STATISTICS

Dalarna University

One year Master's Thesis 2008

Can young skiers perform well both in sprint and endurance races?

Author: Qi CAO

Registered Nr: 831211 6893

Supervisor: Kenneth Carling

Date: 2008-06-12

Abstract

The skiers participate in two races in Falu Winter Games which are held on Saturday and Sunday. The race on Saturday is an individual race while on Sunday it is a relay race (team sprint). The races are divided into different classes by age and gender.

The individual race requires endurance whereas the team sprint requires strength and speed. This thesis is to determine whether the young skiers can perform well both in the endurance race (Saturday) and in the sprint race (Sunday).

I will use linear model analysis and apply the method on data from the individual races and the team sprints collected in February 2008.

I conclude that the relationship between young skiers' performance in endurance and sprint races is not so strong and if they want to perform well in both races, they need more practice to improve some special abilities.

Key words:

Falu Winter Games, Endurance race, Team sprint

List of Contents

1 Introduction

1.1 Cross-Country Skiing in Dalarna

Cross-Country Skiing (also known as XC skiing) is a winter sport popular in many countries with large snowfields, primarily Northern Europe, Canada and Alaska. The popularity of the sport has been quickly growing in the United States and in Australia. [1]

As one of the major countries in Northern Europe, the Cross-Country Skiing is also very popular in Sweden. Visit Sweden states that Sweden's mountain world has an unlimited range to offer all skiing enthusiasts, "Taking a trip" for Swedes means travelling on your Cross-Country Skiing from point a to b – and much more.

Obviously, Sweden has a very good environment to unfold Cross-Country Skiing, which is also fine in the province of Dalarna.

There are many Cross-Country Skiing tracks in Dalarna such as Orsa Grönklitt, Mora Parken, Gyllbergen in Borlänge and Sörskog in Falun. And one of the most famous races in Cross-Country Skiing is traditionally held in Dalarna—Vasaloppet.

Vasaloppet runs from Sälen to Mora in Dalarna and it is a long distance cross-country ski race held annually on the first Sunday of March. Wikipedia Vasaloppet states that it is the oldest, one of the longest, and in popular consideration the biggest cross-country ski race in the world. A total of more than 40,000 participate in one of the seven different races held during the first week of March. The race was started in 1922 being inspired by the run that the King to be Gustav Vasa made in 1520.

1.2 About Falu Winter Games[1]

The Falu Winter Games consists of two races which are held on Saturday and Sunday and is open for skiers in the age of 7-20 years. The competition on Saturday is a traditional one in classical style. All competitors are grouped according to their age. Skiers of the same age race the same distance and they start individually and their race time is clocked. The winner is the skier with the fastest time. The competitions

[1] More details can be found in the appendix 1 which is shown in the last part of this thesis

on Sunday are relay races which are divided into different groups by age-interval and gender. In every age group, there are several teams varying from 10 to 30. In each team there are two team members who race several laps on the race track. Each member needs to ski 3 laps, which may be 400 or 600 meters depending on age group. 400 meters distance for the skiers age 9-12, skiers whose age is greater than 12, the race distance is 600 meters.

Before the formal relay race, there is a prerun race which means the members in the team ski one lap each with individual starts. The starting position in the following relay race is arranged according to the resulting rank of each team in the prerun. [4]

Generally speaking, some competitors take part only in Saturday's endurance race, while other competitors take part only in Sunday's team sprint race. However, it is fairly easy to find competitors who participate in both races.

Table 1.1 Examples of female skiers participating in both races

Name	Gender	Club	Age	Distance
Ella Halvarsson	Female	Stora Tuna IK SK	9	2000
Sofia Eriksson	Female	Stora Tuna IK SK	9	2000
Jenny Nilsson	Female	Töcksfors IF	9	2000
Amanda Eriksson	Female	Töcksfors IF	9	2000
Kristina Axelsson	Female	Sya SK	9	2000

Name	Gender	Club	Age	Prerun Distance
Ella Halvarsson	Female	Stora Tuna IK SK	9	400
Sofia Eriksson	Female	Stora Tuna IK SK	9	400
Jenny Nilsson	Female	Töcksfors IF	9	400
Amanda Eriksson	Female	Töcksfors IF	9	400
Kristina Axelsson	Female	Sya SK	9	400

Source: http://www.racetimer.se/sv/race/show/99?layout=marathon

So some interesting problems appear, for instance: What are their ranking in these two different types of races? Actually, the main factor which is taken to differentiate the race between Saturday’s and Sunday’s is the distance. Long distance asks for endurance and short distance asks for speed. In short, it is useful to find out some general guide lines about the endurance and speed of youth when doing cross-country skiing.

1.3 The endurance and speed of young skiers

Endurance can be defined quite simply as one's ability to withstand fatigue or the ability to control the functional aptitude of movement in lieu of external stress. [5]. A Reader’s Digest Condensed History of Cross-Country Skiing states that “of all the skiing disciplines, it is cross-country only that is practiced by "endurance athletes."”. Thus, young Cross-Country skiers can be similar treated as young endurance athletes, in general.

The objective of endurance training is to develop the energy production systems to meet the demands of the event. [7] The energy production systems, however, contain some specific parts which relate to adenosine triphosphate (ATP) —the chemical compound that supplies energy for muscular contraction. The endurance includes four types: Aerobic Endurance, Anaerobic endurance, Speed endurance and Strength endurance. Generally speaking, the heavier a young athlete is, the less endurance he (she) is likely to be.

Speed is the quickness of movement of a limb, whether this is the legs of a runner or the arm of the shot putter. Speed is an integral part of every sport and can be expressed as any one of, or combination of, the following: maximum speed, elastic strength (power) and speed endurance. Speed is influenced by the athlete's mobility, special strength, strength endurance and technique. [8]

All in all, there are two factors which both influence on endurance and speed of young athletes, as well as young skiers: Speed endurance and Strength endurance. It is accessible that if the young skiers train well both in Speed endurance and Strength

endurance, they can perform well both in endurance race and in sprint race. In another word, as a matter of experience training method study, young skiers can have well performance both in sprint and in endurance race.

1.4 Aims of this thesis

My aim is to use the race result from the year 2008 of Falu Winter Games and build statistical models to find the relationship between the skiers' performance in endurance race and the skiers' performance in sprint race. Then, at the final part of this thesis, I will give some recommendations about not only the skiers, but also the organization of Falu Winter Games.

2 Data

2.1 Selected sample

Data are used to answer questions, to make decisions, and to gain a deeper understanding of some phenomena [9]. As a thesis of statistics, data analyzing part is very important and need be paid lots of attention to.

Specifically to this thesis, the data background is the result of Falu Winter Games which was held in Lugnet, Falun, Sweden on Feb 9th and Feb 10th 2008. I focus on the question "Can young skiers perform well both in sprint and endurance races?" So I select skiers who take part not only in the endurance race (Saturday's race) but also in the sprint race (Sunday's race). There are total of 313 skiers belong to this group of which 47.3% are female and 52.7% are male. These young skiers all belong to the age group 9 to 21.[2]

Table 2.1: The frequency of skiers from different age groups

Age Group	Gender	Counts
9~12	Female	79
9~12	Male	77
13~21	Female	69
13~21	Male	88

Data Source: www.racetimer.se

From table 2.1, it can be clearly seen that if one divides the age groups in two parts (9~12, 13~21), the number of skiers who are under 13 years old are similar as those from 13 to 21 years old.

2.2 Explanation of variables

As I have mentioned before, I just pay attention to the skiers who take part not only in

[2] 8 years old skiers also participate in the race, but consider about their age and the aim for them is just for fun, so their result are not read and calculated. Hence, they are excluded from the sample.

the Saturday's race but also in Sunday's race. To reach the convenient to analyze, it is useful to put these data into two separate sets (like different sheets in Excel). The first is shown in Table 2.2 and it interprets the Saturday's race.

Table 2.2: Example of records in Saturday's dataset

Number	Name	Gender	Club	Age	Distance	Time	Velocity	Ranking	TR
92	Sara	0	Mora	9	2000m	831s	2.4m/s	18	1.1948
255	Erik	1	Falu	11	2000m	661s	3.0m/s	23	1.1573
556	Linnea	0	Mån	17	6000m	1279s	4.7m/s	14	1
682	Martin	1	Bore	19	9000m	1522s.	5.9m/s	1	0.9223

Data source: www.racetimer.se

There are some explanations to go with this table Under the column "Gender", a dummy variable appears where "0" denotes when the skier is a girl and "1" denotes when the skier is a boy. .The variable "Time" is the skiers' racing time converted into seconds. "Velocity" is simply the ratio between Distance and Time. The variable "Ranking" refers to the resulting position of the skier with in her racing class. The skiers which are from different Gender and Age groups have different distance to complete in Saturday's race; and the distances are tabulated in Table 2.3.

Table 2.3 Distance in Saturday's race D refers to girls and H to boys

Distance	Gender and Age groups
2000m	D9-12, H9-12
3000m	D13-14, H13-14
5000m	D15, H15
6000m	D16-20, H16
9000m	D21, H17-20
15000m	H21

The column TR stands for Time Ratio which is the ratio between the skier and the median time of Saturday's race. For example, if the total number of participant skiers in the group was 31, the median time is the skier's time who reached number 16. The function of TR is shown below:

TR=Tp/Tm

T_m denotes the time of the median position skier and T_p denotes the time of the observed skier. When the skier ranks before the median position, the value of TR is smaller than 1 because $T_m > T_p$, and vice versa. The higher position the skier have, the smaller value his (her) TR is. If the observed skier is also the median position skier, his (her) TR becomes 1 obviously which is the condition of the third skier in table 2.2.

Table 2.4: Examples of records in Sunday's dataset (the same selected skier as Table 2.2)

Number	Name	Gender	Club	Age	Distance	Time 1	Time 2	SP
25	Sara	0	Mora	9	0	88.7	87.2	0
135	Erik	1	Falu	11	0	74.4	73.7	0
220	Linnea	0	Mån	17	1	81.1	78.2	0
256	Martin	1	Bore	19	1	70.3	65.9	0

Data source: www.racetimer.se

Table 2.4 shows the set from Sunday's race (the same four skiers as in Table 2.2 are shown as an example).

The skier's number is different from the number in Table 2.2 because this number is the number of the sprint team and the same number is used to the skier's teammate. The number is the same in both the prerun and final race.

The distance here is a dummy variable here where "0" means that the skier does the distance 400m in the prerun and 2*3*400m in the final race, whereas "1" means that the distance is 600m in the prerun and 2*3*600m in the final race.

"Time 1" denotes the time of skier's prerun time and Time 2 denotes the skier's time in the final race. It can be noted that Time 2 is smaller than Time 1 for these four skiers which might be a consequence of the prerun where, the skier appears alone on the track and hence it means that he (she) lacks of competition with other skiers. While, when the final race was held, the skier has more feeling about risk of incidents and competes with others, so he (she) is stimulated to speed up.

SP stands for Starting Position of the skier. It determines if the skier's position in Sunday's race is the first or the second. "0" denotes that the skier is in the first position and thereby would run the first, third and fifth lap in the final race. "1" denotes the skier is in the second position and would run the second, fourth and sixth lap.

Figure 2.1: A stretch of the track and reading points at the Sunday's race.

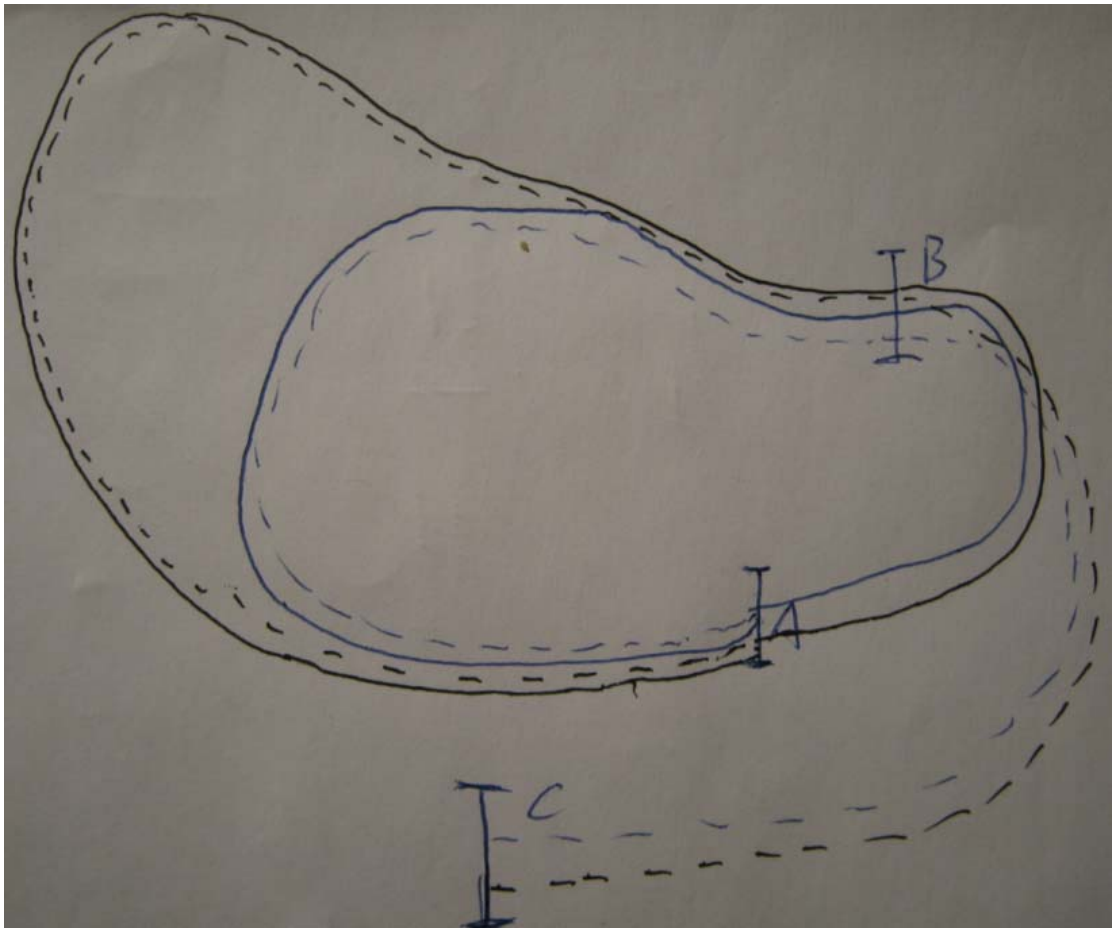

Figure 2.1 shows the skier's prerun map of the two laps which is also the first and the sixth lap in the final race. In my thesis, I am concerned about if the skier can perform well both in endurance race and sprint race. Actually, it is unnecessary to put every lap's time into the dataset of Sunday because the key point is just the short distance. Under this condition, even the existences of teams of Sunday are no emulated but the individual's performance.

From Figure 2.1, the big circle denotes the distance of skier whose age is from 13 to 21. The small circle, at the same time, denotes the distance of skier whose age is from 9 to 12. A is the starting point and B is time reading point. The first skier of the team (the skier belongs to "1" under SP column of Table 2.4) skies from A to B and then has one time recorded which is the time —Time 1. And then at A the first skier change to the second skier of the team (the skier belongs to "2" under SP column of Table 2.4) who skies from A to the final point C. However, the time 1 to the second

skier is measured from point B to C which is also the time under Time 2 column and apparently this distance (from B to C) is the same between the skiers from 9-12 and the skiers from 13-21.

2.3 Handling non-response in Sunday's race

In many censuses and sample surveys, some of the units contacted do not respond to at least some items being asked. Such non-response, which we will call survey non-response whether it arises from a census or a sample survey, is common in practice whenever the population consists of units such as individual people, households, or businesses. [10]

In Sunday's sprint race, because the skiers' velocity is high and many competitors race simultaneously, it is difficult for the time reader to read every skier's time. Under this condition, non-response of time might appear.

Table 2.5: Example of three kinds of non-response in the Sunday's race

Number	Name	Gender	Club	Age	Distance	Time 1	Time 2	SP
21	Ida	0	Töck	9	0	75	missing	0
48	Jacob	1	Två	10	2[3]	missing	53.4	1
120	Victor	0	Beng	12	2	missing	missing	1

Data source: www.racetimer.se

From Table 2.5, there are totally three kinds of non-response which is: non-response of Time2, non-response of Time1 and non-response both Time1 and Time2. In Sunday's race, the sample is 313 skiers which is the same number as in Saturday's race. There are 18 skiers belonging to the non-response of Time 2, 8 skiers belonging to the non-response of Time1 and 6 skiers belonging to the non-response of both Time1 and Time2.

The method I use to handling non-response is to add one column T1-T2 to calculate the difference between T1 and T2. Then, calculate the expected value of T1-T2 and

[3] A dummy variable which denotes distance of the skier who has the second position to start race and from Figure 2.1 it is easily to know that this distance is the same between all skiers.

take this expected value into non-response of Time2 and non-response of Time1 to calculate the other time. For example, if E (T1-T2) =0.5, then Ida's Time2 in Table 2.5 becomes T2=T1-0.5=74.5. After this calculation, the problem of non-response of Time1 and Time2 can be solved. For the non-response both Time1 and Time2, because there are just 6 cases which contains only 6/313=1.9% of the sample, so I decided to delete these 6 skiers' information from the whole data. Now the new sample equals 307.

After calculation, the expected value of T1-T2 is -0.8345, so the problem of non-response is solved and the new dataset have the sample of 307 both in Saturday's race and Sunday's race.

3 Model and results

3.1 Method selection

My aim is to examine if the skier can perform well both in endurance race and in sprint race. So I select TR from Table 2.2 which denotes the skiers' performance in Saturday's endurance race and set it as the response variable.

There are two kinds of variables which can be one of the independent variable. The first one is Time1 in Table 2.4 which denotes the pre-run time for the skier. No matter which starting position the skier has, he (she) will complete the pre-run race individually. So the first model can be expressed as:

$$TR = \alpha_1 + \beta_0 T1 + C \qquad 3.1$$

The independent variable in the second model is Time2 in Table2.4 which denotes the final race time for the skier. This race was held among many skiers and the skier will complete the race by competing with others fiercely. The second model can be expressed as:

$$TR = \alpha_2 + \gamma_0 T2 + D \qquad 3.2$$

C and D in these two models are the other independent variables, such as gender, age, etc.

I choose GLM which stands for General Linear Models (with dummy variables) to analyze the relationship between skiers' performance in endurance race and in sprint race and then calculate how strong the relationship is.

3.2 Model fitting

3.2.1 Model selection

Models where the response variable is considered to be continuous are common in many application areas. [11] In this problem, to make the variables have small difference, I standardize TR, T1 and T2 respectively and define them as TRST, T1ST and T2ST.

The variables I choose to put into model 1 which is similar with 3.1 are:

1) TRST as the response variable

2) T1ST which is the standardized prerun time of skiers in Sunday's race

3) SP which stands for Starting Position. It means if the skier's position in Sunday's race is first or second. "0" denotes the skier is the first position to ski and would run the first, third and fifth lap in the final race. "1" denotes the skier is the second position to ski and would run the second, fourth and sixth lap in the final race.

4) Agea which is a dummy variable: "0" denotes the skier is from age group 9 to 12,and "1" denotes the skier is from age group 13 to 21.

5) Gender which is a dummy variable: "0" denotes the female skier and "1" denotes the male skier.

I put 1 to 5 into a fully model with all interactions with T1ST take into account. While in model 2 the only different variable is the independent variable T2ST which denotes the standardized formal race time of skiers in Sunday's race.

After seeing the ANOVA table for the fully models, I find that the gender is not a significant variable to the model and the staring position plays an important role. In another word, whatever gender the skier has is not so important for the whole trend. I decide to choose the three most significant variables as independent variables and then rebuild the fully models as below.[4]

$$TRST = \alpha + \beta_1 T_1 ST + \beta_2 T_1 ST * Agea + \beta_3 SP + \varepsilon_1 \quad 3.3$$

$$TRST = \alpha + \gamma_1 T_2 ST + \gamma_2 T_2 ST * Agea + \gamma_3 SP + \varepsilon_2 \quad 3.4$$

From the models above, Agea is slope dummy variable while SP is intercept dummy.

[4] The three most significant variables of these two fully models can be found in Appendix

3.2.2 Results

Table 3.1 Estimates of model 3.3

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.69166	0.09279	-7.454	9.55e-13 ***
T1ST	1.21561	0.12644	9.614	<2e-16 ***
SP	1.53517	0.17689	8.679	2.51e-16***
T1ST:Agea	-0.55517	0.11269	-4.927	1.38e-06 ***
Signif. codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1			

Table 3.2 Estimates of model 3.4

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.57253	0.08215	-6.969	1.99e-11 ***
T2ST	1.04320	0.10351	10.079	<2e-16 ***
SP	1.26570	0.14985	8.447	1.27e-15***
T2ST:Agea	-0.44357	0.10617	-4.178	3.85e-05 ***
Signif. codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1			

Table 3.1 and Table 3.2 show the coefficients in model 3.3 and model 3.4 respectively. From these two tables, it is apparent that the standardized difference of results in Saturday’s endurance race which is explained by TRST will increase with increasing of the skiers’ time after standardization. That is to say, the performance of skiers’ has positive relationship between endurance and sprint races

The coefficient of T1ST is larger than that of T2ST. Due to the fierce competition in Sunday’s final race; the skier need to pay more attention to speed up and catch up the others , while in the pre-run, the race type is more similar as the one in Saturday though at a shorter distance.

Table 3.3 Correlation table between endurance race (TRST) and sprint race (T1ST) in prerun.

	First starter	Second starter
9-12 years	0.5930232	0.4454894
13-21 years	0.5595389	0.5962833

Table 3.4 Correlation table between endurance race (TRST) and sprint race (T2ST) in the final race.

	First starter	Second starter
9-12 years	0.650353	0.4172922
13-21 years	0.6325162	0.5557469

It shows that the trend between performance in endurance and sprint races within different age group and SP are similar. Specifically from above tables, the relationships are not strongly either in pre-run or in formal race.

4 Conclusion and recommendation

Firstly, due to the correlation showed in table 3.3 and 3.4, the relationship between younger skiers' performance in endurance and sprint races are not so high, with the range is just between 0.41 to 0.65, so young skiers can not perform well both in sprint and endurance races.

Secondly, if the young skiers want to perform well both in individual and in sprint races, I suggest them practice more mainly in speed endurance and strength endurance as I have mentioned before

My recommendation to the organization of Falu Winter Games is that they can give more kinds of prizes to the most outstanding skiers who are older than 12 years old which can stimulate them to show the best on the track[5]. For the skiers from 9 to 12, the most important thing is their happiness.

[5] This year(2008) the BjörnLindpriset was given to the most outstanding teams which belong to age groups 17-21.

References

[1]*Wikipedia Cross-CountrySkiing* [online].

Available from: http://en.wikipedia.org/wiki/Cross-country_skiing

[2] *Visit Sweden* [online]

Available from: http://www.visitsweden.com/VSTemplates/Page___9651.aspx

[3] *Wikipedia Vasa loppet* [online]

Available from: http://en.wikipedia.org/wiki/Vasaloppet

[4] Huan Liu and Yan Wu, 2007. *Does randomly created relay teams perform worse than self-created?-------- An Empirical research on Falu Ski Winter Game, 5-6*

[5] *Endurance and the young athlete* [online]

Available from: http://www.brianmac.co.uk/articles/scni16a1.htm

[6] *A Reader's Digest Condensed History of Cross-Country Skiing* [online]

Available from: http://home.hia.no/~stephens/skihis.htm

[7] *Endurance training* [online]

Available from: http://www.brianmac.co.uk/enduranc.htm

[8] *Speed training* [online]

Available from: http://www.brianmac.co.uk/speed.htm

[9] Ajit C. Tamhane and Dorothy D. Dunlop 2000, *Statistics and Data Analysis------from Elementary to Intermediate*, Prentice-Hall, Inc. Upper Saddle River, NJ 07458, 3, 82

[10] Donald B. Rubin, 1987, *Multiple Imputation For Nonresponse In Surveys*, John Wiley& Sons Inc. 1(1), 1

[11] Ulf Olsson, 2003, *Generalized Linear Models------An Applied Approach*, Printed in United States of America 2003, Lightning Source, 2(1(2)), 32

Appendix

I Sample of the interview of the former race leader Björn Helgåsen

Interview to the former leader of Falu Winter Games: Björn Helgåsen

Time: March 25th 2007 6~8 pm

Place: Lugnet Sports Center, Falun

Attended Person: Yan WU, Huan LIU, Qi CAO, Björn Helgåsen

Main Aim: to have more information of Falu Winter Games

Some Major Ask and Answer: （W：we three B: Björn Helgåsen）

(1) W: Can you introduce the development of Falu Winter Games briefly?

B: In 1990 we have 1500 young players who all are cross-country skiers. But every year the amount of players was falling down. Now we have about 650.who attended the match on Feb 17th and 18th 2007. Last three years the amount is doubles and I hope it can going up. We also have visitors from Norway. But the new model of this competition is to let young people to make friends each other. That is why I organized the Falu Winter Games.

(2) W: Which competition do you have?

B: We have alpin skiing, ski jumping, Nordic combine and Cross-Country skiing and together we have about 1000 young skiers here in Falun.

(3) W: Can you tell us the specific competition mode on Sunday?

B: Yes. We have 170 pairs started on Sunday and we have a new model in this competition which called pre-match. The model of pre-match likes the F1 match,

that is to say, the pre-match give them the start position for the final. Every pair started in this pre-match and the time together gives their position in the start final. The first 10 starting at the first line and then 5 meters to the next line and we have about 30 pairs in every year in the final and it was the first time we have this competition here. We think the result was very successful and very positive response from clubs

(4) W: Are there any problem in the person combination in pairs?

B: Yes. It happened sometimes that when it was on Sunday morning and some competitors of the group were ill and couldn't attend the competition. As a matter of fact, the other ones have to join another group and then have a mixed group. When they were younger competitors, they often up to the elder groups.

(5) W: What is their objective to attend this competition?

B: Well, the result and the prize of the competition are not very important for the competitor .They attends this game for fun and the game also offers a good chance for every young amateur to communicate or make friends with each other and know more about ski. They can learn how to cooperate with others. And it is also good for ski coach to find out some young skiers who have ski talent.

II Three of the most significant variables in ANOVA of the fully models

These two tables show the way to select variables from the fully models clearly. I decided to use the variables which have most significance property.

Analysis of Variance Table

Response: TRST

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
SP	1	51.616	51.616	69.6735	2.712e-15***
T1ST	1	9.870	9.870	13.3233	0.0003098 ***
T1ST:Agea	1	19.233	19.233	25.9614	6.217e-07 ***
Residuals	296	219.284	0.741		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Analysis of Variance Table

Response: TRST

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
SP	1	55.754.	55.754	79.1489	< 2.2e-16 ***.
T2ST	1	19.860	19.860	28.1938	2.159e-07 ***
T2ST:Agea	1	13.399	13.399	19.0210	1.787e-05 ***
Residuals	296	208.508	0.704		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1