Linear Models for experimental designs

E. Papageorgiou, G. Katsouleas

University of West Attica

June 19, 2024

1. Introduction: AnOVa as as completely randomized design

2. The randomized complete block design

3. Experiment with two or more factors

4. The repeated measures design
    - Two-Factor repeated measures design

AnOVa as completely randomized design

AnOVa as completely randomized design

- An outcome variable is represented by the set of measured values that result from an experiment or some other statistical process.
- An explanatory variable, on the other hand, is a variable that is useful for predicting the value of the outcome variable.
- A linear model is any model that is linear in the parameters that define the model. We can represent such models generically in the form:

  $$Y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \cdots + \beta_k X_{kj} + \epsilon_j,$$

  In this equation, β_j represent the coefficients in the model and ϵ_j represents random error (due to extraneous variables). Therefore, any model that can be represented in this form, where the coefficients are constants and the algebraic order of the model is one, is considered a linear model.
- In the context of analysis of variance, the predictor variables are classification variables used to define factors of interest (e.g., differentiating between a control group and a treatment group–treatment variables), and in the context of correlation and linear regression the predictor variables are most often continuous variables, or at least variables at a higher level than nominal classes.
- Question: Do the different “values” of the treatment variable result in differences, on the average, in the response variable?

AnoVa as completely randomized design (2)

- The one-way analysis of variance model may be written as follows:

$$x_{ij} = \mu + \tau_j + \epsilon_{ij} \quad i = 1, 2, \ldots, n_j, \quad j = 1, 2, \ldots, k.$$

  Here:

  - x_{ij} represents the i–th observation resulting from the j–th treatment of a total of k treatments.
  - μ represents the mean of all k population means and is called the grand mean.
  - τ_j represents the difference between the mean of the j–th population and the grand mean and is called the treatment effect.
  - ϵ_{ij} represents the amount by which an individual measurement differs from the mean of the population to which it belongs and is called the error term.

- Using the means comparison notation in the previous set of slides, we clearly have $\mu_j \equiv \mu + \tau_j$, i.e., the mean of the j–th population.
- In most situations we are interested only in the k treatments represented in our experiment. Any inferences that we make apply only to these treatments. We do not wish to extend our inference to any larger collection of treatments. When we place such a restriction on our inference goals, we refer to our model as the fixed-effects model.
- The experiment is designed in such a way that the treatments of interest are assigned completely at random to the subjects or objects on which the measurements to determine treatment effectiveness are to be made. For this reason the design is called the completely randomized experimental design.

AnoVa: Assumptions in the context of fixed-effects model

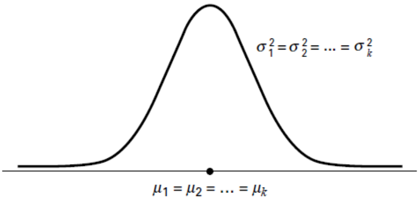
- Assumptions.
  1. The k sets of observed data constitute k independent random samples from the respective populations.
  2. Each of the populations from which the samples come is normally distributed with mean μ_j and variance σ_j^2.
  3. Each of the populations has the same variance. That is, $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma^2$ the common variance.
  4. The τ_j are unknown constants and $\sum_{j=1}^{k} \tau_j = 0$ since the sum of all deviations of the μ_j from their mean, μ, is zero.
  5. The ϵ_{ij} have a mean of 0, since the mean of x_{ij} is μ_j.
  6. The ϵ_{ij} have a variance equal to the variance of the x_{ij}, since the ϵ_{ij} and x_{ij} differ only by a constant; that is, the error variance is equal to σ^2, the common variance specified above.
  7. The ϵ_{ij} are normally (and independently) distributed.

AnoVa: Assumptions in the context of fixed-effects model

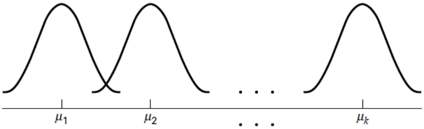
- Assumptions.
  1. The k sets of observed data constitute k independent random samples from the respective populations.
  2. Each of the populations from which the samples come is normally distributed with mean μ_j and variance σ_j^2.
  3. Each of the populations has the same variance. That is, $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma^2$ the common variance.
  4. The τ_j are unknown constants and $\sum_{j=1}^{k} \tau_j = 0$ since the sum of all deviations of the μ_j from their mean, μ, is zero.
  5. The ϵ_{ij} have a mean of 0, since the mean of x_{ij} is μ_j.
  6. The ϵ_{ij} have a variance equal to the variance of the x_{ij}, since the ϵ_{ij} and x_{ij} differ only by a constant; that is, the error variance is equal to σ^2, the common variance specified above.
  7. The ϵ_{ij} are normally (and independently) distributed.
- Hypotheses:

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \cdots = \mu_k, \\ H_a : \text{ not all } \mu_j \text{ are equal.} \end{cases} \quad \Leftrightarrow \quad \begin{cases} H_0 : \tau_1 = \tau_2 = \cdots = \tau_k = 0, \\ H_a : \text{ not all } \tau_j = 0. \end{cases}$$

Implications of the assumptions


$\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$
$\mu_1 = \mu_2 = \dots = \mu_k$

- Picture of the populations represented in a completely randomized design when H_0 is true and the assumptions are met.
- If the populations are all normally distributed with equal variances the distributions will be identical, so that in drawing their pictures each is superimposed on each of the others, and a single picture sufficiently represents them all.

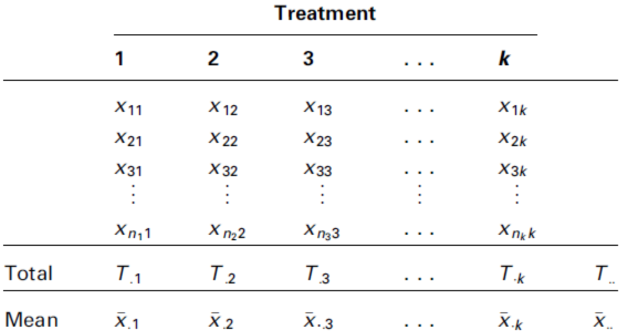


- Picture of the populations represented in a completely randomized design when the assumptions of equal variances and normally distributed populations are met, but H_0 is false because none of the population means are equal.

AnoVa: Why not use a number of independent stamples *t*–tests instead?

- When interested in testing the null hypothesis of no difference among several population means one might be inclined to suggest that all possible pairs of sample means be tested separately by means of the Student *t*–test.
- Suppose there are five populations involved. The number of possible pairs of sample means is $\binom{5}{2} = \frac{5!}{2! \cdot (5-2)!} = 10$.
- As the amount of work involved in carrying out this many *t*–tests is substantial, it would be worthwhile if a more efficient alternative for analysis were available. A more important consequence of performing all possible *t*–tests, however, is that it is very likely to lead to a false conclusion.
- Suppose we draw five samples from populations having equal means.
  - As we have seen, there would be 10 tests if we were to do each of the possible tests separately. If we select a significance level of $\alpha = 0.05$ for each test, the probability of failing to reject a hypothesis of no difference in each case would be 0.95.
  - By the multiplication rule of probability, if the tests were independent of one another, the probability of failing to reject a hypothesis of no difference in all 10 cases would be $\alpha = 0.95^{10} = 0.5987$.
  - The probability of rejecting at least one hypothesis of no difference, then, would be $1 - 0.5987 = 0.4013$. Since we know that the null hypothesis is true in every case in this illustrative example, rejecting the null hypothesis constitutes the committing of a type I error.
- In the long run, then, in testing all possible pairs of means from five samples, we would commit a type I error 40 percent of the time. The problem becomes even more complicated in practice, since three or more *t*–tests based on the same data would not be independent of one another.
- It becomes clear, then, that some other method for testing for a significant difference among several means is needed. Analysis of variance provides such a method.

Sample Values for the Completely Randomized Design



	Treatment					
	1	**2**	**3**	. . .	***k***	
	x_{11} x_{21} x_{31} $\vdots$ $x_{n_1 1}$	x_{12} x_{22} x_{32} $\vdots$ $x_{n_2 2}$	x_{13} x_{23} x_{33} $\vdots$ $x_{n_3 3}$	. . . . $\vdots$ . . .	x_{1k} x_{2k} x_{3k} $\vdots$ $x_{n_k k}$	
Total	$T_{.1}$	$T_{.2}$	$T_{.3}$	. . .	$T_{.k}$	$T_{..}$
Mean	$\bar{x}_{.1}$	$\bar{x}_{.2}$	$\bar{x}_{.3}$	. . .	$\bar{x}_{.k}$	$\bar{x}_{..}$

Here:

- x_{ij} represents the i–th observation resulting from the j–th treatment of a total of k treatments ($i = 1, 3, \ldots, n_j$, $j = 1, 2, \ldots, k$).
- $T_{\cdot j} = \sum_{i=1}^{n_j} x_{ij}$ represents the total of the j–th treatment ($j = 1, 2, \ldots, k$).
- $\bar{x}_{\cdot j} = \frac{T_{\cdot j}}{n_j}$ represents the mean of the j–th treatment ($j = 1, 2, \ldots, k$).
- $T_{\cdot\cdot} = \sum_{j=1}^{k} T_{\cdot j} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} x_{ij}$ represents the total of all observations.
- $\bar{x}_{\cdot\cdot} = \frac{T_{\cdot\cdot}}{N}$, where $N = \sum_{j=1}^{k} n_j$.

The randomized complete block design

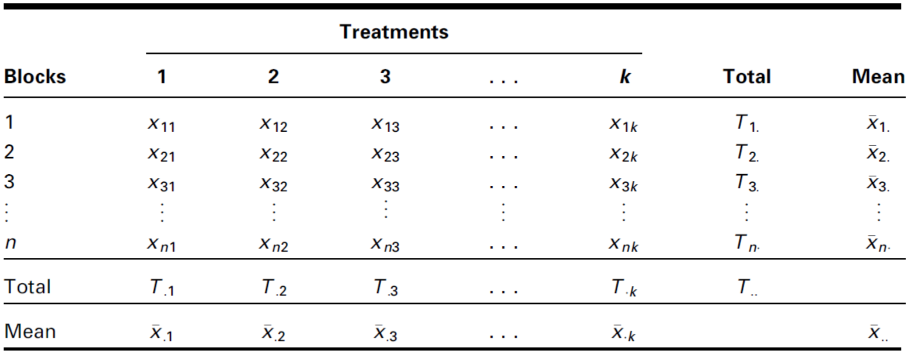
The randomized complete block design

- The randomized complete block design is a design in which the units (called experimental units) to which the treatments are applied are subdivided into homogeneous groups called blocks, so that the number of experimental units in a block is equal to the number (or some multiple of the number) of treatments being studied.
- The treatments are then assigned at random to the experimental units within each block.
- It should be emphasized that each treatment appears in every block, and each block receives every treatment.
- Objective: The objective in using the randomized complete block design is to isolate and remove from the error term the variation attributable to the blocks, while assuring that treatment means will be free of block effects.
- The effectiveness of the design depends on the ability to achieve homogeneous blocks of experimental units.
- The ability to form homogeneous blocks depends on the researcher's knowledge of the experimental material.
- When blocking is used effectively, the error mean square in the ANOVA table will be reduced, the Variance Ratio will be increased, and the chance of rejecting the null hypothesis will be improved.

The randomized complete block design (2)

- In animal experiments, the breed of animal may be used as a blocking factor. Litters may also be used as blocks, (an animal from each litter receives a treatment).
- In experiments involving human beings, if it is desired that differences resulting from age be eliminated, then subjects may be grouped according to age so that one person of each age receives each treatment.
- The randomized complete block design also may be employed effectively when an experiment must be carried out in more than one laboratory (block) or when several days (blocks) are required for completion.
- The random allocation of treatments to subjects is restricted in the randomized complete block design. That is, each treatment must be represented an equal number of times (one or more times) within each blocking unit.
- In practice, this is generally accomplished by assigning a random permutation of the order of treatments to subjects within each block.
- For example, if there are four treatments representing three drugs and a placebo (drug A, drug B, drug C, and placebo P), then there are $4! = 24$ possible permutations of the four treatments: (A, B, C, P) or (A, C, B, P) or (C, A, P, B), and so on. One permutation is then randomly assigned to each block.
- Note that the paired comparisons analysis is a special case of the randomized complete block design. Indeed, the two points in time (before & after, for instance, Pre-op & Post-op) are the treatments and the individuals on whom the measurements were taken are the blocks.

Table of Sample Values for the Randomized Complete Block Design



	Treatments						
Blocks	**1**	**2**	**3**	**. . .**	**k**	**Total**	**Mean**
1	x_{11}	x_{12}	x_{13}	$\dots$	x_{1k}	$T_{1.}$	$\bar{x}_{1.}$
2	x_{21}	x_{22}	x_{23}	$\dots$	x_{2k}	$T_{2.}$	$\bar{x}_{2.}$
3	x_{31}	x_{32}	x_{33}	$\dots$	x_{3k}	$T_{3.}$	$\bar{x}_{3.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	x_{n1}	x_{n2}	x_{n3}	$\dots$	x_{nk}	$T_{n\cdot}$	$\bar{x}_{n\cdot}$
Total	$T_{.1}$	$T_{.2}$	$T_{.3}$	$\dots$	$T_{\cdot k}$	$T_{..}$	
Mean	$\bar{x}_{.1}$	$\bar{x}_{.2}$	$\bar{x}_{.3}$	$\dots$	$\bar{x}_{\cdot k}$		$\bar{x}_{..}$

Here:

- $T_{i\cdot} = \sum_{j=1}^{k} x_{ij}$ represents the total of the i–th block $(i = 1, 2, \dots, n)$.
- $\overline{x}_{i\cdot} = \frac{T_{i\cdot}}{k}$ represents the mean of the i–th block $(i = 1, 2, \dots, n)$.
- $T_{\cdot\cdot} = \sum_{j=1}^{k} T_{\cdot j} = \sum_{i=1}^{n} T_{i\cdot}$ represents the total of all observations.

Two-way AnoVa

- Two-way AnoVa. The technique for analyzing the data from a randomized complete block design is called two-way analysis of variance since an observation is categorized on the basis of two criteria—the block to which it belongs as well as the treatment group to which it belongs.
- The two-way analysis of variance model may be written as follows:

$$x_{ij} = \mu + \beta_i + \tau_j + \epsilon_{ij} \quad i = 1, 2, \ldots, n, \qquad j = 1, 2, \ldots, k.$$

  Here:

  - μ represents the mean of all k population means and is called the grand mean.
  - β_i represents a block effect reflecting the fact that the experimental unit fell in the i–th block.
  - τ_j represents a treatment effect, reflecting the fact that the experimental unit received the j–th treatment.
  - ϵ_{ij} is a residual component representing all sources of variation other than treatments and blocks.

- Hypotheses:

$$\begin{cases} H_0 : \tau_j = 0, \quad j = 1, 2, \ldots, k \quad \text{vs.} \\ H_a : \ \text{not all } \tau_j = 0. \end{cases}$$

Two-way AnoVa (2)

- Hypotheses:

$$\begin{cases} H_0 : \tau_j = 0, \quad j = 1, 2, \ldots, k \;\; \text{vs.} \\ H_a : \;\text{not all}\; \tau_j = 0. \end{cases}$$

- A hypothesis test regarding block effects is not usually carried out under the assumptions of the fixed-effects model for two reasons:
  1. First, the primary interest is in treatment effects, the usual purpose of the blocks being to provide a means of eliminating an extraneous source of variation.
  2. Second, although the experimental units are randomly assigned to the treatments, the blocks are obtained in a nonrandom manner.

Two-way AnoVa: test statistic

- Hypotheses:

$$\begin{cases} H_0 : \tau_j = 0, \quad j = 1, 2, \ldots, k \quad \text{vs.} \\ H_a : \text{ not all } \tau_j = 0. \end{cases}$$

- Analysis of Variance: It can be shown that the total sum of squares for the randomized complete block design can be partitioned into three components, one each attributable to blocks ($SSBl$), treatments ($SSTr$), and error (SSE). That is,

$$SST = SSBl + SSTr + SSE,$$

  where

  - $SST = \sum_{j=1}^{k} \sum_{i=1}^{n} \left(x_{ij} - \overline{x}_{\cdot\cdot}\right)^2,$
  - $SSBl = \sum_{j=1}^{k} \sum_{i=1}^{n} \left(\overline{x}_{i\cdot} - \overline{x}_{\cdot\cdot}\right)^2,$
  - $SSTr = \sum_{j=1}^{k} \sum_{i=1}^{n} \left(\overline{x}_{\cdot j} - \overline{x}_{\cdot\cdot}\right)^2,$
  - $SSE = SST - SSBl - SSTr.$

- Degrees of freedom:

$$\overbrace{kn-1}^{\text{Total}} = \overbrace{n-1}^{\text{Blocks}} + \overbrace{k-1}^{\text{Treatments}} + \overbrace{(n-1)(k-1)}^{\text{Residual}}$$

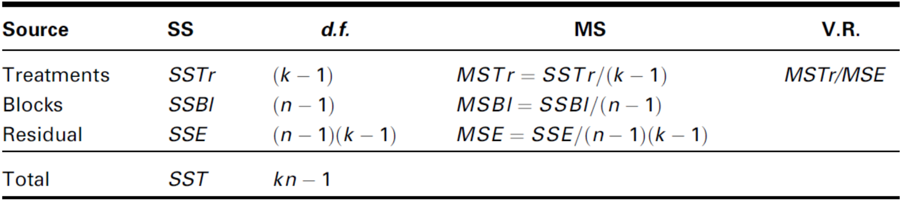
- Test statistic: $MSTr/MSE \sim F_{k-1,(n-1)(k-1)}$

ANOVA Table for the Randomized Complete Block Design

- Hypotheses:

$$\begin{cases} H_0 : \tau_j = 0, \quad j = 1, 2, \ldots, k \quad \text{vs.} \\ H_a : \ \text{not all } \tau_j = 0. \end{cases}$$

- Test statistic: $MSTr/MSE \sim F_{k-1,(n-1)(k-1)}$



Source	**SS**	***d.f.***	**MS**	**V.R.**
Treatments	$SSTr$	$(k-1)$	$MSTr = SSTr/(k-1)$	$MSTr/MSE$
Blocks	$SSBl$	$(n-1)$	$MSBl = SSBl/(n-1)$	
Residual	SSE	$(n-1)(k-1)$	$MSE = SSE/(n-1)(k-1)$	
Total	SST	$kn-1$		

Randomized Complete Block Design: Assumptions

- Assumptions.

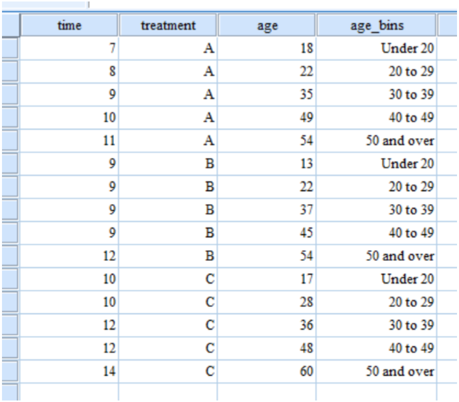
1. Each x_{ij} that is observed constitutes a random independent sample of size 1 from one of the kn populations represented.
2. Each of these kn populations is normally distributed with mean μ_{ij} and the same variance s^2. This implies that the ϵ_{ij} are independently and normally distributed with mean 0 and variance s^2.
3. The block and treatment effects are additive. This assumption may be interpreted to mean that there is no interaction between treatments and blocks. In other words, a particular block-treatment combination does not produce an effect that is greater or less than the sum of their individual effects. It can be shown that when this assumption is met,

$$\sum_{j=1}^{k} \tau_j = \sum_{i=1}^{n} \beta_i = 0.$$

The consequences of a violation of this assumption are misleading results. One need not become concerned with the violation of the additivity assumption, unless the largest mean is more than 50 percent greater than the smallest.

When these assumptions hold true, the τ_j and β_i are a set of fixed constants, and we have a situation that fits the fixed-effects model.

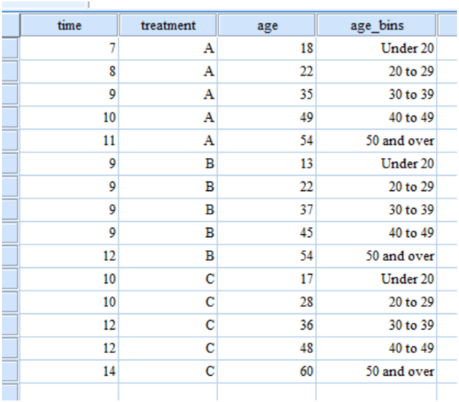
Example: Days Time Required to Learn the Use of a Certain Prosthetic Device



time	treatment	age	age_bins
7	A	18	Under 20
8	A	22	20 to 29
9	A	35	30 to 39
10	A	49	40 to 49
11	A	54	50 and over
9	B	13	Under 20
9	B	22	20 to 29
9	B	37	30 to 39
9	B	45	40 to 49
12	B	54	50 and over
10	C	17	Under 20
10	C	28	20 to 29
12	C	36	30 to 39
12	C	48	40 to 49
14	C	60	50 and over

- A physical therapist wished to compare three methods for teaching patients to use a certain prosthetic device.
- He felt that the rate of learning would be different for patients of different ages and wished to design an experiment in which the influence of age could be taken into account.
- Data. Three patients in each of five age groups were selected to participate in the experiment, and one patient in each age group was randomly assigned to each of the teaching methods.
- The methods of instruction constitute our three treatments, and the five age groups are the blocks.

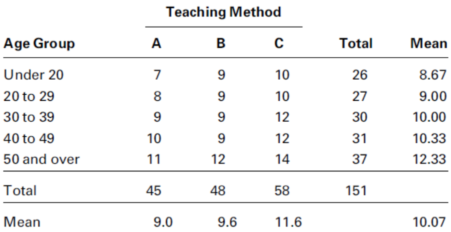
Example: Days Time Required to Learn the Use of a Certain Prosthetic Device (2)



time	treatment	age	age_bins
7	A	18	Under 20
8	A	22	20 to 29
9	A	35	30 to 39
10	A	49	40 to 49
11	A	54	50 and over
9	B	13	Under 20
9	B	22	20 to 29
9	B	37	30 to 39
9	B	45	40 to 49
12	B	54	50 and over
10	C	17	Under 20
10	C	28	20 to 29
12	C	36	30 to 39
12	C	48	40 to 49
14	C	60	50 and over

- Assumptions. We assume that each of the 15 observations constitutes a simple random sample of size 1 from one of the 15 populations defined by a block-treatment combination.
- For example, we assume that the number 7 in the table constitute s a randomly selected response from a population of responses that would result if a population of subjects under the age of 20 received teaching method A.
- We assume that the responses in the 15 represented populations are normally distributed with equal variances.

Calculation of test statistic



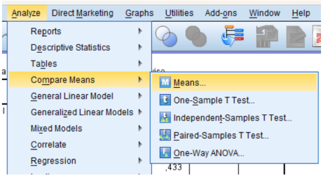
	Teaching Method				
Age Group	A	B	C	Total	Mean
Under 20	7	9	10	26	8.67
20 to 29	8	9	10	27	9.00
30 to 39	9	9	12	30	10.00
40 to 49	10	9	12	31	10.33
50 and over	11	12	14	37	12.33
Total	45	48	58	151	
Mean	9.0	9.6	11.6		10.07

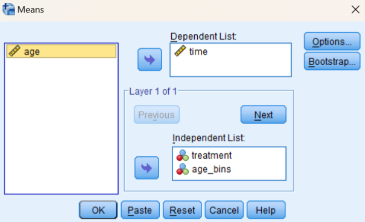
- We compute the following sums of squares:

$$
\begin{aligned}
SST &= (7-10.07)^2 + (8-10.07)^2 + \cdots + (14-10.07)^2 = 46.933,\\
SSBl &= 3\left[(8.67-10.07)^2 + (9-10.07)^2 + \cdots + (12.33-10.07)^2\right] = 24.933,\\
SSTr &= 5\left[(9-10.07)^2 + (9.6-10.07)^2 + (11.6-10.07)^2\right] = 18.533,\\
SSE &= 46.933 - 24.933 - 18.533 = 3.467.
\end{aligned}
$$

- Degrees of freedom.
  (a.) Total: $3 \times 5 - 1 = 14$,
  (b.) Blocks: $5 - 1 = 4$,
  (c.) Treatments: $3 - 1 = 2$,
  (d.) Residual (Error): $(3-1) \times (5-1) = 8$.
- Variance Ratio $= MSTr/MSE = 21.385$

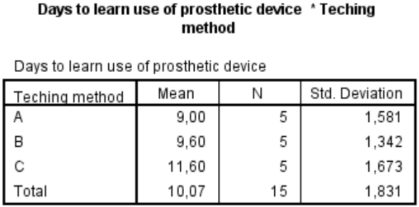
Marginal means in SPSS





Dependent List: time

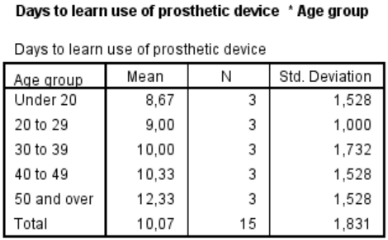
Independent List: treatment, age_bins



Days to learn use of prosthetic device * Teching method

Days to learn use of prosthetic device

Teching method	Mean	N	Std. Deviation
A	9,00	5	1,581
B	9,60	5	1,342
C	11,60	5	1,673
Total	10,07	15	1,831

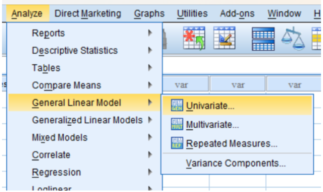


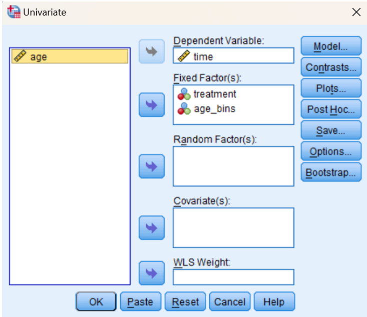
Days to learn use of prosthetic device * Age group

Days to learn use of prosthetic device

Age group	Mean	N	Std. Deviation
Under 20	8,67	3	1,528
20 to 29	9,00	3	1,000
30 to 39	10,00	3	1,732
40 to 49	10,33	3	1,528
50 and over	12,33	3	1,528
Total	10,07	15	1,831

Two-way AnoVa in SPSS





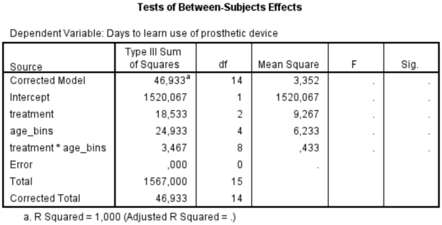
Dependent Variable: time

Fixed Factor(s): treatment, age_bins

Random Factor(s):

Covariate(s):

WLS Weight:



Tests of Between-Subjects Effects

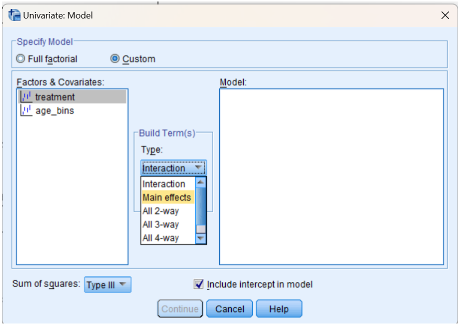
Dependent Variable: Days to learn use of prosthetic device

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	46,933[a]	14	3,352	.	.
Intercept	1520,067	1	1520,067	.	.
treatment	18,533	2	9,267	.	.
age_bins	24,933	4	6,233	.	.
treatment * age_bins	3,467	8	,433	.	.
Error	,000	0	.		
Total	1567,000	15			
Corrected Total	46,933	14			

a. R Squared = 1,000 (Adjusted R Squared = .)

- Note that Interaction effects are included in this output.
- Resultingly, the Variance Ratio cannot be computed.
- We could compute relevant ratio by hand (compare with previous "Calculation of test statistic" slide), or..

Two-way AnoVa in SPSS: Model selection


Univariate: Model

Specify Model
○ Full factorial ● Custom

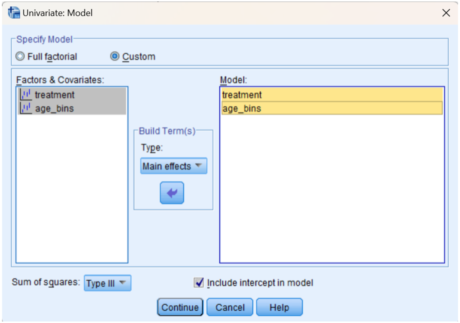
Factors & Covariates:
treatment
age_bins

Build Term(s)
Type: Interaction

Model:

Sum of squares: Type III

☑ Include intercept in model


Univariate: Model

Specify Model
○ Full factorial ● Custom

Factors & Covariates:
treatment
age_bins

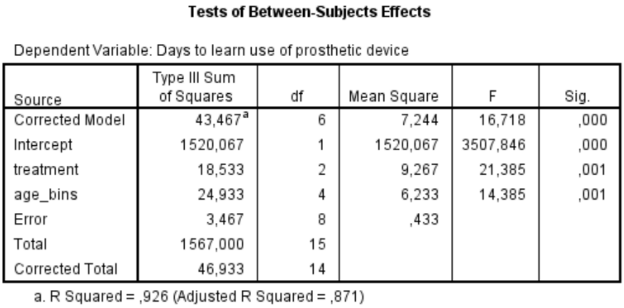
Build Term(s)
Type: Main effects

Model:
treatment
age_bins

Sum of squares: Type III

☑ Include intercept in model

Two-way AnoVa in SPSS: Output without interaction effects



Tests of Between-Subjects Effects

Dependent Variable: Days to learn use of prosthetic device

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	43,467[a]	6	7,244	16,718	,000
Intercept	1520,067	1	1520,067	3507,846	,000
treatment	18,533	2	9,267	21,385	,001
age_bins	24,933	4	6,233	14,385	,001
Error	3,467	8	,433		
Total	1567,000	15			
Corrected Total	46,933	14			

a. R Squared = ,926 (Adjusted R Squared = ,871)

- Statistical decision. Since our computed variance ratio, 21.385, is greater than the critical value 4.46 (F(2,8)), we reject the null hypothesis of no treatment effects on the assumption that such a large V.R. reflects the fact that the two sample mean squares are not estimating the same quantity.
  The only other explanation for this large V.R. would be that the null hypothesis is really true, and we have just observed an unusual set of results. We rule out the second explanation in favor of the first.

Experiment with two or more factors

The factorial experiment

- In the experimental designs that we have considered up to this point, we have been interested in the effects of only one variable—the treatments. Frequently, however, we may be interested in studying, simultaneously, the effects of two or more variables.
- We refer to the variables in which we are interested as factors. The experiment in which two or more factors are investigated simultaneously is called a factorial experiment.
- The different designated categories of the factors are called levels.
  - Suppose, for example, that we are studying the effect on reaction time of three dosages of some drug. The drug factor, then, is said to occur at three levels.
  - Suppose the second factor of interest in the study is age, and it is thought that two age groups, under 65 years and 65 years and older, should be included. We then have two levels of the age factor.

  In general, we say that factor A occurs at a levels and factor B occurs at b levels.
- In a factorial experiment we may study not only the effects of individual factors but also, if the experiment is properly conducted, the interaction between factors.

Example: No interaction

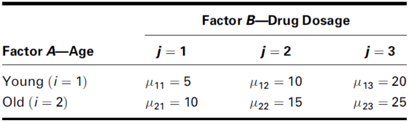
- Suppose, in terms of effect on reaction time, that the true relationship between three dosage levels of some drug and the age of human subjects taking the drug is known.
- Suppose further that age occurs at two levels—"young" (under 65) and "old" (65 and older). If the true relationship between the two factors is known, we will know, for the three dosage levels, the mean effect on reaction time of subjects in the two age groups. Let us assume that effect is measured in terms of reduction in reaction time to some stimulus.

	Factor *B*—Drug Dosage		
Factor *A*—Age	$j = 1$	$j = 2$	$j = 3$
Young ($i = 1$)	$\mu_{11} = 5$	$\mu_{12} = 10$	$\mu_{13} = 20$
Old ($i = 2$)	$\mu_{21} = 10$	$\mu_{22} = 15$	$\mu_{23} = 25$

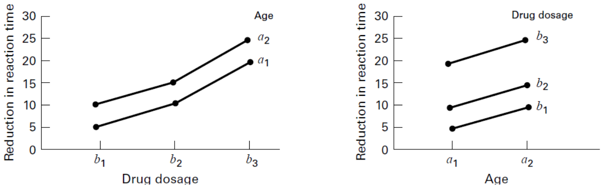
1. For both levels of factor A the difference between the means for any two levels of factor B is the same. That is, for both levels of factor A, the difference between means for levels $j = 1$ and $j = 2$ is 5, for levels $j = 2$ and $j = 3$ the difference is 10, and for levels $j = 1$ and $j = 3$ the difference is 15.
2. For all levels of factor B the difference between means for the two levels of factor A is the same. In the present case, the difference is 5 at all three levels of factor B.

Example: No interaction (2)

1. For both levels of factor A the difference between the means for any two levels of factor B is the same. That is, for both levels of factor A, the difference between means for levels $j = 1$ and $j = 2$ is 5, for levels $j = 2$ and $j = 3$ the difference is 10, and for levels $j = 1$ and $j = 3$ the difference is 15.
2. For all levels of factor B the difference between means for the two levels of factor A is the same. In the present case, the difference is 5 at all three levels of factor B.
3. A third characteristic is revealed when the data are plotted. We note that the curves corresponding to the different levels of a factor are all parallel.



	Factor *B*—Drug Dosage		
Factor *A*—Age	$j = 1$	$j = 2$	$j = 3$
Young ($i = 1$)	$\mu_{11} = 5$	$\mu_{12} = 10$	$\mu_{13} = 20$
Old ($i = 2$)	$\mu_{21} = 10$	$\mu_{22} = 15$	$\mu_{23} = 25$



Example: Interaction present

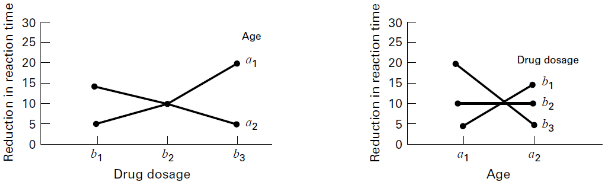
- The presence of interaction between two factors can affect the characteristics of the data in a variety of ways depending on the nature of the interaction. To illustrate:

	Factor *B*—Drug Dosage		
Factor *A*—Age	$j = 1$	$j = 2$	$j = 3$
Young ($i = 1$)	$\mu_{11} = 5$	$\mu_{12} = 10$	$\mu_{13} = 20$
Old ($i = 2$)	$\mu_{21} = 15$	$\mu_{22} = 10$	$\mu_{23} = 5$

1. The difference between means for any two levels of factor B is not the same for both levels of factor A. Note, for example, that the difference between levels $j = 1$ and 2 of factor B is -5 for the young age group and $+5$ for the old age group.
2. The difference between means for both levels of factor A is not the same at all levels of factor B. The differences between factor A means are -10, 0, and 15 for levels $j = 1, 2$ and 3, respectively, of factor B.
3. The factor level curves are not parallel.

Example: Interaction present (2)

	Factor B—Drug Dosage		
Factor A—Age	$j = 1$	$j = 2$	$j = 3$
Young ($i = 1$)	$\mu_{11} = 5$	$\mu_{12} = 10$	$\mu_{13} = 20$
Old ($i = 2$)	$\mu_{21} = 15$	$\mu_{22} = 10$	$\mu_{23} = 5$

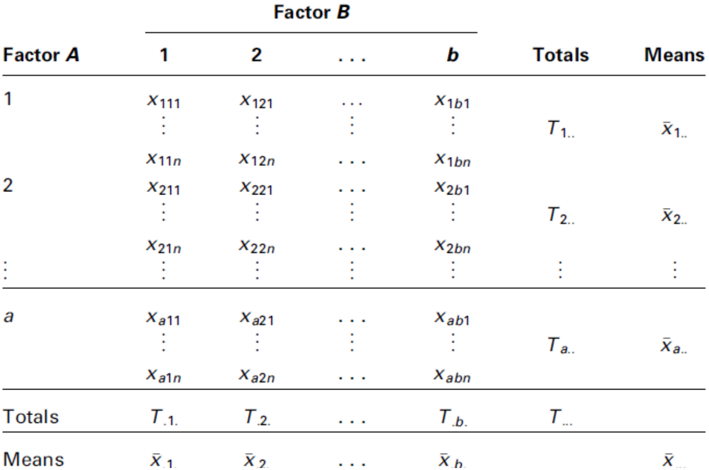


- In summary, then, we can say that there is interaction between two factors if a change in one of the factors produces a change in response at one level of the other factor different from that produced at other levels of this factor.

Two-Factor Completely Randomized Experiment: Advantages

- The interaction of the factors may be studied.
- There is a saving of time and effort.
  In the factorial experiment all the observations may be used to study the effects of each of the factors under investigation. The alternative, when two factors are being investigated, would be to conduct two different experiments, one to study each of the two factors. If this were done, some of the observations would yield information only on one of the factors, and the remainder would yield information only on the other factor. To achieve the level of accuracy of the factorial experiment, more experimental units would be needed if the factors were studied through two experiments. It is seen, then, that 1 two-factor experiment is more economical than 2 one-factor experiments.
- Because the various factors are combined in one experiment, the results have a wider range of application.

Sample Data from a Two-Factor Completely Randomized Experiment



	Factor ***B***					
Factor *A*	**1**	**2**	**. . .**	***b***	**Totals**	**Means**
1	x_{111}	x_{121}	. . .	x_{1b1}		
	⋮	⋮	⋮	⋮	$T_{1..}$	$\bar{x}_{1..}$
	x_{11n}	x_{12n}	. . .	x_{1bn}		
2	x_{211}	x_{221}	. . .	x_{2b1}		
	⋮	⋮	⋮	⋮	$T_{2..}$	$\bar{x}_{2..}$
	x_{21n}	x_{22n}	. . .	x_{2bn}		
⋮	⋮	⋮	⋮	⋮	⋮	⋮
a	x_{a11}	x_{a21}	. . .	x_{ab1}		
	⋮	⋮	⋮	⋮	$T_{a..}$	$\bar{x}_{a..}$
	x_{a1n}	x_{a2n}	. . .	x_{abn}		
Totals	$T_{.1.}$	$T_{.2.}$	. . .	$T_{.b.}$	$T_{...}$	
Means	$\bar{x}_{.1.}$	$\bar{x}_{.2.}$	. . .	$\bar{x}_{.b.}$		$\bar{x}_{...}$

Sample Data from a Two-Factor Completely Randomized Experiment (2)

- Here we have *a* levels of factor A, *b* levels of factor B, and *n* observations for each combination of levels. Each of the *ab* combinations of levels of factor A with levels of factor B is a treatment.
- In addition to the totals and means shown in the Table, we note that the total and mean of the *ij*-th cell are

$$T_{ij\cdot} = \sum_{k=1}^{n} x_{ijk} \text{ and } \overline{x}_{ij\cdot} = T_{ij\cdot}/n \ \ (i = 1, \ldots, a, j = 1, \ldots, b).$$

- We consider that each combination of factor levels is a treatment and that we have *n* observations for each treatment.
- Total number of observations: *nab*.
- The factorial experiment, in order that the experimenter may test for interaction, requires at least two observations per cell, whereas the randomized complete block design (note the similarity of the Tables) requires only one observation per cell. We use two-way analysis of variance to analyze the data from a factorial experiment of the type presented here.

The factorial experiment

- The model for the two-factor repeated measures design must represent the fact that there are two factors, A and B, and they have a potential interaction:

$$x_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \quad i = 1, 2, \ldots, a, \quad j = 1, 2, \ldots, b, \quad k = 1, 2, \ldots, n.$$

  - α_j represents the main effect of factor A,
  - β_k represents the main effect of factor B,
  - $(\alpha\beta)_{jk}$ represents the interaction effect of factor A and factor B,
  - ϵ_{ijk} is a residual component representing all sources of variation other than treatments and blocks (experimental error).

- Assumptions:
  - The observations in each of the *ab* cells constitute a random independent sample of size *n* drawn from the population defined by the particular combination of the levels of the two factors.
  - Each of the *ab* populations is normally distributed.
  - The populations all have the same variance.

Hypotheses

1. $$\begin{cases} H_0 : \alpha_i = 0, \quad i = 1, 2, \dots, a, \\ H_a : \text{not all } \alpha_i = 0. \end{cases}$$

2. $$\begin{cases} H_0 : \beta_j = 0, \quad j = 1, 2, \dots, b, \\ H_a : \text{not all } \beta_j = 0. \end{cases}$$

3. $$\begin{cases} H_0 : (\alpha\beta)_{ij} = 0, \quad i = 1, 2, \dots, a, \quad j = 1, 2, \dots, b, \\ H_a : \text{not all } \ (\alpha\beta)_{ij} = 0. \end{cases}$$

- Before collecting data, the researchers may decide to test only one of the possible hypotheses.
- In this case they select the hypothesis they wish to test, choose a significance level α, and proceed in the familiar, straightforward fashion. This procedure is free of the complications that arise if the researchers wish to test all three hypotheses.
- When all three hypotheses are tested, the situation is complicated by the fact that the three tests are not independent in the probabilistic sense.
- If we let α be the significance level associated with the test as a whole, and α'; α''; and α''' the significance levels associated with hypotheses 1, 2, and 3, respectively, we find

$$\alpha < 1 - \left(1 - \alpha'\right)\left(1 - \alpha''\right)\left(1 - \alpha'''\right).$$

Hence, If $\alpha' = \alpha'' = \alpha''' = 0.05$, then $\alpha < 1 - 0.95^3 = 0.143$. This means that the probability of rejecting one or more of the three hypotheses is less than 0.143 when a significance level of 0.05 has been chosen for the hypotheses and all are true.

Calculation of the test statistic

- By an adaptation of the procedure used in partitioning the total sum of squares for the completely randomized design, it can be shown that the total sum of squares under the present model can be partitioned into two parts as follows:

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(x_{ijk}-\overline{x}_{...})^2=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(\overline{x}_{ij.}-\overline{x}_{...})^2+\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(x_{ijk}-\overline{x}_{ij.})^2,$$

i.e.,

$$SST=SSTr+SSE,$$

where the sum of squares for treatments can be partitioned into three parts as follows:

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(\overline{x}_{ij.}-\overline{x}_{...})^2=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(\overline{x}_{i..}-\overline{x}_{...})^2+\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(\overline{x}_{.j.}-\overline{x}_{...})^2+$$

$$+\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(\overline{x}_{ij.}-\overline{x}_{i..}-\overline{x}_{.j.}+\overline{x}_{...})^2,$$

i.e.,

$$SSTr=SSA+SSB+SSAB.$$

Analysis of Variance Table for a Two-Factor Completely Randomized Experiment (Fixed-Effects Model)

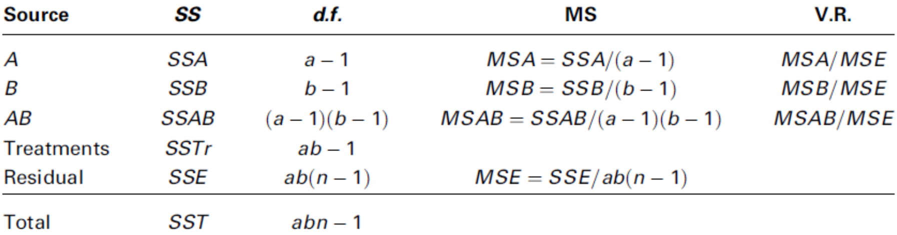
- It can be shown that

$$SST = SSTr + SSE,$$

where the sum of squares for treatments can be partitioned into three parts as follows:

$$SSTr = SSA + SSB + SSAB.$$

- Test statistic: Variance ratios, according to the following AnOVa Table (following F distributions with the indicated degrees of freedom, respectively):

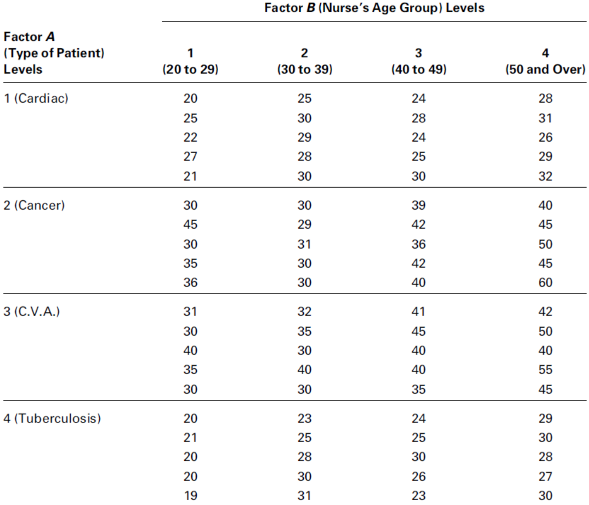


Source	***SS***	***d.f.***	**MS**	**V.R.**
A	SSA	$a-1$	$MSA = SSA/(a-1)$	MSA/MSE
B	SSB	$b-1$	$MSB = SSB/(b-1)$	MSB/MSE
AB	$SSAB$	$(a-1)(b-1)$	$MSAB = SSAB/(a-1)(b-1)$	$MSAB/MSE$
Treatments	$SSTr$	$ab-1$		
Residual	SSE	$ab(n-1)$	$MSE = SSE/ab(n-1)$	
Total	SST	$abn-1$		

Two-Factor Completely Randomized Experiment (Fixed-Effects Model): Application

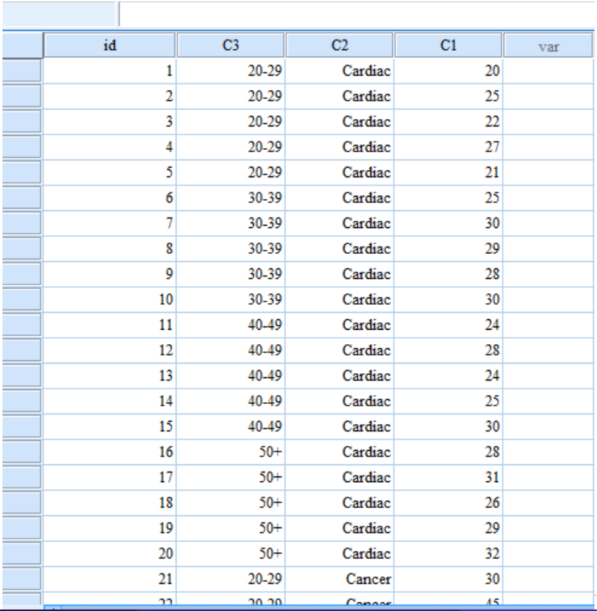
- In a study of length of time spent on individual home visits by public health nurses, data were reported on length of home visit, in minutes, by a sample of 80 nurses. A record was made also of each nurse's age and the type of illness of each patient visited.
- The researchers wished to obtain from their investigation answers to the following questions:
  1. Does the mean length of home visit differ among different age groups of nurses?
  2. Does the type of patient affect the mean length of home visit?
  3. Is there interaction between nurse's age and type of patient?

Length of Home Visit in Minutes by Public Health Nurses by Nurse's Age Group and Type of Patient



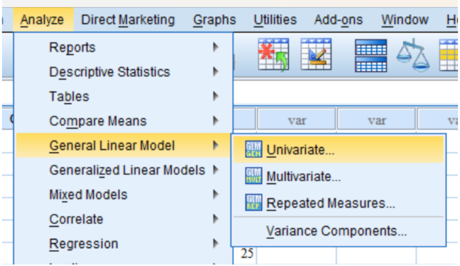
Factor *A* (Type of Patient) Levels	Factor *B* (Nurse's Age Group) Levels			
	1 (20 to 29)	2 (30 to 39)	3 (40 to 49)	4 (50 and Over)
1 (Cardiac)	20	25	24	28
	25	30	28	31
	22	29	24	26
	27	28	25	29
	21	30	30	32
2 (Cancer)	30	30	39	40
	45	29	42	45
	30	31	36	50
	35	30	42	45
	36	30	40	60
3 (C.V.A.)	31	32	41	42
	30	35	45	50
	40	30	40	40
	35	40	40	55
	30	30	35	45
4 (Tuberculosis)	20	23	24	29
	21	25	25	30
	20	28	30	28
	20	30	26	27
	19	31	23	30

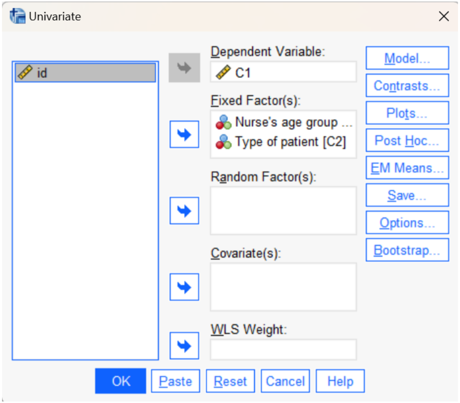
Data in SPSS



id	C3	C2	C1	var
1	20-29	Cardiac	20	
2	20-29	Cardiac	25	
3	20-29	Cardiac	22	
4	20-29	Cardiac	27	
5	20-29	Cardiac	21	
6	30-39	Cardiac	25	
7	30-39	Cardiac	30	
8	30-39	Cardiac	29	
9	30-39	Cardiac	28	
10	30-39	Cardiac	30	
11	40-49	Cardiac	24	
12	40-49	Cardiac	28	
13	40-49	Cardiac	24	
14	40-49	Cardiac	25	
15	40-49	Cardiac	30	
16	50+	Cardiac	28	
17	50+	Cardiac	31	
18	50+	Cardiac	26	
19	50+	Cardiac	29	
20	50+	Cardiac	32	
21	20-29	Cancer	30	
22	20-29	Cancer	45	

The procedure in SPSS





Dependent Variable: C1

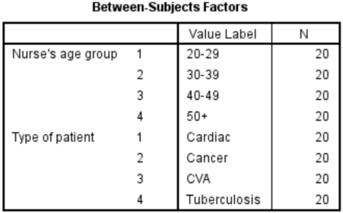
Fixed Factor(s): Nurse's age group ...; Type of patient [C2]

Random Factor(s):

Covariate(s):

WLS Weight:

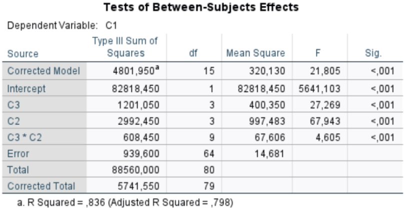
SPSS Output: Tests of Between-Subjects Effects



Between-Subjects Factors

		Value Label	N
Nurse's age group	1	20-29	20
	2	30-39	20
	3	40-49	20
	4	50+	20
Type of patient	1	Cardiac	20
	2	Cancer	20
	3	CVA	20
	4	Tuberculosis	20

- We consider here the case where the number of observations in each cell is the same. When the number of observations per cell is not the same for every cell, the analysis becomes more complex. In such cases, the design is said to be unbalanced. Software packages such as SPSS accommodates unequal cell sizes.



Tests of Between-Subjects Effects

Dependent Variable: C1

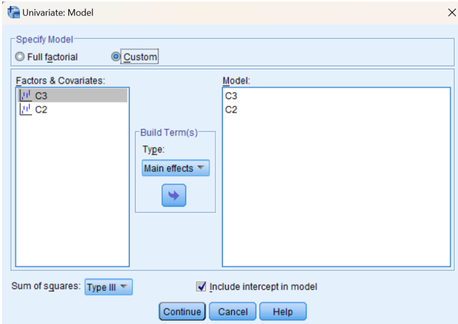
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4801,950[a]	15	320,130	21,805	<,001
Intercept	82818,450	1	82818,450	5641,103	<,001
C3	1201,050	3	400,350	27,269	<,001
C2	2992,450	3	997,483	67,943	<,001
C3 * C2	608,450	9	67,606	4,605	<,001
Error	939,600	64	14,681		
Total	88560,000	80			
Corrected Total	5741,550	79			

a. R Squared = ,836 (Adjusted R Squared = ,798)

- $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$: Variance ratio is $997.5/14.7 = 67.94 \rightarrow H_0$ is rejected (differences in the average amount of time spent in home visits with different types of patients).
- $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$: Variance ratio is $400.4/14.7 = 27.27 \rightarrow$ differences in the average amount of time spent on home visits among the different nurses when grouped by age.
- H_0 : all $(\alpha\beta)_{ij} = 0$: Variance ratio is $67.6/14.7 = 4.61 \rightarrow$ different combinations of levels of the two factors produce different effects.
- If the interaction term turns out to be not significant in the model – or if the effect size $\eta^2 < 0.14$) – it might be preferable to adjust your model, removing the interaction term and leaving only main effects.

SPSS Output: Main effects (to remove interaction effect, if so desired)

- This option is available from the Model Tab in the main interface, in case the interaction term turns out to be not significant and respecification of the model is desired, leaving only main effects of the Factors & Covariates:



Univariate: Model

Specify Model

○ Full factorial ● Custom

Factors & Covariates:

C3
C2

Model:

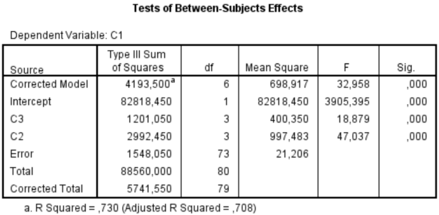
C3
C2

Build Term(s)

Type: Main effects

Sum of squares: Type III

☑ Include intercept in model



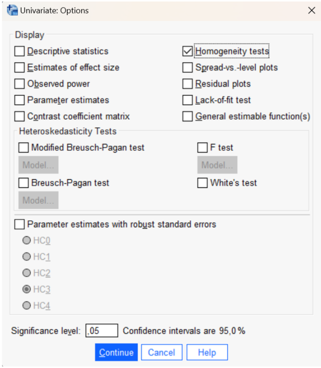
Tests of Between-Subjects Effects

Dependent Variable: C1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4193,500[a]	6	698,917	32,958	,000
Intercept	82818,450	1	82818,450	3905,395	,000
C3	1201,050	3	400,350	18,879	,000
C2	2992,450	3	997,483	47,037	,000
Error	1548,050	73	21,206		
Total	88560,000	80			
Corrected Total	5741,550	79			

a. R Squared = ,730 (Adjusted R Squared = ,708)

Levene's test of equality of error variances



Univariate: Options

Display

- ☐ Descriptive statistics
- ☐ Estimates of effect size
- ☐ Observed power
- ☐ Parameter estimates
- ☐ Contrast coefficient matrix
- ☑ Homogeneity tests
- ☐ Spread-vs.-level plots
- ☐ Residual plots
- ☐ Lack-of-fit test
- ☐ General estimable function(s)

Heteroskedasticity Tests

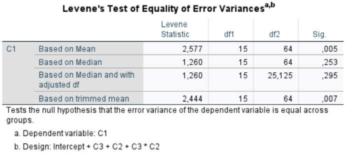
- ☐ Modified Breusch-Pagan test
- ☐ F test
- ☐ Breusch-Pagan test
- ☐ White's test

☐ Parameter estimates with robust standard errors

- ○ HC0
- ○ HC1
- ○ HC2
- ● HC3
- ○ HC4

Significance level: ,05 Confidence intervals are 95,0 %

- The previous output should only be interpreted under the assumption of homogeneity of error variances across cells.
- To verify whether this assumption is met or not, Levene's test should be considered.



Levene's Test of Equality of Error Variances[a,b]

		Levene Statistic	df1	df2	Sig.
C1	Based on Mean	2,577	15	64	,005
	Based on Median	1,260	15	64	,253
	Based on Median and with adjusted df	1,260	15	25,125	,295
	Based on trimmed mean	2,444	15	64	,007

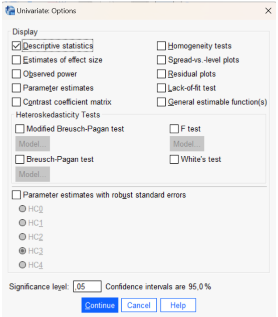
Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Dependent variable: C1

b. Design: Intercept + C3 + C2 + C3 * C2

- Levene's test can be accessed, using the Options Tab in the main interface and flagging "Homogeneity Tests".
- The null hypothesis of this test involves equality of error variances, hence a Sig. value greater than $\alpha = 0.05$ is desired, so that the H_0 cannot be rejected.
- Its classical version is the one 'based on mean', the validity of which can be affected by the presence of outliers/non-normality.
- Three modifications of Levene's test are also provided which are more robust, hence preferable, in such instances.
- Here, the classical test is significant (p=0.005), while the more robust modifications are not ($p > 0.05$), hence indicative that the assumption of homogeneity of variances is met.

SPSS Output: Descriptive Statistics



Univariate: Options

Display

☑ Descriptive statistics ☐ Homogeneity tests

☐ Estimates of effect size ☐ Spread-vs.-level plots

☐ Observed power ☐ Residual plots

☐ Parameter estimates ☐ Lack-of-fit test

☐ Contrast coefficient matrix ☐ General estimable function(s)

Heteroskedasticity Tests

☐ Modified Breusch-Pagan test ☐ F test

Model... Model...

☐ Breusch-Pagan test ☐ White's test

Model...

☐ Parameter estimates with robust standard errors

○ HC0

○ HC1

○ HC2

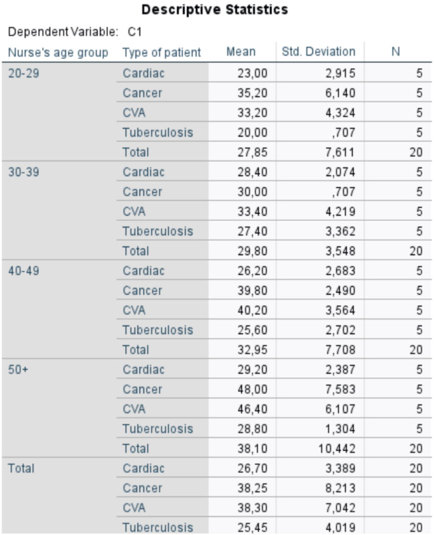
● HC3

○ HC4

Significance level: ,05 Confidence intervals are 95,0 %

Continue Cancel Help

- When the Anova table in significant, it is desirable to report differing means.
- Using the Options Tab in the main interface and flagging "Descripitve Statistics", we have immediate access to cell means and standard deviations, along with respective cell sizes.
- More detailed means information (including marginal means and the respective Confidence Intervals) are obtainable via the EM Means Tab in the main interface.

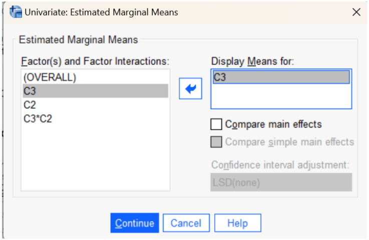


Descriptive Statistics

Dependent Variable: C1

Nurse's age group	Type of patient	Mean	Std. Deviation	N
20-29	Cardiac	23,00	2,915	5
	Cancer	35,20	6,140	5
	CVA	33,20	4,324	5
	Tuberculosis	20,00	,707	5
	Total	27,85	7,611	20
30-39	Cardiac	28,40	2,074	5
	Cancer	30,00	,707	5
	CVA	33,40	4,219	5
	Tuberculosis	27,40	3,362	5
	Total	29,80	3,548	20
40-49	Cardiac	26,20	2,683	5
	Cancer	39,80	2,490	5
	CVA	40,20	3,564	5
	Tuberculosis	25,60	2,702	5
	Total	32,95	7,708	20
50+	Cardiac	29,20	2,387	5
	Cancer	48,00	7,583	5
	CVA	46,40	6,107	5
	Tuberculosis	28,80	1,304	5
	Total	38,10	10,442	20
Total	Cardiac	26,70	3,389	20
	Cancer	38,25	8,213	20
	CVA	38,30	7,042	20
	Tuberculosis	25,45	4,019	20

Estimated marginal means



Univariate: Estimated Marginal Means

Estimated Marginal Means

Factor(s) and Factor Interactions:
(OVERALL)
C3
C2
C3*C2

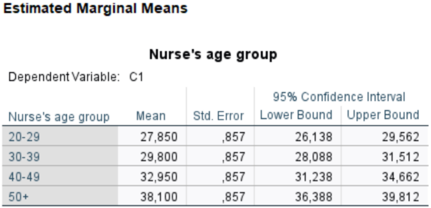
Display Means for:
C3

□ Compare main effects

□ Compare simple main effects

Confidence interval adjustment:
LSD(none)

- Marginal Means for various factors are obtainable via the EM Means (Estimated Marginal Means) Tab in the main interface.
- Here, we consider Marginal Means with respect to the groups defined by Variable C3 (Nurse Age Group).



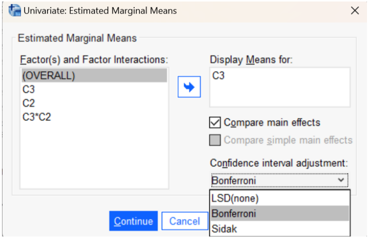
Estimated Marginal Means

Nurse's age group

Dependent Variable: C1

			95% Confidence Interval	
Nurse's age group	Mean	Std. Error	Lower Bound	Upper Bound
20-29	27,850	,857	26,138	29,562
30-39	29,800	,857	28,088	31,512
40-49	32,950	,857	31,238	34,662
50+	38,100	,857	36,388	39,812

Estimated marginal means (compare main effects)



Univariate: Estimated Marginal Means

Estimated Marginal Means

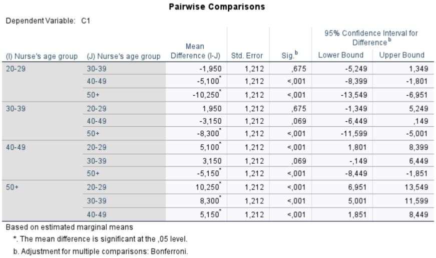
Factor(s) and Factor Interactions:
(OVERALL)
C3
C2
C3*C2

Display Means for:
C3

☑ Compare main effects
□ Compare simple main effects

Confidence interval adjustment:
Bonferroni
LSD(none)
Bonferroni
Sidak

- To obtain pairwise comparisons among the mean times across the different groups defined by by Variable C3 (Nurse Age Group), flag "compare mean effects".
- Here, we consider Marginal Means with respect to the groups defined by Variable C3 (Nurse Age Group).
- Since multiple tests will be conducted, it is advisable to opt for a more conservative testing approach using the Bonferroni modification, which adjusts the significance level α to α/r, where r is the number of pairwise tests being carried out.



Pairwise Comparisons

Dependent Variable: C1

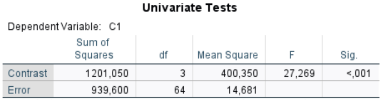
					95% Confidence Interval for Difference[b]	
(I) Nurse's age group	(J) Nurse's age group	Mean Difference (I-J)	Std. Error	Sig.[b]	Lower Bound	Upper Bound
20-29	30-39	-1,950	1,212	,675	-5,249	1,349
	40-49	-5,100*	1,212	<,001	-8,399	-1,801
	50+	-10,250*	1,212	<,001	-13,549	-6,951
30-39	20-29	1,950	1,212	,675	-1,349	5,249
	40-49	-3,150	1,212	,069	-6,449	,149
	50+	-8,300*	1,212	<,001	-11,599	-5,001
40-49	20-29	5,100*	1,212	<,001	1,801	8,399
	30-39	3,150	1,212	,069	-,149	6,449
	50+	-5,150*	1,212	<,001	-8,449	-1,851
50+	20-29	10,250*	1,212	<,001	6,951	13,549
	30-39	8,300*	1,212	<,001	5,001	11,599
	40-49	5,150*	1,212	<,001	1,851	8,449

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

- Significant differences are flagged.
- Note that for significant differences (i.e., with p-values (Sig.)<0.05), the difference 0 lies within the corresponding 95%-C.I. (confidence interval).
- The last portion of the output contains an ANOVA table, replicating the relevant portion from the initial "Tests of Between-Subjects Effects".

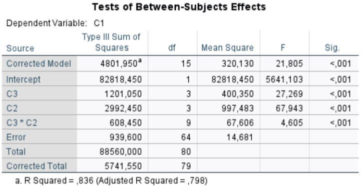


Univariate Tests

Dependent Variable: C1

	Sum of Squares	df	Mean Square	F	Sig.
Contrast	1201,050	3	400,350	27,269	<,001
Error	939,600	64	14,681		

Estimated marginal means (compare main effects) - cont'd



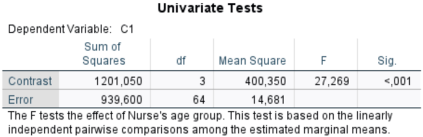
Tests of Between-Subjects Effects

Dependent Variable: C1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4801,950[a]	15	320,130	21,805	<,001
Intercept	82818,450	1	82818,450	5641,103	<,001
C3	1201,050	3	400,350	27,269	<,001
C2	2992,450	3	997,483	67,943	<,001
C3 * C2	608,450	9	67,606	4,605	<,001
Error	939,600	64	14,681		
Total	88560,000	80			
Corrected Total	5741,550	79			

a. R Squared = ,836 (Adjusted R Squared = ,798)

- The last part of the output includes an ANOVA table, replicating the relevant portion from the initial "Tests of Between-Subjects Effects".



Univariate Tests

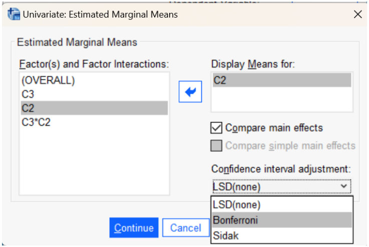
Dependent Variable: C1

	Sum of Squares	df	Mean Square	F	Sig.
Contrast	1201,050	3	400,350	27,269	<,001
Error	939,600	64	14,681		

The F tests the effect of Nurse's age group. This test is based on the linearly independent pairwise comparisons among the estimated marginal means.

Marginal means for Type of Patient

- The analogue procedure to obtain Marginal Means and compare main effects with respect to the groups defined by Variable C2 (Type of Patient).
- Again, since multiple tests will be conducted, it is advisable to opt for a more conservative testing approach using the Bonferroni modification.



Univariate: Estimated Marginal Means

Estimated Marginal Means

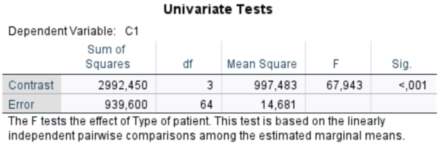
Factor(s) and Factor Interactions:
(OVERALL)
C3
C2
C3*C2

Display Means for:
C2

☑ Compare main effects
☐ Compare simple main effects

Confidence interval adjustment:
LSD(none)

LSD(none)
Bonferroni
Sidak

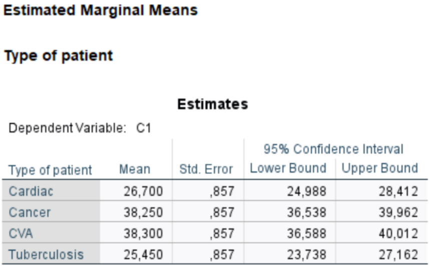


Univariate Tests

Dependent Variable: C1

	Sum of Squares	df	Mean Square	F	Sig.
Contrast	2992,450	3	997,483	67,943	<,001
Error	939,600	64	14,681		

The F tests the effect of Type of patient. This test is based on the linearly independent pairwise comparisons among the estimated marginal means.



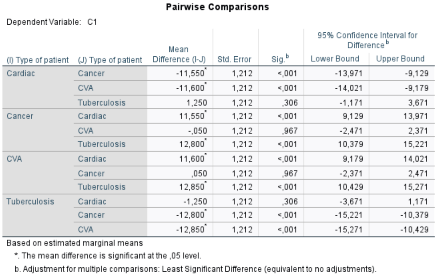
Estimated Marginal Means

Type of patient

Estimates

Dependent Variable: C1

Type of patient	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Cardiac	26,700	,857	24,988	28,412
Cancer	38,250	,857	36,538	39,962
CVA	38,300	,857	36,588	40,012
Tuberculosis	25,450	,857	23,738	27,162



Pairwise Comparisons

Dependent Variable: C1

(I) Type of patient	(J) Type of patient	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
Cardiac	Cancer	-11,550*	1,212	<,001	-13,971	-9,129
	CVA	-11,600*	1,212	<,001	-14,021	-9,179
	Tuberculosis	1,250	1,212	,306	-1,171	3,671
Cancer	Cardiac	11,550*	1,212	<,001	9,129	13,971
	CVA	-,050	1,212	,967	-2,471	2,371
	Tuberculosis	12,800*	1,212	<,001	10,379	15,221
CVA	Cardiac	11,600*	1,212	<,001	9,179	14,021
	Cancer	,050	1,212	,967	-2,371	2,471
	Tuberculosis	12,850*	1,212	<,001	10,429	15,271
Tuberculosis	Cardiac	-1,250	1,212	,306	-3,671	1,171
	Cancer	-12,800*	1,212	<,001	-15,221	-10,379
	CVA	-12,850*	1,212	<,001	-15,271	-10,429

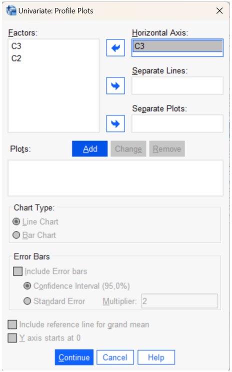
Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Graphing options (marginal means for nurse age group)

- From the "Profile Plots" Tab in the main interface, we may visualize the corresponding marginal means.
- To graph marginal mean times for the different nurse age groups, select the corresponding variable (variable C3) and place it the Horizontal Axis tab.


Univariate: Profile Plots

Factors: C3, C2

Horizontal Axis: C3

Separate Lines:

Separate Plots:

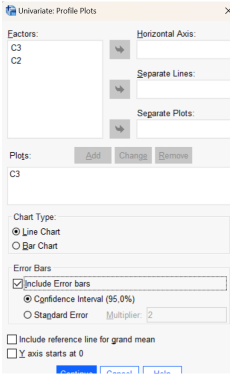
Plots:

Chart Type: ● Line Chart ○ Bar Chart

Error Bars: ☐ Include Error bars; ● Confidence Interval (95,0%) ○ Standard Error Multiplier: 2

☐ Include reference line for grand mean

☐ Y axis starts at 0


Univariate: Profile Plots

Factors: C3, C2

Horizontal Axis:

Separate Lines:

Separate Plots:

Plots: C3

Chart Type: ● Line Chart ○ Bar Chart

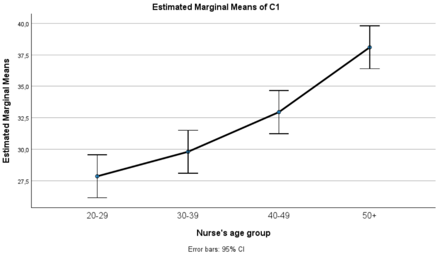
Error Bars: ☑ Include Error bars; ● Confidence Interval (95,0%) ○ Standard Error Multiplier: 2

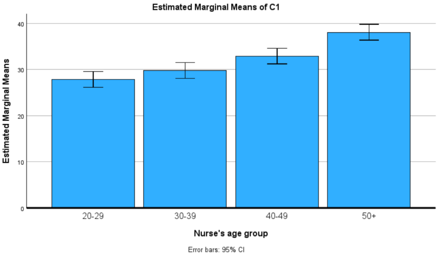
☐ Include reference line for grand mean

☐ Y axis starts at 0

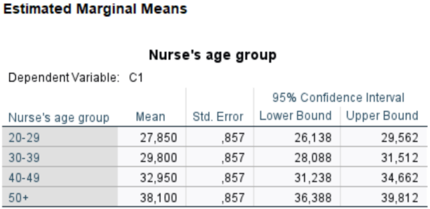
- Line Chart and Bar Chart options are provided (prefer Line chart myself..).
- Also, option to include Error Bars in the resulting chart are provided (although that may clutter the resulting graph).

Charts for marginal means for nurse age group





- Line Chart
- Bar Chart



Estimated Marginal Means

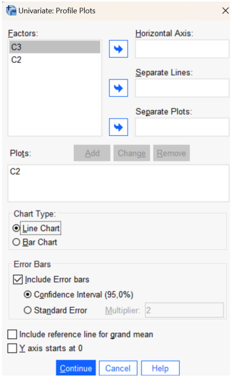
Nurse's age group

Dependent Variable: C1

Nurse's age group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
20-29	27,850	,857	26,138	29,562
30-39	29,800	,857	28,088	31,512
40-49	32,950	,857	31,238	34,662
50+	38,100	,857	36,388	39,812

Graphing options (marginal means for type of patient)

- To graph marginal mean times for type of patient groups, select the corresponding variable (variable C2) and place it the Horizontal Axis tab.



Univariate: Profile Plots

Factors:
C3
C2

Horizontal Axis:

Separate Lines:

Separate Plots:

Plots: Add Change Remove

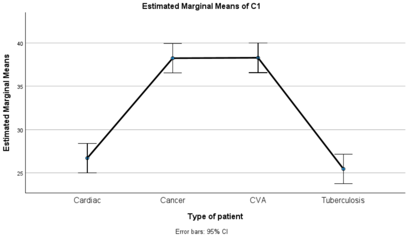
C2

Chart Type:
● Line Chart
○ Bar Chart

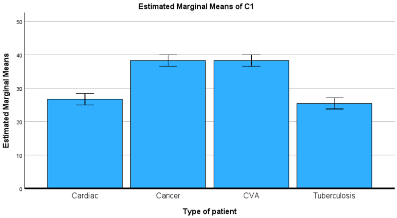
Error Bars
☑ Include Error bars
● Confidence Interval (95,0%)
○ Standard Error Multiplier: 2

□ Include reference line for grand mean
□ Y axis starts at 0

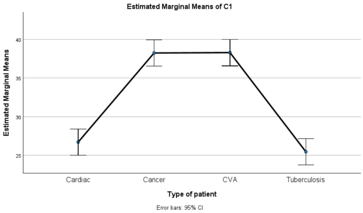
Continue Cancel Help

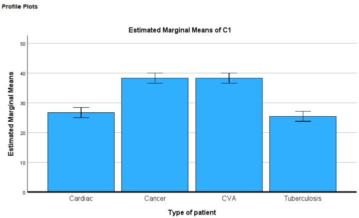


Profile Plots

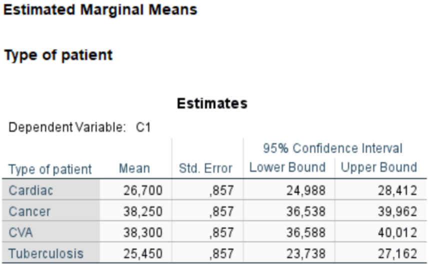


Charts for marginal means for type of patient groups





- Line Chart
- Bar Chart



Estimated Marginal Means

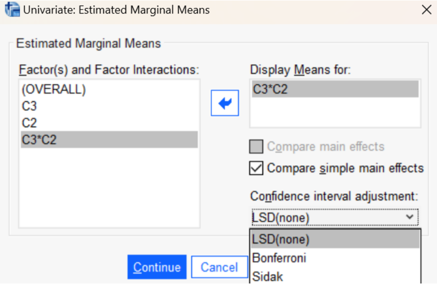
Type of patient

Estimates

Dependent Variable: C1

Type of patient	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Cardiac	26,700	,857	24,988	28,412
Cancer	38,250	,857	36,538	39,962
CVA	38,300	,857	36,588	40,012
Tuberculosis	25,450	,857	23,738	27,162

Means for each of the 16 cells of the experiment



Univariate: Estimated Marginal Means

Estimated Marginal Means

Factor(s) and Factor Interactions:

(OVERALL)
C3
C2
C3*C2

Display Means for:

C3*C2

□ Compare main effects

☑ Compare simple main effects

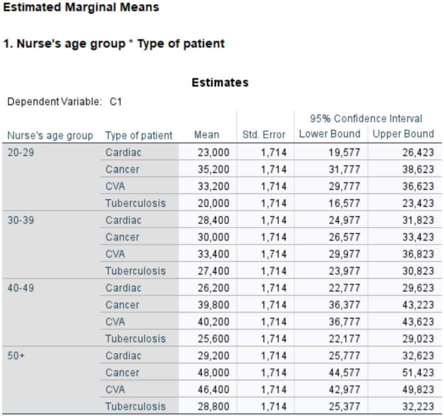
Confidence interval adjustment:

LSD(none)

LSD(none)
Bonferroni
Sidak

- Flagging "Simple effects tests" in the EM Means Tab, we investigate pairwise differences in mean length of home visit among different nurse age groups for different types of patients, i.e., among the 16 cells.
- Since multiple tests will be conducted, it is advisable to opt for a more conservative testing approach using the Bonferroni modification, which adjusts the significance level α to α/r, where r is the number of pairwise tests being carried out.

Means for each of the 16 cells of the experiment



Estimated Marginal Means

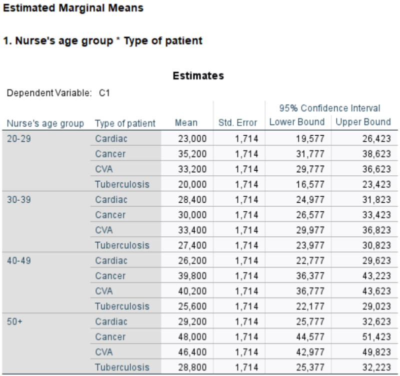
1. Nurse's age group * Type of patient

Estimates

Dependent Variable: C1

Nurse's age group	Type of patient	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
20-29	Cardiac	23,000	1,714	19,577	26,423
	Cancer	35,200	1,714	31,777	38,623
	CVA	33,200	1,714	29,777	36,623
	Tuberculosis	20,000	1,714	16,577	23,423
30-39	Cardiac	28,400	1,714	24,977	31,823
	Cancer	30,000	1,714	26,577	33,423
	CVA	33,400	1,714	29,977	36,823
	Tuberculosis	27,400	1,714	23,977	30,823
40-49	Cardiac	26,200	1,714	22,777	29,623
	Cancer	39,800	1,714	36,377	43,223
	CVA	40,200	1,714	36,777	43,623
	Tuberculosis	25,600	1,714	22,177	29,023
50+	Cardiac	29,200	1,714	25,777	32,623
	Cancer	48,000	1,714	44,577	51,423
	CVA	46,400	1,714	42,977	49,823
	Tuberculosis	28,800	1,714	25,377	32,223

Means for each of the 16 cells of the experiment - Simple main effects comparisons (1st Version - partial output)



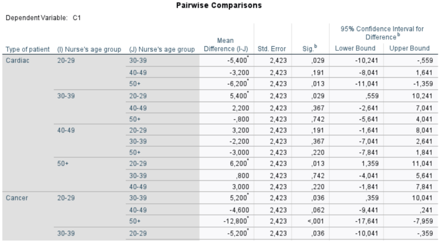
Estimated Marginal Means

1. Nurse's age group * Type of patient

Estimates

Dependent Variable: C1

Nurse's age group	Type of patient	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
20-29	Cardiac	23,000	1,714	19,577	26,423
	Cancer	35,200	1,714	31,777	38,623
	CVA	33,200	1,714	29,777	36,623
	Tuberculosis	20,000	1,714	16,577	23,423
30-39	Cardiac	28,400	1,714	24,977	31,823
	Cancer	30,000	1,714	26,577	33,423
	CVA	33,400	1,714	29,977	36,823
	Tuberculosis	27,400	1,714	23,977	30,823
40-49	Cardiac	26,200	1,714	22,777	29,623
	Cancer	39,800	1,714	36,377	43,223
	CVA	40,200	1,714	36,777	43,623
	Tuberculosis	25,600	1,714	22,177	29,023
50+	Cardiac	29,200	1,714	25,777	32,623
	Cancer	48,000	1,714	44,577	51,423
	CVA	46,400	1,714	42,977	49,823
	Tuberculosis	28,800	1,714	25,377	32,223



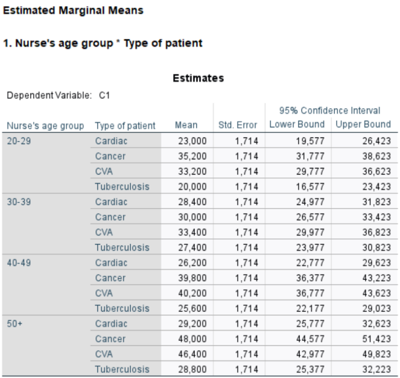
Pairwise Comparisons

Dependent Variable: C1

Type of patient	(I) Nurse's age group	(J) Nurse's age group	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
						Lower Bound	Upper Bound
Cardiac	20-29	30-39	-5,400*	2,423	,029	-10,241	-,559
		40-49	-3,200	2,423	,191	-8,041	1,641
		50+	-6,200*	2,423	,013	-11,041	-1,359
	30-39	20-29	5,400*	2,423	,029	,559	10,241
		40-49	2,200	2,423	,367	-2,641	7,041
		50+	-,800	2,423	,742	-5,641	4,041
	40-49	20-29	3,200	2,423	,191	-1,641	8,041
		30-39	-2,200	2,423	,367	-7,041	2,641
		50+	-3,000	2,423	,220	-7,841	1,841
	50+	20-29	6,200*	2,423	,013	1,359	11,041
		30-39	,800	2,423	,742	-4,041	5,641
		40-49	3,000	2,423	,220	-1,841	7,841
Cancer	20-29	30-39	5,200*	2,423	,036	,359	10,041
		40-49	-4,600	2,423	,062	-9,441	,241
		50+	-12,800*	2,423	<,001	-17,641	-7,959
	30-39	20-29	-5,200*	2,423	,036	-10,041	-,359

- Like previously, significant differences are flagged.

Means for each of the 16 cells of the experiment - Simple main effects comparisons (2nd Version - partial output)



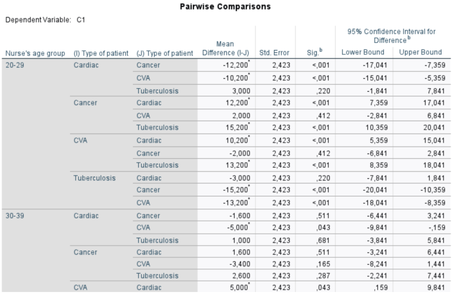
Estimated Marginal Means

1. Nurse's age group * Type of patient

Estimates

Dependent Variable: C1

Nurse's age group	Type of patient	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
20-29	Cardiac	23,000	1,714	19,577	26,423
	Cancer	35,200	1,714	31,777	38,623
	CVA	33,200	1,714	29,777	36,623
	Tuberculosis	20,000	1,714	16,577	23,423
30-39	Cardiac	28,400	1,714	24,977	31,823
	Cancer	30,000	1,714	26,577	33,423
	CVA	33,400	1,714	29,977	36,823
	Tuberculosis	27,400	1,714	23,977	30,823
40-49	Cardiac	26,200	1,714	22,777	29,623
	Cancer	39,800	1,714	36,377	43,223
	CVA	40,200	1,714	36,777	43,623
	Tuberculosis	25,600	1,714	22,177	29,023
50+	Cardiac	29,200	1,714	25,777	32,623
	Cancer	48,000	1,714	44,577	51,423
	CVA	46,400	1,714	42,977	49,823
	Tuberculosis	28,800	1,714	25,377	32,223



Pairwise Comparisons

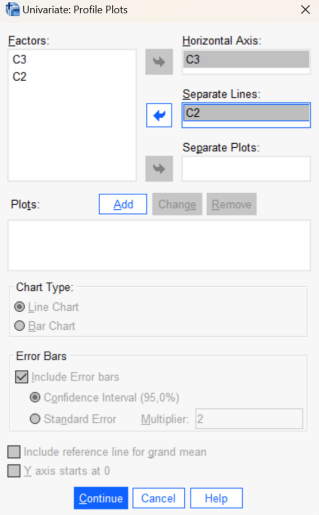
Dependent Variable: C1

Nurse's age group	(I) Type of patient	(J) Type of patient	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
						Lower Bound	Upper Bound
20-29	Cardiac	Cancer	-12,200*	2,423	<,001	-17,041	-7,359
		CVA	-10,200*	2,423	<,001	-15,041	-5,359
		Tuberculosis	3,000	2,423	,220	-1,841	7,841
	Cancer	Cardiac	12,200*	2,423	<,001	7,359	17,041
		CVA	2,000	2,423	,412	-2,841	6,841
		Tuberculosis	15,200*	2,423	<,001	10,359	20,041
	CVA	Cardiac	10,200*	2,423	<,001	5,359	15,041
		Cancer	-2,000	2,423	,412	-6,841	2,841
		Tuberculosis	13,200*	2,423	<,001	8,359	18,041
	Tuberculosis	Cardiac	-3,000	2,423	,220	-7,841	1,841
		Cancer	-15,200*	2,423	<,001	-20,041	-10,359
		CVA	-13,200*	2,423	<,001	-18,041	-8,359
30-39	Cardiac	Cancer	-1,600	2,423	,511	-6,441	3,241
		CVA	-5,000*	2,423	,043	-9,841	-,159
		Tuberculosis	1,000	2,423	,681	-3,841	5,841
	Cancer	Cardiac	1,600	2,423	,511	-3,241	6,441
		CVA	-3,400	2,423	,165	-8,241	1,441
		Tuberculosis	2,600	2,423	,287	-2,241	7,441
	CVA	Cardiac	5,000*	2,423	,043	,159	9,841

- Like previously, significant differences are flagged.

Means plots (1st Version)

- To construct means plot with respect to Nurse's age group:



Univariate: Profile Plots

Factors:
C3
C2

Horizontal Axis: C3

Separate Lines: C2

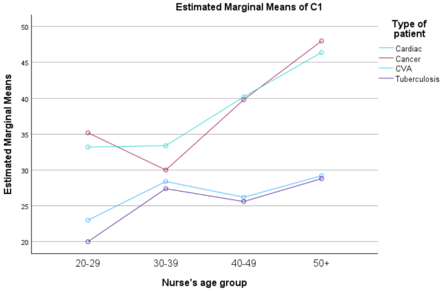
Separate Plots:

Plots: Add Change Remove

Chart Type:
● Line Chart
○ Bar Chart

Error Bars
☑ Include Error bars
● Confidence Interval (95,0%)
○ Standard Error Multiplier: 2

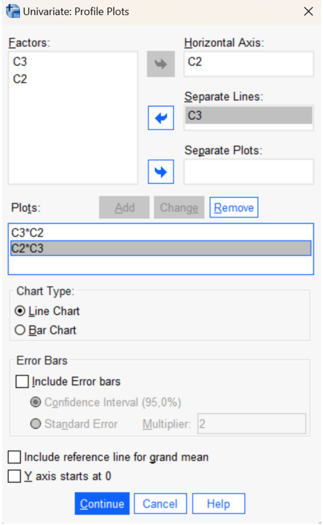
□ Include reference line for grand mean
□ Y axis starts at 0



- When no interaction is present, we would expect the line connecting the means for different nurse age groups to be roughly parallel across levels of the type of patient factor.

Means plots (2nd Version)

- To construct means plot with respect to Type of Patient group:



Univariate: Profile Plots

Factors: C3, C2

Horizontal Axis: C2

Separate Lines: C3

Separate Plots:

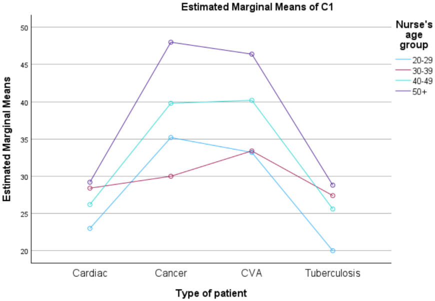
Plots: C3*C2, C2*C3

Chart Type: ● Line Chart ○ Bar Chart

Error Bars: □ Include Error bars ● Confidence Interval (95,0%) ○ Standard Error Multiplier: 2

□ Include reference line for grand mean

□ Y axis starts at 0



- When no interaction is present, we would expect the line connecting the means for different types of patient groups to be roughly parallel across levels of nurse age.

Repeated measures design

The repeated measures design

- A repeated measures design is one in which measurements of the same variable are made on each subject on two or more different occasions.
- The different occasions during which measurements are taken may be either points in time or different conditions such as different treatments
- Motivation. Desire to control for variability among subjects. In such a design, each subject serves as its own control.
- When measurements are taken on only two occasions, we have the paired means comparison design.
- Most frequent use. Situation in which the investigator is concerned with responses over time.
- Advantages.
  - Ability to control for extraneous variation among subjects. Since the variability in the error term due to individual differences is removed (as we are “blocking on each subject”), this generally makes these designs more powerful than randomized designs, where subjects are randomly assigned to the different treatments.
  - Also, fewer subjects are needed than for a design in which different subjects are used for each occasion on which measurements are made. Suppose, for example, that we have four treatments (in the usual sense) or four points in time on each of which we would like to have 10 measurements $\rightarrow$ 40 subjects vs. 10 subjects required in repeated measures.

  This can be a very attractive advantage if subjects are scarce or expensive to recruit.

The repeated measures design: Disadvantages

- A major potential problem to be on the alert for is what is known as the carry-over effect. When two or more treatments are being evaluated, the investigator should make sure that a subject's response to one treatment does not reflect a residual effect from previous treatments.
- This problem can frequently be solved by allowing a sufficient length of time between treatments.
- Another possible problem is the position effect. A subject's response to a treatment experienced last in a sequence may be different from the response that would have occurred if the treatment had been first in the sequence.
- In certain studies, such as those involving physical participation on the part of the subjects, enthusiasm that is high at the beginning of the study may give way to boredom toward the end.
- A way around this problem is to randomize the sequence of treatments independently for each subject. Otherwise, time and the order of administration of stimuli will be confounded.

Single-Factor Repeated Measures Design

- The repeated measures design in which one factor (additionally to the already present treatment variable) is introduced into the experiment is called a single-factor repeated measures design. The reason for introducing this additional variable is to measure and isolate its contribution to the total variability among the observations.
- We refer to the additional factor as subjects (“blocking on each subject”). In the single-factor repeated measures design, each subject receives each of the treatments. The order in which the subjects are exposed to the treatments, when possible, is random, and the randomization is carried out independently for each subject.

Single-Factor Repeated Measures Design: Assumptions

- Assumptions.
  1. The subjects under study constitute a simple random sample from a population of similar subjects.
  2. Each observation is an independent simple random sample of size 1 from each of kn populations, where n is the number of subjects and k is the number of treatments to which each subject is exposed.
  3. The kn populations have potentially different means, but they all have the same variance.
  4. The k treatments are fixed; that is, they are the only treatments about which we have an interest in the current situation. We do not wish to make inferences to some larger collection of treatments.
  5. There is no interaction between treatments and subjects; that is, the treatment and subject effects are additive.

Single-Factor Repeated Measures Design: Assumptions (2)

- Additionally, in a repeated measures experiment there is a presumption that correlations should exist among the repeated measures. That is, measurements at time 1 and 2 are likely correlated, as are measurements at time 1 and 3, 2 and 3, and so on. This is expected because the measurements are taken on the same individuals through time.
- An underlying assumption of the repeated-measures ANOVA design is that all of these correlations are the same, a condition referred to as compound symmetry. This assumption, coupled with assumption 3 concerning equal variances, is referred to as sphericity. Violations of the sphericity assumption can result in an inflated type I error.
- Most computer programs provide a formal test for the sphericity assumption along with alternative estimation methods if the sphericity assumption is violated.

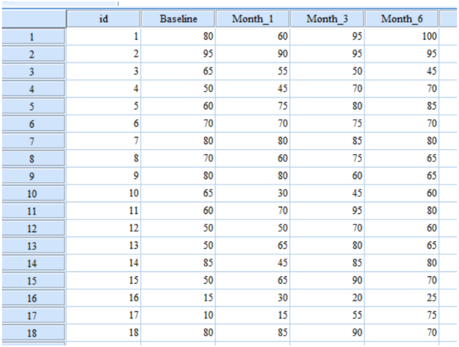
Single-Factor Repeated Measures Design

- The model for the fixed-effects additive single-factor repeated measures design may be written as follows:

$$x_{ij} = \mu + \beta_i + \tau_j + \epsilon_{ij} \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, k.$$

- This model is completely analogous to the model for the randomized complete block design. The subjects are the blocks.
- Consequently, the notation, data display, and hypothesis testing procedure are the same as for the randomized complete block design as presented earlier.

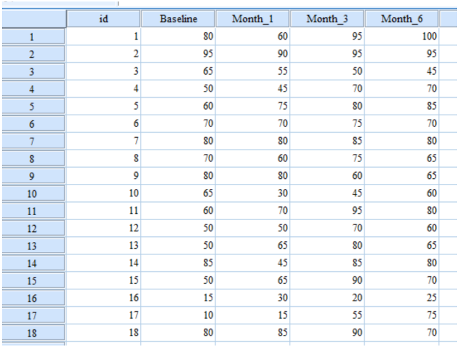
Health Condition Scores at Four Different Points in Time



	id	Baseline	Month_1	Month_3	Month_6
1	1	80	60	95	100
2	2	95	90	95	95
3	3	65	55	50	45
4	4	50	45	70	70
5	5	60	75	80	85
6	6	70	70	75	70
7	7	80	80	85	80
8	8	70	60	75	65
9	9	80	80	60	65
10	10	65	30	45	60
11	11	60	70	95	80
12	12	50	50	70	60
13	13	50	65	80	65
14	14	85	45	85	80
15	15	50	65	90	70
16	16	15	30	20	25
17	17	10	15	55	75
18	18	80	85	90	70

- Subjects with chronic, nonspecific low back pain.
- 18 of the subjects completed a survey questionnaire assessing physical functioning at baseline, and after 1, 3, and 6 months.
- Data for those subjects who received a sham treatment that appeared to be genuine osteopathic manipulation. Higher values indicate better physical functioning.
- The goal of the experiment was to determine if subjects would report improvement over time even though the treatment they received would provide minimal improvement.
- We wish to know if there is a difference in the mean survey values among the four points in time.

Health Condition Scores at Four Different Points in Time (2)

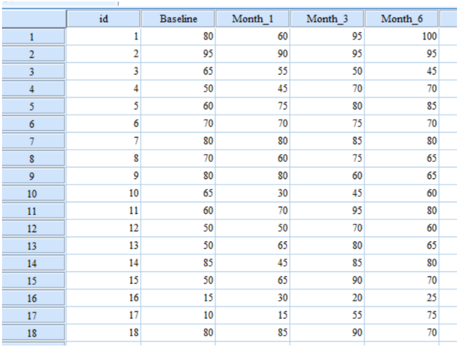


	id	Baseline	Month_1	Month_3	Month_6
1	1	80	60	95	100
2	2	95	90	95	95
3	3	65	55	50	45
4	4	50	45	70	70
5	5	60	75	80	85
6	6	70	70	75	70
7	7	80	80	85	80
8	8	70	60	75	65
9	9	80	80	60	65
10	10	65	30	45	60
11	11	60	70	95	80
12	12	50	50	70	60
13	13	50	65	80	65
14	14	85	45	85	80
15	15	50	65	90	70
16	16	15	30	20	25
17	17	10	15	55	75
18	18	80	85	90	70

- The goal of the experiment was to determine if subjects would report improvement over time even though the treatment they received would provide minimal improvement.
- We wish to know if there is a difference in the mean survey values among the four points in time.
- Hypotheses.

$$\begin{cases} H_{\mathbf{0}} : \mu_B = \mu_{M_{\mathbf{1}}} = \mu_{M_{\mathbf{3}}} = \mu_{M_{\mathbf{6}}}, \\ H_a : \text{ not all } \mu\text{'s are equal.} \end{cases}$$

Health Condition Scores at Four Different Points in Time (3)



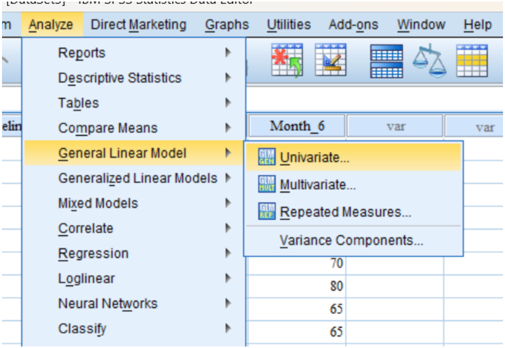
	id	Baseline	Month_1	Month_3	Month_6
1	1	80	60	95	100
2	2	95	90	95	95
3	3	65	55	50	45
4	4	50	45	70	70
5	5	60	75	80	85
6	6	70	70	75	70
7	7	80	80	85	80
8	8	70	60	75	65
9	9	80	80	60	65
10	10	65	30	45	60
11	11	60	70	95	80
12	12	50	50	70	60
13	13	50	65	80	65
14	14	85	45	85	80
15	15	50	65	90	70
16	16	15	30	20	25
17	17	10	15	55	75
18	18	80	85	90	70

- Hypotheses.

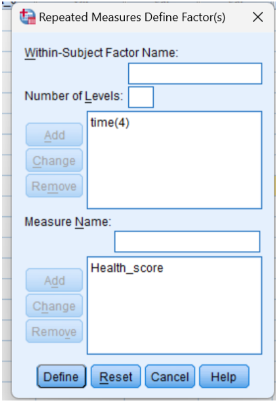
$$\begin{cases} H_0 : \mu_B = \mu_{M_1} = \mu_{M_3} = \mu_{M_6}, \\ H_a : \text{ not all } \mu\text{'s are equal.} \end{cases}$$

- Test statistic. Variance ratio = Treatment MS/Error MS $\sim F_{(4-1),(71-3-17)} = F_{3,51}$.

Single-Factor Repeated Measures in SPSS



Single-Factor Repeated Measures in SPSS (2)



Repeated Measures Define Factor(s)

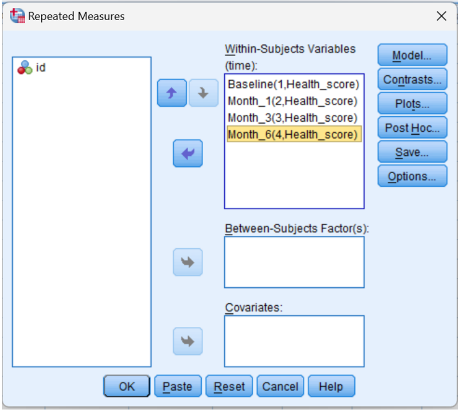
Within-Subject Factor Name:

Number of Levels:

time(4)

Measure Name:

Health_score



Repeated Measures

id

Within-Subjects Variables (time):

Baseline(1,Health_score)
Month_1(2,Health_score)
Month_3(3,Health_score)
Month_6(4,Health_score)

Between-Subjects Factor(s):

Covariates:

Single-Factor Repeated Measures in SPSS: Output

If no further options are activated, SPSS Output provides the following Tables:

(a.) Multivariate Tests

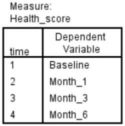
(b.) Mauchly's Test of Sphericity

(c.) Test of Within-Subjects Effects

(d.) Test of Within-Subjects Contrasts

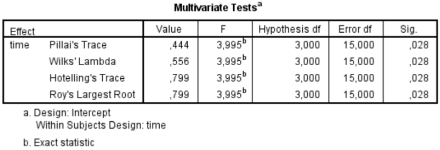
(e.) Test of Between-Subjects Effects

Multivariate Tests vs. Test of Within-Subjects Effects



Measure:
Health_score

time	Dependent Variable
1	Baseline
2	Month_1
3	Month_3
4	Month_6



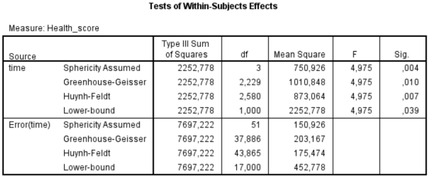
Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.
time	Pillai's Trace	,444	3,995[b]	3,000	15,000	,028
	Wilks' Lambda	,556	3,995[b]	3,000	15,000	,028
	Hotelling's Trace	,799	3,995[b]	3,000	15,000	,028
	Roy's Largest Root	,799	3,995[b]	3,000	15,000	,028

a. Design: Intercept
Within Subjects Design: time

b. Exact statistic

- The test of overall mean differences in the repeated measures design can be carried out in two ways:
  1. using either the multivariate test approach (see above), or
  2. the univariate approach (see below).

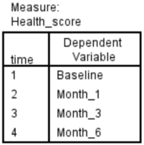


Tests of Within-Subjects Effects

Measure: Health_score

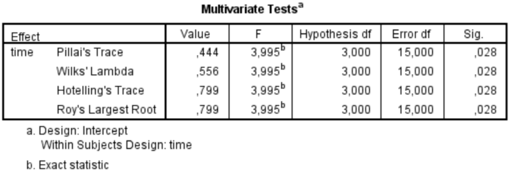
Source		Type III Sum of Squares	df	Mean Square	F	Sig.
time	Sphericity Assumed	2252,778	3	750,926	4,975	,004
	Greenhouse-Geisser	2252,778	2,229	1010,848	4,975	,010
	Huynh-Feldt	2252,778	2,580	873,064	4,975	,007
	Lower-bound	2252,778	1,000	2252,778	4,975	,039
Error(time)	Sphericity Assumed	7697,222	51	150,926		
	Greenhouse-Geisser	7697,222	37,886	203,167		
	Huynh-Feldt	7697,222	43,865	175,474		
	Lower-bound	7697,222	17,000	452,778		

Multivariate Tests



Measure:
Health_score

time	Dependent Variable
1	Baseline
2	Month_1
3	Month_3
4	Month_6



Multivariate Tests[a]

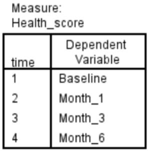
Effect		Value	F	Hypothesis df	Error df	Sig.
time	Pillai's Trace	,444	3,995[b]	3,000	15,000	,028
	Wilks' Lambda	,556	3,995[b]	3,000	15,000	,028
	Hotelling's Trace	,799	3,995[b]	3,000	15,000	,028
	Roy's Largest Root	,799	3,995[b]	3,000	15,000	,028

a. Design: Intercept
Within Subjects Design: time

b. Exact statistic

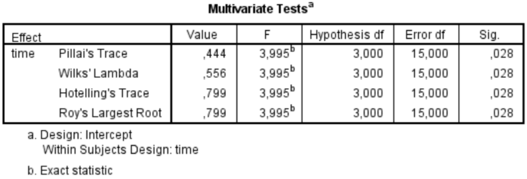
- Assumptions. The multivariate test assumes independence of observations and multivariate normality.
- A benefit of this approach is that it does not require one of the assumptions necessary for the univariate approach (via Test of Within-Subjects Effects Table); namely, sphericity.
- There are times when the multivariate test may be more powerful than the univariate test. However, when sphericity is assumed, the univariate approach tends to be more powerful than the multivariate test.

Multivariate Tests (2)



Measure:
Health_score

time	Dependent Variable
1	Baseline
2	Month_1
3	Month_3
4	Month_6



Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.
time	Pillai's Trace	,444	3,995[b]	3,000	15,000	,028
	Wilks' Lambda	,556	3,995[b]	3,000	15,000	,028
	Hotelling's Trace	,799	3,995[b]	3,000	15,000	,028
	Roy's Largest Root	,799	3,995[b]	3,000	15,000	,028

a. Design: Intercept
Within Subjects Design: time

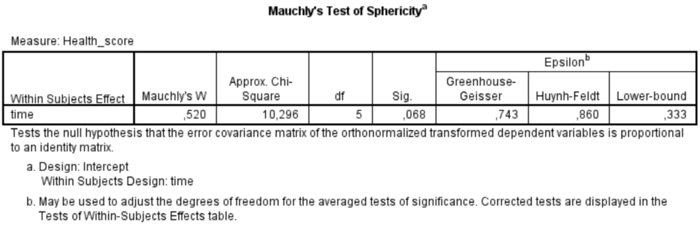
b. Exact statistic

- According to this Table, we have:

$$\text{Wilks' lambda} = 0.556, \quad F(3,15) = 3.995, \quad p = 0.028.$$

Hence, we conclude significant differences in means on the Health Score across time periods.

Univariate approach and Sphericity assumption



Mauchly's Test of Sphericity[a]

Measure: Health_score

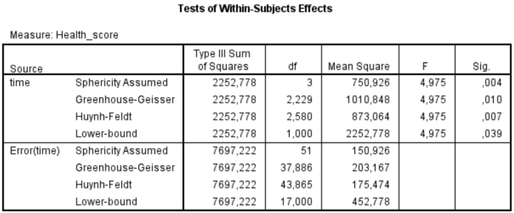
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
time	,520	10,296	5	,068	,743	,860	,333

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: time

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

- The standard univariate repeated measures ANOVA (Test of Within-Subjects Effects Table below) assumes a condition called sphericity.
- When sphericity is violated, there is increased risk of committing Type 1 error. To evaluate whether that condition is met, we consider the information contained in the table above.
- Problems with violating sphericity (or with compound symmetry for that matter) tend to arise when the time elapsed between measurement occasions are not equal.

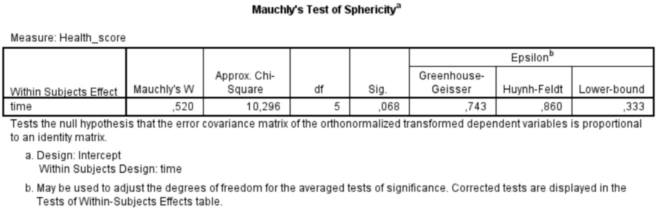


Tests of Within-Subjects Effects

Measure: Health_score

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
time	Sphericity Assumed	2252,778	3	750,926	4,975	,004
	Greenhouse-Geisser	2252,778	2,229	1010,848	4,975	,010
	Huynh-Feldt	2252,778	2,580	873,064	4,975	,007
	Lower-bound	2252,778	1,000	2252,778	4,975	,039
Error(time)	Sphericity Assumed	7697,222	51	150,926		
	Greenhouse-Geisser	7697,222	37,886	203,167		
	Huynh-Feldt	7697,222	43,865	175,474		
	Lower-bound	7697,222	17,000	452,778		

Ascertaining the Sphericity Assumption: Epsilon (ϵ) parameters



Mauchly's Test of Sphericity[a]

Measure: Health_score

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
time	,520	10,296	5	,068	,743	,860	,333

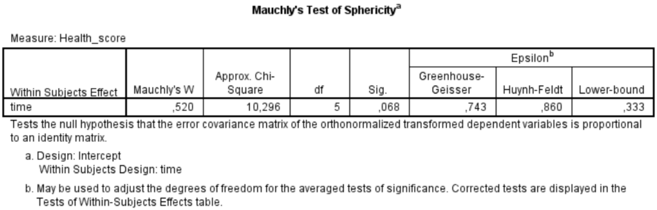
Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: time

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

- The sphericity assumption may be evaluated using the Greenhouse-Geisser epsilon (ϵ) parameter and/or Mauchly's test.
  - When $\epsilon = 1$, this is considered an indicator that sphericity is met. Values < 1 indicate departure from sphericity.
  - In the table above, the Greenhouse-Geisser $\epsilon = 0.743$. This parameter is used to adjust the degrees of freedom of the Greenhouse-Geisser repeated measures ANOVA results in the table containing the 'Tests of within-subjects effects'.
- Huynh-Feldt also defined an ϵ parameter that can used to adjust degrees of freedom in the repeated measures analysis (see Table containing 'Tests of Within-subjects effects').
- The G-G epsilon tends to underestimate the degree to which sphericity is met (making it a more conservative estimate of sphericity), while the H-F epsilon tends to overestimate the degree of sphericity (i.e., it is a more liberal estimate of sphericity).

Ascertaining the Sphericity Assumption: Mauchley's test



Mauchly's Test of Sphericity[a]

Measure: Health_score

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
time	,520	10,296	5	,068	,743	,860	,333

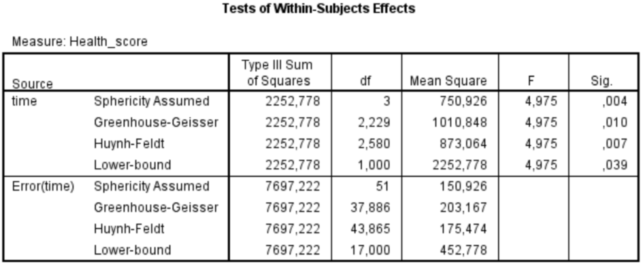
Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: time

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

- Mauchley's test provides a test of sphericity. If significant, then we assume sphericity is not met as the matrix of difference scores differs significantly from a diagonal matrix.
- In our case, p=0.068, which suggests sphericity is met.
- Note. There will be no test of sphericity (and corresponding Sig.=.) and the Greenhouse-Geisser epsilon parameter will be 1 if there are only two levels of the repeated factor. The issue of sphericity is a non-issue in this case.
- Disadvantages:
  - Mauchley's test is sensitive to multivariate nonnormality.
  - The power of the test will be impacted by sample size (i.e., less powerful for detecting a violation in smaller samples versus overpowered in larger samples).
  - Many analysts suggest Mauchly's test is unnecessary since the Greenhouse-Gessier test incorporates the degree to which the data depart from sphericity into the test results. Hence, when there is some minor deviation from sphericity, a minor adjustment to the degrees of freedom is performed and when there is greater deviation from sphericity, a more substantial adjustment to the degrees of freedom is made.

Test of Within-Subjects Effects: Choosing between the different alternatives



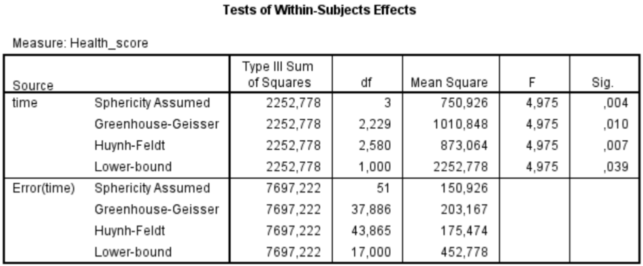
Tests of Within-Subjects Effects

Measure: Health_score

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
time	Sphericity Assumed	2252,778	3	750,926	4,975	,004
	Greenhouse-Geisser	2252,778	2,229	1010,848	4,975	,010
	Huynh-Feldt	2252,778	2,580	873,064	4,975	,007
	Lower-bound	2252,778	1,000	2252,778	4,975	,039
Error(time)	Sphericity Assumed	7697,222	51	150,926		
	Greenhouse-Geisser	7697,222	37,886	203,167		
	Huynh-Feldt	7697,222	43,865	175,474		
	Lower-bound	7697,222	17,000	452,778		

- Since the ϵ parameter computed using G-G computation can be overly conservative (thereby making the repeated measures ANOVA too conservative in terms of rejecting the null), the Huynh-Feldt test provides a less conservative alternative to testing for differences in means.
- As a general "rule of thumb": If the Greenhouse-Geisser $\epsilon < 0.75$, then use the Greenhouse-Geisser test. Otherwise, if you determine sphericity is violated (or at least are seeking a more conservative alternative to the standard 'sphericity assumed test'), then use the Huynh-Feldt test (when the G-G $\epsilon \in [0.75, 1.0]$).

Test of Within-Subjects Effects: Conclusion



Tests of Within-Subjects Effects

Measure: Health_score

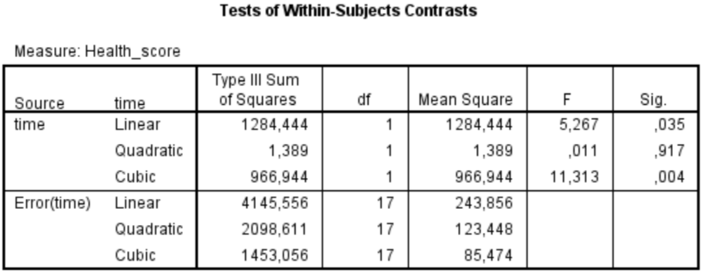
Source		Type III Sum of Squares	df	Mean Square	F	Sig.
time	Sphericity Assumed	2252,778	3	750,926	4,975	,004
	Greenhouse-Geisser	2252,778	2,229	1010,848	4,975	,010
	Huynh-Feldt	2252,778	2,580	873,064	4,975	,007
	Lower-bound	2252,778	1,000	2252,778	4,975	,039
Error(time)	Sphericity Assumed	7697,222	51	150,926		
	Greenhouse-Geisser	7697,222	37,886	203,167		
	Huynh-Feldt	7697,222	43,865	175,474		
	Lower-bound	7697,222	17,000	452,778		

- For our data, the G-G $\epsilon = 0.743 (< 0.75)$ suggests the use of the Greenhouse-Geisser test.
- The univariate repeated measures ANOVA using the Greenhouse-Geisser correction indicated there were significant differences in scores over time:

$$F(2.229, 37.886) = 4.975, \quad p = 0.010.$$

- Note that the assumption of sphericity was not violated for these data (marginally), but the decision rule did not change, since all of the p–values were less than $\alpha = 0.05$.

Test of Within-Subjects Contrasts



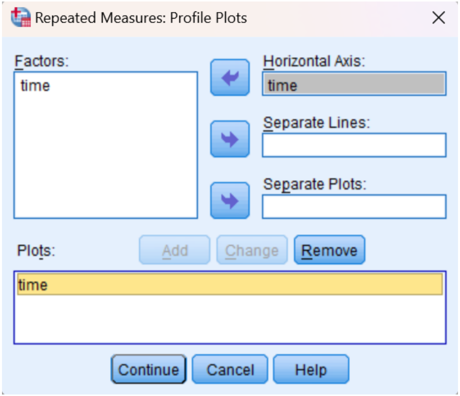
Tests of Within-Subjects Contrasts

Measure: Health_score

Source	time	Type III Sum of Squares	df	Mean Square	F	Sig.
time	Linear	1284,444	1	1284,444	5,267	,035
	Quadratic	1,389	1	1,389	,011	,917
	Cubic	966,944	1	966,944	11,313	,004
Error(time)	Linear	4145,556	17	243,856		
	Quadratic	2098,611	17	123,448		
	Cubic	1453,056	17	85,474		

- One may wonder whether there is evidence of trending over time with respect to the means of the repeated measurements.
- The 'Tests of Within-subjects contrasts' Table above can be useful in this regard.
  - A linear trend implies that the change on the repeated measure will be the same between each pair of adjacent measurement occasions.
  - A quadratic trend implies change in the change over time, and will give the appearance of a "bowl" shape as there is one "bend" in the line.
  - A cubic trend assumes two bends in the line.
  - The highest possible trend is equal to $k-1$ (i.e., # of repeated measurements minus 1). When $k=2$, the highest order polynomial trend is linear. When $k=3$, the highest order polynomial trend that is possible is quadratic. When $k=4$ (as we have here), the highest order trend that is possible is cubic.
- When pondering such questions, it is instructive to provide profile plots for illustration/comparison.

Option to obtain Profile Plots



Repeated Measures: Profile Plots

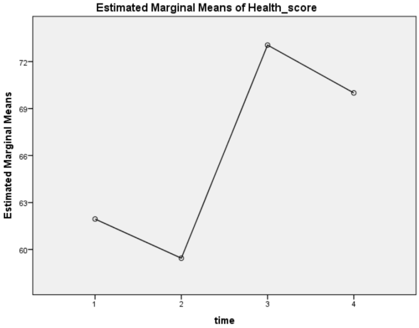
Factors: time

Horizontal Axis: time

Separate Lines:

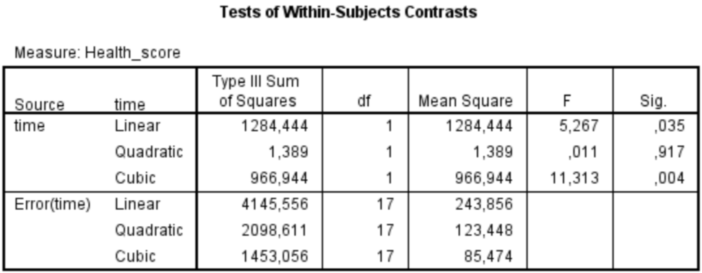
Separate Plots:

Plots: time



- When performing a trend analysis, we need to look at the highest-order polynomial terms that are significant, consider the added explanatory power that results from the addition of terms, and also consider the shape of change itself (e.g., inspection of the profile plot).
- Although one rule of thumb might be to simply go with the highest order polynomial terms that are significant, it is also important to consider the value-added of adding in those terms and whether the loss of parsimony is worth the cost of added complexity in terms of your ability to interpret the results.

Test of Within-Subjects Contrasts



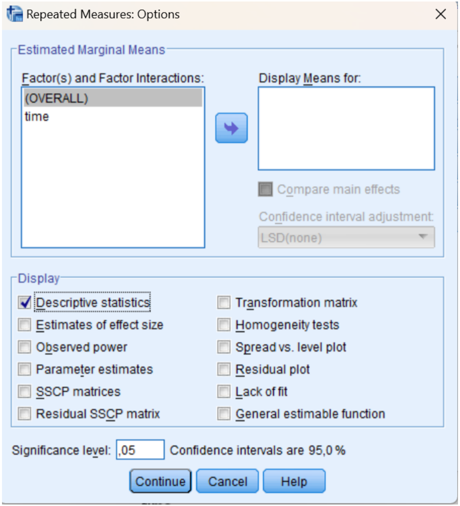
Tests of Within-Subjects Contrasts

Measure: Health_score

Source	time	Type III Sum of Squares	df	Mean Square	F	Sig.
time	Linear	1284,444	1	1284,444	5,267	,035
	Quadratic	1,389	1	1,389	,011	,917
	Cubic	966,944	1	966,944	11,313	,004
Error(time)	Linear	4145,556	17	243,856		
	Quadratic	2098,611	17	123,448		
	Cubic	1453,056	17	85,474		

- When performing a trend analysis, we need to look at the highest-order polynomial terms that are significant, consider the added explanatory power that results from the addition of terms, and also consider the shape of change itself (e.g., inspection of the profile plot).
- Although one rule of thumb might be to simply go with the highest order polynomial terms that are significant, it is also important to consider the value-added of adding in those terms and whether the loss of parsimony is worth the cost of added complexity in terms of your ability to interpret the results.
- Here, we could say the trend is cubic ($p = 0.004$).

Additional Options: Descriptives



Repeated Measures: Options

Estimated Marginal Means

Factor(s) and Factor Interactions:

(OVERALL)
time

Display Means for:

☐ Compare main effects

Confidence interval adjustment: LSD(none)

Display

☑ Descriptive statistics

☐ Estimates of effect size

☐ Observed power

☐ Parameter estimates

☐ SSCP matrices

☐ Residual SSCP matrix

☐ Transformation matrix

☐ Homogeneity tests

☐ Spread vs. level plot

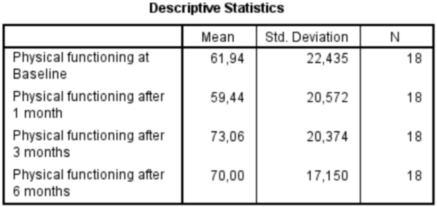
☐ Residual plot

☐ Lack of fit

☐ General estimable function

Significance level: ,05 Confidence intervals are 95,0 %

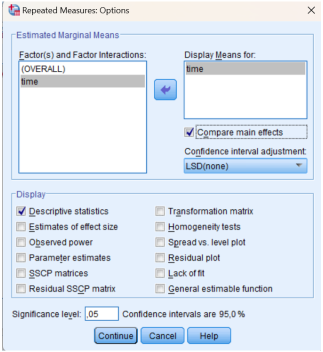
Continue
Cancel
Help



Descriptive Statistics

	Mean	Std. Deviation	N
Physical functioning at Baseline	61,94	22,435	18
Physical functioning after 1 month	59,44	20,572	18
Physical functioning after 3 months	73,06	20,374	18
Physical functioning after 6 months	70,00	17,150	18

Additional Options: Estimated Marginal Means


Repeated Measures: Options

Estimated Marginal Means

Factor(s) and Factor Interactions:
(OVERALL)
time

Display Means for:
time

☑ Compare main effects

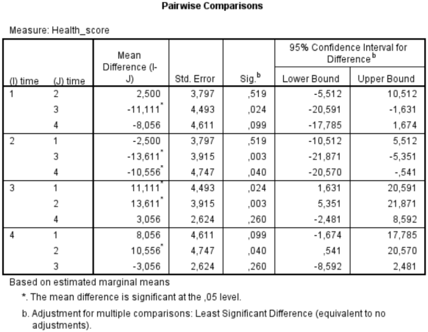
Confidence interval adjustment: LSD(none)

Display

☑ Descriptive statistics	□ Transformation matrix
□ Estimates of effect size	□ Homogeneity tests
□ Observed power	□ Spread vs. level plot
□ Parameter estimates	□ Residual plot
□ SSCP matrices	□ Lack of fit
□ Residual SSCP matrix	□ General estimable function

Significance level: ,05 Confidence intervals are 95,0 %

Continue
Cancel
Help


Pairwise Comparisons

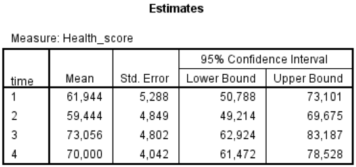
Measure: Health_score

(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	2,500	3,797	,519	-5,512	10,512
	3	-11,111*	4,493	,024	-20,591	-1,631
	4	-8,056	4,611	,099	-17,785	1,674
2	1	-2,500	3,797	,519	-10,512	5,512
	3	-13,611*	3,915	,003	-21,871	-5,351
	4	-10,556*	4,747	,040	-20,570	-,541
3	1	11,111*	4,493	,024	1,631	20,591
	2	13,611*	3,915	,003	5,351	21,871
	4	3,056	2,624	,260	-2,481	8,592
4	1	8,056	4,611	,099	-1,674	17,785
	2	10,556*	4,747	,040	,541	20,570
	3	-3,056	2,624	,260	-8,592	2,481

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).


Estimates

Measure: Health_score

time	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	61,944	5,288	50,788	73,101
2	59,444	4,849	49,214	69,675
3	73,056	4,802	62,924	83,187
4	70,000	4,042	61,472	78,528

- These are paired t-tests with p-values adjusted for multiple comparisons.
- Significant pairwise differences in scores among the time periods are flagged.
- No significant differences are observed here.

Two-Factor repeated measures design

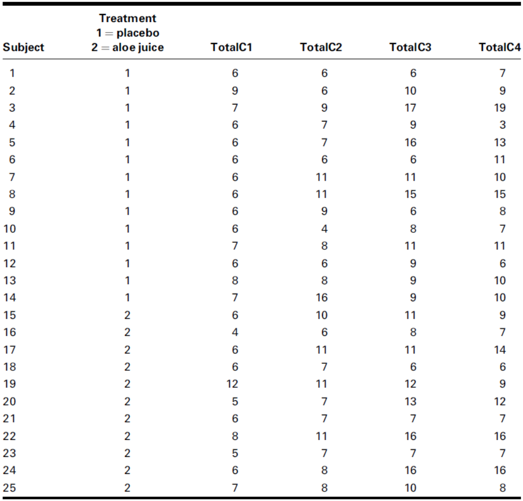
Two-Factor Repeated Measures Design

- Repeated measures ANOVA is not useful just for testing means among different observation times. The analyses are easily expanded to include testing for differences among times for different treatment groups.
- This approach can be used when testing whether individuals react the same or differently across levels of a repeated factor (for example, different stimuli for which a person is exposed) and a grouping variable.
- As an example, a clinic may wish to test a placebo treatment against a new medication treatment. Researchers will randomly assign patients to one of the two treatment groups and will obtain measurements through time for each subject. In the end they are interested in knowing if there were differences between the two treatments on subjects that were measured multiple times.
- The model for the two-factor repeated measures design must represent the fact that there are two factors, A and B, and they have a potential interaction:

$$x_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \quad i = 1, 2, \ldots, a, \quad j = 1, 2, \ldots, b, \quad k = 1, 2, \ldots, n.$$

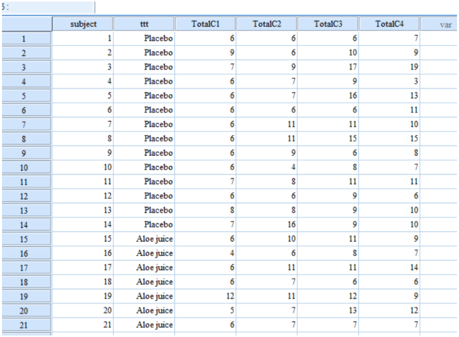
  - α_j represents the main effect of factor A,
  - β_k represents the main effect of factor B,
  - $(\alpha\beta)_{jk}$ represents the interaction effect of factor A and factor B,
  - ϵ_{ijk} is a residual component representing all sources of variation other than treatments and blocks.

Oral Health Condition Scores at Four Different Points in Time Under Two Treatment Conditions



Subject	Treatment 1 = placebo 2 = aloe juice	TotalC1	TotalC2	TotalC3	TotalC4
1	1	6	6	6	7
2	1	9	6	10	9
3	1	7	9	17	19
4	1	6	7	9	3
5	1	6	7	16	13
6	1	6	6	6	11
7	1	6	11	11	10
8	1	6	11	15	15
9	1	6	9	6	8
10	1	6	4	8	7
11	1	7	8	11	11
12	1	6	6	9	6
13	1	8	8	9	10
14	1	7	16	9	10
15	2	6	10	11	9
16	2	4	6	8	7
17	2	6	11	11	14
18	2	6	7	6	6
19	2	12	11	12	9
20	2	5	7	13	12
21	2	6	7	7	7
22	2	8	11	16	16
23	2	5	7	7	7
24	2	6	8	16	16
25	2	7	8	10	8

Oral Health Condition Scores at Four Different Points in Time Under Two Treatment Conditions (2)



5 :

	subject	ttt	TotalC1	TotalC2	TotalC3	TotalC4	var
1	1	Placebo	6	6	6	7	
2	2	Placebo	9	6	10	9	
3	3	Placebo	7	9	17	19	
4	4	Placebo	6	7	9	3	
5	5	Placebo	6	7	16	13	
6	6	Placebo	6	6	6	11	
7	7	Placebo	6	11	11	10	
8	8	Placebo	6	11	15	15	
9	9	Placebo	6	9	6	8	
10	10	Placebo	6	4	8	7	
11	11	Placebo	7	8	11	11	
12	12	Placebo	6	6	9	6	
13	13	Placebo	8	8	9	10	
14	14	Placebo	7	16	9	10	
15	15	Aloe juice	6	10	11	9	
16	16	Aloe juice	4	6	8	7	
17	17	Aloe juice	6	11	11	14	
18	18	Aloe juice	6	7	6	6	
19	19	Aloe juice	12	11	12	9	
20	20	Aloe juice	5	7	13	12	
21	21	Aloe juice	6	7	7	7	

- Examination of 25 subjects with neck cancer with outcome variable an oral health condition score.
- Random division into two treatment groups → placebo treatment (treatment 1) and an aloe juice group (treatment 2).
- Cancer health was measured at baseline and at the end of 2, 4, and 6 weeks of treatment.
- The goal was to discern if there was any change in oral health condition over the course of the experiment and to see if there were any differences between the two treatment conditions.

Hypotheses

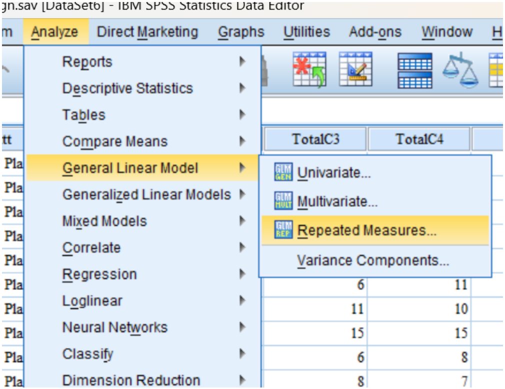
1. $$\begin{cases} H_0 : \alpha_i = 0, \quad i = 1, 2, \ldots, a, \\ H_a : \text{not all } \alpha_i = 0. \end{cases}$$

2. $$\begin{cases} H_0 : \beta_j = 0, \quad j = 1, 2, \ldots, b, \\ H_a : \text{not all } \beta_j = 0. \end{cases}$$

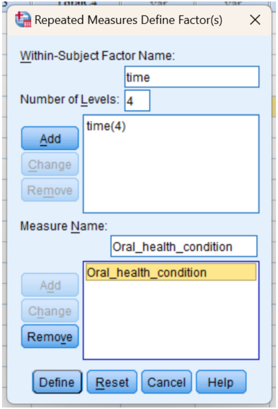
3. $$\begin{cases} H_0 : (\alpha\beta)_{ij} = 0, \quad i = 1, 2, \ldots, a, \quad j = 1, 2, \ldots, b, \\ H_a : \text{not all } (\alpha\beta)_{ij} = 0. \end{cases}$$

- Test statistic. Distributed as F with:
  - Within-subject effects: $4 - 1 = 3$ numerator degrees of freedom and $(4 - 1)(25 - 2) = 69$ denominator degrees of freedom for the time factor.
  - Within-subject effects: $(4 - 1)(2 - 1) = 3$ numerator degrees of freedom for the interaction factor and $(4 - 1)(25 - 2) = 69$ denominator degrees of freedom for the interaction factor.
  - Between-subject factor: $2 - 1 = 1$ numerator degrees of freedom and $25 - 2 = 23$ denominator degrees of freedom .
- If the assumptions, specifically of sphericity, are not met, then the computer program will alter the degrees of freedom and hence the critical value for comparisons.

Two-Factor Repeated Measures in SPSS



Two-Factor Repeated Measures in SPSS (2)



Repeated Measures Define Factor(s)

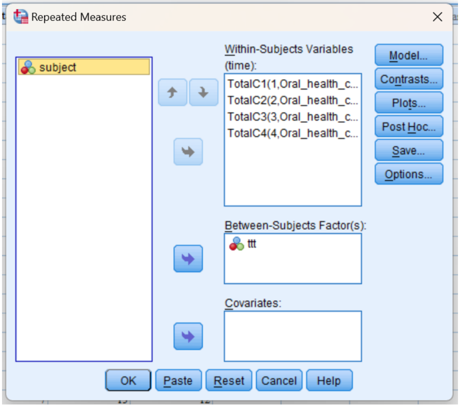
Within-Subject Factor Name: time

Number of Levels: 4

time(4)

Measure Name: Oral_health_condition

Oral_health_condition



Repeated Measures

subject

Within-Subjects Variables (time):

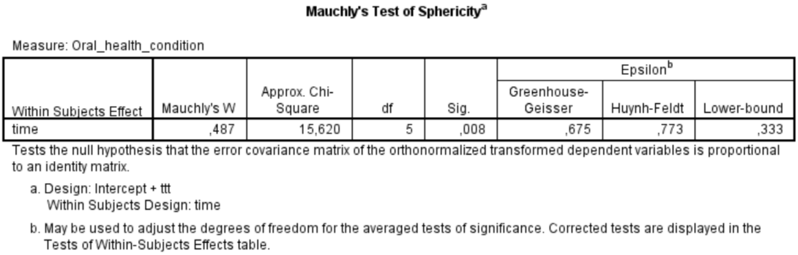
TotalC1(1,Oral_health_c...
TotalC2(2,Oral_health_c...
TotalC3(3,Oral_health_c...
TotalC4(4,Oral_health_c...

Between-Subjects Factor(s):

ttt

Covariates:

Tests of Sphericity



Mauchly's Test of Sphericity[a]

Measure: Oral_health_condition

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
time	,487	15,620	5	,008	,675	,773	,333

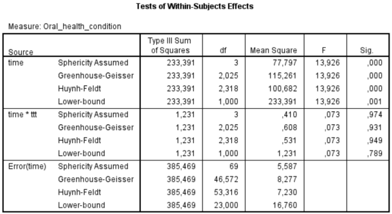
Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + ttt
Within Subjects Design: time

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

- The sphericity assumption is required for all univariate main effects tests and interaction tests. Given Mauchly's test is impacted by non-normality and by sample size, it is not highly recommended when evaluating whether the sphericity condition has been met. We would reject the null for this test, according to the output p–value (p=0.008).
- A Greenhouse-Geisser epsilon (ϵ) value $< .75$, suggests using the Greenhouse-Geisser adjustment with the univariate test of mean differences (see table of "Tests of within-subjects effects"), whereas a value falling between .75 and 1 suggests the use of the Huynh-Feldt adjustment with the univariate tests. [$\epsilon = 1$ is consistent with sphericity]. The sphericity assumed test can be used if you determine sphericity is not violated.
- The Lower-Bound test is generally overly conservative and is not typically used.
- Following the considerations above, we will proceed, referring to the G-G modification of the degrees of freedom in the "Tests of within-subjects effects" Anova Table.

Tests of within subjects effects in SPSS: Output



Tests of Within-Subjects Effects

Measure: Oral_health_condition

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
time	Sphericity Assumed	233,391	3	77,797	13,926	,000
	Greenhouse-Geisser	233,391	2,025	115,261	13,926	,000
	Huynh-Feldt	233,391	2,318	100,682	13,926	,000
	Lower-bound	233,391	1,000	233,391	13,926	,001
time * ttt	Sphericity Assumed	1,231	3	,410	,073	,974
	Greenhouse-Geisser	1,231	2,025	,608	,073	,931
	Huynh-Feldt	1,231	2,318	,531	,073	,949
	Lower-bound	1,231	1,000	1,231	,073	,789
Error(time)	Sphericity Assumed	385,469	69	5,587		
	Greenhouse-Geisser	385,469	46,572	8,277		
	Huynh-Feldt	385,469	53,316	7,230		
	Lower-bound	385,469	23,000	16,760		

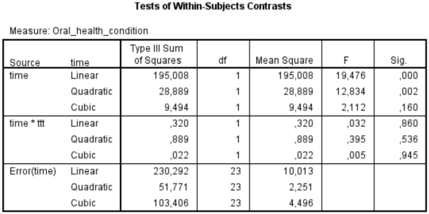
- All three test results yield the same conclusions with respect to the main and interaction effects.
- The main effect of time on oral condition scores is statistically significant, according to the G-G modification. Variance ratio:

  $$F(2.025, 46.572) = 13.926, p < 0.001.$$

  Hence, we reject the null hypothesis concerning changes through time.
- Not significant time X group interaction effect:

  $$F(2.025, 46.572) = 0.073, p > 0.05.$$

  Hence, we fail to reject the null hypothesis concerning the interaction of time and treatment.



Tests of Within-Subjects Contrasts

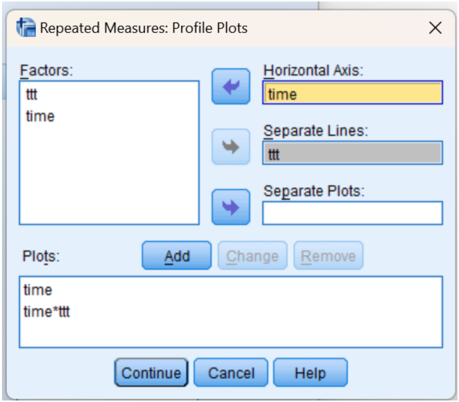
Measure: Oral_health_condition

Source	time	Type III Sum of Squares	df	Mean Square	F	Sig.
time	Linear	195,008	1	195,008	19,476	,000
	Quadratic	28,889	1	28,889	12,834	,002
	Cubic	9,494	1	9,494	2,112	,160
time * ttt	Linear	,320	1	,320	,032	,860
	Quadratic	,889	1	,889	,395	,536
	Cubic	,022	1	,022	,005	,945
Error(time)	Linear	230,292	23	10,013		
	Quadratic	51,771	23	2,251		
	Cubic	103,406	23	4,496		

- Although the test of the linear component of the trend is significant (p<0.001), the higher-order quadratic component was also significant [F(1,23)=12.834, p=0.002]. This suggests that across groups, the mean oral health score exhibited a quadratic trend over the four measurement occasions. This is further suggested by examining the profile plot of the means.
- Also, the test of the interaction between the linear (also quadratic etc.) component of the trend and treatment group is not significant [F(1,23)=0.320, p=.860].

Plotting the mean scores by time and by time and treatment group

- Though the previous output can be valuable for statistical interpretation, it is often useful to examine plots to obtain a visual interpretation of the results:



Repeated Measures: Profile Plots

Factors:
ttt
time

Horizontal Axis: time

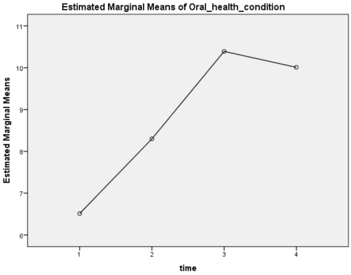
Separate Lines: ttt

Separate Plots:

Plots:
time
time*ttt

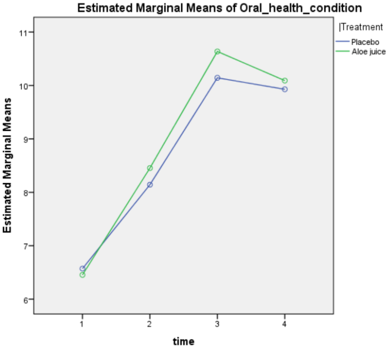
Assessment of trending over time

Assessment of trending over time (irrespective of group membership)



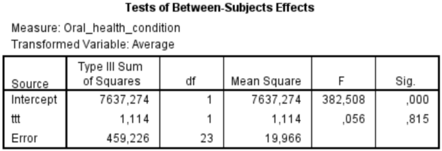
- We observe that across groups, the mean level of oral condition scores exhibited a quadratic trend over the four measurement occasions.
- It is evident that changes in oral condition did occur through time, but that the two treatments were very similar, as can be seen by the close proximity of the two curves in the differential trending plot on the right:

Testing for differential trending across groups



- Plot of marginal means against time, with lines representing each of the treatments.
- Looking at the profile plot of means, we see that the curvatures of the lines for the two Treatments are not that different. Since these trends are roughly parallel, it is no surprise the test of the time X group interaction is not significant.

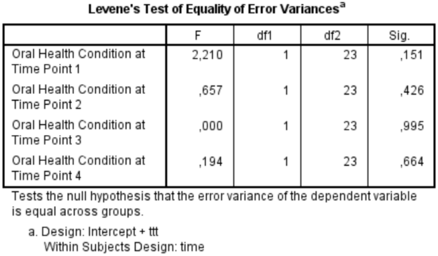
Tests of between subjects effects in SPSS: Output



Tests of Between-Subjects Effects

Measure: Oral_health_condition
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	7637,274	1	7637,274	382,508	,000
ttt	1,114	1	1,114	,056	,815
Error	459,226	23	19,966		



Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
Oral Health Condition at Time Point 1	2,210	1	23	,151
Oral Health Condition at Time Point 2	,657	1	23	,426
Oral Health Condition at Time Point 3	,000	1	23	,995
Oral Health Condition at Time Point 4	,194	1	23	,664

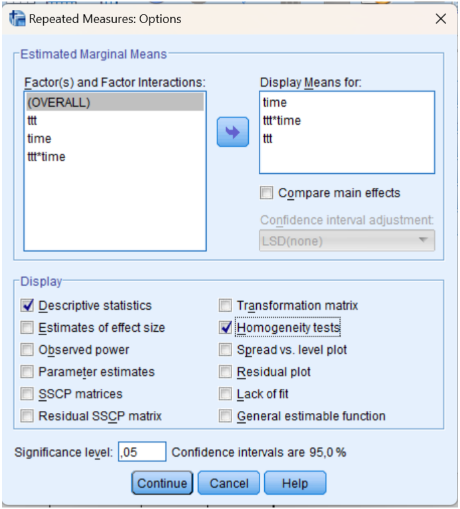
Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + ttt
Within Subjects Design: time

- The Tests of Between-subjects Effects is a test of the main effect of the grouping variable on scores on the repeated measure averaged over time. The result presented here is simply a test of group differences on the average of oral health condition scores (i.e., those scores averaged over time for each person).
- The main effect of treatment group on the average oral health condition score across time is not statistically significant, F(1, 23)=.056, p=0.815>0.05. Hence, we fail to reject the null hypothesis concerning differences between treatments.
- The Levene's test results involve tests of differences in variances at each time point, an assumption of the univariate ANOVA (for the Tests of Between-subjects effects). It turns out that the standard Levene's tests (and robust tests, based on median, etc.) are non-significant for all Times periods.
- Nevertheless, in general, a potential violation of this assumption is less of an issue with roughly equivalent sample sizes (where largest n/smallest $n < 1.5$).

Levene's Test of Equality of Error Variances in SPSS

To get the output for Levene's Test of Equality of Error Variances:



Repeated Measures: Options

Estimated Marginal Means

Factor(s) and Factor Interactions:
- (OVERALL)
- ttt
- time
- ttt*time

Display Means for:
- time
- ttt*time
- ttt

□ Compare main effects

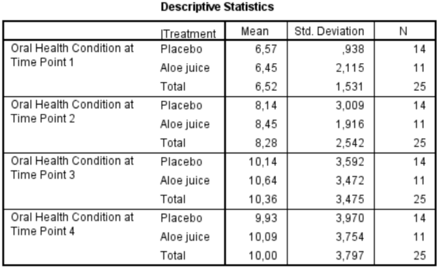
Confidence interval adjustment: LSD(none)

Display

☑ Descriptive statistics
□ Estimates of effect size
□ Observed power
□ Parameter estimates
□ SSCP matrices
□ Residual SSCP matrix
□ Transformation matrix
☑ Homogeneity tests
□ Spread vs. level plot
□ Residual plot
□ Lack of fit
□ General estimable function

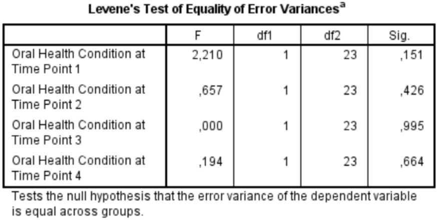
Significance level: ,05 Confidence intervals are 95,0 %

Continue
Cancel
Help



Descriptive Statistics

	Treatment	Mean	Std. Deviation	N
Oral Health Condition at Time Point 1	Placebo	6,57	,938	14
	Aloe juice	6,45	2,115	11
	Total	6,52	1,531	25
Oral Health Condition at Time Point 2	Placebo	8,14	3,009	14
	Aloe juice	8,45	1,916	11
	Total	8,28	2,542	25
Oral Health Condition at Time Point 3	Placebo	10,14	3,592	14
	Aloe juice	10,64	3,472	11
	Total	10,36	3,475	25
Oral Health Condition at Time Point 4	Placebo	9,93	3,970	14
	Aloe juice	10,09	3,754	11
	Total	10,00	3,797	25



Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
Oral Health Condition at Time Point 1	2,210	1	23	,151
Oral Health Condition at Time Point 2	,657	1	23	,426
Oral Health Condition at Time Point 3	,000	1	23	,995
Oral Health Condition at Time Point 4	,194	1	23	,664

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.