Written Descriptions of Quantitative Data Analyses (…and some tables)

Dr. Peter Miksza

This is a collection of general models/templates that I use when introducing students to writing about foundational quantitative analyses typically found in music education research. These sample write-ups and tables are generated from data that I use as examples in the Quantitative Research Methods courses I teach at the Indiana University Jacobs School of Music as well as the data that are provided with the book written by myself and Dr. Ken Elpus, *Design and Analysis for Quantitative Research in Music Education.* I aggregated all of these examples into one document so that my students wouldn't have to search through powerpoints for each topic to find them. I'm sharing this as MS Word format as opposed to PDF so that tables can be cut and pasted and used as templates.

If you find these useful, you are welcome to use these as examples your own teaching …

…just please let me know you have done so.

Please also let me know if you see any problems or mistakes. I am sure there are some that got by me…

Contents

Descriptive Analyses	
Interval/Ratio Data	
Central Tendency, Dispersion, Distribution properties	On average, the participants reported 7.91 years of experience singing in an ensemble. The most frequently reported amount of years was 9 years of experience. There was relatively little variation across participants given the standard deviation value of 1.61 and the 25th percentile, median, and 75th percentile of 7, 8, and 9, respectively. However, the range of responses was somewhat large with 3 years being the least amount of experience reported and 10 years being the most. Participants' reports of years singing in an ensemble were slightly skewed (-.92) with very little kurtosis (.31). However, the skewness and kurtosis values fall within the range of -1 and +1, and therefore do not suggest extreme departures from normality.
Ordinal Data [Interval/Ratio data that fails to meet assumptions]	
Frequencies, Percentages, Proportions	Most participants indicated a preference for rehearsal two times per month (40.9%), with the next most popular response being once per week (25.2%). A substantial number of participants also preferred rehearsal once per month (18.3%) and twice per week (11.3%). Very few participants reported a preference for more than two rehearsals per week (4.3%).
Nominal Data	
Frequencies, Percentages, Proportions	All participants indicated a sex type. A slight majority of participants were male (53.9%), only 2 reported intersex (1.7%), with the remaining participants reporting female (44.3%).

Correlation Analyses	
Interval/Ratio Data	
Pearson coefficient	A significant, moderate, positive correlation was found between sight-reading scores and etude performance scores, $r(463) = .62$, $p < .001$, 95% CI [.56, .67]. Participants who tended to have higher sight-reading scores also tended to have higher etude performance scores.
Ordinal Data [Interval/Ratio data that fails to meet assumptions]	
Spearman coefficient	No significant correlation was found between sight-reading scores and motivation scores, $r_s(463) = .04$, $p > .05$, 95% CI [-.05, .13]. Participants' sight-reading scores were not associated with their degree of motivation.
Nominal Data	
Phi (2 X 2); Cramer's V (k X k)	A significant, moderate correlation was found between participants' sex type and instrument, Cramer's $V = .51$, $p < .001$. Standardized residuals indicate the largest disproportionalities were that men were overrepresented among percussionists and underrepresented among brass and string instrumentalists.

Comparing two independent samples	
Interval/Ratio Data	
Independent-samples t-test	A significant difference between mean control group (M = 11.93) and treatment group (M= 22.95) sight-singing scores was found, $t(38) = 2.07$, $p < .05$. On average, the treatment group performed better the control group. Cohen's d indicates that the difference represents a moderate effect ($d = .65$). The 95% confidence interval of the difference between means indicates a good deal of uncertainty in the findings with the lower bound being .26 and the upper bound being 20.45.
Ordinal Data [Interval/Ratio data that fails to meet assumptions]	
Mann-Whitney U Test	A significant difference between the control (Md = 2.00) and treatment group's (Md = 4.00) frequency of music making at home was found, Mann-Whitney U = 169.50, $p < .05$. Participants in the treatment group tended to report a greater frequency of home music making.
Nominal Data	
Chi Square Test	Although there were more students on the honor roll in the treatment group, the difference in the proportions of honor roll students in the treatment and control group was not significant, $\chi^2 = 3.33$, $p > .05$.

Comparing two dependent samples	
Interval/Ratio Data	
Dependent-samples t-test	A significant difference between mean pretest (*M* = 8.21) and posttest (*M*= 22.95) sight-singing scores for the treatment group was found, $t(19) = 3.61$, $p < .05$. On average, the posttest scores were higher than the pretest scores. Cohen's *d* indicates that the difference represents a large effect ($d = .80$). The 95% confidence interval of the difference between means indicates a good deal of uncertainty in the findings with the lower bound being 5.92 and the upper bound being 22.23.
Ordinal Data [Interval/Ratio data that fails to meet assumptions]	
Wilcoxon-Signed Rank Test	A significant difference between the pretest (*Md* = 2.00) and posttest (*Md* = 4.00) frequency of music making at home was found, $Z = 212.00$, $p < .05$. Participants reported significantly greater frequency of home music making at posttest.
Nominal Data	
McNemar's Test	Although six students who initially did not intend to join HS choir in September changed their minds by May, no significant difference in the consistency of responses from September to May was found overall, $\chi^2 = .44$, $p > .05$.

<table>
<tr><th colspan="2">Comparing more than two independent samples</th></tr>
<tr><td colspan="2">Interval/Ratio Data</td></tr>
<tr><td>Oneway ANOVA</td><td>A oneway ANOVA was performed with sightsinging scores serving as the dependent variable and experimental condition as the independent variable. Sightsinging scores varied significantly as a function of experimental condition, $F(2, 147) = 20.66$, $p < .05$. The eta squared value of .22 indicated this was a large effect. Post-hoc Tukey tests indicated that the control group ($M = 40.90$) scored significantly ($p < .05$) lower than the weak ($M = 49.22$) or strong ($M = 53.53$) treatment groups. However, the weak and strong treatment group means were not significantly different from each other.</td></tr>
<tr><td></td><td>Analysis of Variance Summary Table for the Effects of Experimental Treatment on Sightsinging
<table>
<tr><th>Source</th><th>df</th><th>SS</th><th>MS</th><th>F</th><th>p</th><th>η^2</th></tr>
<tr><td>Model: Experimental Group</td><td>2</td><td>4116.28</td><td>2058.14</td><td>20.66</td><td><.001</td><td>.22</td></tr>
<tr><td>Residual</td><td>147</td><td>14643.56</td><td>99.62</td><td></td><td></td><td></td></tr>
<tr><td>Total</td><td>149</td><td>18759.84</td><td></td><td></td><td></td><td></td></tr>
</table></td></tr>
<tr><td colspan="2">Ordinal Data [Interval/Ratio data that fails to meet assumptions]</td></tr>
<tr><td>Kruskal-Wallis Test</td><td>There was a statistically significant difference in pre-service teachers’ interest in teaching in urban settings as a function of experimental group, $H(2) = 7.78$, $p < .05$. Dunn-Bonferroni follow-up comparisons indicated that the group that observed in an urban school ($Md = 4$) reported being significantly more likely to teach in an urban school than those with no experience ($Md = 3$) ($p < .05$). No other pairwise comparisons were significant.</td></tr>
<tr><td colspan="2">Nominal Data</td></tr>
<tr><td>Chi Square Test</td><td>A Chi-square test of independence was calculated comparing the frequency of elementary-vocal and instrumental emphasis students in each experimental condition. A significant effect was found, $\chi^2(2) = 6.43$, $p < .05$. The Cramer’s V value of .38 indicated a strong effect. Standardized residuals indicated that the largest disproportionality was due to the lack of instrumental emphasis students in the practicum experimental group.</td></tr>
</table>

Comparing more than two dependent samples

Interval/Ratio Data

Repeated Measures ANOVA

A repeated measures ANOVA with experimental condition serving as the independent variable and flexibility scores serving as the dependent variable was conducted. The assumption of sphericity was not met and therefore, the Greenhouse-Geisser correction was applied, revealing a significant difference among conditions, $F(2, 58) = 34.26$, $p < .05$, η^2=.54. The eta squared value indicated this was a large effect. Helmert within-subjects contrasts showed that the control group mean was significantly lower than the combined routine A and routine B group's mean ($p < .05$), and that the mean for routine A was significantly lower than the mean for routine B, ($p < .05$). The test of the interaction of experimental condition and order of presentation found that no order effect was present.

Repeated-measures ANOVA examining the differences among flexibility routines

Source	*df*	*SS*	*MS*	*F*	*p*	η^2
Flexibility Routine	2	2914.57	1457.28	34.25	<.001	.54
Residual	58	2467.24	42.54			

Ordinal Data [Interval/Ratio data that fails to meet assumptions]

Friedman 2-way Analysis of Variance

A Friedman Two-Way analysis of variance revealed a significant difference among the participants reported likelihood of teaching in an urban setting over time, $\chi^2(2) = 18.13$, $p < .05$. Dunn-Bonferroni follow-up comparisons indicated that participants indicated being significantly ($p < .05$) more likely to teach in an urban setting in May as compared to August. However, no other pairwise comparisons were significant.

Nominal Data

Cochran's Q Test

A significant change in participants' sense of whether they planned to be a teacher occurred over time, $\chi^2(2) = 18.13$, $p < .05$. Dunn-Bonferroni follow-up comparisons indicated that participants were significantly more likely to say "no" in May as compared to August or December($p < .05$). However, no other pairwise comparisons were significant.

Factorial ANOVA: Only Between-subjects Independent Variables

A two-way ANOVA was performed with the presence/absence of finger tape and the presence/absence of harmonic accompaniment serving as the independent variables. The dependent variable was participants' intonation rating. The main effects and interaction effect were all significant, $p < .05$. The interaction effect indicated that finger tapes were only effective for improving intonation for those that did not also play with a harmonic accompaniment, $F(1, 64) = 11.56$, $p = .001$. The eta squared value of .15 indicated this was a large effect.

2 X 2 ANOVA Summary

Source	*df*	*SS*	*MS*	*F*	*p*	*eta*^{2}
Finger Tape	1	4.99	4.99	10.03	.002	.14
Harmonic Context	1	5.44	5.44	10.94	.002	.15
Finger Tape X Harmonic Context	1	5.75	5.75	11.56	.002	.15
Error	64	31.85	.50			
Total	68					

Mixed-Design ANOVA: Between- and Within-subjects Independent Variables

A mixed design two-way ANOVA was performed with active vs. passive instruction serving as the between-subjects independent variable and musical characteristic (pitch vs. rhythm/beat) serving as the within-subjects independent variable. The main effects were each significant, $p < .05$, however, the interaction effect was not. Results showed that pitch scores were significantly higher than rhythm/beat scores, $F(1, 33) = 26.36$, $p < .001$. A partial eta squared value of .44 indicated this was a large effect. In addition, scores from the active music instruction group were significantly higher than those of the passive music instruction group, $F(1, 33) = 6.15$, $p < .05$. A partial eta squared value of .16 indicated that this was also a strong effect.

2 X 2 Mixed Design ANOVA Summary

Source	*df*	*SS*	*MS*	*F*	*p*	eta^2
Between subjects						
Listening Mode	1	98.25	98.25	6.15	.018	.16
Error 1	33	527.07	15.97			
Within subjects						
Musical Characteristic	1	300.38	399.38	26.35	<.001	.44
Musical Characteristic X Listening Mode	1	25.41	25.41	1.69	.20	.05
Error 2	33	500.08	15.15			

Fully Repeated-Measures ANOVA: Only Within-subjects Independent Variables

A fully-repeated two-way ANOVA was performed with audio vs. audio visual observation and matched vs. mismatched instrument type serving as the within-subjects independent variables and error identification scores serving as the dependent variable. The main effects were each significant, $p < .05$, but the interaction effect was not. Results showed that error identification scores were significantly higher in the audio-visual condition than the audio only condition, and significantly higher when participants were making judgements in the matched observation condition vs. the mismatched condition. Partial eta squared values of .64 and .84, respectively, indicated that these were very strong effects.

Fully-repeated Design ANOVA Summary

Source	*df*	*SS*	*MS*	*F*	*p*	eta^2
1. Audio vs. Audio-Visual Observation	1	257.14	257.14	50.55	<.001	.64
Error 1	29	147.51	5.09			
2. Matched vs. Mismatched Observation	1	853.01	853.01	153.31	<.001	.84
Error 2	29	69.31	2.39			
Interaction 1 X 2	1	.32	.32	.06	.804	.002
Error 1 X 2	29	149.38	5.15			

ANCOVA

A oneway ANCOVA was run with instrument condition serving as a the independent variable and trait anxiety scores serving as the covariate. Participants' state anxiety scores served as the dependent variable. The covariate explained significant variation in participants' state anxiety scores, $F(1, 296) = 16.64$, $p < .001$, $eta^2 = .05$. The main effect of instrument condition was significant and large, $F(2, 296) = 45.82$, $p < .001$, $eta^2 = .24$. The covariate-adjusted means indicated that the strings condition yielded the lowest state anxiety score ($M = 48.50$), followed successively by the piano ($M = 58.03$) and marimba ($M = 63.27$) conditions. Pairwise comparisons of the covariate-adjusted means showed that each mean was significantly different ($p < .05$) from each other.

Oneway ANCOVA Summary

Source	*df*	*SS*	*MS*	*F*	*p*	eta^2
Covariate: Trait Anxiety	1	2033.48	2033.48	16.64	<.001	.05
Listening condition	2	11196.71	5598.36	45.82	<.001	.24
Error	296	122.18				
Total	300					

MANOVA

A multivariate analysis of variance was performed with listening condition serving as the independent variable and self-evaluation ratings of tone, pitch, rhythm, and dynamics as the combined dependent variable. The multivariate test revealed a significant effect, $F(8, 204) = 8.62$, $p < .05$. The partial eta^2 value of .25 indicated that this was a large effect. Examination of univariate ANOVA analyses for each of the dependent variables shows significant differences by listening condition for tone, pitch, and rhythm ratings ($p < .01$), but not dynamics. A conservative Bonferroni correction was applied for the univariate ANOVA analyses (.05/4 = .0125). Post-hoc Tukey tests showed…

MANOVA Analysis of Self-evaluation Variables

Source	*Wilks' Lambda*	*df*	*df error*	*F*	*p*	*eta*^{2}
Listening Condition	.56	8	204	8.62	< .001	.25

Follow-up Univariate ANOVA Analyses

	df	*SS*	*MS*	*F*	*p*	*eta*^{2}
Tone						
Listening Condition	2	20.01	10.01	10.60	< .001	.17
Error	105	99.13	.94			
Total	108	1311.31				
Pitch						
Listening Condition	2	11.99	5.93	6.54	.002	.11
Error	105	96.25	.92			
Total	108	1454.97				
Rhythm						
Listening Condition	2	37.54	18.77	26.38	<.001	.33
Error	105	74.71	.71			
Total	108	1359.11				
Dynamics						
Listening Condition	2	6.06	3.03	3.39	.038	.06
Error	105	93.98	.89			
Total	108	1184.17				

Simple Linear Regression

Interval/Ratio Independent Variable

A linear regression analysis was conducted with performance test scores serving as the dependent variable and months of practicing as the independent variable. The model *F* test was significant, $F(1, 48) = 56.12$, $p < .001$, and a large amount of variance in the outcome was explained by the model ($R^2 = .54$). The significant beta coefficient for months of practice ($p < .001$) indicated that a 4.15 point increase of performance test score would be predicted with each month increase in practicing. The confidence interval for the coefficient ranged from 3.04 to 5.27.

Regression of Performance Test Scores on Months of Practice

	b	*SE*	*t*	95% CI	*p*
Intercept	2.05	6.72	.31	[-11.46, 15.57]	.76
Months of Practice	4.15	.55	7.49	[3.04, 5.27]	<.001

Note. $F(1, 48) = 56.12$, $p < .001$; $R^2 = .54$

Categorical/Dummy Independent Variable

A linear regression analysis was conducted with performance test scores serving as the dependent variable and experimental grouping as the independent variable. The model *F* test was significant, $F(1, 48) = 9.32$, $p < .05$, and a moderate amount of variance in the outcome was explained by the model ($R^2 = .16$). The significant beta coefficient for experimental grouping ($p < .05$) indicated a 13.92 point difference between the control and experimental group performance test score means. The confidence interval for the coefficient ranged from 4.75 to 23.09.

Regression of Performance Test Scores on Experimental Group

	b	*SE*	*t*	95% CI	*p*
Intercept	43.84	3.22	13.60	[37.36, 50.32]	<.001
Experimental Group	13.92	4.56	3.05	[4.75, 23.09]	.004

Note. $F(1, 48) = 9.32$, $p < .05$; $R^2 = .16$

Multiple Linear Regression

Scalar and Categorical/Dummy Independent Variables

A linear regression analysis was conducted with performance test scores serving as the dependent variable and experimental grouping and months of practice as the independent variables. The model *F* test was significant, $F(1, 47) = 32.81$, $p < .05$, and a relatively large amount of variance in the outcome was explained by the model (adjusted $R^2 = .56$). Beta coefficients for both predictors were significant ($p < .05$). The coefficients indicated a significant effect of the treatment. Holding months of practice constant, those in the treatment group scored 7.50 points higher, on average, than those in the control group. However, the standardized beta coefficients suggested that months of practice was a stronger predictor of playing test scores than experimental grouping.

Regression of Performance Test Scores on Experimental Group and Months of Practice

	b	*SE*	*β*	*t*	95% CI	*p*
Intercept	2.28	6.46		.35	[-10.72, 15.29]	.726
Experimental Group	7.50	3.38	.22	2.22	[.69, 14.31]	.031
Months of Practice	3.82	.55	.67	8.88	[2.70, 4.94]	<.001

Note. $F(1, 47) = 32.81$, $p < .05$; adjusted $R^2 = .56$

Multiple Linear Regression	
Scalar and Categorical/Dummy Independent Variables	A series of linear regression models were estimated to determine the collection of predictors that would explain the most variation in first grade math test scores. Model 1 included music lessons as a predictor variable, model 2 included music lessons and student urbanicity, and model 3 included music lessons, student urbanicity, and student SES. Model comparison tests indicated that model 3 yielded the best fit to the data, $F_{change}(1, 994) = 132.20$, $p < .001$. Model 3 explained 13.6% of the variation in first grade math scores, whereas model 1 and 2 explained 1.3% and 2.2%, respectively. Only the Beta coefficients for the dummy indicator of rural urbanicity and student SES were significant ($p < .05$). The ceofficients indicated that students from a rural urbanicities score higher on math test scores than those from a city urbanicity and that students with higher SES tended to score higher on math test scores. The standardized beta coefficients suggested that SES was the stronger predictor.

Regression of First Grade Math Test Scores on Music Lessons, Urbanicity, and SES

	b	*SE*	*β*	*t*	95% CI	*p*
Intercept	24.05	1.73		13.86	[20.64, 27.56]	< .001
Music Lessons	1.71	1.10	.05	1.56	[-.44, 3.86]	.119
Suburban	.01	.83	.001	.02	[-1.61, 1.64]	.987
Town	1.53	1.27	.04	1.21	[-.96, 4.02]	.227
Rural	2.03	.91	.08	2.23	[.25, 3.81]	.026
SES	3.82	.03	.35	11.50	[.33, .47]	< .001

Note. $F(5, 994) = 32.32$, $p < .05$; adjusted $R^2 = .136$

Logistic Regression	
Scalar and Categorical/Dummy Independent Variables	Hierarchical logistical regression analyses were conducted to determine the degree to which participants' audition and interview scores would predict their acceptance decision regarding application to music school and whether this would change after controlling for whether participants took private lessons or not. Model Chi square tests, the Hosmer-Lemeshow test, and examinations of pseudo-R^2 values indicated that a model with private lessons serving as a covariate yielded the best fit to the data. The classification accuracy of this model was quite high at 86.9%. The beta coefficients indicated that all predictors were positively related to the outcome, in that those with relatively high audition and interview scores and those with private lessons were more likely to be accepted to music school. The odds ratios for the coefficients suggested a strong effect for private lessons, with those who have had private lessons being approximately 18 times more likely to be accepted than those who did not. Predicted probabilities of acceptance were calculated for participants with interview scores of 50 and 60, who did or did not have private lessons while holding audition scores constant at the mean. The probability of acceptance for those with interview scores of 50 and 60, who did not have private lessons was .10 and .27, respectively. The probability of acceptance for those with interview scores of 50 and 60, who did have private lessons was .66 and .87, respectively.

Hierarchical Logistic Regression Analyses Predicting Acceptance to Music School

	Model I				*Model II*			
	B	*SE*	*Wald*	*OR*	*B*	*SE*	*Wald*	*OR*
Constant	-14.42***	3.18	20.57		-14.38***	3.61	15.84	
Interview	.13***	.03	17.41	1.14	.12***	.04	10.65	1.13
Audition	.12***	.03	12.32	1.13	.16***	.04	9.43	1.12
Private Lessons					2.89***	.81	12.88	18.08
Deviance	68.78				51.90			

*$p < .05$, **$p < .01$, ***$p < .001$