Extra exercise 18

(Continuation of previous supplementary exercises 7.102-104 with data for added groups)

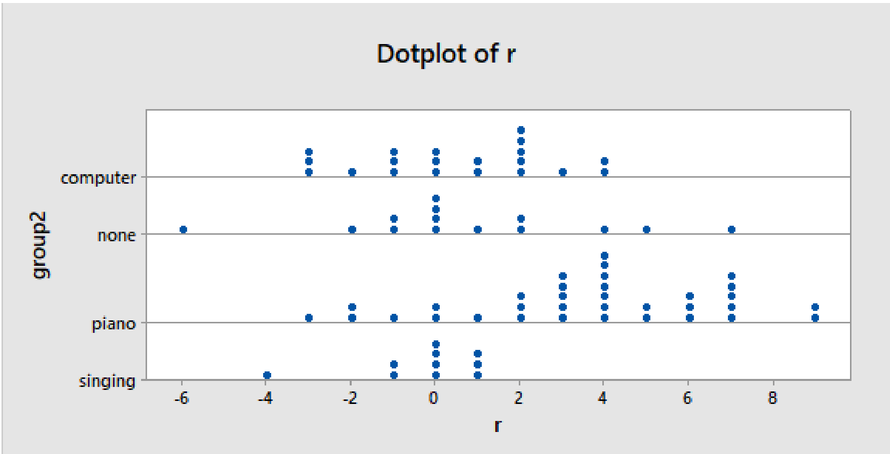
Data: 4 samples of changes (improvements, differences after-before) in spatial-temporal reading test scores for 34 children attending six months of piano lessons and a total of 44 children in three groups (computer, singing, none) considering as a single control group in previous analyses.

We start with some descriptive statistics and a data display by a dotplot.

Descriptive Statistics: r

Statistics

Variable	group2	N	Mean	SE Mean	StDev	Minimum	Median	Maximum	Skewness	Kurtosis
r	computer	20	0.450	0.495	2.212	-3.000	0.500	4.000	-0.12	-0.93
	none	14	0.786	0.853	3.191	-6.000	0.000	7.000	0.01	1.02
	piano	34	3.618	0.524	3.055	-3.000	4.000	9.000	-0.36	-0.28
	singing	10	-0.300	0.473	1.494	-4.000	0.000	1.000	-1.86	4.26



The descriptive statistics show the 4 groups to differ in several ways. Looking at the means, the piano groups is clearly highest. Looking at the standard deviations, the singing group is clearly lowest, and the IPS guideline for equal variances is violated ($3.191/1.494 = 2.14 > 2$). However, testing homogeneity of variances does not give clear significance with any of the two tests (see next page). Therefore it might be ok to use the variance homogeneity assumption, with a small reservation perhaps. As to the normal distribution, only the singing group presents some problems ($P = 0.015$); next page. The low value of -4 falls outside the others (and if it was removed, the standard deviation would drop dramatically). It is fair to say that we may proceed with the analysis, with a small reservation for the singing group. If results are not clear, we may want to rerun the analysis without that group.

Model: If we denote the change in scores for child j in activity group i (where $i = 1, 2, 3, 4$ and $j = 1, \ldots, n_i$, with $(n_1, n_2, n_3, n_4) = (20, 14, 34, 10)$), the statistical model is that the X_{ij}'s are normally distributed $\mathrm{N}(\mu_i, \sigma)$, and all observations are independent.

Test for Equal Variances: r versus group2

Method

Null hypothesis	All variances are equal
Alternative hypothesis	At least one variance is different
Significance level	α = 0.05

95% Bonferroni Confidence Intervals for Standard Deviations

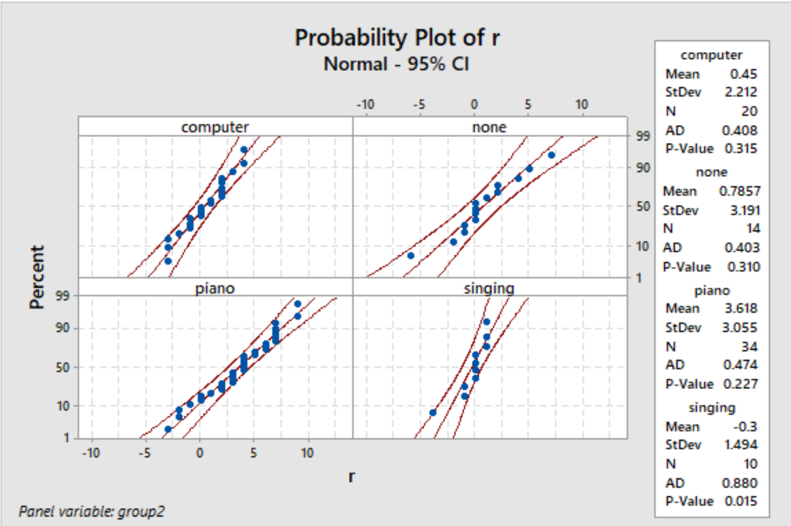
group2	N	StDev	CI
computer	20	2.21181	(1.66370, 3.36012)
none	14	3.19082	(1.80321, 6.87230)
piano	34	3.05520	(2.32745, 4.32846)
singing	10	1.49443	(0.45553, 6.53495)

Individual confidence level = 98.75%

Tests

Method	Test Statistic	P-Value
Multiple comparisons	—	0.130
Levene	1.73	0.169

Test for Equal Variances: r vs group2



We start with the basic 1-way ANOVA, including only the overall test for homogeneity between the groups in the ANOVA table, and the confidence intervals for groups.

One-way ANOVA: r versus group2

Method

Null hypothesis	All means are equal
Alternative hypothesis	Not all means are equal
Significance level	α = 0.05

Equal variances were assumed for the analysis.

Analysis of Variance

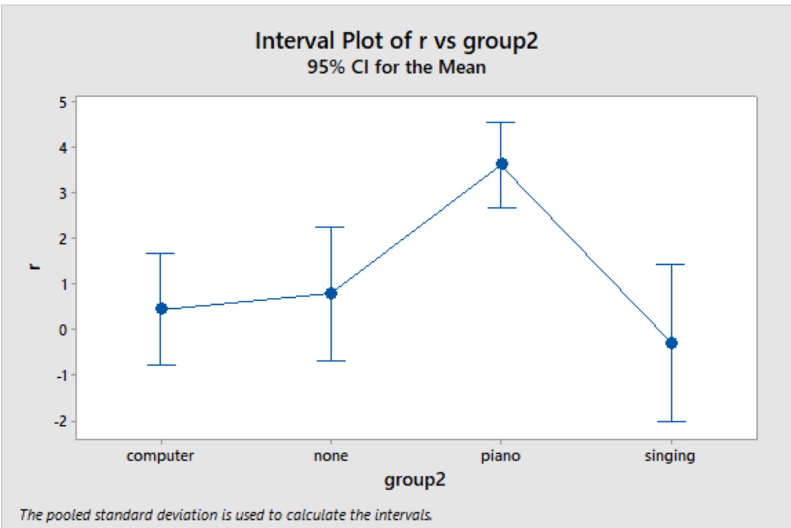
Source	DF	Adj SS	Adj MS	F-Value	P-Value
group2	3	207.3	69.094	9.24	0.000
Error	74	553.4	7.479		
Total	77	760.7			

Means

group2	N	Mean	StDev	95% CI
computer	20	0.450	2.212	(-0.768, 1.668)
none	14	0.786	3.191	(-0.671, 2.242)
piano	34	3.618	3.055	(2.683, 4.552)
singing	10	-0.300	1.494	(-2.023, 1.423)

Pooled StDev = 2.73475

Interval Plot of r vs group2



Our statistical testing is for the hypothesis

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \quad \text{versus} \quad H_a : \text{ some of the } \mu\text{'s differ},$$

and with a test statistic of $F = 9.24$ corresponding to $P < 0.0005$ in the $F(3, 74)$-distribution we conclude the test to be clearly significant and to strongly indicate that some differences exist between groups. That is, the type of lesson has some impact on the changes in test scores. The estimated means and confidence intervals suggest that the piano group has higher score changes than the other groups, which have more or less equal score changes. Actually, if we were satisfied with inference not adjusted for multiple comparisons, we could directly infer this conclusion from the groupwise confidence intervals (based on the rules from slide 10−12).

We supplement the analysis with an additional Minitab listing for multiple comparisons. Among 4 groups there are $4 \cdot 3/2 = 6$ comparisons, so for the Bonferroni method the individual error rate should be $0.05/6 = 0.0083333$ (which needs to be entered as the error level for Fisher comparisons, when using Minitab's `One-Way ANOVA` menu).

Fisher Pairwise Comparisons

Grouping Information Using the Fisher LSD Method and 99.1667% Confidence

group2	N	Mean	Grouping	
piano	34	3.618	A	
none	14	0.786		B
computer	20	0.450		B
singing	10	-0.300		B

Means that do not share a letter are significantly different.

Fisher Individual Tests for Differences of Means

Difference of Levels	Difference of Means	SE of Difference	99.1667% CI	T-Value	Adjusted P-Value
none - computer	0.336	0.953	(-2.248, 2.919)	0.35	0.726
piano - computer	3.168	0.771	(1.078, 5.257)	4.11	0.000
singing - computer	-0.75	1.06	(-3.62, 2.12)	-0.71	0.481
piano - none	2.832	0.868	(0.478, 5.186)	3.26	0.002
singing - none	-1.09	1.13	(-4.16, 1.98)	-0.96	0.341
singing - piano	-3.918	0.984	(-6.585, -1.251)	-3.98	0.000

Simultaneous confidence level = 95.93%

None of the confidence intervals for comparisons with the piano group contain zero, and significant differences at the overall 5% error level therefore exist between the piano and all other groups. The 3 remaining confidence intervals all contain zero; there are no significant differences between the other groups. Therefore, we conclude that there is evidence for the piano group to be above all others, but there is no evidence to distinguish the other groups. This conclusion is represented by the letter-coding graphic above the comparisons: the piano group has letter A, and all the other groups have letter B. Note that contrary to what the listing says, the P-values are *not* adjusted for multiple comparisons, but if we assess them at a significance level of $\alpha = 0.0083333$ we will get significance assessments corresponding to the Bonferroni method for an overall significance level of 0.05.