

Tel: (814) 863-0281

Fax: (814) 863-7114

The Pennsylvania State University

323 Thomas Building

University Park, Pa 16802

Statistical Consulting Center

Department of Statistics

Eberly College of Science

**SCCID**: N/A

**Client**: Mr. Ag Researcher, Address: PA, USA

**Consultant**: Daljeet Singh

**Date**: 3/10/2018

**Research Field:** Psychology

**Project Title**: Effects of Word Distractors and Word Type on Memory Retention

**Project Description:**

We have recognized that we are dealing with a 2-factor repeated measure design. Students from a pshychology class with different majors participate in a repeated measures study where their memory recall was tested.

**Variables of Interest**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Major | Student | Test | Wordlist | Distracter | Treatment Code | Score |

**Study Diagram**

**Factors:**

The study has the following cross factors where Treatment Codes (TR1 – TR4) are an artificially created variable by combining the factors that were administered in the study namely Wordlist and Distractor.

**Wordlist**: Fixed factor with two levels – abstract and concrete.

**Distractor**: Fixed factor with two levels – poetry and math.

**Major**: Fixed factor with four levels: Math, CS, English, and History.

|  |  |  |
| --- | --- | --- |
| **Treatment Code** | **Wordlist** | **Distractor** |
| TR1 | abstract | poetry |
| TR2 | abstract | math |
| TR3 | concrete | poetry |
| TR4 | concrete | math |

The treatments are applied after a wash-out period to avoid cross-over.

Rep (Student) are selected randomly within a fixed major.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Major: Math** | |  |  |  |  |  | | --- | --- | --- | --- | --- | | REP | TR1 | TR2 | TR3 | TR4 | | 1 |  |  |  |  | | 2 |  |  |  |  | | 3 |  |  |  |  | |
| **Major: CS** | |  |  |  |  |  | | --- | --- | --- | --- | --- | | REP | TR1 | TR2 | TR3 | TR4 | | 4 |  |  |  |  | | 5 |  |  |  |  | | 6 |  |  |  |  | |
| **Major: Hist** | |  |  |  |  |  | | --- | --- | --- | --- | --- | | REP | TR1 | TR2 | TR3 | TR4 | | 7 |  |  |  |  | | 8 |  |  |  |  | | 9 |  |  |  |  | |
| **Major: Eng** | |  |  |  |  |  | | --- | --- | --- | --- | --- | | REP | TR1 | TR2 | TR3 | TR4 | | 10 |  |  |  |  | | 11 |  |  |  |  | | 12 |  |  |  |  | |

Rep is a student.

| **Class Level Information** | | |
| --- | --- | --- |
| **Class** | **Levels** | **Values** |
| **Student** | 12 | 1 2 3 4 5 6 7 8 9 10 11 12 |
| **TreatmentCode** | 4 | TR1 TR2 TR3 TR4 |
| **Major** | 4 | CS Eng Hist Math |

**Research and Statistical Questions**

* Request for help in analyzing the memory data.

The null hypothesis: states that there are no differences in memory recall (or that they are all equal) among the group means.

H0: μ1=μ2= μ3=μ4

Ha: One of the means is different

**EDA**

The structure of the data indicates that we have the following type of research:

* The study is a split plot in CRD with repeated measures.
* We have a repeated measures study here with the main plot treatment being the major. We have four majors with three students in each major. Each student is given the four treatments (2 X 2 crossed factor corresponding to wordlist and distractor combinations)
* The main plot treatment (Major) is applied to the Student that is the main plot.
* The subplots treatments are the treatments. The treatments are crossed factors of Wordlist and Distractor
* We will consider students will be homogenous since all of them are from the psychology class. Their variation is assumed to be negligible and contained within the random error.
* Major is considered to be fixed effect and not a blocking factor. Major will have effect on the memory study and therefore should not be removed and included in the treatment.
* Since this is a split plot study we don’t want to have 3-way interactions. Therefore we will create a new variable (TreatmentCode) that codifies the crossed factors (Wordlist and Distractor)

## Box plots of Major vs Score

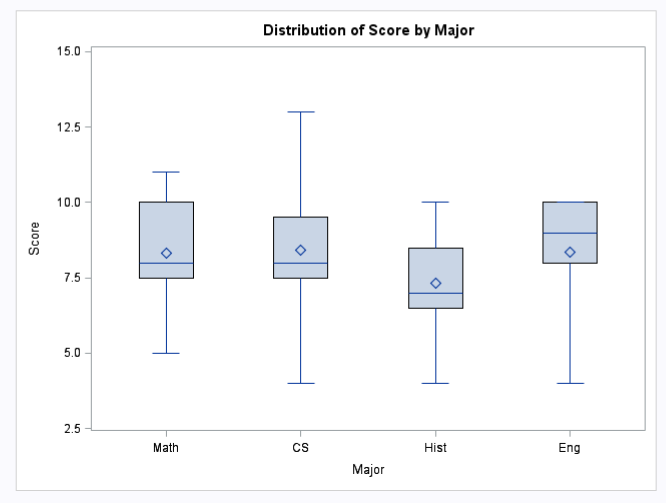
When plotting the box plot of scored grouped by Major, we use the code:

PROC BOXPLOT data=memory;

plot Score\*Major;

RUN;

This specific code yields a boxplot for the Score variable organized by the Major for the memory data. The boxplot obtained from this code is:



This boxplot reveals to us that the means are fairly consistent at all four majors of the memory data. There does not seem to be any outliers or oddities about the plot.

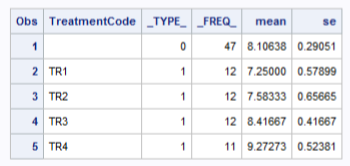
## Data Transformation

We add a new computed column called TreatmentCode. The computation of the column is as follows:

|  |  |  |
| --- | --- | --- |
| **Treatment Code** | **Wordlist** | **Distrator** |
| TR1 | abstract | poetry |
| TR2 | abstract | math |
| TR3 | concrete | poetry |
| TR4 | concrete | math |

## Means: Grand and Treatment specific

Let’s look at the various means of the scores grouped by treatment:



# Statistical Analysis

We have the option of using Split-Plot or Repeated Measures in this scenario. Let’s look at both of these.

## Split-Plot

We use this code:

title 'Split-Plot approach';

proc mixed data=memory method=type3;

class Student TreatmentCode Major;

model Score = TreatmentCode Major TreatmentCode\*Major / ddfm=kr;

random Student(Major);

run;

The results are as follows:

| **Type 3 Tests of Fixed Effects** | | | | |
| --- | --- | --- | --- | --- |
| **Effect** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **TreatmentCode** | 3 | 24 | 8.79 | 0.0004 |
| **Major** | 3 | 8 | 0.39 | 0.7647 |
| **TreatmentCode\*Major** | 9 | 24 | 6.16 | 0.0002 |

We see that the interaction term is significant since the p-value < alpha (assume 0.05) and therefore we reject the null hypothesis. We have sufficient evidence to reject the null hypothesis that the means are the same and conclude the alternative.

Since the interaction is significant we need to perform mean comparisons and interpret the treatment combinations. Should NOT do mean comparisons or interpretation of either the Major or TreatmentCode main effects because the interaction is significant.

## Repeated Measures – Co-variance summary

As we discussed above, we are dealing here with a repeated measures study. Our analysis relies on a computed column named Treatment Code that combines the crossed factors Wordlist and Distractor.

We start off with a repeated measures analysis and consider the Unstructured, Compound Symmetry, and Variable Co-variance based covariance structures. We get the following FIT statistics:

|  |
| --- |
| Covariance Summary |

| **Obs** | **Descr** | **CS** | **VC** | **UN** |
| --- | --- | --- | --- | --- |
| **1** | -2 Res Log Likelihood | 166.4 | 134.1 | 123.0 |
| **2** | AIC (Smaller is Better) | 200.4 | 186.1 | 127.0 |
| **3** | AICC (Smaller is Better) | 220.8 | 253.0 | 127.5 |
| **4** | BIC (Smaller is Better) | 208.6 | 198.7 | 128.0 |

We clearly see that the smallest value of AICC is obtained for the unstructured co-variance. Therefore, we will use this structure for further analysis.

### Repeated measure – Unstructured Co-variance

When using the following code:

title 'Unstructured'; run;

proc mixed data=memory method=ml covtest

plot=all;

class Student Major TreatmentCode;

model Score = TreatmentCode Major TreatmentCode\*Major / ddfm=kr;

repeated / subject=Student type=un rcorr;

store final\_model;

run;

We see the following results:

| **Type 3 Tests of Fixed Effects** | | | | |
| --- | --- | --- | --- | --- |
| **Effect** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **TreatmentCode** | 3 | 10 | 13.41 | 0.0008 |
| **Major** | 3 | 12 | 0.58 | 0.6378 |
| **Major\*TreatmentCode** | 9 | 12.8 | 15.94 | <.0001 |

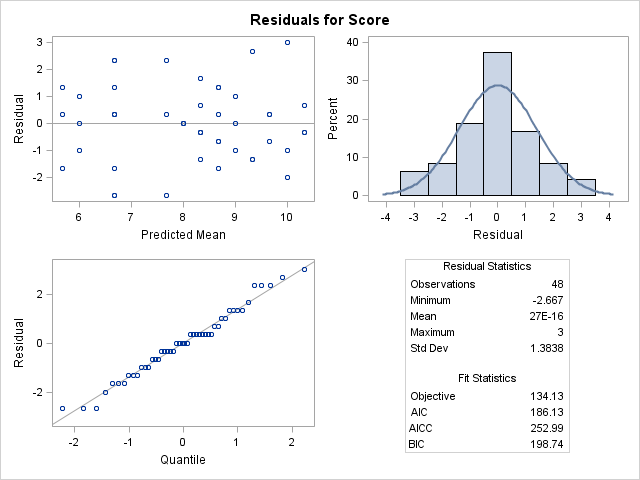
As also indicated in the Split-Plot results, here too we find significant results for the interaction between Major and Treatments.

The covariance estimates are:

| **Covariance Parameter Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Cov Parm** | **Subject** | **Estimate** | **Standard Error** | **Z Value** | **Pr Z** |
| **UN(1,1)** | **Student** | 1.6667 | 0.6804 | 2.45 | 0.0072 |
| **UN(2,1)** | **Student** | 1.5556 | 0.7460 | 2.09 | 0.0371 |
| **UN(2,2)** | **Student** | 2.5556 | 1.0433 | 2.45 | 0.0072 |
| **UN(3,1)** | **Student** | 0.6944 | 0.5212 | 1.33 | 0.1827 |
| **UN(3,2)** | **Student** | 1.3611 | 0.7137 | 1.91 | 0.0565 |
| **UN(3,3)** | **Student** | 1.6667 | 0.6804 | 2.45 | 0.0072 |
| **UN(4,1)** | **Student** | 1.1111 | 0.5715 | 1.94 | 0.0519 |
| **UN(4,2)** | **Student** | 1.3056 | 0.6965 | 1.87 | 0.0609 |
| **UN(4,3)** | **Student** | 1.1667 | 0.5807 | 2.01 | 0.0445 |
| **UN(4,4)** | **Student** | 1.6111 | 0.6577 | 2.45 | 0.0072 |

We see significant correlation for many entries.

Further looking at the residuals, we find:



This indicates the following:

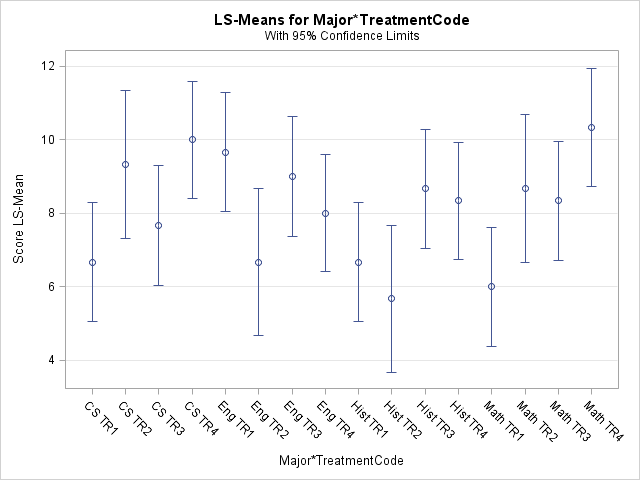
* The plot of the residuals mostly indicates that there’s no visible trend in the residuals, no outliers, and in general, no changing variance.
* The Q-Q plot is a normal probability plot – The plot is almost linear and the assumption of normally distributed residuals holds.

We see desirable results from the residual analysis.

## Mean comparison for Major\* TreatmentCode

|  |  |  |  |
| --- | --- | --- | --- |
| Conservative Tukey-Kramer Grouping for Major\*TreatmentCode Least Squares Means (Alpha=0.05) | | | |
| LS-means with the same letter are not significantly different. | | | |
| Major | **TreatmentCode** | **Estimate** |  |
| Math | **TR4** | 10.3333 | A |
|  |  |  | A |
| CS | **TR4** | 10.0000 | A |
|  |  |  | A |
| Eng | **TR1** | 9.6667 | A |
|  |  |  | A |
| CS | **TR2** | 9.3333 | A |
|  |  |  | A |
| Eng | **TR3** | 9.0000 | A |
|  |  |  | A |
| Math | **TR2** | 8.6667 | A |
|  |  |  | A |
| Hist | **TR3** | 8.6667 | A |
|  |  |  | A |
| Hist | **TR4** | 8.3333 | A |
|  |  |  | A |
| Math | **TR3** | 8.3333 | A |
|  |  |  | A |
| Eng | **TR4** | 8.0000 | A |
|  |  |  | A |
| CS | **TR3** | 7.6667 | A |
|  |  |  | A |
| CS | **TR1** | 6.6667 | A |
|  |  |  | A |
| Hist | **TR1** | 6.6667 | A |
|  |  |  | A |
| Eng | **TR2** | 6.6667 | A |
|  |  |  | A |
| Math | **TR1** | 6.0000 | A |
|  |  |  | A |
| Hist | **TR2** | 5.6667 | A |
| The LINES display does not reflect all significant comparisons. | | | |

**The following additional pairs are significantly different: (Math TR4,Math TR1), (CS TR4,CS TR3), (CS TR4,CS TR1), (Eng TR1,Eng TR2), (CS TR2,CS TR1), (Math TR2,Math TR1), (Hist TR3,Hist TR2).**



**Recommendations**: We have seen that this case study required a repeated measures analysis with unstructured covariance structure. We also saw that the interaction between the treatments and major of the student are significant.

**Considerations**:

Some recommendations based on our analysis are as follows:

* We are not entirely sure whether Major is a random or fixed effect. This should be further investigated with the client.
* The wash-out period issue is glossed over and may cause significant challenge. This type of a repeated measures study is a cross-over design. Cross-over designs need to use a wash-out period ¬ between treatment applications to prevent (or minimize) carry-over effects. Carry-over effects occur when the application of one treatment affects the response of the next treatment applied in the cross-over design. The coding for analysis of cross-over designs are very similar to repeated measures in time, with the addition of a ‘sequence’ variable added initially to our model to test for the presence of carry-over effects.

**Appendix A – SAS Code**

|  |  |
| --- | --- |
| **SAS Code with treatment code (used above)** | **SAS Code without using treatment code** |
| data memory;  input Major $ Student Test Wordlist $ Distracter $ TreatmentCode $ Score;  datalines;  Math 1 1 abstract poetry TR1 6  Math 1 2 abstract math TR2 8  Math 1 3 concrete poetry TR3 8  Math 1 4 concrete math TR4 11  Math 2 5 abstract poetry TR1 5  Math 2 6 abstract math TR2 8  Math 2 7 concrete poetry TR3 8  Math 2 8 concrete math TR4 10  Math 3 9 abstract poetry TR1 7  Math 3 10 abstract math TR2 10  Math 3 11 concrete poetry TR3 9  Math 3 12 concrete math TR4 10  CS 4 13 abstract poetry TR1 9  CS 4 14 abstract math TR2 12  CS 4 15 concrete poetry TR3 10  CS 4 16 concrete math TR4 13  CS 5 17 abstract poetry TR1 4  CS 5 18 abstract math TR2 8  CS 5 19 concrete poetry TR3 8  CS 5 20 concrete math TR4 9  CS 6 21 abstract poetry TR1 7  CS 6 22 abstract math TR2 8  CS 6 23 concrete poetry TR3 5  CS 6 24 concrete math TR4 8  Hist 7 25 abstract poetry TR1 8  Hist 7 26 abstract math TR2 7  Hist 7 27 concrete poetry TR3 9  Hist 7 28 concrete math TR4 10  Hist 8 29 abstract poetry TR1 7  Hist 8 30 abstract math TR2 6  Hist 8 31 concrete poetry TR3 10  Hist 8 32 concrete math TR4 8  Hist 9 33 abstract poetry TR1 5  Hist 9 34 abstract math TR2 4  Hist 9 35 concrete poetry TR3 7  Hist 9 36 concrete math TR4 7  Eng 10 37 abstract poetry TR1 10  Eng 10 38 abstract math TR2 9  Eng 10 39 concrete poetry TR3 10  Eng 10 40 concrete math TR4 8  Eng 11 41 abstract poetry TR1 9  Eng 11 42 abstract math TR2 4  Eng 11 43 concrete poetry TR3 9  Eng 11 44 concrete math TR4 8  Eng 12 45 abstract poetry TR1 10  Eng 12 46 abstract math TR2 7  Eng 12 47 concrete poetry TR3 8  Eng 12 48 concrete math TR4 8  ;  /\* Split-Plot Approach \*/  title 'Split-Plot approach';  proc mixed data=memory method=type3;  class Student TreatmentCode Major;  model Score = TreatmentCode Major TreatmentCode\*Major / ddfm=kr;  random Student(Major);  run;  title 'Repeated measures approach';  proc mixed data=memory method=ml covtest;  class Student Major TreatmentCode;  model Score = TreatmentCode Major TreatmentCode\*Major / ddfm=kr;  repeated / subject=Student type=CS rcorr;  run;  ods output FitStatistics=FitCS (rename=(value=CS))  FitStatistics=FitCSp;  title 'Compound Symmetry'; run;  proc mixed data=memory method=ml covtest;  class Student Major TreatmentCode;  model Score = TreatmentCode Major TreatmentCode\*Major / ddfm=kr;  repeated / subject=Student;  run;  ods output FitStatistics=FitVC (rename=(value=VC))  FitStatistics=FitVCp;  title 'Variance Components'; run;  proc mixed data=memory method=ml covtest;  class Student Major TreatmentCode;  model Score = TreatmentCode Major TreatmentCode\*Major / ddfm=kr;  repeated / subject=Student type=un rcorr;  run;  ods output FitStatistics=FitUN (rename=(value=UN))  FitStatistics=FitUNp;  title 'Unstructured'; run;  title 'Covariance Summary'; run;  data fits;  merge FitCS FitVC FitUN;  run;  ods listing; proc print data=fits; run;  ods graphics on;  ods html style=statistical sge=on;  title 'Unstructured'; run;  proc mixed data=memory method=ml covtest  plot=all;  class Student Major TreatmentCode;  model Score = TreatmentCode Major TreatmentCode\*Major / ddfm=kr;  repeated / subject=Student type=un rcorr;  store final\_model;  run;    proc plm restore=final\_model;  lsmeans TreatmentCode\*Major / adjust=tukey plot=meanplot cl lines;  /\*lsmeans TreatmentCode / adjust=tukey plot=meanplot cl lines;  lsmeans Major / adjust=tukey plot=meanplot cl lines;  \*/  ods exclude diffs diffplot;  run; | data memory;  input Major $ Student Test Wordlist $ Distracter $ TreatmentCode $ Score;  datalines;  Math 1 1 abstract poetry TR1 6  Math 1 2 abstract math TR2 8  Math 1 3 concrete poetry TR3 8  Math 1 4 concrete math TR4 11  Math 2 5 abstract poetry TR1 5  Math 2 6 abstract math TR2 8  Math 2 7 concrete poetry TR3 8  Math 2 8 concrete math TR4 10  Math 3 9 abstract poetry TR1 7  Math 3 10 abstract math TR2 10  Math 3 11 concrete poetry TR3 9  Math 3 12 concrete math TR4 10  CS 4 13 abstract poetry TR1 9  CS 4 14 abstract math TR2 12  CS 4 15 concrete poetry TR3 10  CS 4 16 concrete math TR4 13  CS 5 17 abstract poetry TR1 4  CS 5 18 abstract math TR2 8  CS 5 19 concrete poetry TR3 8  CS 5 20 concrete math TR4 9  CS 6 21 abstract poetry TR1 7  CS 6 22 abstract math TR2 8  CS 6 23 concrete poetry TR3 5  CS 6 24 concrete math TR4 8  Hist 7 25 abstract poetry TR1 8  Hist 7 26 abstract math TR2 7  Hist 7 27 concrete poetry TR3 9  Hist 7 28 concrete math TR4 10  Hist 8 29 abstract poetry TR1 7  Hist 8 30 abstract math TR2 6  Hist 8 31 concrete poetry TR3 10  Hist 8 32 concrete math TR4 8  Hist 9 33 abstract poetry TR1 5  Hist 9 34 abstract math TR2 4  Hist 9 35 concrete poetry TR3 7  Hist 9 36 concrete math TR4 7  Eng 10 37 abstract poetry TR1 10  Eng 10 38 abstract math TR2 9  Eng 10 39 concrete poetry TR3 10  Eng 10 40 concrete math TR4 8  Eng 11 41 abstract poetry TR1 9  Eng 11 42 abstract math TR2 4  Eng 11 43 concrete poetry TR3 9  Eng 11 44 concrete math TR4 8  Eng 12 45 abstract poetry TR1 10  Eng 12 46 abstract math TR2 7  Eng 12 47 concrete poetry TR3 8  Eng 12 48 concrete math TR4 8 ;  /\* Split-Plot Approach \*/  title 'Split-Plot approach';  proc mixed data=memory method=ml covtest;  class Student Wordlist Distracter Major;  /\*model Score = Wordlist|Distracter|Major / DDFM=KR residual;\*/  model Score = Wordlist Distracter Wordlist\*Distracter Major Wordlist\*Major Distracter\*Major Wordlist\*Distracter\*Major / ddfm=kr;  random Student(Major);  run;    /\* Repeated Measures Approach \*/  /\* Fitting covariance structures: \*/  /\* Note: the code beginning with "ods output...." for each  run of the Mixed procedure generates an output that  is tablulated at the end to enable comparison of  the candidate covariance structures  \*/  title 'Repeated measures approach';  proc mixed data=memory method=ml covtest;  class Student Wordlist Distracter Major;  model Score = Wordlist Distracter Wordlist\*Distracter Major Wordlist\*Major Distracter\*Major Wordlist\*Distracter\*Major / ddfm=kr;  repeated / subject=Student(Major) type=CS rcorr;  run;  ods output FitStatistics=FitCS (rename=(value=CS))  FitStatistics=FitCSp;  title 'Compound Symmetry'; run;  proc mixed data=memory method=ml covtest;  class Student Wordlist Distracter Major;  model Score = Wordlist Distracter Wordlist\*Distracter Major Wordlist\*Major Distracter\*Major Wordlist\*Distracter\*Major / ddfm=kr;  repeated / subject=Student(Major);  run;  ods output FitStatistics=FitVC (rename=(value=VC))  FitStatistics=FitVCp;  title 'Variance Components'; run;  proc mixed data=memory method=ml covtest;  class Student Wordlist Distracter Major;  model Score = Wordlist Distracter Wordlist\*Distracter Major Wordlist\*Major Distracter\*Major Wordlist\*Distracter\*Major / ddfm=kr;  repeated / subject=Student(Major) type=un rcorr;  run;  ods output FitStatistics=FitUN (rename=(value=UN))  FitStatistics=FitUNp;  title 'Unstructured'; run;  title 'Covariance Summary'; run;  data fits;  merge FitCS FitVC FitUN;  by descr;  run;  ods listing; proc print data=fits; run; |