An example of an APA-style write-up for

the Repeated Measures Analysis of Variance and Multivariate Analysis

of Variance lab example

by

Michael Chajewski

Fordham University

Department of Psychology, Psychometrics

An example of an APA-style write-up for the

Repeated Measures Analysis of Variance lab example

Within the many branches of the social and behavioral sciences the repeated measures model is one of the most frequently used and applied designs. This stems largely from the general practice of wanting to observe a certain behavior over a given period of time. For many developmental psychologists the longitudinal research design is somewhat of a holy grail, which lends itself to gathering insight into the progressive nature of the human psyche. For many biologists, sociologists, anthropologists and many other fields a temporal component to their analysis allows for a much greater explication of observed complexities.

The example provided in class was aimed at addressing two distinct tasks: 1) to give students the opportunity to demonstrate their understanding and usage of the repeated measures Analysis of Variance (ANOVA) and Multivariate Analysis of Variance (MANOVA) procedures and 2) familiarize students with the appropriate reporting style for statistical results (used in both report and publication writing) as delineated by the American Psychological Association (APA).

Method

Participants

The participants analyzed in this exercise consisted of 125 males (50.20%) and 124 females (49.80%) students from an upstate New York high school. This high school was consolidated from two other schools in an effort to relocate resources to a central institution. These students ranged in their ages from 13 to 19 during the cumulative duration of the study (across the three time points). Tree distinct academic tracks were monitored, *Vocational Educational Training* (VET), *College Preparatory* (CP) and *Advanced Placement* (AP). Each track had 83 students in it.

Materials

A psychological battery assign stress interference was administered to every student at the high school at four distinct time points in their high school careers, at the inception of their high school education, in the middle of their sophomore, junior and senior years. Further, for the educational benefit of the graduate students analyzing the dataset an instructional guide to the assignment was handed out.

Design and Procedures

Each student's stress level was measured once a year and recorded by the school's psychological staff. Students were assessed at the beginning of their freshmen year one month into the inception of the academic semester. In subsequent measures were taken around late November in the students' subsequent sophomore, junior and senior years. Data was compiled following the four year repeated measures study.

The purpose of the study was to investigate whether there are any significant differences in the coping mechanisms of students between different academic tracks. Further, the local School Board wanted to investigate individual differences between genders and their interactive nature with the corresponding academic tracks students were enrolled in. A particular believe delineated by the Board was that students in the *Vocational Education Training* track would experience significantly more stress as the time of graduate from high school approached.

Results

The results of the example were twofold. The first part of the analysis took a univariate approach that is most commonly recognized, that of the Repeated Measures Analysis of Variance (RM-ANOVA). The second part focused on the multivariate generalization applying a MANOVA approach to the same data. The purpose of the exercise was to demonstrate the versatility and broader application of both tests and the utility associated with one over the other. While the general approaches are fairly similar there are fundamental differences between the assumptions as well as the subsequent follow-up tests needed to be conducted.

A first investigation of the data revealed the separate means on the stress measure for each of the factorial groups (here *Gender* by *Track*). The six groups are summarized in Table 1.

Table 1. Univariate statistics for variables in analysis (standard errors)

		Means			
Group	N	Base	T1	T2	T3
Male VET	39	6.23 (.40)	5.95 (.30)	5.79 (.30)	5.97 (.42)
Male CP	42	6.71 (.38)	5.48 (.29)	5.90 (.29)	6.57 (.34)
Male AP	44	6.00 (.47)	5.73 (.31)	5.54 (.38)	9.18 (.37)
Female VET	44	6.54 (.34)	7.34 (.28)	8.41 (.30)	5.95 (.30)
Female CP	41	7.15 (.40)	7.41 (.32)	9.22 (.32)	10.00 (.41)
Female AP	39	6.64 (.53)	8.20 (.36)	9.18 (.30)	10.15 (.41)

From Table 1 it follows that female students in the CP track experienced the highest amount of stress at the beginning of high school. This, however, is not significantly greater as compared to this groups corresponding female and male other-track-groups.

Also, a simple bivariate correlation was computed between all of the stress scores (pooled across all of the groups) for each of the four possible time point pairings. Table 2 shows that the lowest correlation was observed between the first year (also referred to as the baseline) and the second year of study (this correlation was also not significant). Conversely, the highest correlation was between the sophomore and the junior year, $r = .62$, $p < .05$.

Table 2. Variable correlations (*p*-value)

	Freshmen	Sophomore	Junior	Senior
Freshmen (Base)	1.00 (--)			
Sophomore (T1)	.06 (.36)	1.00 (--)		
Junior (T2)	.13 (.04)	.62 (<.001)	1.00 (--)	
Senior (T3)	.12 (.05)	.58 (<.001)	.59 (<.001)	1.00 (--)

It is a feature of longitudinal designs that time points closer together are more highly correlated than those further apart. This observation however does not seem to hold in this study, given that the lowest correlation is between two adjacent time points. Also, it can be inferred from Table 2 that all of the correlations between the baseline measure (the stress level at the beginning of the freshmen year) and all of the later time points are the lowest in the dataset.

Multivariate normality of the data was investigated using information from two sources. First, the multivariate interrelationship between all response variables (stress scores at each time point) was assessed using individual factorial group by-case computed leverage values. Critical cut-off values for these were computed based on the corresponding Mahalonobis Distance critical chi-square values with the appropriate group sample size and an alpha level of .01. As can be seen in Table 3 none of the group's maximal leverage values exceeded the critical cut-off. From this we can infer that given the data there are no multivariate outliers in the dataset. Secondly, the multivariate skewness and kurtosis were investigated using the SAS macro *multnorm2* which produces Mardia's skewness, kurtosis and the chi-square Q-Q plot test statistic Henze-Zirkler *T* which assesses whether the dataset follows an expected multivariate normal distribution.

Table 3. Assessment of leverage values and multivariate normality (*p*-values)

	Leverage		Mardia's Coefficients		
Group	Max	Critical	Skewness	Kurtosis	Henze-Zirkler *T*
Male VET	.35	.38	24.42 (.22)	-.39 (.69)	-1.23 (.22)
Male CP	.30	.35	23.63 (.26)	-0.60 (.55)	1.42 (.16)
Male AP	.27	.33	13.79 (.84)	-1.56 (.12)	0.34 (.73)
Female VET	.30	.33	17.21 (.64)	-0.45 (.65)	-0.80 (.42)
Female CP	.25	.36	16.93 (.66)	-1.11 (.26)	1.12 (.26)
Female AP	.38	.38	26.65 (.14)	0.29 (.77)	-1.05 (.29)

Again, the three groups showed no deviation from an assumed multivariate normal distribution. None of the distributional coefficients were significant suggesting a multivariate normal distribution of the data. With this particular assessment it was appropriate to proceed with a multivariate analysis.

RM-ANOVA

The RM-ANOVA was conducted jointly with the multivariate test in the statistical software package SAS ®. One of the core underlying assumptions in the univariate RM-ANOVA procedure is that of sphericity. Sphericity, a special case of circularity assumptions, checks whether the variance/covariance matrix of the observed data follows a particular pattern. This pattern is usually identified as one with equal variances in the diagonal, and equal covariance in the off-diagonal elements. Given the earlier discussed nature of longitudinal data it is highly unlikely that this assumption will hold. Nonetheless, if sphericity is observed the RM-ANOVA procedure provides a powerful test about repeated measures.

In order to test sphericity we inspected Mauchly's Test which tests for the equivalence of the hypothesized and the observed variance/covariance patterns. The test was highly significant, $W = .49$, χ^2 (5) = 171.54, $p < .001$, suggesting that the observed matrix does not have approximately equal variances and equal covariances. This suggests that using an uncorrected RM-ANOVA F-test would result in a likely inflation of Type I Errors, rejecting the null hypothesis while it was true more often that generally accepted. Several corrections have been proposed, most notably the Greenhouse-Geisser and Huynh-Feldt epsilon corrections. These do not affect the computed *F*-statistic, but instead raise the critical *F* value needed to reject the null hypothesis. For our data these corresponding corrective coefficients were: Greenhouse-Geisser ε = .66 and Huynh-Feldt ε = .68.

Table 4 summarizes the results of the RM-ANOVA analysis. The column labeled *F* gives the *F* value of the test followed by three columns of significance values. The last two columns represent the corrected significance levels for the observed statistic given the above reported corrective coefficients. It follows that there is a significant change in the stress scores across time, F (3, 729) = 29.03, $p < .05$. Moreover, both *Gender*, academic *Track* type and their interaction was found to also be significant across the time points. In order to investigate the temporal relationships with the two categorical variables appropriate follow-up contrasts were investigated comparing all time points against the initial score (baseline).

Table 4. Repeated measures Analysis of Variance

Effect	*MS*	*df*	*F*	*p*	Greenhouse-Geisser	Huynh-Feldt
Time	107.07	3	29.03	< .001	< .001	< .001
Time x Gender	79.53	3	21.56	< .001	< .001	< .001
Time x Track	69.99	6	18.98	< .001	< .001	< .001
Time x Gender x Track	15.28	6	4.14	< .001	< .01	< .01
Error	3.69	729				

Three contracts produced significant results. The comparison between Freshmen and Sophomores student's scores, for *Gender* only, was significant, F (1, 243) = 12.20, $p < .001$. Similarly the contrast between Freshmen and Juniors stress scores was also significant for *Gender*, F (1, 243) = 43.49, $p < .001$. The comparison of Freshmen and Seniors stress scores revealed to be significant for all three investigated effects, *Gender*, *Track* and *Gender x Track* interaction, producing the significant F values F (1, 243) = 5.37, $p < .05$, F (1, 243) = 25.57, $p < .001$, F (1, 243) = 5.59, $p < .01$ respectively.

However, since the original assumption of sphericity was not met it would be advisable to either continue the analysis with a multivariate approach given that the dataset is relatively large. If the dataset is limited the multivariate approach would be discouraged and the RM-ANOVA with the appropriate corrections would be preferred. In this particular case there is sufficient data for a multivariate analysis to be carried out, in addition to having a balanced design with an approximately equal number of students in each crossed *Gender* by *Track* group.

MANOVA

Since the multivariate approach analyses the repeated measures data similarly as though it would compute a regular MANOVA other assumptions than those observed in the RM-ANOVA procedure apply. We have already demonstrated that the data follow a multivariate normal distribution, however, one of the assumption for MANOVA is the equality of variance/covariance matrices of the different groups analyzed. A Bartlett's test was conducted investigating this assumption and was found to be not significant, $\chi^2(50) = 53.14$, $p = .35$, *n.s.* This indicates that the six groups analyzed have roughly equal variances/covariances.

All of the multivariate tests (here Wilk's lambdas) were significant (Table 5). This suggests that the stress scores across the four time points have at least one mean vector pairing which produced a significant difference.

Table 5. Multivariate tests (all significant at $p < .001$)

Effect	Λ	*F*	df_1	df_2
Time	.66	40.72	3	241
Time x Gender	.75	27.23	3	241
Time x Track	.53	30.05	6	484
Time x Gender x Track	.84	7.30	6	484

Though the multivariate test informs us of the significance of at least one mean pairing it is unclear from the multivariate test for which individual comparison (for which contrast between groups) the observed mean difference is significant. In order to determine the significance of these differences a series of univariate ANOVAs are conducted. From Table 6 we can deduce that there were no significant differences between the delineated groups on their average stress level in the freshmen year of high school. However, there were observed *Gender* differences in stress scores in the subsequent sophomore and junior years. Only in the last, senior year, did both *Gender* and *Track* as well as their interaction produce a significant difference between groups. Since the overall focus was on the cross classification of genders and academic tracks the interaction (factorial group) would be of most interest.

Table 6. Significant *F*-tests for univariate follow up tests (all significant at $p < .001$)

Time point	Effect	*MS*	*F*	df_1	df_2
Sophomore	Gender	232.78	58.24	1	243
Junior	Gender	630.97	149.11	1	243
Senior	Gender	132.41	22.58	1	243
	Track	289.64	49.40	2	243
	Gender x Track	65.30	11.14	2	243

Much like with the multivariate omnibus tests, the univariate ANOVA follow-ups do not provide specific mean difference, but rather overall group effects for any time point by any on of the effects of interest. In order to investigate the specific mean difference individual *t*-tests about the mean difference need to be conducted. Further, in order to prevent alpha inflation at this level of the analysis a Bonferroni correction for multiple comparisons will be applied. Also, simultaneous 95% confidence intervals were computed in order to provide further insight into the variability of plausible mean differences between the observed groups. Table 7 summarizes the significant findings for the individual *Gender*, *Track* and *Gender* by *Track* group comparisons for the three time points with significant ANOVAs. It follows that the largest mean difference was observed between females in the VET track and females in the AP track in the last year of high school. The smallest significant difference was between the CP and AP tracks (pooled across genders) during the senior year.

Table 7. Significant mean difference *t*-tests (all significant at $p < .001$)

				95% Simultaneous confidence interval	
Time point	Effect	Compare	Mean difference	Lower	Upper
Sophomore	Gender	M-F	-1.94	-2.44	-1.44
Junior	Gender	M-F	-3.19	-3.70	-2.67
Senior	Gender	M-F	-1.46	-2.06	-.85
	Track	VET-CP	-2.32	-3.23	-1.41
		VET-AP	-3.70	-4.61	-2.80
		CP-AP	-1.38	-2.29	-.48
	Gender x Track	MVET-MAP	-3.21	-4.79	-1.63
		MVET-FCP	-4.02	-5.63	-2.42
		MVET-FAP	-4.18	-5.80	-2.55
		MCP-MAP	-2.61	-4.16	-1.06
		MCP-FCP	-3.43	-5.00	-1.85
		MCP-FAP	-3.58	-5.18	-1.98
		MAP-FVET	3.23	1.70	4.76
		FVET-FCP	-4.04	-5.60	-2.49
		FVET-FAP	-4.20	-5.78	-2.62

Note. A Bonferroni correction has been applied for multiple comparisons

Discussion

Investigating the mean differences observed in the multivariate analysis of repeated measures, it is not entirely clear what patter those means produce. In order to have better insight into the six factorial groups a means plot was produced depicting each group's stress score mean for each of the four time points.

Figure 1. Comparative means plot for the six factorial groups across the time points

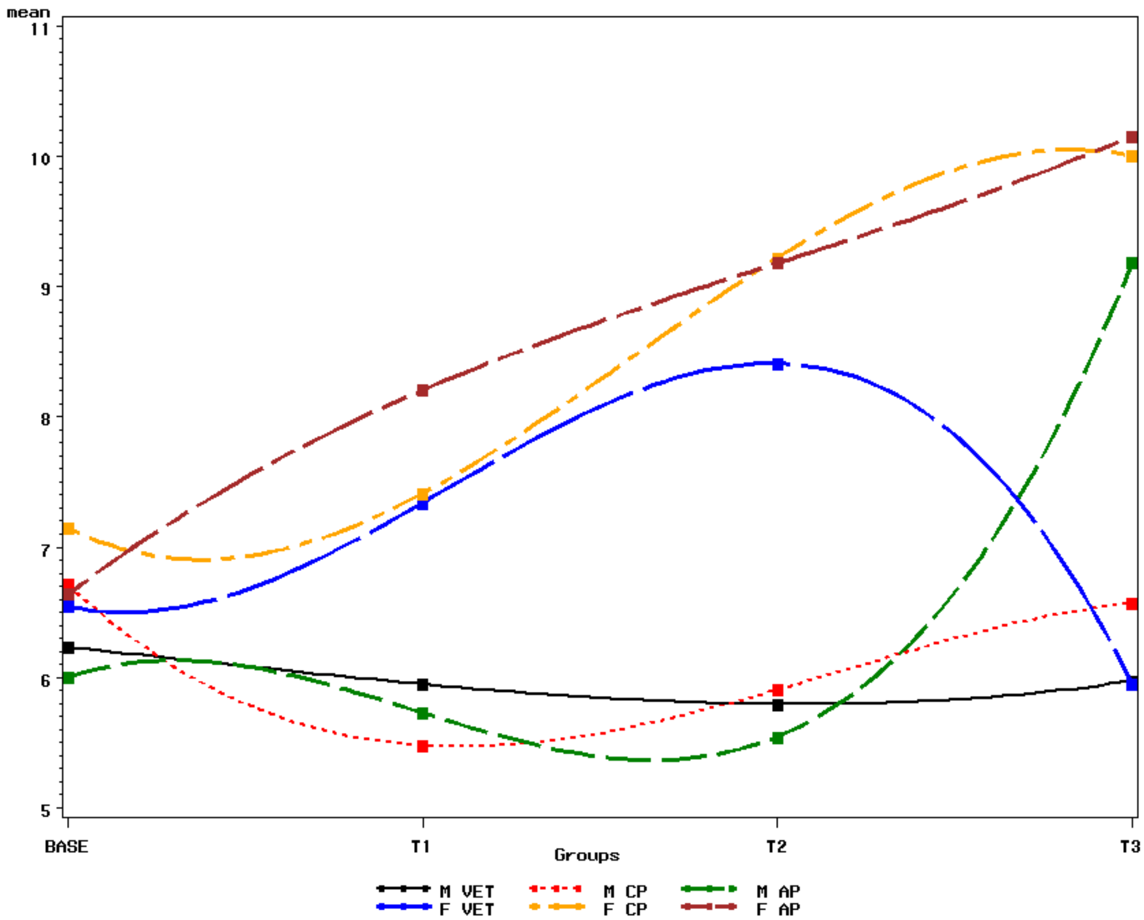

Figure 1 further supports the absence of any significant mean differences for the freshmen year, seeing as the mean scores for all six groups are distributed within a single point (between 6 and 7) from one another. As time progresses the means in stress scores become more divergent, most apparently so for *Gender*. Both the RM-ANOVA and the MANOVA approach detected a significant change in mean scores across the time points related to gender differences. Seemingly it would appear that females in this high school experience greater stress then their male counterparts in the second and third years of their high school education.

However, the most interesting observation is in the senior year when it would appear the interaction between *Gender* and *Track* takes effect. Here we can see that females in the VET track drop significantly in their stress levels (joining their male VET counterparts), whereas the male AP track students show a significant increase in stress.

Following from the School Boards original hypothesis we can conclude that their original assumptions about the gender and academic track interaction was not supported. Alternatively, it would seem that the exact opposite has taken effect, where students in the Vet track (both male and female) experience less stress. This could be due to the fact that perhaps mid academic year these students are not yet faced with the need to make projective choices, whereas students in the CP and AP tracks that may want to enter institutions of higher learning at this point already need to have taken several placement examinations, college entrance examinations and filed/selected their applications to colleges of preference. It would further seem that females are generally more stressed then males. Except for the VET track females have the highest stress scores their senior year in high school. In fact males in the CP track do not differ significantly from their Vet program counterparts. This could suggest that he real effect lies in the AP program, which culminates at the end of the year with a standardized national advanced subject test, students must pass in order to receive college credit.

Generally, stress ought to be monitored first and foremost in females. Secondly, those enrolled in an AP program may also be at a higher risk of experiencing stress. Furthermore, there may be several confounds not controlled for in this study. Prior research has demonstrated that many VET program high school students go on to receive advanced educational degrees. Perhaps the true effect in this study was of a different personality or familial nature rather than that of a academic track as hypothesized by the local School Board. Further investigation should be conducted in order to crystallize plausible interpretations of this school's student body stress levels across their high school education.

References

Cody, R. P, & Smith, J. K. (2005). *Applied statistics and the SAS programming language*. Upper Saddle River, NJ: Prentice Hall.

Johnson, R. A. & Wichern, D. W. (2007). *Applied multivariate statistical analysis*. Upper Saddle River, NJ: Prentice Hall.

Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). *Applied multivariate research*. Thousand Oaks, CA: Sage Publications, Inc.

Raykov, T. & Marcoulides, G. A. (2008). *An introduction to applied multivariate analysis*. New York, NY: Routledge.