Università degli Studi di Padova

Facoltà di Scienze Statistiche

Corso di laurea in Statistica e Tecnologie Informatiche

Aggression Questionnaire, un applicazione per valutare gli effetti delle attività sportive

(Aggression Questionnaire, an application to evaluate effects of sport activities)

Relatore: Prof.ssa Laura Ventura

Correlatore: Prof. Attilio Carraro

Laureando: Michele Gaiotto

Anno Accademico 2009/2010

To my family and friends

Un Ringraziamento speciale alla prof. Laura Ventura

per il tempo, la pazienza e la grande disponibilità

Preface

In psychology, as well as other social behavioural sciences, aggression refers to behaviour between members of the same species that is intended to cause pain or harm. Aggression takes a variety of forms among human beings, and it can be physical, mental, or verbal.

In the last century many psychologists have been wondering about "how to measure aggression?" and some of them have successfully developed methods for this purpose (Buss&Durkee,1982, Buss&Perry, 1991, and Bryant&Smith, 2001). The widely used methods are the "AQ"s, "Aggression Questionnaire", that is composed of various items (questions) capable of determining the scale of aggression in a single person (Buss & Durkee, 1991). In this thesis we will investigate the use of this questionnaire, particularly the "12-item AQ", explained in the following chapters, with reference to an experiment on teenagers (13-14 years old) before and after sport activities within school physical education lessons. The experiment was carried out by Professor Attilio Carraro and his team from the University of Padua.

The outline of this thesis is as follows: the first part will focus on the structure of the "Aggression questionnaire", both long (29 item) and short (12 item). In Chapter two the statistical analysis of the data set is discussed. Finally, in Chapter three, there is a brief presentation of the "Mixed Effect Linear Model" with the fitting of this model to our dataset. Finally, conclusions about the experiments are presented.

INDEX

1 Chapter one: Introduction

1.1 *Aggression*

In psychology, as well as other social and behavioural sciences, aggression refers to behaviour between members of the same species that is intended to cause pain or harm. Aggression takes a variety of forms among human beings, and it can be physical, mental, or verbal. It can be influenced by several factors. For instance Culture is a distinctly human factor that plays a role on aggression. Empirical cross-cultural research has found differences in the level of aggression between cultures. For example, within America's culture, southerners were shown to become more aroused and to respond more aggressively then northerners. Another factor that is strongly correlated with aggression, is Media. Watching and imitating the behaviour of others can learn behaviours like aggression. A considerable amount of evidence suggests that watching violence on television increases the likelihood of short-term aggression in children. A long-term study of over 700 families found a "significant association" between the amounts of time spent watching violent television as a teenager and the likelihood to commit acts of aggression later in life. Situational factors like alcohol are strongly correlated to aggression. Others factors can be frustration, pain and discomfort and also the hot temperatures. It's very curious that in various studies hot temperatures have been implicated as a factor. One study completed in 1979 demonstrated that students were more aggressive and irritable after taking a test in a hot classroom. An important role is played by the gender. Males are historically believed to be generally more physically aggressive then females. It has been shown that males are quicker to aggression and more likely than females to express their aggression physically. Females express aggression by using a variety of non-physical means. Aggression is a personality characteristic relevant for both clinicians and researchers. In adults' subjects, previous studies showed a significant association between aggression and severe personality disorders. A significant association has also been shown in adolescent subjects between aggression and reckless behaviour, potentially injurious to self and others. Several studies have demonstrated that aggression is a stable personality trait, and that aggressiveness assessed early in life is predictive of later adult criminal behaviour.

1.2 Sports and behaviours

In this study we are interested in the effects that sports have on aggression, more particularly, what type of sports helps individuals to improve their behaviour. In 1996, the qualification and curriculum authority's PE and School sport (PESS) investigation concluded that physical activity has a motivational impact on children, increasing their self-esteem and general well-being. It has also found that PE and sport help children to develop essential social skills such as concentration, self-discipline cooperation and an awareness of the need to think of things and people other than themselves. Despite this general study we wondered: Do all the sports act like this, or some more than others? Do all the sports act fine in terms of aggression? At this point, a distinction had to be made. We divided the sport in two categories. The first one, non-contact sport, and a second one, contact sport. With contact sports we mean sports like fighting and with non-contact sport we mean sports like volleyball. Can contact sports like Judo, all-in wrestling be linked to a positive behaviour more then other sports? In other terms, how do these two different categories of sport act in terms of aggression? At this point we needed a system that allowed us to measure the aggression in a subject.

1.3 The aggression measurement

Historically, aggression has been measured in a variety of ways. At much younger ages such as preschool or lower elementary levels, the methods to measure aggression are principally observations. Adolescents and adults are usually studied by means of self-report questionnaires or laboratory methods. Among these several measures of aggression, the hostility inventory developed by Buss and Durkee (1991) was one of the most frequently used questionnaires. Despite the wide use of this questionnaire, Buss and Perry felt dissatisfied with its psychometric[1] properties and developed the Aggression Questionnaire (AQ 29-item, Buss & Perry, 1992), which represented a psychometrically improved and updated version of the mentioned Buss-Durkee Hostility Inventory. Recently, Bryant and Smith (2001) developed a better-fitting, shortened version of the AQ 29-item. The 12-item AQ not only preserved the conceptual content of the original model but was also psychometrically superior to the original 29-item measure. They demonstrated that across three independent samples, the refined four-factor model explained an acceptable proportion of common variance. In the next two chapters both AQ 29-Item and AQ 12-Item will be shortly explained.

---

[1] Psychometrics is the field of study concerned with the theory and technique of educational and psychological measurement, which includes the measurement of knowledge, abilities, attitudes and personality traits.

1.4 The "AQ 29-Item"

In the previous section, we told about how aggression is measured in social sciences. The hostility inventory, which was the first real method, still remains one of most used questionnaires on aggression, with a lot of citations in literature. However, this questionnaire had some issues: the fundamental problem was that this type of questionnaire did not meet modern psychometric standards. Buss and Perry (1992) developed a new self-report questionnaire. It retained the major virtue of the older inventory, but, met the psychometric standards of that time. The new questionnaire was divided into various components of aggression: Physical Aggression, Verbal Aggression, Anger, Indirect aggression, Resentment and Suspicion. Some items were borrowed intact from the earlier Hostility Inventory. A number of items from earlier inventory were not used because, on reflection, they were judged to be ambiguous or troublesome. For instance an indirect aggression item, such as "Since the age of ten, I have never had a temper tantrum", denies anger and requires that the respondent have an excellent memory. Other items were rewritten in the interest of greater clarity, and many new items were added to form an initial pool of 52. To assess the functionality of the questionnaire, the entire set was administered to large groups of college students. Almost all of them were 18 to 20 age. There were 612 men and 614 women, a total of 1253 subjects. They were asked to rate each item on a scale of 1 (extremely uncharacteristic of me) to 5 (extremely characteristic of me). In the next factor analysis only four of the supposed six item emerged: Physical Aggression, Verbal Aggression, Anger and Hostility, with hostility being a combination of resentment and suspicion. Indirect aggression did not appear as a separate factor and the items related to this component were eliminated from the questionnaire. After further studies in the questionnaire (for a complete description of the questionnaire see Buss & Perry), the researchers eliminated more items from the questionnaire remaining with 29 items.

At the end of the questionnaire's analysis there was a clear evidence that personality trait of aggression consists of four subtraits. Physical and Verbal Aggression, which involve hurting or harming others, represent the instrumental or motor component of behavior. Anger, which involves psyiological arousal and preparation for aggression, represents the emotional or affective component of behavior. Hostility, which consists of feelings of ill will and injustice, represents the cognitive component of behavior. Studies also demonstrated how physical and verbal aggression correlated strongly and how, not surprisingly, that these two subtraits only weakly correlated with the cognitive compontent, Hostility. What was really unexpected was the fact that Anger correlated strongly with the other three subtraits. Partial correlation suggested that anger was a kind of psichological bridge between the instrumental components and the cognitive component. Anger is,

in fact, a prelude to aggression, and we need not document here that people more likely to aggress when angry than when not angry. The sex differences in aggression are of some interest. Men are know to be more aggressive than women, and, as expected, men had a higher total score than women. The individual scales provide more detailed information. Men were much more physically aggressive than women, somewhat more verbally aggressive, and just a little more hostile. There were no sex differences for anger. One way of integrating these data is to suggest that inhibition may be at work here: Women become just as angry as men but inhibit expression of this anger by means of instrumental aggression. The reason for constructing a new questionnaire of aggression was the need for an instrument that, like the previous Hostility inventory, assessed various components but still met the current psychometric standards. The questionnaire led to important information about the components of aggression, specifically, where the sex differences are and which components correlate with particular personality traits. Despite these powerful characteristic of this questionnaire, it was rivisited and adjusted by Bryant and Smith(2001).

1.5 The "AQ 12-item"

Recently, Bryant and Smith (2001) developed a better-fitting and shortened version of Buss and Perry (1992) 29-item measure. The 12-item AQ not only preserved the conceptual content of the original model but was also psychometrically superior to the original 29-item AQ measure and replicated its four factor structure. Bryant and Smith(2001) demonstrated that across three independent samples, the refined four-factor model explained an acceptable proportion of common variance. The refined model's root mean-square error of approximation also reflected reasonably close fit across all three samples (RMSEA s^2=.062). It appears that the existing measurement models in the literature applicable for the original 29-item AQ were also relevant for the 12-item short version. Two main differences between the 29-item and the 12-item AQ are worth exploring. First, the short 12-item version of the AQ has fewer items than the original 29-item version. Given that inter-item correlations are about the same, Cronbach-α [3]approaches 1 as the number of items approach infinity. The reliability of the scores from a particular scale should increase as a result of adding items that measure the same phenomenon. Hence, it is possible that one consequence of shortening the scale is that of lowering its Cronbach-α estimate. Second, the original 29-item

---

[2]   The RMSEA (or Root Mean Square Error of Approximation) is a measure based on the non-centrality parameter. It's formula

is $\sqrt{\dfrac{(\frac{\chi^2}{df}-1)}{N-1}}$ where n it the sample size, df the degrees of freedom of the model, χ^2 is the chi-square test for fitting. Good models

have an RMSEA of .05 or more. Models whose RMSEA is .10 or more have poor fit.

[3]   The Cronbach alpha is a commonly used measure of the internal consistency or reliability of a psychometric test score. See §2.4.1.

version of the AQ contained both positive items and negative items for aggression necessitating reverse scoring for the negative items prior to obtaining a total AQ score, on the contrary the 12-item version does not contain negative items.

1.6 *Our problem, an application of short version*

At this point the reader should have all the information needed to understand the aim of our problem. Bryant and Smith (2001) gave us an important instrument to valuate the aggression of a person whenever we want. Thus, we decided to make an experiment in order to evaluate if the scale of aggression has improved after giving some subjects a treatment. More particularly, we decided to measure the aggression scale before and after this treatment due to evaluate if this treatment has really worked positively (scale of aggression lowest after the treatment), worked negatively (scale of aggression highest after the treatment) or worked neutrally (scale of aggression unchanged). The experiment took place between November 2008 and February 2009, in two suburban secondary schools of Vicenza (Italy). The goal of the experiment was evaluating if "contact sport lessons"[4] during physical education lessons in a period of one month were positive for aggression. More particularly, in the month of November, students compiled a short version of the AQ 12-item. Those students were split into two groups: in the first one students were given of "contact sport lesson"; in the second one, students practiced sports like volleyball, were contact was absolutely absent. Right after this period, aggression were evaluated again in both samples, again using the AQ-12 item. Note that this experiment could be referred to a "Case-Control Study"[5] where cases are students of group one (contact activities) and controls, group two, are a kind of "placebo" (contact absent). At this point, the effect of the two treatments to aggression could be evaluated and for this purpose in the followings chapters a complete statistical analysis will be presented.

---

[4]   Contact sports are sports like Judo, Karate and all-in wrestling.

[5]   A wide used type of experiments in which cases are given of one type of treatment and controls the other one. Most of the times controls are not given of anything (placebo) in order to valuate the neutrality.

2 Chapter two: The Variables

In this chapter we present the questionnaire used for the experiment. The structure is a "12-item AQ" where each item has five scales (1=not at all like me, 5=completely like to me). Together with a global aggression measure the questionnaire allows us to valuate the single substraits of this characteristic. As said, the four components are *physical aggression, verbal aggression, anger and hostility*. For example, items number three, eight and eleven are used to determine physical aggression. Each subscale has three items per subscale. In the following summary we report the composition of the scale (numbers before the phrases refers to question number of the questionnaire):

-Physical Aggression

3: I have threatened people I know

8: Given enough provocation, I may hit another person

11: There are people who pushed me so far that we came to blows

-Verbal Aggression

1: I often find myself disagreeing with people

6: My friends say that I'm somewhat argumentative

9: I can't help getting into arguments when people disagree with me

-Anger

5: I have trouble controlling my temper

7: I flare up quickly but get over it quickly

12: Sometimes I fly over the handle for no good reason

-Hostility

2: At times I fell I have gotten a raw deal out of life

4: I wonder why sometimes I feel so bitter about things

10: Other people always seem to get the breaks

Each measure of the subscale can be obtained by summing the ratings to questions appertaining to each subscale. For example, if a subject answered to question number three with "2", to question eight with 2 and to question eleven with 2 the final measure for Physical Aggression will be "2+2+2=6". The total measure of aggression is given by the summing of all components:

" Aggression = Physical Aggression + Verbal Aggression + Anger + Hostility "

Two main parts make the questionnaire. In the first one students were asked to give personal details, but being the questionnaire made to be fully anonymous, only information regarding age, gender and classroom were requested. The second and main part, being the body of the questionnaire, students had to answer the questions. The questionnaire is given in the Appendix.

2.1 Collecting the data

n this section we will briefly explaining how we collected the data. The students were part of two different schools and were 13 to 14 years old. Five classes from both schools were randomly selected. In the classes male and females were almost equally divided (90 females and 123 males), for every class we made a sheet where we wrote their names and then we proceed to a draw where each class has the same chance of getting chosen. After the selection of classes to be part of the experiment, three of the previously selected classes from school one were assigned, together with two classes of school two, to group one (the group that will be attending the contact sport lessons) and the remaining to group two (the group that will be attending noncontact sport lessons). After selecting the groups, the questionnaires were given to students of both groups. During the following month students from group one were given the lessons of fighting sports and the students from group two were given of noncontact sport lessons. Right after the treatments the questionnaire were given again to all the students.

2.2 The dataset

Before getting inside a pure analysis of the data, preliminary considerations have to be made. The dataset is divided in three parts. The first one contains the data referred to the first measure of aggression, made before the treatments in November 2009. The second part contains the data referred to the measure that took place in February 2010 (after the treatments). The third contains measure of likeness to the activity taught in the months before. This part of the experiment won't be considered in the analysis, mostly because its purpose is out of our main objective. Though, just for curiosity, students whom compiled the questionnaire on sport likeness didn't like those activities.

Let's get inside those parts and describe them. The first one contains the data referred to the first round of the test. It is composed of twenty-eight variables: the age of the subjects, the gender (1 for male, 2 for female), the group (1 for sample one, 2 for sample two), the school (1 for school one, 2 for school two), AQ-x (where x from 1 to 12, are the answers to the different items of the AQ, in a scale from 1 to 5), AQ-tot (the total measure of aggression of the subjects), PhysAgg (the Physical aggression measure of a single subject), VerbAgg (the verbal aggression measure of a single subject), Anger (the anger measure of a single subject) and finally Hostility (Hostility measure of a single subject). As the first part, the second contains the same variables, taken after the treatments.

3 Chapter three: Statistical Analysis

In this chapter an analysis of the dataset is discussed. First, a brief presentation of the data is given. Second, the consistency of the samples is tested. Third, differences between groups are investigated, using the appropriate statistic test. Finally, the correlations between the various variables involved in the experiment are studied. The data has been analysed in R 2.10.1(http://www.r-project.org/).

3.1 Some summaries

In the previous chapter described how the dataset is made, without giving any general measure of it. However in order to have a better idea of what we are treating, we computed some exploratory statistics.

We recall that sample size is n=104. Students are from two school: 105 students coming from school one and 102 from school two. The summary of the gender is given in Table 3.1.

Gender	Frequency	%
Male	121	58.45%
Females	86	41.55%

Tab. 3.1 Males and Females in the study

The students are 13 or 14 years old, with thirteen's being more than fourteen's (see Table 3.2)

Age	Frequency	%
13 yo	156	75.36%
14 yo	51	24.64%

Tab. 3.2 Males and Females in the study

Table 3.3 gives the summaries of the variables pre and post treatment in group one (the post treatment variables have a "R." in front).

Object	Min.	First Q.	Median	Mean (SD)	Third Q.	Max
Age	13	13	13	13.25	13	14
Aq_tot	16	25	31	30,75(7,16)	36	47
PhysAgg	3	4	7	6,91(3,10)	9	14
VerbAgg	3	5	7	7,13(2,26)	9	14
Anger	3	7	8	8,4(2,46)	10	14
Hostility	3	7	9	8,29(2,63)	10	13
R.AQ_tot	12	20,5	25	24,94(6,31)	28,5	44
R.PhysAgg	3	3	5	5,35(2,38)	7	13
R.VerbAgg	3	4	6	5,87(1,99)	7	13
R.Anger	3	5	7	6,91(2,33)	8	14
R.Hostility	3	5	7	6,81(2,29)	9	12

Tab.3. 3 Summary for every variable in group one

Table 3.4 refers the summaries for group two:

Object	Min.	First Q.	Median	Mean (SD)	Third Q.	Max
Age	13	13	13	13.25	13	14
Aq_tot	14	22	26,5	27,44(7,17)	32	45
PhysAgg	3	4	7	6,54(3,27)	9	15
VerbAgg	3	4,75	6	6,06(2,10)	8	12
Anger	3	5	7	7,13(2,70)	9	14
Hostility	3	6	7	7,62(2,71)	9,25	14
R.AQ_tot	17	23,75	28	28,88(6,26)	33	46
R.PhysAgg	3	4	7	7,15(3,20)	9	15
R.VerbAgg	3	5	6	6,41(2,05)	8	12
R.Anger	3	6	7	7,42(2,25)	9	12
R.Hostility	3	6	8	7,89(2,43)	9	15

Tab.3. 4 Summary for every variable in group two

Recalling that group one is the group that has been given of contact sport lessons, we can say how every variable shows improvements after the treatments. Section 3.3 verifies the differences.

3.2 Reliability estimates

Once we collected the sample we have to determine its internal consistency. In statistics and research, **internal consistency** is typically a measure based on the correlations between different items on the same test (or the same subscale on a larger test). It measures whether several items that propose to measure the same general construct produce similar scores. In order to evaluate the reliability of our measures we use the "*Cronbach-α*"(Cronbach (1951), a measure of the internal consistency or reliability of a test score. It can take any values less then or equal to *1*. Higher values of α are more desirable. The "*Cronbach-α*" can be obtained by this formula:

$$\alpha = \frac{K}{K-1}\left(1 - \frac{\sum \sigma_{Y_i}^2}{\sigma_X^2}\right), \tag{1}$$

where K is the total number of components (items), σ_X^2 the variance of the observed total test scores, and $\sigma_{Y_i}^2$ the variance of component *i* for the current sample of persons.

In R, the *Cronbach-α* can be computed with function `cronbach(items).` On first sample the *Cronbach-α* alpha is *0,645*, which is considered to be a good measure. On the second sample it is equal to *0,592*. Both values approach *0,6* wich means good reliability.

3.3 Comparison pre-post treatment

In this section we analyse the effect of the different treatment on the two groups. Depending on the assumption of normality (Shapiro-Wilk's test) and on the result of homogeneity of variance (F test), we used parametric (Student's t-test) or nonparametric tests (Wilcoxon test, Mann-Whitney test) to study differences between pre-treatment and post-treatment.

3.3.1 General Aggression

In this subsection we treat the general measure of aggression (AQ_tot), first by giving some summary measure and then by testing within and between differences.
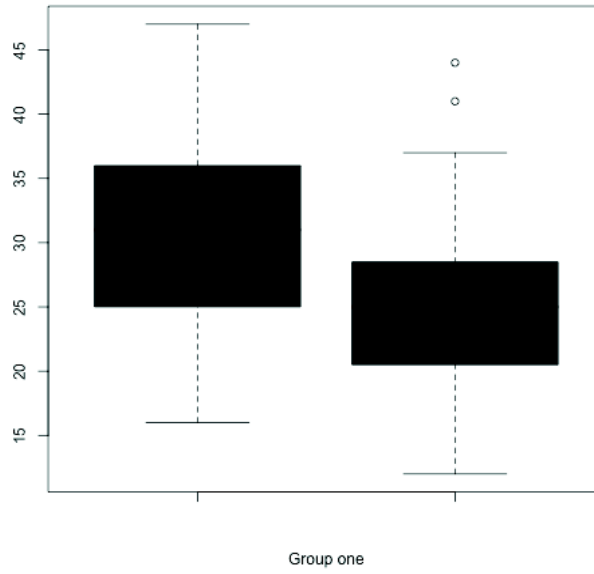
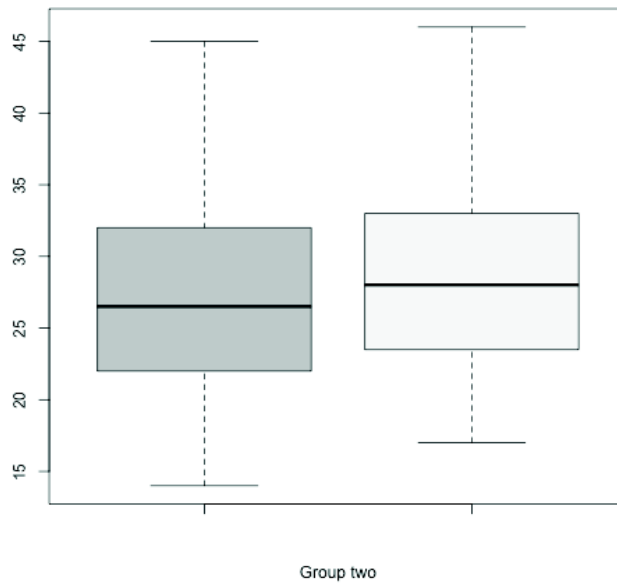Fig. 3.1 Aggression in group one pre-post treatment

Fig. 3.2 Aggression in group two pre-post treatment

	GROUP ONE	GROUP TWO
PRE-TREATMENT	30,74(7,61)	27,44(7,19)[c]
POST-TREATMENT	24,94(6,31)[a]	28,88(6.26)[c]

[a]: significant difference within groups (p < .05)
[b]: significant difference between genders in group (p< .05)
[c]: significant difference between groups (p< .05)

Tab.3. 5 General Aggression differences. Means (and SD) are given.

The main purpose of our experiment was to assess if students treated with contact sport lessons had an improvement in their behaviour. From this first investigation, reported in Table 3.5 a significant improvement (p<.05) has been found in the first group on the post treatment. On the contrary, in group two, the tests show no evidence of differences between the pre and post. The box-plots, both for group one (Fig. 3.1) and group two (Fig 3.2), agree with our theories. Testing the differences between the two groups before and after the treatments suggests how the groups are different before the starting of the experiment and also after the end. No evidences of difference between genders exist.

3.3.2 Physical Aggression

This subsection explains the measure of physical Aggression (PhysAgg). The outline of the table is the same as the previous analysis.
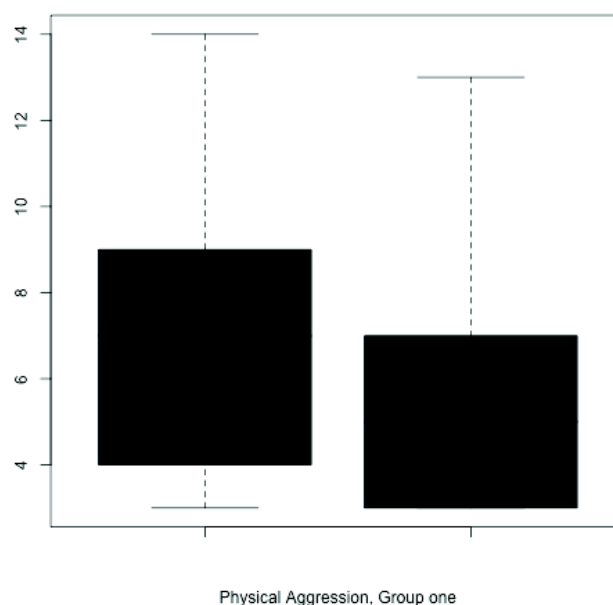

Physical Aggression, Group one

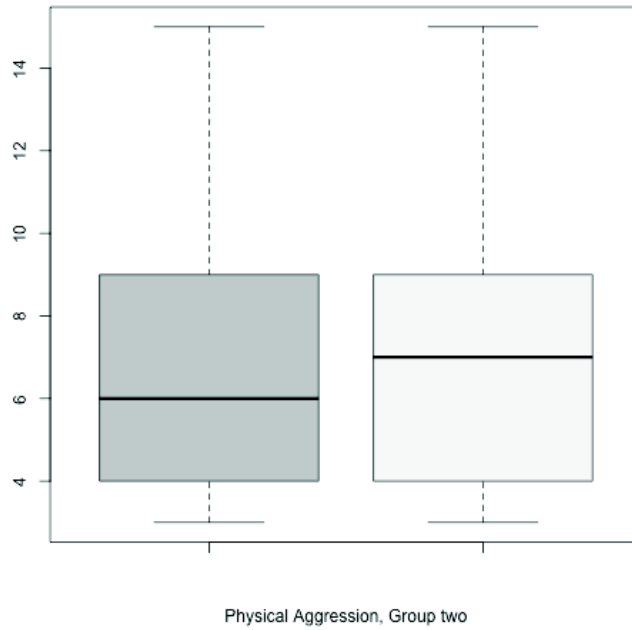Fig. 3.3 Physical Aggression in group one pre-post treatment

Fig. 3.4 Physical Aggression in group two pre-post treatment

	GROUP ONE	GROUP TWO
PRE-TREATMENT	6,91(3,10)[b]	6,54(3,27)[bc]
POST-TREATMENT	5,34(2,38)[ba]	7,13(3,26)[bc]
[a]: significant difference within groups (p < .05) [b]: significant difference between genders in each group (p< .05) [c]: significant difference between groups (p< .05)		

Tab. 3.6 Physical Aggression differences. Means(and SD) are given.

Analysing the physical aggression in groups we can observe a difference in the group one between the pre and post treatments. Again, in group two, the test rejects the hypothesis of difference suggesting an unchanging situation after the experiment. Previously (see §1.1), we explained how several studies demonstrated how females show their aggression more with verbal meanings rather then physical acts. Indeed, testing the differences between genders in each group, the statements found agreements. Yet another time, a difference between the two groups is present.
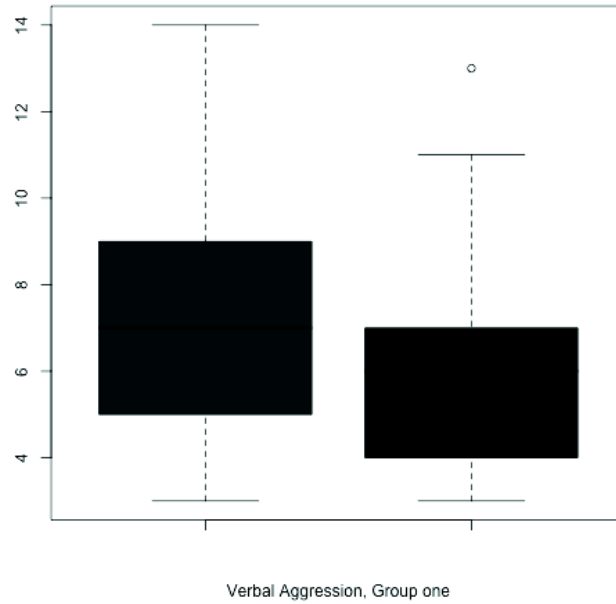
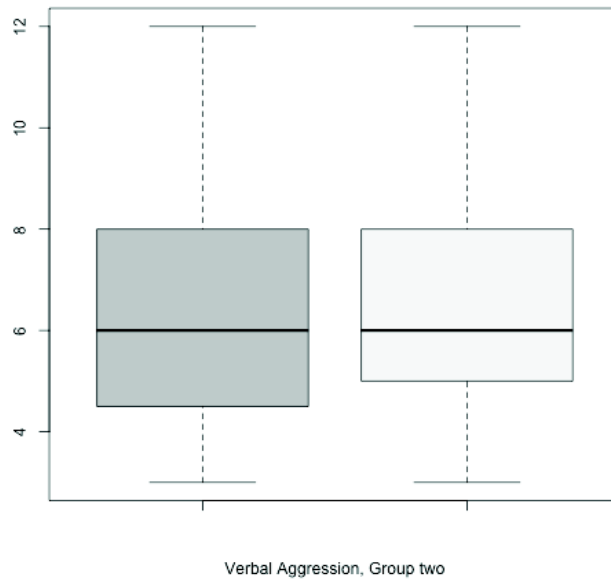### *3.3.3* **Verbal Aggression**

Verbal Aggression, Group one

Fig. 3.5 Verbal Aggression in group one pre-post treatment

Verbal Aggression, Group two

Fig. 3.6 Verbal Aggression in group two pre-post treatment

	GROUP ONE	GROUP TWO
PRE-TREATMENT	7,13(2,24)	6,06(2,10)[c]
POST-TREATMENT	5,87(1,99)[a]	6,41(2,05)[c]
[a]: significant difference within groups (p < .05) [b]: significant difference between genders in each group (p< .05) [c]: significant difference between groups (p< .05)		

Tab.3. 7 Verbal Aggression differences. Means(and SD) are given.

A part from the fact that another time a difference between the groups exist before and after the treatments, regarding verbal aggression significant improvements (p<.05) in group one after contact sport lessons are found. Group two shows no differences over time. No evidences of gender difference exist.

3.3.4 **Anger**

Fig. 3.7 Anger in group one pre-post treatment

Fig. 3.8 Anger in group two pre-post treatment

	GROUP ONE	GROUP TWO
PRE-TREATMENT	8,41(2,46)	7,13(2,70)[c]
POST-TREATMENT	6,91(2.29)[a]	7,42(2,25)[c]
[a]: significant difference within groups (p < .05) [b]: significant difference between genders in each group (p< .05) [c]: significant difference between groups (p< .05)		

Tab. 3.8 Anger differences

Same considerations made for verbal aggression can be made for anger.

3.3.5 **Hostility**

Fig. 3.9 Hostility in group one pre-post treatment

Fig. 3.10 Hostility in group two pre-post treatment

	GROUP ONE	GROUP TWO
PRE-TREATMENT	8,29(2,63)	7,69(2,71)[b]
POST-TREATMENT	6,80(2,33)[a]	7,89(2,43)[c]

[a]: significant difference within groups (p < .05)
[b]: significant difference between genders in each group (p< .05)
[c]: significant difference between groups (p< .05)

Tab.3. 9 Hostility differences. Means(and SD) are given.

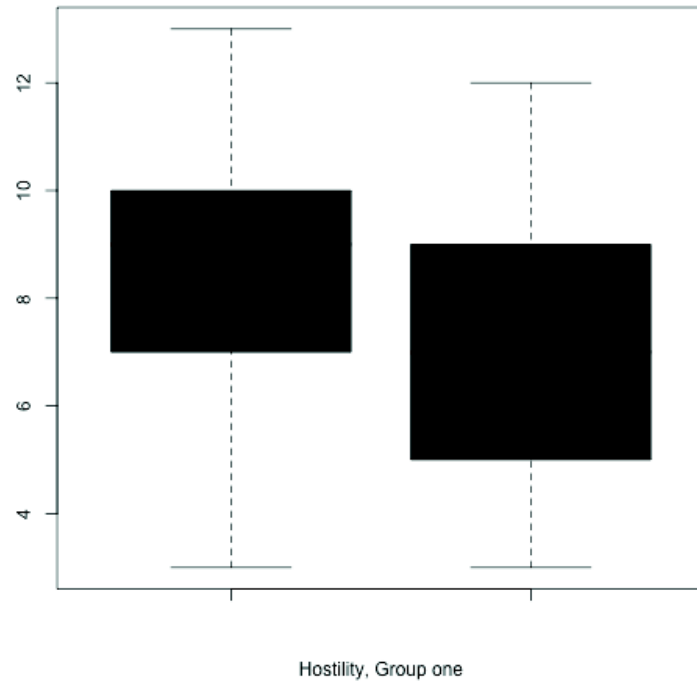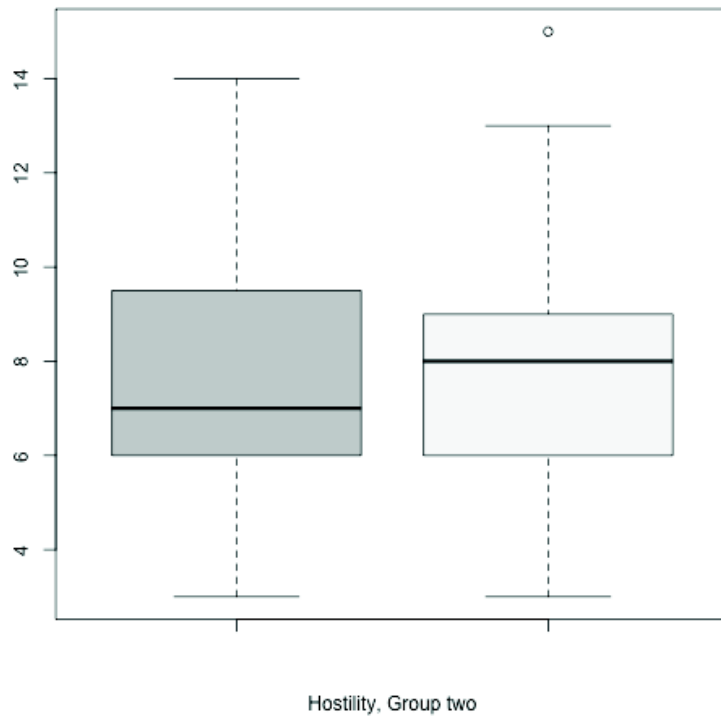Again, same considerations made for verbal aggression and anger can be made. In this case, we found a difference between the genders in group two in the pre treatment.

In several studies, especially in social science, where the normality assumption were rejected and the sample size were large enough, there's the convention to treat the data as normal distributed even when they aren't. In testing the normality in our experiment, we had that this assumption may be rejected and we decided to use both parametric and nonparametric tests and, not surprisingly, we obtained the same results. Almost the same p-values were obtained for both type of tests. This can be due to the large sample size.

3.4 Investigating the correlations

We are now investigating the correlations between the various variables interested in our experiment. The same investigation has been done by Buss & Perry (1992) and Bryant & Smith (2001) in developing the structure of their AQ's (§1.4). The same analysis can be carried out on our data. Starting with the correlation in group one before and after the treatments:

First variable	Second variable	Spearman correlation	p-value
Physical Aggression	Verbal Aggression	0,33[0,22]	0,0006103[0,0050]*
Physical Aggression	Hostility	0,27[0,18]	0,005804[0,023]*
Physical Aggression	Anger	0,34[0,27]	0,0003145[0,004]*
Verbal Aggression	Hostility	0,50[0,51]	4,57e-08[3,059e-08]*
Verbal Aggression	Anger	0,57[0,40]	1,56e-10[2,19e-05]*
Anger	Hostility	0,27[0,36]	0,005041[0,00017]*

Tab. 3.10 Correlation for Group one, pre treatment [Post Treatment].

All the variables in group one demonstrate a strong correlation (p<.05) between them, both in the pre and post treatment. Buss & Perry (1992) found that anger was strictly correlated to the other components. The reasons may be found in the fact that anger is a prelude to aggression, and is known that people are more likely to aggress when angry than when not angry (§1.4).

First variable	Second variable	Spearman correlation	p-value
Physical Aggression	Verbal Aggression	0,23[0,18]	0,014[0,05]*
Physical Aggression	Hostility	0,018[-0,09]	0,85[0,3198]
Physical Aggression	Anger	0,38[0,36]	0,000041[0,004]*
Verbal Aggression	Hostility	0,21[0,17]	0,026*[0,16]
Verbal Aggression	Anger	0,53[0,38]	5,56e-09[6,57e-05]*
Anger	Hostility	0,20[0,23]	0,05[0,016]

Tab. 3.11 Correlation for Group two, pre treatment [post treatment].

Inside group two, correlation between the components assumes a different form. Another time, anger is correlated to every other component. The hostility component seems not to be linked to any other subtrait except anger. Indeed, we can observe how the correlations between this variable and the others is not so significant (p<.05). This factor can in fact be considered as strongly connected to the individuals (Buss & Perry, 1992). Each person got different reaction to ill will or injustice; we can obtain high scores in physical or verbal aggression with either a high or low score in hostility. Anger, representing the bridge (Buss & Perry, 1992, §1.4) between the cognitive component and the instrumental components (verbal, physical aggressions), is correlated to hostility.

4 Chapter four:
Mixed Effect Linear Model

In many applications of practical interest, we may be interested on a model that evaluates the effects of treatments over time. We can estimate a normal multiple regression model on our data. In our study, however, the measurements are made on the same subjects over two different times. The normal linear model is not appropriate in this situation. A normal linear model can be expressed as

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{1p} + e_i, \quad i = 1, \ldots, n$$

$$e_i \sim N(0, s^2) \tag{2}$$

where $\beta = (\beta_1, \ldots, \beta_p)$ are the unknow parameters and $y_i = (y_1, \ldots, y_n)$ are the responses. The previous model has one random component, the error term e. Usually $x_{iJ} = 1$ for $i = 1, \ldots, n$ and β_1 is the intercept. The model can also be represented in a matrix form, as

$$y = X\beta + e$$

$$e \sim N(0, s^2 I_n) \tag{3}$$

where $y = (y_1, \ldots, y_n)$ is the response vector. X is the (*pxn*) model matrix. $\beta = (\beta_1, \ldots, \beta_p)$ is the vector of regression coefficients.

In the presence of repeated measurements, model (2) can be generalised to the "Linear Mixed-effect model" (Laird & Ware, 1982). The so-called mixed effect models include additional random-effects terms, and are often appropriate for representing clustered, and therefore dependent, data arising, for example, when data are collected hierarchically, i.e. **when observations are taken on related individuals, or when data are gathered over time on the same individuals.**

Linear mixed models can be expressed in different but equivalent forms. In the social behavioural sciences, it is common to express such model in hierarchical form, as explained below.

$$y_{ii} = \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{1p} + e_{ii} + \alpha_i$$

$$e_i \sim N(0, s^2) \tag{4}$$

$$\alpha_i \sim N(0, \Psi^2)$$

where $\beta_i = (\beta_1, \ldots, \beta_p)$, $y_i = (y_1, \ldots, y_n)$, $\alpha_i = (\alpha_1, \ldots, \alpha_x)$ are the random effects.

Suppose that repeated measurements at time t time points are obtained from s groups of subjects. Let n_h denote the number of subjects in group h (h=1,...s), and let $n = \sum n_k$. Let y_{hij} denote the response at time j from *i*th subject in group h for $h = 1,...,s$, $i = 1,...,n_k$ and $j = 1,...,t$.

We can express a model for this situation, given by

$$y_{hij} = \mu + \gamma_h + \tau_j + (\gamma\tau)_{hj} + \pi_{i(h)} + e_{hij} \qquad (5)$$

In equation (5), μ is the overall mean and γ_h is the fixed effect of group h, with $\sum \gamma_k = 0$. In addition, τ_j, is the fixed effect of time j, with $\sum \tau_j = 0$, and $(\gamma\tau)_{hj}$ is the fixed effect for the interaction of the *h*th group with the *j*th time. The constraints on the interaction parameters are

$$\sum (\gamma\tau)_{hj} = \sum (\tau\gamma)_{hj} = 0.$$

The parameters $\pi_{i(h)}$ are random effects for the *i*th subject in the *h*th group. The $\pi_{i(h)}$ are assumed to be independently normally distributed with mean zero and variance σ^2_π. Finally, the e_{ihj} parameters are independent random error terms, with $e_{hij} \sim N(0,\sigma^2_e)$.

The following table displays the sum of squares (SS) , the degrees of freedom (df), and expected mean squares(MS) for each source of variation. In the column of expected mean squares, the quantities labeled Dg, Dt and Dgt represent differences among groups, differences among time points, and the group x time interaction, respectively.

Source	Sum of squares(SS)	Degrees of freedom	E(MS)
Group	SS_G	(s-1)	$s_e^2 + ts_p^2 + D_G$
Subjects	$SS_{S(G)}$	(n-s)	$s_e^2 + ts_p^2$
Time	SS_T	(t-1)	$s_e^2 + D_T$
Group x Time	SS_{GT}	(s-1)(t-1)	$s_e^2 + D_{GT}$
Residual	SS_R	(n-s)(t-1)	$s_e^2 + D_{GT}$

The sum of squares in the table are based on the following decomposition of the deviations of each observation about the overall mean:

$$y_{hij} - \bar{y}_K = (\bar{y}_{h..} - \bar{y}_K) + (\bar{y}_{hi.} - \bar{y}_{h..}) + (\bar{y}_{..j} - \bar{y}_K) + (\bar{y}_{h.j} - \bar{y}_{h..} - \bar{y}_{..j} + \bar{y}_K) + (\bar{y}_{hij} - \bar{y}_{h.j} - \bar{y}_{hi.} + \bar{y}_{h..})$$

where

$$\bar{y}_{..} = \frac{\sum\limits_{k=1}^{s}\sum\limits_{i=1}^{n_k}\sum\limits_{j=1}^{t} y_{hij}}{nt}$$

is the overall mean,

$$\bar{y}_{k..} = \frac{\sum\limits_{i=1}^{n_k}\sum\limits_{j}^{t} y_{hij}}{n_h t}$$

is the mean for group h,

$$\bar{y}_{..j} = \frac{\sum\limits_{h=1}^{s}\sum\limits_{i=1}^{n_h} y_{hij}}{n}$$

is the mean at time j,

$$\bar{y}_{k.j} = \frac{\sum\limits_{i=1}^{n_h} y_{hij}}{n_k}$$

is the mean for group h at time j, and

$$\bar{y}_{ki.} = \frac{\sum\limits_{j=1}^{t} y_{hij}}{t}$$

is the mean for the ith subject in group h.

The sums of squares are then defined as follows:

$$SS_G = \sum_{h=1}^{s}\sum_{i=1}^{n_k}\sum_{j=1}^{t}(\bar{y}_{k.} - \bar{y}_{..})^2 = t\sum_{h=1}^{s} n_h(\bar{y}_{h..} - \bar{y}_{..})^2,$$

$$SS_{S(G)} = \sum_{h=1}^{s}\sum_{i=1}^{n_k}\sum_{j=1}^{t}(\bar{y}_{k..} - \bar{y}_{h..})^2 = t\sum_{h=1}^{s}\sum_{i=1}^{n_h}(\bar{y}_{hi.} - \bar{y}_{h..})^2,$$

$$SS_T = \sum_{h=1}^{s}\sum_{i=1}^{n_h}\sum_{j=1}^{t}(\bar{y}_{..j} - \bar{y}_{..})^2 = n\sum_{h=1}^{s}(\bar{y}_{..j} - \bar{y}_{..})^2,$$

$$SS_{GT} = \sum_{h=1}^{s}\sum_{i=1}^{n_h}\sum_{j=1}^{t}(\bar{y}_{h.j} - \bar{y}_{h..} - \bar{y}_{..j} + \bar{y}_{..})^2,$$

$$SS_R = \sum_{h=1}^{s}\sum_{i=1}^{n_k}\sum_{j=1}^{t}(\bar{y}_{hij} - \bar{y}_{h.j} - \bar{y}_{hi.} + \bar{y}_{h..})^2.$$

Note that SS_G, SS_T, and SS_{GT} are equal to the sums of squares from a two-factor ANOVA model (assuming that all the *nt* observations are independent) with effect for groups, time and group x time interaction. The residual sum of squares SSR is due to the subject effect nested within the cross classification of group x time. The F-statistic for testing for differences among groups is given by:

$$F = \frac{MS_G}{MS_{S(G)}} = \frac{SS_G/(s-1)}{SS_{S(G)}/(n-s)}$$

with *s-1* and *n-s* df. This test requires the assumption that the within-group covariance are equal. In general, this assumption is required for all tests of between-subjects effects.

The F-statistic for testing differences among time points is given by:

$$F = \frac{MS_T}{MS_R} = \frac{SS_T/(t-1)}{SS_R/(n-s)(t-1)}$$

with *t-1* and *(n-s)(t-1)* degrees of freedom.

Similarly the F-statistic for testing the significance of the group x time interaction is given by:

$$F = \frac{MS_{GT}}{MS_R} = \frac{SS_{GT}/[(s-1)(t-1)]}{SS_R/[(n-s)(t-1)]}$$

with *(s-1)(t-1)* and *(n-s)(t-1)* degrees of freedom. Both of these test require the assumption that the within-group covariance are equal and that the sphericity (Mauchly,1940) condition is satisfied.

This type of ANOVA, called "repeated measures ANOVA" is very suitable for our problem. A standard ANOVA, in fact, is not applicable since we don't have two independent samples.

4.1 Mixed effect linear model using "R"

In R the linear mixed effect model is estimated by using the following function:

```
lme(response ~ x*y, random= ~(x or y)|Subj)
```

where x is the time and y the group.

A full explanation of this command can be found using the command help(lme) in "R.11.0".

For our purpose, the *x* variable refers to the time, the *y* to the groups and the *subj* to the single subject.

The casual effect in the model is due to subject and time and can expressed by the terms:
```
random=~time|subject.
```

Before estimating the model, we had to modify the dataset by inserting the variables time and subj. Time, referred to the time we made the measure (0=pre-treatment, 1=post-treatment) and subj, referred to the single subject (1 to 207). The remaining measures are still the same used as before (AQ_tot, PhysAgg, VerbAgg, Anger, Hostility)

The dataset assumed this form:

	Subj	AQ	PhysAgg	VerbAgg	Anger	Hostility	Time	Group
1	1	16	3	4	5	4	0	1
2	2	16	3	3	4	6	0	1
3	3	17	3	6	5	3	0	1
4	4	17	5	4	3	5	0	1
5	5	17	3	3	5	6	0	1
6	6	17	3	4	6	4	0	1
7	7	18	4	4	4	6	0	1
8	8	18	3	5	6	4	0	1
9	9	19	7	4	4	4	0	1
10	10	19	3	3	7	6	0	1

The estimated model for the general measure of aggression is:

```
fit<-lme(AQ~Time*Group,random=~Time|Subj)

> summary(fit)
Linear mixed-effects model fit by REML
Data: NULL
AIC BIC logLik
2593.762 2625.891 -1288.881
Random effects:
Formula: ~Time | Subj
Structure: General positive-definite, Log-Cholesky parametrization
StdDev Corr
(Intercept) 7.038265 (Intr)
Time 3.576229 -0.556
Residual 2.290990
Fixed effects: AQ ~ Time * Group
```

```
              Value      Std.Error  DF  t-value   p-value
(Intercept) 34.05284   1.6292301 205 20.901184 0.0000
Time        -13.05396   1.0621900 205 -12.289664 0.0000
Group        -3.30527   1.0289255 205 -3.212346 0.0015
Time:Group   7.24813   0.6708165 205 10.804942 0.0000
Correlation:
(Intr) Time Group
Time -0.539
Group -0.949 0.511
Time:Group 0.511 -0.949 -0.539
Standardized Within-Group Residuals:
Min Q1 Med Q3 Max
-2.94965809 -0.27504805 -0.01421883 0.27035928 2.64744281
Number of Observations: 414

Number of Groups: 207
```

Obtaining a repeated measure ANOVA is relatively easy. First, we need to fit a mixed effect model using `lme(...)`, then, by using the function `aov(fittedmodel)` we obtain the ANOVA table with every measure we need (sum of squares, F-test, degrees of freedom ecc.)

The "repeated measure ANOVA" can be applied to our data; we start with general aggression measurement (tab. 4.2)

	Df	Sum of Sq	Mean Sq	F-Value	Pr(>F)
Time	1	484,08	484,08	10,28	0,001454*
Group	1	10,5	10,52	0,22	0,63
Time:Group	1	1359,3	1359,32	28,81	1,338e-07*
Residuals	410	19343,4	47,18		

Tab 4.2 Repeated Measure ANOVA for general aggression (AQ_tot)

	Df	Sum of Sq	Mean Sq	F-Value	Pr(>F)
Time	1	23,2	23,2	2,54	0,11
Group	1	53,6	53,6	5,89	0,01*
Time:Group	1	121,7	121,71	13,36	0,00002892*
Residuals	410	3732,9	9,11		

Tab 4.3 Repeated Measure ANOVA for Physical Aggression (PhysAgg)

	Df	Sum of Sq	Mean Sq	F-Value	Pr(>F)
Time	1	21,34	21,34	4,83	0,0285*
Group	1	7,24	7,24	1,63	0,2
Time:Group	1	66,93	66,92	15,15	0,00001159*
Residuals	410	1811,21	4,42		

Tab 4.4 Repeated Measure ANOVA for Verbal Aggression (VerbAgg)

	Df	Sum of Sq	Mean Sq	F-Value	Pr(>F)
Time	1	37,14	37,14	6,24	0,0128*
Group	1	15,05	15,05	2,53	0,11
Time:Group	1	82,31	82,31	13,85	0,00002255*
Residuals	410	2436,59	5,94		

Tab 4.5 Repeated Measure ANOVA for Anger (Anger)

	Df	Sum of Sq	Mean Sq	F-Value	Pr(>F)
Time	1	42,09	42,09	6,54	0,0107*
Group	1	6,2	6,2	0,96	0,32
Time:Group	1	73,67	73,66	11,45	0,00007828*
Residuals	410	2637,37	6,43		

Tab 4.6 Repeated Measure ANOVA for Hostility (Hostility)

The analysis of variances confirms what we found with the difference tests in the previous Chapter (§3.3). Table 4.2 shows the ANOVA for the general measure aggression; it suggests how time and the interaction between group and time (the treatments for each group) influence our investigation (p<.05). It also explains that groups seem not to have any effect. Recalling that we found that the groups before the treatment were different (§3.4), this analysis tell us that this difference does not imply any consequence on our experiment. In fact, F test states that groups do not have any effect in our experiment. The other tables confirm what we have previously stated. In every subtrait of aggression the effect of treatments and groups is very strong; it confirms us how the treatments worked differently for each group.

CONCLUSIONS

The purpose of our experiment was to determine if students given of contact sport lessons within school had an improvement in term of aggression. By using the Aggression Questionnaire (Bryant & Smith, 2001, §1.5), the aggression was measured before and after this treatment. The data were then analysed (§3.1). First, we noticed how a starting difference between the two groups existed; this difference could led us to a misunderstanding of the final results, but, the analysis of variance (Chap. 4) told us how groups had no effect on the investigation. Using statistical tests for differences (§ 3.3) we found how the first group of students had significant improvements after the treatment and, on the contrary, how group two had the situation unchanged. As a further investigation, we investigate differences between genders, finding that in physical aggression males and females show strong differences.

In the final section (Chap. 4) we estimated a model that could fit our data and test the effect of groups, treatments and their interaction with the "repeated measure ANOVA". This final analysis confirms what we'd found in testing the differences.

To wind up it can be stated that the contact sport lessons had an effect on students behaviours. Making some general considerations on experiment, we could say that these results can hardly be taken in a long-term view. In fact, the data were taken right after the treatment and, to assess the real effects, it should be taken again after a longer time and compared to the previous analysis.

The used questionnaire (in italian):

ISTRUZIONI:

Valuta in che misura il comportamento indicato nelle seguenti espressioni è vero o falso per te, mettendo una crocetta su uno dei punteggi a lato (da 1 a 5)

Genere: M ☐ F ☐ Classe: Data di nascita: mese…. anno…. Data rilevazione:	ASSOLUTAMENTE FALSO PER ME	ABBASTANZA FALSO PER ME	NE' VERO NE' FALSO	ABBASTANZA VERO PER ME	ASSOLUTAMENTE VERO PER ME
1. Spesso mi trovo in disaccordo con gli altri	1	2	3	4	5
2. A volte mi sembra di aver subito un trattamento ingiusto dalla vita	1	2	3	4	5
3. Ho minacciato delle persone che conosco	1	2	3	4	5
4. Mi chiedo perché talvolta mi sento così amareggiato per qualcosa	1	2	3	4	5
5. Ho difficoltà a controllare la mia rabbia	1	2	3	4	5
6. I miei amici dicono che sono piuttosto polemico	1	2	3	4	5
7. Mi arrabbio facilmente ma mi calmo in fretta	1	2	3	4	5
8. Se sono provocato posso picchiare qualcuno	1	2	3	4	5
9. Non riesco a trattenermi dal litigare quando qualcuno non è d'accordo con me	1	2	3	4	5
10. Mi sembra che agli altri capitino sempre delle opportunità	1	2	3	4	5
11. Ci sono persone che mi hanno provocato ad un punto tale da arrivare alle mani	1	2	3	4	5
12. A volte perdo le staffe senza una buona ragione	1	2	3	4	5

BIBLIOGRAPHY

Arnold H.Buss and Mark Perry (1991), Personality and Individual Differences, The Aggression Questionnaire, University of Texas Austin

Bland Martin (2000), An Introduction to Medical Statistic, Oxford University Press

Domenico Piccolo (2000), Statistica, Il Mulino

Fred B. Bryant and Bruce D. Smith (2001), Refining the Architecture of Aggression: A Measurement Model for the Buss_Perry Aggression Questionnaire, University of Texas Austin

Lee J. Cronbach (1951), Coefficent alpha and the internal structure of tests, Psychometrika, Vol. 16, n° 3 (pg 297-334)

Nan M. Laird & James H. Ware (1982), Random-effect models for longitudinal data, Harvard School of Public Health, Biometrics 38

P. Armitage, G.Berry, J.N.S. Matthews (1994), Statistical Methods in Medical Research, ,Fourth edition, BlackWell Publishing.