CHAPTER 4

DATA DESCRIPTION AND ANALYSIS

4.1 INTRODUCTION

The evidence presented to us indicates that the implementation of interventions should have a positive effect on the number of line-breaks that a team achieves in a specific rugby match or season. However, it is important to give alternative support to the analysis through the use of statistical methods. The author will start off by giving some descriptive statistics on the data to form an initial expectation on the relationship between the specific variables. Thereafter, the author will specifically look at the 2002 season and investigate what specific running lines had the most significant impact on the total number of linebreaks achieved during the course of the season. This is followed by an analysis on which interventions are used in creating these specific running lines in order to assess the level of effectiveness of the interventions. Finally, the author will perform a hypothesis test to indicate whether or not the average linebreaks made during the 2002 season were significantly higher than the average linebreaks made during the 2001 season.

4.2 SOURCES OF DATA AND THE DATA SAMPLE

Analysing video types of matches played by the Blue Bulls U21 team during the 2001 and 2002 seasons obtained the data used in the analysis. A total of nine and ten games were played in the 2001 and 2002 seasons respectively. However, because of data problems only nine of the ten games in the 2002 season were considered in the analysis. For each game played in the 2001 and 2002 seasons the total number of linebreaks achieved in a match was calculated. In addition the total number of linebreaks achieved in the 2002 season was further subdivided into the specific categories of intervention in order to determine which intervention had the biggest impact on the total number of linebreaks achieved.

Our sample therefore consists of nine matches played in each of the 2001 and 2002 seasons, where the data of the 2002 season is further divided into relative subgroups.

4.3 DESCRIPTIVE STATISTICS

In order to form an initial expectation on the impact of the interventions we will view some descriptive statistics regarding the data. The author will start off by analysing the total linebreaks made during the 2001 and 2002 season when playing against various opposition teams. These values are given in Table 2.

Table 2: Comparisons of the total number of linebreaks made in the 2001 and 2002 seasons

2001		**2002**	
Opposition	**Total linebreaks**	**Opposition**	**Total linebreaks**
NW	11	**BB A**	19
Pumas	25	**Border**	36
Falcons	12	**Falcons**	30
Lions	9	**Lions**	12
WP	12	**WP**	14
Boland	9	**Natal**	21
Natal	7	**FS**	20
FS	6	**WP**	43
WP	6	**Lions**	29
Total	97	**Total**	225

From Table 2 it is clear that the total number of linebreaks achieved during the 2002 season is much higher than the total number achieved during the 2001 season. Without exception a comparison between similar teams played during both seasons indicates that the total number of linebreaks achieved during the 2002 season is much higher than when the team competed against the same opposition during the 2001 season. It seems as if the aggregate numbers indicate a significant increase in linebreaks from the 2001 to 2002 season.

Table 2 has clearly indicated that there is a significant improvement in the number of linebreaks made from the 2001 to 2002 seasons. The question arises therefore what has caused this rapid increase in the number of linebreaks being made? To answer this question, we proceed to take a closer look at what factors played the determining role in the linebreaks being made in the 2002 season.

In Table 3 the total number of line-breaks made is divided into a specific type of running line. The major components of total linebreaks are the "Overs" and "Unders" running lines with respectively 39% and 30% of total linebreaks being made up by these components. Together they make up more than 65% of total linebreaks, which indicates that they are a very important component of linebreaks being made in a match or season.

Table 3: Type of linebreak as a percentage of total linebreaks being made during the 2002 Bankfin U21 competition

Type of linebreak	Number of accuracy	As percentage of total
"Overs"	88	39%
"Unders"	67	30%
"1,1"	27	12%
"Happy"	14	6%
"X"	12	5%
"Shark"	5	2%
"ACT"	7	3%
"Slap chips"	3	1%
"DSP"	1	0.4%
"OI" Strike	1	0.4%

Although the "1,1", "Happy" and "X" running lines have also made significant contributions towards the total linebreaks being made, they did not play such a dominating role as when compared with the previous two variables of "overs" and "unders" running lines. The remaining five variables had even less of a significant impact. This shows that the increased number of linebreaks experienced within the 2002 season was dominated by only a few variables.

Although it may seem as if these variables are the most effective, it may very well be that the other less significant variables were indeed the foundation or building blocks from which the more significant variables have developed. It could be the case that the development of the "less important" variables can lead to a huge beneficial increase in linebreaks experienced. Although the aim of this study is not to see whether this is the case, it may indeed be fruitful to engage in such further research activities.

A further analysis was done in order to determine which intervention had the most significant impact on the various running lines. Table 4 lists the specific running lines and their respective or underlining interventions. Column one shows the specific linebreak variable while column two presents the specific intervention that played the major role in the success of the variable in column one. The figure presented in parenthesis in column two indicates the percentage of times the specific intervention variable was partially responsible for the success of the linebreak.

It is clear from Table 4 that a variety of interventions were responsible for the success of the linebreaks. This is an indication that the development of the interventions would be beneficial for all running lines, which shows their significant importance in achieving linebreaks.

Table 4: Most significant interventions in determining specific linebreaks during the 2002 season

Running Line	Most significant intervention
"Overs"	Speed (83%) Decoy inside (42%)
"Unders"	Momentum advantage (67%) Power (60%) Expanded attack (42%)
"1,1"	Power (63%) Momentum advantage (63%) Change in initial starting position (33%)
"Happy"	Speed (79%) Change in speed of movement (71%) Decoy outside (57%)
"X"	Speed (84%) Expanded attack (67%)
"ACT"	Expanded attack (100%) Decoy outside (100%) Decoy inside (100%)
"Shark'	Expanded attack (100%) Decoy outside (100%) Decoy inside (100%)
"Slap chips'	N.A
"DSP"	N.A.
"O,I" Strike	N.A.

It is clear that the total number of linebreaks achieved during the 2002 season was mainly determined by specific running lines, which in turn were based on multiple interventions that were brought into the preparation or training sessions of the 2002 season. It can therefore be concluded that the major determinants of the increased linebreaks experienced in the 2002 season were indeed affected by the intervention, which was implemented into the specific running lines.

4.4 HYPOTHESIS TESTING

The descriptive data analysis has clearly indicated that interventions were the major cause of linebreaks experienced in the 2002 season. According to Table 2 it seems as if there is a significant difference between the number of linebreaks achieved in the 2002 season and the previous season due to the implementation of specific running lines and intervention. However, it is important to statistically prove that there is a significant difference between the line-breaks achieved in the two periods.

The author will therefore perform a hypothesis test to indicate that on average the linebreaks achieved in the 2002 season is statistically greater than that achieved in the 2001 season.

Hypothesis testing regarding averages can be divided into two main groups. Inference regarding the average of a single sample and inference on the averages of two or more samples. The second group of inference would be the one in which the most interest will be shown. Given the fact that two separate teams with different players have participated in the two seasons, the assumption will be made that the two samples are independent from each other, therefore the linebreaks achieved in the 2001 season will have no correlation with the linebreaks achieve in the 2002 season.

The hypothesis test regarding the averages of two independent samples is a simple T-test, however, there is an important distinction to be made between two samples for which the variance (σ_i^2) is equal and two samples for which the variances (σ_i^2) is not

equal. Therefore, an applied F-test will be done to determine whether the two samples have equal variances or not.

The null and alternative hypotheses to test for equal variances are presented in equation 4.1. Under the null hypothesis we assume that the variances of the two samples are equal, while the alternative states that the two samples have different variances. The test statistic is presented in equation 4.2. A value for the test statistic that is greater than the critical value will lead to a rejection of the null hypothesis.

$$\begin{aligned} H_0 &: \sigma_1^2 = \sigma_2^2 \\ H_A &: \sigma_1^2 \neq \sigma_2^2 \end{aligned} \tag{4.1}$$

$$F = \frac{S_1^2}{S_2^2} \sim F_{(n_1-1, n_2-1)} \tag{4.2}$$

where S_1^2 and S_2^2 represent the two sample variances.

The test statistic was calculated as in equation 4.3 and evaluated against the $F_{(8,8)} = 2.59$ critical value on a 5% level of significance.

$$F = \left(\frac{103.5}{6.5} \right) = 15.921 \tag{4.3}$$

The value of 15.921 is greater than the critical value of 2.59 and we can therefore not accept the null hypothesis, concluding that the two samples do not have equal variances. We can now proceed and test whether the 2002 average linebreaks are significantly higher than the average linebreaks achieved in the 2001 season.

The null and alternative hypotheses are given in equation 4.4. Under the null hypothesis the two sample averages are equal. Under the alternative the 2002 average is higher than the 2001 average.

$$H_0 : \mu_{2002} = \mu_{2001}$$
$$H_A : \mu_{2002} > \mu_{2001} \tag{4.4}$$

The appropriate test statistic is given by equation 4.5. In contrast to normal T-tests, this specific test is a one-sided upper or right hand test due to the fact that we are testing whether the one average is greater and not equal to the other. Therefore, we would only reject the null hypothesis of equal sample averages if the test statistic were greater than the appropriate critical value.

$$T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_{\alpha,\nu} \tag{4.5}$$

where: $\nu = \dfrac{(S_1^2/n\ + S_2^2/n_2)^2}{\dfrac{(S_1^2/n_1)^2}{n_1 - 1} + \dfrac{(S_2^2/n_2)^2}{n_2 - 1}}$

The calculated test statistic is given in equation 4.6 and was evaluated against the $t_{0.05,9} = 1.833$ critical value.

$$T = \frac{(25 - 9.33)}{\sqrt{\frac{10.17349^2}{9} + \frac{2.54951^2}{9}}} = 4.4827 \tag{4.6}$$

Once again we cannot accept the null hypothesis. Therefore we can conclude that the average of the total linebreaks made during the 2002 season is statistically greater than the average of the total linebreaks made during the 2001 season.

4.5 CONCLUSION

The objective of this chapter was to assess whether the implementation of the intervention had a significant positive effect on the total number of linebreaks made during the 2002 season. To reach a conclusion the first step was to compare the number of linebreaks made in various matches played in both the 2001 and 2002 seasons. From these figures it was immediately apparent that the number of linebreaks made in the 2002 season was significantly higher than the linebreaks made for matches played in the 2001 season. The following step was to look at the factors that had the most significant impact on the increased linebreaks made during the 2002 season and from these observations it was concluded that specific running lines based on the newly implemented interventions were the major contributor towards the increased linebreaks.

The final aspect of the evaluation was to end off by doing a hypothesis test to see whether the average linebreaks in the 2002 is statistically greater than the average linebreaks made in the 2001 season. The results of the hypothesis test did indeed show that statistically the 2002 number of linebreaks were higher than the 2001 number of linebreaks.

It can therefore be concluded that the implementation of the interventions through various running lines did indeed lead to a significant increase in the total number of linebreaks made.