[16:29]) #GPID2
> mean(Postcomp); sd(Postcomp); mean (Precomp) # Grand Mean

> mean(Postcomp[1:15]); sd(Postcomp[1:15]); mean(Precomp[1:15]) #GPID1
[1] 12.2            # Postcomp mean - group 1 males
[1] 2.908117
[1] 11.06667
>
```

```
> mean(Postcomp[16:29]); sd(Postcomp[16:29]); mean(Precomp[16:29]) #GPID2
[1] 13.78571        # Postcomp mean - group 2 females
[1] 4.56034
[1] 10.71429
>
> mean(Postcomp); sd(Postcomp); mean (Precomp)
[1] 12.96552        # Postcomp grand mean - total sample
[1] 3.812412
[1] 10.89655


# Posthior Descriptive Statistics

> mean(Posthior[1:15]); sd(Posthior[1:15]); mean(Precomp[1:15]) #GPID1
> mean(Posthior[16:29]); sd(Posthior[16:29]); mean(Precomp[16:29]) #GPID2
> mean(Posthior); sd(Posthior); mean (Precomp)

> mean(Posthior[1:15]); sd(Posthior[1:15]); mean(Precomp[1:15]) #GPID1
[1] 5               # Posthior mean - group 1 males
[1] 3.722518
[1] 11.06667

> mean(Posthior[16:29]); sd(Posthior[16:29]); mean(Precomp[16:29]) #GPID2
[1] 10.5            # Posthior mean - group 2 females
[1] 5.374441
[1] 10.71429

> mean(Posthior); sd(Posthior); mean (Precomp)
[1] 7.655172 # Posthior grand mean - total sample
[1] 5.306841
[1] 10.89655
```

The descriptive statistics for the two dependent variables for each group can now be summarized together. Table 5.1 presents the original dependent variable means and the adjusted dependent variable means.

The separate ANCOVA results indicated that the pretest related differently with each of the dependent variables. The correlation between *Postcomp* and *Precomp* was $r = .764$, which was statistically significant. The correlation between *Posthior* and *Precomp* was $r = -.1496$, which was not statistically significant. To obtain the different correlations between the covariate variable and each dependent variable use the following R commands.

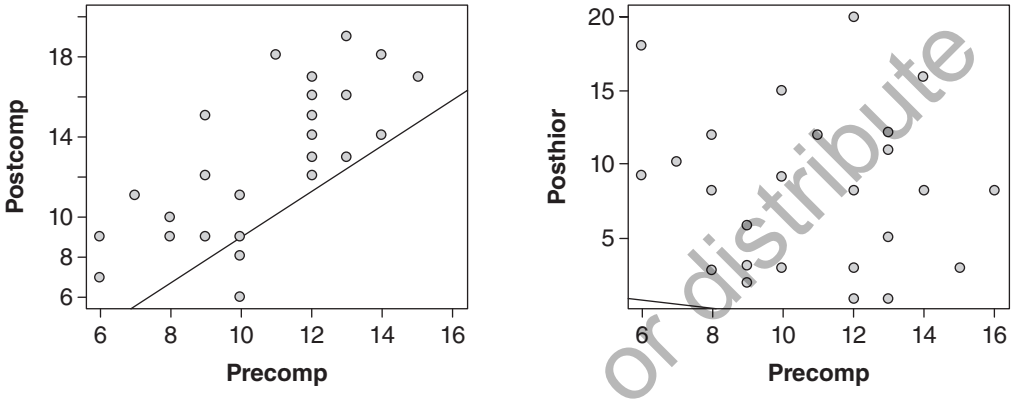
Table 5.1 Original and Adjusted Dependent Variable Means

Dependent Variables	Original Mean	Adjusted Mean
Postcomp		
Group 1: Males	12.20	12.00
Group 2: Females	13.79	13.99
Posthior		
Group 1: Males	5.00	5.04
Group 2: Females	10.50	10.45

```
> cor.test(Precomp,Postcomp)
> cor.test(Precomp,Posthior)

> cor.test(Precomp,Postcomp)

  Pearson's product-moment correlation

data: Precomp and Postcomp

t = 6.1573, df = 27, p-value = 0.000001399
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5524739 0.8833240
sample estimates:
 cor
0.7642338

> cor.test(Precomp,Posthior)

  Pearson's product-moment correlation

data: Precomp and Posthior

t = -0.7865, df = 27, p-value = 0.4384
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.4893260 0.2294296
sample estimates:
  cor
-0.1496606
```

A graph of the relation between the covariate and each dependent variable can be viewed using the following R commands.

```
> par(mfrow=c(2,2),usr=c(0,20,0,20))   # Dimension 2 by 2 graph and axis from
                                                                    0 to 20
> plot(Precomp,Postcomp)
> abline(modelA)
> plot(Precomp,Posthior)
> abline(modelB)
```



The covariate (*Precomp*) is significantly correlated with the *Postcomp* variable ($r = .76$, $p < .001$), but not the *Posthior* variable ($r = -.15$, $p = .44$). The graphs visually display the pretest scores relation with the *Postcomp* and *Posthior* scores. From a design perspective, this could be a mismatched situation. Each dependent variable would normally have its own pretest measure, so *Precomp* would not be used to adjust means for *Posthior*.

Reporting and Interpreting

The MANCOVA technique should meet all the assumptions of the MANOVA technique and report that the additional assumptions for the MANCOVA technique were met. A basic write-up for reporting MANCOVA results would be as follows:

The dependent variables were continuous, linear, and normally distributed variables with equal variance–covariance matrices between the groups; thus, met the MANOVA assumptions. In addition, the dependent and covariate variables were linear, and the two groups had parallel lines with homoscedasticity, thus had equal slopes and variances, which met the additional assumptions for MANCOVA. This was indicated by a nonsignificant group by pretest interaction ($F = 1.90$, $df = 2, 24$, $p = .17$). The covariate variable was statistically significant ($F = 22.79$, $df = 2, 24$, $p < .001$), which indicates that the groups were initially different on the pretest, thus requiring adjustment to the posttest means. The groups were statistically different on the adjusted posttest means ($F = 6.76$, $df = 2, 24$, $p = .004$). Females had higher dependent variable posttest means than males.

The Stevens (2009) data set was chosen because it points out the difficulty in meeting the MANCOVA assumptions, which are in addition to the MANOVA assumptions (not shown). The example showed the importance of conducting univariate *F* tests for each dependent variable and covariate variable. The results indicated that *Precomp* was correlated with *Postcomp*, but not with *Posthior*; groups were different on *Postcomp* scores and *Precomp* was a significant pretest; and groups were different on *Posthior* scores, but no significant pretest was indicated, thus the two univariate analyses had different results. The multivariate analysis combines the individual variable effects; thus, sometimes it can mask the different univariate results.

It is difficult to meet the ANCOVA assumptions, yet researchers continue to use the technique despite violating the assumptions. On the surface, the statistical control for pretest differences falls short. Researchers have sought other methods when unable to conduct an experimental design with random assignment to control for threats to internal validity (Campbell & Stanley, 1966). Matching or blocking on key variables has been recommended, which aids in the selection of similar subjects for a comparison group.

Critics of ANCOVA point out drawbacks to making statistical adjustments to means over random assignment of subjects to groups. Two issues cited were that the inclusion of covariate variables changes the criterion variable (dependent variable) such that the adjusted means change the construct (Tracz, Nelson, Newman, & Beltran, 2005), and the adjusted means technique does not match the research question of interest, but propensity score analysis with unadjusted posttest scores will (Fraas, Newman, & Pool, 2007). I therefore turn my attention to the propensity score method.

△ Propensity Score Matching

In experimental research designs, random assignment would control for bias in subject selection and other threats to internal validity; however, in nonexperimental research designs, matching subjects on the covariate variable(s) is generally recommended rather than statistical adjustment to the means. Propensity score methods have been advocated in place of previous matching or blocking methods (D'Agostino, 1918). Propensity score matching (PSM) uses covariate variables to obtain a matched sample of subjects (Ho, Imai, King, & Stuart, 2007). There are different PSM methods, so a researcher should exercise care in using PSM (Schumacker, 2009). The R software has propensity score packages available (McCaffrey, Ridgeway, & Morral, 2004—R *twang* package with **mnps()** function; Ho, Imai, King, & Stuart, 2007—R *MatchIt* package with **matchit()** function to run various types of propensity score methods).

An SPSS data set with freshman students at a southern university was used to select a matching sample (International Baccalaureate Organization [IBO], 2014). The data consisted of entering freshman students in 2007 and included gender, race, ethnicity, graduation status, and grade point averages for the 2007 to 2010 academic years. The researcher wanted to test GPA (grade point average) mean difference between AP (Advanced Placement) and IB (International Baccalaureate) students across the 2007 to 2010 academic years, however, the number of AP students outnumbered the IB students at the university. Specifically in 2007, there were $n = 279$ IB freshman students compared with $n = 6{,}109$ AP freshman students at the university.

Propensity score analysis was conducted to select a matching group of 279 AP freshman students at the university (Austin, 2011; Guo & Fraser, 2014; Holmes, 2014). In the study, gender, race, and graduation status were used as covariates when selecting a matching group of AP freshman students. R software was used with the *MatchIt* package using the "nearest neighbor" selection criteria with the covariates (http://www.r-project.org/). The R script to read in the SPSS data file, select a matching group of students, then write out the IDs to a file is given below. The file of IDs were then used in SPSS to select the matching AP students. The total number of freshman students was $N = 558$ (IB = 279 students; AP = 279 students). The R script file commands were as follows:

```
# Propensity score matching - nearest neighbor matching

> install.packages("MatchIt")
> install.packages("Hmisc")
> library(MatchIt)
> library(Hmisc)

# Data is from SPSS file with selected covariate variables
# status, gender, race, graduation
# Save SPSS dataset in transport format
# Use SAVE AS command in SPSS
# OR export outfile='c:\propensity.por'.

# Read in SPSS data set in R with value labels - last option converts value labels to R factors

> mydata = spss.get("C:/propensity.por", use.value.labels=FALSE)
> mydata

# Missing values corrected in SPSS file or finaldata = as.data.frame(na.omit(mydata))
# Matching is performed using propensity scores with covariate variables
# Data set must not have missing values and Y variable must be 0 or 1 coded
```

```
> m.out = matchit(STATUS~GENDER+RACE+GRADUATE, method="nearest", data=mydata, + ratio = 1)
> m.out

# Final matched data saved as final.data

> final.data = match.data(m.out)

# Set directory to save file

> setwd("C:/")

# Write out the data file

> write.table(final_data,file="match1",sep=" ", row.names=TRUE, col.names = TRUE,
+ quote=FALSE)
```

In PSM, it is important to check that the two samples are equivalent on the covariate variables used in the matching process: Did PSM achieve similar numbers of AP students across gender, race, and graduation completion? A chi-square analysis of *status* by *gender*, *race*, and *graduation* are presented in Tables 5.2, 5.3, and 5.4, respectively. Table 5.2 indicates the cross-tabulation of AP and IB students with *gender* ($\chi^2 = 1.54$, $p = .21$). Table 5.3 indicates the cross-tabulation of AP and IB students with *race* ($\chi^2 = 5.27$, $p = .15$). Table 5.4 indicates the cross-tabulation of AP and IB students with *graduation* ($\chi^2 = .23$, $p = .62$). The chi-square statistics for all the propensity score analyses were nonsignificant, which indicated that the PSM did provide a matching number of AP to IB freshman students across the covariate variables.

The ability to obtain a matched sample of subjects permits statistical analysis of mean differences on dependent variables without having to meet the assumptions in ANCOVA. It also doesn't change the construct or test the wrong hypothesis by using adjusted means. Overall, the matching of subjects provides a sound research design option that does not involve statistical adjustments to means.

Table 5.2 Status by Gender

Group	Male	Female	Total
AP	91	188	279
IB	105	174	279

Note: $\chi^2 = 1.54$, $p = .21$.

Table 5.3 Status by Race

Group	White	Black	Asian	Other	Total
AP	163	19	41	56	279
IB	149	29	54	47	279

Note: $\chi^2 = 5.27$, $p = .15$.

Table 5.4 Status by Graduation

Group	No	Yes	Total
AP	19	260	279
IB	22	257	279

Note: $\chi^2 = .23$, $p = .62$.

SUMMARY

MANCOVA combines the approach of testing mean differences with the multiple regression approach of estimating slope, or rate of change. Basically, the dependent variable means are adjusted based on the correlation relation of one or more covariate variables. The intent is to statistically adjust for group pretest differences, thus equating groups at the beginning of a research design. This statistical adjustment of the dependent variable means has been scrutinized because it changes the meaning of the dependent variable. In different disciplines, the research design doesn't permit the random selection and assignment to groups due to intact groups; thus, alternative methods have been advocated. Recently, the PSM approach has been advocated to select a matching set of subjects based on the set of similar values on covariate variables. In practice, the random selection and random assignment of subjects to experimental and control groups is the gold standard to control for threats to internal validity.

WEB RESOURCES

Introduction to Propensity Score Matching—UseR! 2013 Conference
http://jason.bryer.org/talks/psaworkshop.html

Software for Propensity Score Matching
http://www.biostat.jhsph.edu/~estuart/propensityscoresoftware.html

Video on Propensity Score Matching Using R
http://www.youtube.com/watch?v=Z8GtYGESsXg

REFERENCES

Anderson, R. L. (1980). William Gemmell Cochran 1909–1980: A personal tribute. *Biometrics, 36,* 574–578.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399–424.

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin.

Cochran, W. G. (1934). The distribution of quadratic forms in a normal system with applications to analysis of covariance. *Proceedings of Cambridge Philosophical Society, 30*(2), 178–191.

D'Agostino, R. B. (1918). Tutorial in biostatistics: Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine, 17,* 2265–2281.

Dempster, A. P., & Mosteller, R. (1981). In Memoriam. William Gemmell Cochran 1909–1980. *The American Statistician, 35*(1), 38.

Fraas, J. W., Newman, I., & Pool, S. (2007). The use of propensity score analysis to address issues associated with the use of adjusted means produced by analysis of covariance. *Multiple Linear Regression Viewpoints, 33*(1), 23–31.

Guo, S., & Fraser, M. W. (2014). *Propensity score analysis: Statistical methods and applications* (2nd ed.). Thousand Oaks, CA: Sage.

Ho, D., Imai, K., King, G., & Stuart, E. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis, 15*(3), 199–236.

Holmes, W. M. (2014). *Using propensity scores in quasi-experimental designs*. Thousand Oaks, CA: Sage.

Huitema, B. E. (1980). *The analysis of covariance and alternatives*. New York, NY: Wiley.

International Baccalaureate Organization. (2014). *Final report: A comparison of IB and non-IB incoming freshman students*. New York, NY: Author.

Lomax, R. G., & Hahs-Vaughn, D. L. (2012). *An introduction to statistical concepts* (3rd ed.). New York, NY: Routledge (Taylor & Francis Group).

McCaffrey, D., Ridgeway, G., & Morral, A. (2004). Propensity score estimation with boosted regression for evaluating adolescent substance abuse treatment. *Psychological Methods, 9*(4), 403–425.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research*: *Explanation and prediction* (3rd ed.). Orlando, FL: Harcourt Brace College.

Schumacker, R. E. (2009). Practical issues to consider before using propensity score analysis. *Multiple Linear Regression Viewpoints, 35*(2), 1–3.

Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). New York, NY: Psychology Press.

Tracz, S. M., Nelson, L. L., Newman, I., & Beltran, A. (2005). The misuse of ANCOVA: The academic and political implications of Type VI errors in studies of achievement and socioeconomic status. *Multiple Linear Regression Viewpoints, 31*(1), 19–24.

Watson, G. S. (1982). William Gemmell Cochran 1909–1980. *Annals of Statistics, 10,* 1–10.