Stockholms universitet

A statistical survey of study results of students at the Department of Mathematics

Ying Liu

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2010:6
Matematisk statistik
Juni 2010

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Bachelor Thesis **2010:6**
http://www.math.su.se

A statistical survey of study results of students at the Department of Mathematics

Ying Liu*

Juni 2010

Abstract

The purpose of this thesis is to investigate the study results (total credit and grade) of students from the course Mathematics I at the Department of Mathematics, Stockholm University. We had observed 149 students that were registered for this course under the fall term 2007 and got a record of total credit until the spring term 2009. The students who had obtained the maximum of total credit, namely 30 credits, would get a passing letter grade, denoted by A, B, C, D and E. We aim to find out the relationships between results and background variables such as age, gender and program. The multiple logistic regression analysis and univariate tests have been used in this study. The results show that: the youngest students receive the highest total credit among all the students and the oldest students receive the lowest total credit; female students have an advantage over male students at obtaining a higher total credit. From the Fisher's and Kruskal-Wallis test, we find that there is a relationship between program and total credit respectively grade. There are no significant difference in grade between age and gender.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail:yili9939@gmail.com . Supervisor: Jan-Olov Persson.

A statistical survey of study results of students at the Department of Mathematics

Ying Liu

May 2010

Abstract

The purpose of this thesis is to investigate the study results (total credit and grade) of students from the course Mathematics I at the Department of Mathematics, Stockholm University. We had observed 149 students that were registered for this course under the fall term 2007 and got a record of total credit until the spring term 2009. The students who had obtained the maximum of total credit, namely 30 credits, would get a passing letter grade, denoted by A, B, C, D and E. We aim to find out the relationships between results and background variables such as age, gender and program. The multiple logistic regression analysis and univariate tests have been used in this study.

The results show that: the youngest students receive the highest total credit among all the students and the oldest students receive the lowest total credit; female students have an advantage over male students at obtaining a higher total credit. From the Fisher's and Kruskal-Wallis test, we find that there is a relationship between program and total credit respectively grade. There are no significant difference in grade between age and gender.

Acknowledgments

I would like to thank Jan-Olov, my supervisor, for his many suggestions and constant support during this research.

I am grateful to my parents for their patience and financial support. Without them this work would never have come into existence. I will also thank Mengyao and Man, for their friendship and understanding.

Contents

Chapter 1 Introduction

In this thesis paper, we will investigate the study results of a group of students at the Department of Mathematics, Stockholm University. The course under investigation is Mathematics I, which consists of 10 different modules with a total of 30 credits. The period under observation is from the fall 2007 to the spring 2009. We have chosen students who were registered for this course in the autumn term 2007 and received a record of the total credit within the observation period. The students' date of birth, gender, total credit, grade and the code of the program were obtained from the study counsellor at the Department of Mathematics. Their ages in 2007 are calculated using the formula: 2007 minus 19xx (the birth year of the students). For instance, if there is a student's date of birth is 850203, then the age of the student=2007-1985=22.

• Objectives

We aim to figure out the following key questions:

- To investigate the relationship between the total credit and age, gender and program.

- To investigate the relationship between grade and age, gender and program.

Chapter 2 Data description

2.1Background

Data was collected from 149 students who were registered in the course Mathematics I at the Department of Mathematics, Stockholm University during the fall term 2007, and had obtained a record of total credit. The course Mathematics I we chose to observe was taught in Swedish and consisted of 10 different modules with a maximum total of 30 credits. The students who had passed all these modules would receive a letter passing grade. Table 2.1 shows the name of each module and credit. The observation period for collecting students' total credits was until the spring term 2009. This meant that students who did not obtain 30 credits in the fall term 2007, could take re-sit exams of the corresponding modules until the spring term 2009. The reason for choosing this observation period is that from the fall term 2007, Stockholm University changed its grading system. The old grading system used VG and G to denote student results. But since the fall term 2007, the grading system began to use letter passing grades on a five-point scale: A, B, C, D and E, where A stands for the highest grade and E stands for the lowest grade.

Table2.1 Modules with codes and their credits

Module with code	Credit
Momentet Algebra räknefärdighet (M101)	1.5
Momentet Algebra polynom (M102)	1.5
Momentet Algebra linjär ekvationssystem (M103)	1.5
Momentet Algebra problemlösning (M104)	7.5
Momentet Analys elementära funktioner (M105)	1.5
Momentet Derivation (M106)	1.5
Momentet Integration (M107)	1.5
Momentet Analys problemlösning (M108)	7.5
Momentet Seminariekurs (M109)	3
Momentet Datorlaborationer (M110)	3

2.2 Descriptions of variables

•Total credit

We have two response variables in this paper: total credit and grade. The total credit is a discrete numerical variable, with outcomes between 0 and 30. There are 7 students who got zero credit and 65 students who got 30 total credits, which is the highest frequency among all the total credit outcomes. The median for total credit is 22.5 and the mean for total credit is 19.6. The frequency distribution of total credit can be seen in Figure A1 in the Appendix. From this figure we see that there are totally 18 different kinds of outcomes of total credit.

•Grade

The grade is an ordinal categorical variable, with outcomes A, B, C, D and E. The total frequency of all the grades is 65. There are 16 students who got A and 5 students who got E. The median for grade is B. The frequency of B is 19, which is the highest frequent among all the outcomes. The frequency distribution of grade can be seen in Figure A2 in Appendix.

•Age

The age of the student in the year 2007 is a numerical variable with outcomes between 18 and 65. The median value of age is 22 and mean is 24.7. The frequency distribution of age can be seen in Figure A3 in Appendix. The age 19 has the highest frequency: 32. The age 20 also has an extremely large frequency, namely 24, which is more than two times larger than other outcomes. The age 65 is the only outcome between age 51 and 65.

•Program

The variable program is a categorical variable. There are totally 14 different programs. Table 2.2 shows each program's name, code and frequency. From this table we see that the Fristående course has the highest frequency of 53, the four programs: Magister of Education and Magister of Science, Specialization in Biology, English, Physics and Chemistry have the lowest frequency of 1.

Table 2.2 Programs with codes and frequencies

Program	Code	Frequency
Fristående course	FRIST	53
Magister of Education and Magister of Science, Specialization in Biology	KB	1
Magister of Education and Magister of Science, Specialization in English	KE	1
Magister of Education and Magister of Science, Specialization in Physics	KF	1
Magister of Education andMagister of Science, Specialization in Chemistry	KK	1
Magister of Education and Magister of Science,Specialization in Mathematics	KM	5
Bachelor's Program in Scientific Computing	NBERK	2
Bachelor's Program in Biomathematics	NBIMK	2
Bachelor's Program in Computer Science	NDATK	7
Bachelor's Program in Physics	NFYSK	36
Bachelor's Program in Mathematics	NMATK	20
Bachelor's Program in Mathematics and Philosophy	NMFIK	3
Master's Program in Medical Physics	NSFPY	15
Bachelor's Program in Mathematics and Economy	SMAEK	2

•Gender.

Variable gender is a categorical variable. There are 55 females and 94 males.

2.3 Relationships between background variables

To describe the relationship between age and gender we make a side-by-side box plot of students' ages for each category of gender, see Figure A4 in Appendix. From this plot we can see that the age range for males is larger than the age range for females since there is a male who is 65 years old which expands the age range for males. No obvious difference between age median for female and male.

In order to describe the relationship between age and program, we make a side-by-side box plot of students' ages for each category of program, see Figure A5 in Appendix. From this plot, we can find the majority of program's age ranges is between 18 and 40. The age range for FRIST is largest of all, since it has an observation which age is 65. The age range for NMATK is also extremely large relative to other programs.

To describe the relationship between gender and program, we perform a contingency table, see Table A1 in Appendix. Male's total frequency is much larger than female's. In program FRIST with frequency 53 there are 15 more males than females and in program NFYSK with frequency 36 there are 10 more males than females. No obvious difference between male and female towards programs.

Chapter 3 Statistical methods

3.1 Wilcoxon-Mann Whitney test and Kruskal-Wallis test

The Wilcoxon Mann-Whitney Test is a powerful of the nonparametric tests for comparing two populations. It is used to test the null hypothesis that two populations have identical distribution functions against the alternative hypothesis that the two distribution functions differ only with respect to location (median), if at all.

The Wilcoxon Mann-Whitney test does not require the assumption that the differences between the two samples are normally distributed.

Initially assume that there are no ties in the two samples: $x_1, x_2, \ldots, x_{n1}$ and $y_1, y_2, \ldots, y_{n2}$.

The steps in the Wilcoxon-Mann Whitney test are as follows.

- Rank all $N=n_1+n_2$ observations in ascending order.
- Sum the ranks of the x's and y's separately. Denote the sums by w_1 and w_2, respectively. Since the ranks range over the integers 1, 2, …, N, we have

$$w_1+w_2=1+2+\ldots+N=\frac{N(N+1)}{2}$$

- Reject H_0 if w_1 is large or equivalently if w_2 is small. [1]

The Kruskal-Wallis test is a generalized form of the Mann-Whitney test method, as it permits two or more groups.

3.2 Spearman rank correlation coefficient

Spearman's rank correlation coefficient, named after Charles Spearman and often denoted by r_s, is a nonparametric measure of correlation; it assesses how well an arbitrary monotonic function could describe the relationship between two variables, without making any other assumptions about the particular nature of the relationship between the variables.

It is also a technique which can be used to summarize the strength and direction (negative or positive) of a relationship between two variables.

•Method-calculating the coefficient

Denote:

- $d_i = x_i - y_i$ = the difference between the ranks of corresponding values X_i and Y_i,
- n = the number of values in each data set (same for both sets).

$r_s=\rho$ is given by

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \qquad [2]$$

3.3 Multiple Logistic Regression model

Logistic regression is a useful way of describing the relationship between one or more independent variables (e.g., age, sex, etc.) and the probability of a binary response variable Y, which has only two possible values, such as Y=0 or Y=1.

Logistic regression does not require normally distributed variables. It does, however, require that observations be independent.

The multiple logistic regression model for p(Y=1) at values x=(x_1, x_2, …. , x_n) of n predictors is

$$P(Y=1) = \frac{\exp(\alpha+\beta 1x1+\beta 2x2+\cdots+\beta nxn)}{1+\exp(\alpha+\beta 1x1+\beta 2x2+\cdots+\beta nxn)}$$

Equivalently, the log odds, called the logit, has the linear relationship

$$\text{logit}\,[p(Y=1)]=\log\frac{p(Y=1)}{1-p(Y=1)} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

α is called the "intercept" and β_1, β_2,…, β_n are called the "regression coefficients" of explanatory variables. The intercept is the value of logit[p(Y=1)] when the value of all independent variables is zero. A positive regression coefficient means that that explanatory variable increases the probability of the outcome, while a negative regression coefficient means that variable decreases the probability of that outcome; a large regression coefficient means that the predictor strongly influences the probability of that outcome; while a near-zero regression coefficient means that that predictor has little influence on the probability of that outcome.

Exponentiating both sides of the logistic regression model shows that the odds are an exponential function of x. This provides a basic interpretation for the magnitude of β: the odds increase multiplicatively by e^{β} for every 1-unit increase in x. In other words, e^{β} is an odds ratio, the odds at X=x+1 divided by the odds at X=x. [3]

3.4 The Hosmer-Lemeshaw Goodness of fit test

The Hosmer-Lemeshaw test is a measure of goodness of fit, typically summarize the discrepancy between observed values and the values expected under the model to describe how well the model fits a set of observations. This test is only available for binary response models. It recommends partitioning the observations into g groups according to their predicted probabilities, where g is between 3 with 10.

The Hosmer-Lemeshow statistic is obtained by calculating the Pearson chi-square statistic with (g-2) degrees of freedom from the $\mathbf{2\times \mathit{g}}$ table of observed and expected frequencies. The row of the table corresponds to the two values of the outcome variable y=1, 0 and the j columns correspond to the groups. Then

$$G_{HL}^2 = \sum_{j=1}^{g} \frac{(Oj-Ej)^{2}}{Ej\left(1-\frac{Ej}{nj}\right)}$$

where

- nj = number of observations in the j^{th} group
- Oj= $\sum_i$ yij = obseved number of cases in the j^{th} group
- Ej= $\sum_i$ pij = expected number of case in the j^{th} group
- j=1,…, g. i=0,1.

Large values of G_{HL}^2or small ***p***-values indicate a lack of fit of the model.

Chapter 4 Analysis and Results

In this chapter we will investigate the relationship between the response variables and background variables by conducting univariate tests at 5% significant level and fitting a multiple logistic regression model with the statistical software SAS.

4.1 Univariate tests

•Total credit versus age

The collected data suggests a fairly weak relationship between age and total credit as shown in the Figure A6 in Appendix.

We will look for a relationship between the total credit and age, where none of these two variables is normally distributed. The Spearman's rank correlation technique is used to see if there is indeed a correlation between age and total credit, and to test the strength of this relationship. We set up H_0: there is no significant relationship between the total credit and age. The calculated Spearman's rho value is -0.224, and P value is 0.006. These results say that there is a negative relationship between total credit and age such as the total credit is decreasing as age increases.

Since age has lots of different outcomes, in order to more simply and clearly investigate the difference among different age brackets, we classify age into four categories: category 1 is from age 18-23; category 2 is from age 24 to 28; category 3 is from age 29 to 38; and category 4 is from age 39 to 65. The frequency of each category is displayed in Table 4.1. The frequency distribution of the four different age categories towards total credit can be seen in Figure A7 in Appendix. From Table 4.1 we can find that age category 18-23 has the highest frequency, i.e. 94. The age category 39-65 has the lowest frequency of 12. The Age categories 24-28 and 29-38 have frequencies of 25 and 18.

Table 4.1 Frequency table of classified age

Age	Frequency
18-23	94
24-28	25
29-38	18
39-65	12

Using this classified age, we want to examine that if there exists a relationship between age categories and total credit. Therefore we perform the Kruskal-Wallis test with null hypothesis: there is no difference among age categories towards total credit, see Table

4.5. The Kruskal-Wallis test shows that there is a statistically significant difference among age categories since p=0.015. Because we do not perform two independent samples tests, so we cannot describe what the exact difference is between 2 out of 4 different age categories. But if we compare mean and median for age categories, see Table 4.5, it shows that it is the youngest of which is different from the others. Age category 18-23 has the highest mean of 21.7 and the highest median of 30 among all the age categories. Age category 39-65 has the lowest mean of 15 and lowest median of 10.5. Mean and median for age category 29-38 are a little larger than age category 24-28.

•**Total credit versus gender**

The frequencies for female and male are 55 and 94 respectively. See Table 4.2 below. The frequency distribution of gender towards total credit can be seen in Figure A8 in Appendix.

Table 4.2 Frequency table of gender

Gender	Frequency
Female	55
Male	94

We want to investigate the difference between male and female towards total credit therefore the Wilcoxon-Mann Whitney test is performed under null hypothesis: $F_1=F_2$, where F_1 denotes the distribution function of female and F_2 denotes the distribution function of male.

The Wilcoxon-Mann Whitney test shows that there is a statistically significant difference between male and female. The total credit mean and median for female are 21.3 and 30. The total credit mean and median for male are 18.6 and 22.5. In other words, females have statistically significantly higher values on total credit than males. See Table 4.5.

•**Total credit versus program**

Table 4.3 displays 14 different kinds of program codes with their frequencies.

Program FRIST has the highest frequency of 53 among all programs. Programs NFYSK and NMATK have the frequencies of 36 and 20 respectively. Programs KB, KE, KF and KK have only one frequency. Programs NBERK, NBIMK and SMAEK have the same frequency of 2. Program NSFPY has the frequency of 15.

Table 4.3 Frequency table of program

Program	Frequency
FRIST	53
KB	1
KE	1
KF	1
KK	1
KM	5
NBERK	2
NBIMK	2
NDATK	7
NFYSK	36
NMATK	20
NMFIK	3
NSFPY	15
SMAEK	2

Because the background variable program has 14 different programs and some of them have very low frequencies, we classify them by major in order to simply investigate the difference of programs. These 14 programs are divided into 4 groups: major in mathematics, major in physics, Fristånde course and Rest which programs belong none of the mentioned majors. Table 4.4 shows these 4 program categories with their program codes and frequencies. The frequency figure of the program categories towards total credit can be seen in Figure A9 in Appendix. Mathematics has the frequency of 32. Fristånde has the frequency of 53. Physics and Rest have the frequency of 37 and 27 respectively.

Table 4.4 Frequency table of classified program

Category	Programs code	Frequency
Mathematics	SMAEK,NMATK,NMFIK,KM, NBIMK	32
Physics	NFYSK KF	37
Fristånde	FRIST	53
Rest	NDATK NBERK KE KB KK KF NSFPY	27

To investigate the relationship between the classified program and total credit, we perform the Kruskal-Wallis test under the null hypothesis: there is no difference among program categories. The result shows that there is not a significant difference among program categories at 5% level because p=0.141. Mathematics has mean and median value of 22.5 and 30. Mean and median for Physics are 21 and 27. Mean values for Fristånde and Rest are 17.7 and 18.1 respectively. See Table 4.5. No obvious difference between program categories.

Table 4.5 Statistics of total credit versus background variables

Variable	H_0:No difference	Category	Frequency	Mean	Median
Age	K-W p= 0.015				
		18-23	94	21.7	30
		24-28	25	16.1	15
		29-38	18	16.4	16.5
		39-65	12	15	10.5
Gender	W-M p=0.046				
		Female	55	21.3	30
		Male	94	18.6	22.5
Program	K-W p=0.141				
		Mathematics	32	22.5	30
		Physics	37	21	27
		Fristånde	53	17.7	22.5
		Rest	27	18.1	21

•**Classified total credit versus background variables**

Now we separate total credit into two groups: those completing 30 credits or those who do not, to better and more simply investigate the relationship between the categories for background variables and total credit. We call this total credit with 2 categories classified total credit. The Fisher's test is performed with H_0: classified total credit is independent of each classified background variable at 5% level. The p values of the Fisher's tests, frequencies and proportions of completing 30 credits for background variables can be seen in Table 4.6.

Obviously there is a significant relationship between classified total credit and age categories, since p value of Fisher's test is 0.006. By comparing each age category's proportion of completing 30 credits, we can find some difference between each age category on classified total credit. The age category 18-23 has the highest proportion of completing 30 credits, namely 54%. Age category 39-65 has the lowest proportion of 17%. The proportions for age categories 24-28 and 29-38 are 24% and 33% respectively.

Classified total credit is dependent of gender, because the p value of the Fisher's test=0.003. The proportion of completing 30 credits for female is 60% and for male is 34%. So females have a statistically significantly higher value on classified total credit than males.

There is a significantly statistical relationship between classified total credit and program categories. We compare proportion of completing 30 credits for every program category, and find that program category Mathematics has the highest proportion of 63%. In contrast, program category Fristånde has the lowest proportion of 34%. Proportions for program categories Physics and Rest are 49% and 42%.

Table 4.6 Classified total credit versus background variables

Variable	Category	H_0: independent	Frequency	Proportion (%)
Age		P=0.006		
	18-23		94	54
	24-28		25	24
	29-38		18	33
	39-65		12	17
Gender		P=0.003		
	Female		55	60
	Male		94	34
Program		P=0.004		
	Mathematics		32	63
	Physics		37	49
	Fristående		53	34
	Rest		27	42

4.2 Multiple logistic regression analysis

Now we investigate how the 3 classified background variables together affect the response variable classified total credit. Since classified total credit is not normally distributed, we cannot set up a normal regression model. We choose logistic regression, which allows us to establish a relationship between a binary outcome variable and a group of predictor variables, and model the logit-transformed probability as a linear relationship with the predictor variables.

Let Y indicate the response variable classified total credit, which has two outcomes: one of two outcomes is denoted by Y=0 if total credit is less than 30 credits, and the other one is that total credit is equal to 30 credits, denoted by Y=1. We set up a multiple logistic regression model for the logit-transformed probability of getting 30 credits, namely p(Y=1) at values x = (A, G, P) for 3 predictors:

$$\text{logit}\,[p(Y=1)]=\alpha+\beta_1 A_i+\beta_2 G_j+\beta_3 P_k$$

where

- A_i=age, and i=1, 2, 3, 4,
- G_j=gender, and j=0, 1.
- P_k=programs, and k=1, 2, 3, 4.

•Interaction

I also look at three background variables' interactions which can have effects on the classified total credit. But none of interactions is significant at 5% level, which may be because we do not have many data to investigate that, so we build our logistic regression model with multiple predictors and no interaction terms.

•Analysis of the effects of predictors

By running the logistic regression model with SAS, we can see each predictor's p value with its degrees of freedom. The predictor program is not significant at 5% level, so we delete program from the model.

We build our logistic regression model with significant predictor variables:

$$\text{logit}[p(Y=1)] = \alpha + \beta_1 A_i + \beta_2 G_j$$

The Wald test's statistics of each predictor variable can be seen in Table 4.6.

Table 4.6 Predictors in the logistic regression model

Variable	DF	Wald Chi-Square	P value
A	3	10.55	0.014
G	1	9.86	0.0017
P	3	4.28	0.23

•Interpret the fitted logistic regression model

The logistic regression model $\text{logit}[p(Y=1)]=\alpha+\beta_1 A_i+\beta_2 G_j$ is fitted by the maximum likelihood estimates, regarding age category18-23 and male as the reference variables. The odds ratio, 95% Wald confidence limits of OR and p values of the maximum likelihood estimates under Wald test are displayed in Table 4.7 below.

Table 4.7 Estimates of the fitted model

Variable	Category	OR	95% CI of OR	p-value
Age1	18-23	1		
Age2	24-28	0.26	0.09 0.74	0.012
Age3	29-38	0.3	0.1 0.92	0.036
Age4	39-65	0.13	0.03 0.68	0.016
Gender 0	Female	3.53	1.66 7.52	0.001
Gender 1	Male	1		

If we look at the p values of Wald test, we can find that the predictor age and gender are significant at 5% level in this multiple logistic regression model.

This fitted model says that, holding gender at a fixed value, the odds of getting 30 credits for age category 24-28 over age category 18-23 is 0.26. In terms of percent change, we can say that the odds for age category 24-28 are 74% lower than the odds for age category 18-23. The odds ratio for age category 29-38 relative to age category

18-23 says that, the age category 29-38 is 70% lower than age category 18-23. The odds ratio for age category 39-65 is 87% lower than the odds ratio for age category 18-23. The result shows that the youngest students get the highest total credit, and the oldest students get the lowest total credit.

Holding age at a fixed value, the odds of getting 30 credits for females over the odds of getting 30 credits for males is 3.53. In terms of percent change we say that the odds ratio for females is 253% higher than the odds ratio for males. This means that female students obtain a higher total credit than male students.

•**Checking Goodness- of- fit**

After fitting the multiple logistic regression model, we need to check its goodness-of-fit. The Hosmer and Lemeshow test is performed here to exam that.
The Chi-Squared p value with 5 degrees of freedom is 0.99. It indicates that our fitted logistic regression model fits observations very well.

4.3 Students' grade

In this section, our aim is to investigate the relationship between the response variable grade and background variables. There are only 65 students who receive a letter passing grade, which is a very small sample group in which to investigate the relationship between grade and background variables. Therefore we direct use the classified age, gender and program as the background variables to investigate their relationship with grade.

•**Descriptions of background variables versus grade**

Age versus grade

To better understand the distribution of the classified age towards grade, we make a frequency table of age towards grade. This table can be seen in Table A2 in Appendix. The frequency of age category 18-23 is 51, which is much higher than the other age categories. The age categories 24-28 and 29-38 have the same frequency of 6. The age frequency 39-65 has the frequency of 2.

Gender versus grade

There are 33 females and 32 males who got a grade. The frequency table of gender towards grade can be seen in Table A3 in Appendix.

Program versus grade

The program categories Physics and Fristående have the frequency of 18, which is two times bigger than Rest's. The program category Mathematics has the highest frequency of 20. See Table A4 in Appendix.

- **Univariate tests**

In order to better investigate the relationship between grade and background variables, we use some numerical data to denote grade: let "1" stands for E, "2" stands for D, ... , "5" stands for A. Then we conduct univariate tests at 5% significant level, to investigate the relationships.

Grade versus age

We compare each age category's mean and median value. The age category 24-28 and age category 39-65 have the same median value 3. Age category 18-23 has the median of 4 and the mean of 3.51. Age category 29-38 has the mean and median value of 2.5. See Table 4.8.

For us to exam that if there is indeed a significant relationship between grade with age, we perform the Kruskal Wallis test with null hypothesis: there is no difference between age categories. Obviously the result is no difference among age categories, since p=0.35.

Grade versus gender

The female mean is 3.27 and male mean is 3.5. The female median is 3 and male median is 4. We conduct the Wilcoxon-Mann Whitney test to examine the difference between male with female, and p=0.46. This indicates that there is not a significant relationship between gender and grade. See Table 4.8.

Grade versus program

To examine the relationship between grade and program categories, the Kruskal-Wallis test is used with H_0: there is no difference among program categories. The K-W test is significant, which indicates that program categories have significant differences. But we cannot find out what the exact difference is between two out of four categories because we do not use two independent samples tests to examine that. But by comparing mean for each program category, we can still see their differences on mean for grade. The program category Fristående has the highest mean=3.72; Physics and Rest have the same mean: 3.67; Mathematics has the lowest mean which is 2.7. See Table 4.8.

Table 4.8 Statistics of grade versus predictor variables

Variables	H_0:no difference	Category	Frequency	Median	Mean
Age	K-W p=0.35	18-23	51	4	3.51
		24-28	6	3	3.33
		29-38	6	2.5	2.5
		39-65	2	3	3
Gender	W-M p=0.91	Female	33	3	3.27
		Male	32	4	3.5
Program	K-W p=0.04	Mathematics	20	2	2.7
		Physics	18	4	3.67
		Fristånde	18	4	3.72
		Rest	9	4	3.67

- **Classified grade versus background variables**

We have got the proportion of students receiving A or B from every category of background variables. To examine the grades more simply we divide grade into two groups: group= 1 if grade is A or B, otherwise group=2 if grade is C, D or E, and call this kind of grade classified grade.

The Fisher's test is performed to test if there is a relationship between the classified grade and background variables at 5% significant level. The proportion of receiving A or B for every background variable is obtained. See Table 4.9.

We set up the null hypothesis as age categories are independent of classified grade. The p value is 0.22 which says that there is not a significant relationship between age categories and classified grade. The age category 18-23 has the highest proportion of receiving A or B, which is 58.82%. The age category 29-38 has the lowest proportion of 16.67%. Proportions for age categories 24-28 and 39-65 are the same: 50%.

We use the classified grade to test the relationship between it and gender. The Fisher's test is performed with null hypothesis: the classified grade is independent of gender. P=0.46, and this says that there is not a relationship between the classified grade and gender. The proportion for female is 48.48%, and for male is 59.38%.

To investigate the relationship between classified grade and program categories, we perform the Fisher's test with H_0: classified grade is independent of program categories. p=0.21, which shows that there is no significant relationship between classified grade and program categories. The proportion for program category Mathematics is 35%, which is the lowest proportion among all the program categories. Both program categories Fristånde and Rest have the highest proportion: 66.67%. The program category Physics has the proportion of 55.56%.

Table 4.9 Classified grade versus background variables

Variable	Category	H_0: independent	Frequency	Proportion (%)
Age		P=0.22		
	18-23		94	58.82
	24-28		25	50
	29-38		18	16.67
	39-65		12	50
Gender		P=0.46		
	Female		55	48.48
	Male		94	59.38
Program		P=0.21		
	Mathematics		32	35
	Physics		37	55.56
	Frist ånde		53	66.67
	Rest		27	66.67

- **Multiple logistic regression analysis**

After investigating the relationship between grade and each background variables separately, we also want to figure out how these background variables affect classified grade together. Because classified grade is not normally distributed, we cannot establish a normal regression model. Therefore we use classified grade with 2 categories: Y=1 for grade=A or B, otherwise Y=0 for grade=C, D and E, and model the logit-transformed probability of classified grade as a linear relationship with predictor variables such as classified age, gender and program.

The multiple logistic regression model with response variable classified grade is built below:

$$\text{logit}\,[p(Y=1)]=\alpha+\beta_1 A_i+\beta_2 G_j+\beta_3 P_k$$

where

A_i=age, and i=1, 2, 3, 4,
G_j=gender, and j=0, 1.
P_k=programs, and k=1, 2, 3, 4.

The p values for predictor classified age, gender and program under the Wald Chi-squared test are 0.36, 0.58 and 0.2 respectively. It is obvious to see that none of predictor variables is significant at 5% level. It means that there is not a significant relationship between classified grade and predictors.

Chapter 5 Discussion

•Age

The fitted logistic regression model shows that the youngest students (from 18 to 23 years old) get an advantage over the older students at receiving a better total credit.

The reason for this phenomenon may be because it is a golden physiology period for these youngest students to study. For instance, the youngest students have a better memory and are more energetic. On the other hand, many of youngest students do not have families or job burdens, so they can utilize more time on studying. The older students have much more burdens from family or society.

• Gender

Our results show that females have a higher total credit than males. This may be caused by a number of reasons. I personally think that these female students maybe work harder than male students.

• Program

The Kruskal-Wallis test and Fisher's exact test show that there exist some differences between program categories towards student results, namely total credit and grade. But we cannot describe what differences are in this study, because we do not perform two independent samples test and set up a logistic regression model for program categories.

•Grade

From the results, we only know that there is a relationship between classified program and grade. We cannot investigate how these three background variables affect classified grade together because none of these variables is significant in the regression analysis. This can be caused by lots of varied reasons, for instance, the data we used to illustrate the relationship between grade and background variables was collected from 65 students who had obtained a grade. It is sparse and can reasonably cause unfitness of regression.

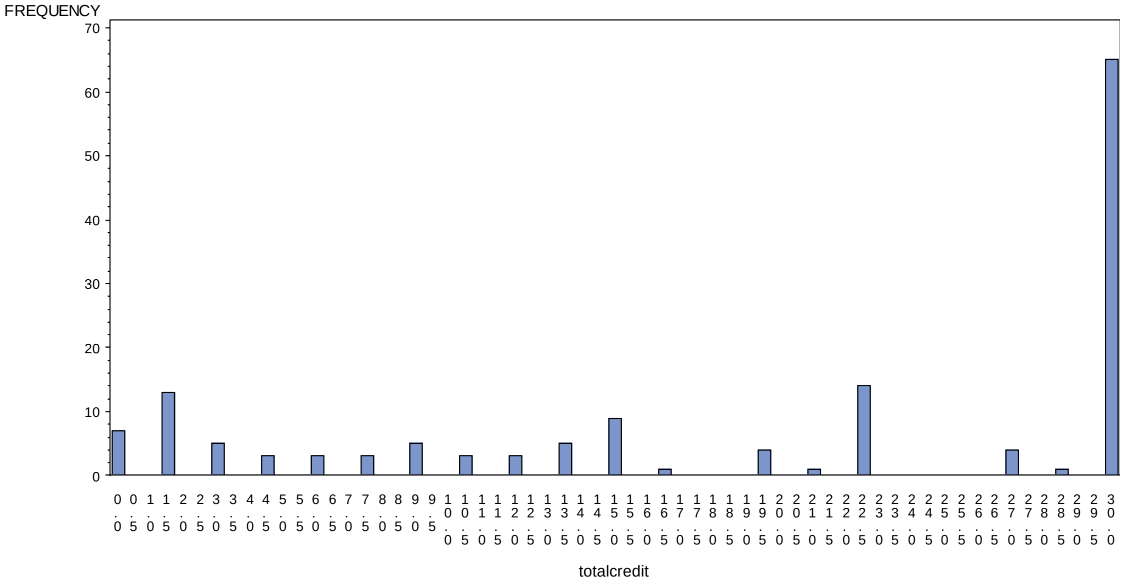
References

[1] Ajit C. Tamhane and Dorothy D. Dunlop: *Statistics and Data Analysis from Elementary to Intermediate.*

[2] http://en.wikipedia.org/wiki/Spearman's_rank_correlation_coefficient

[3] Agresti: *Categorical Data Analysis*. Second Edition (2002)

Appendix



FigureA1 Frequency distribution of total credit

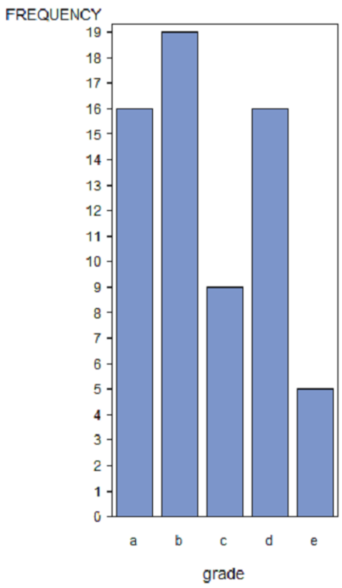


Figure A2 Frequency distribution of grade

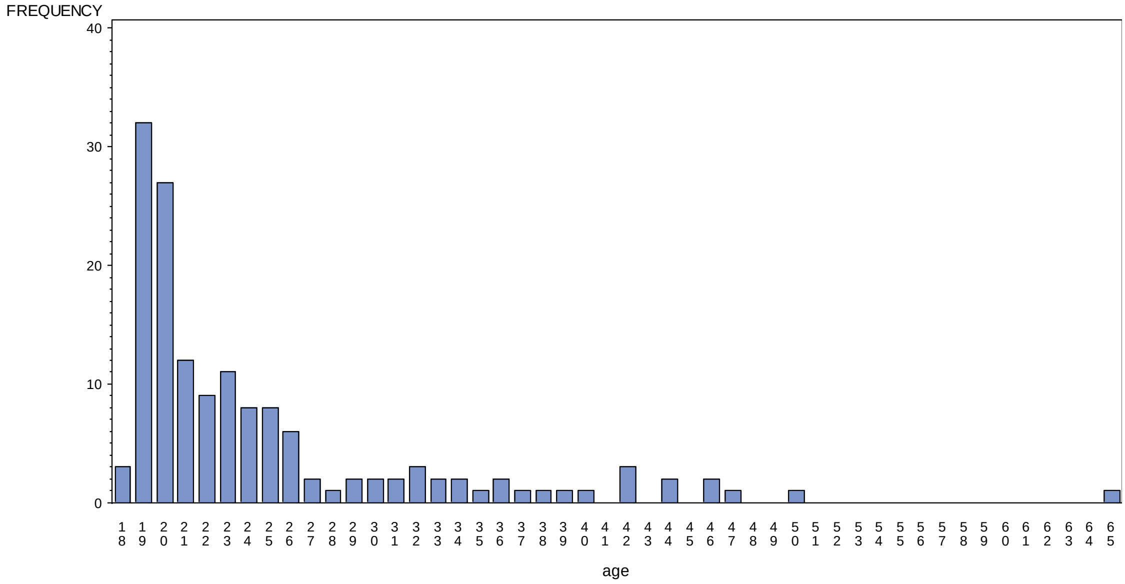


Figure A3 Frequency distribution of age

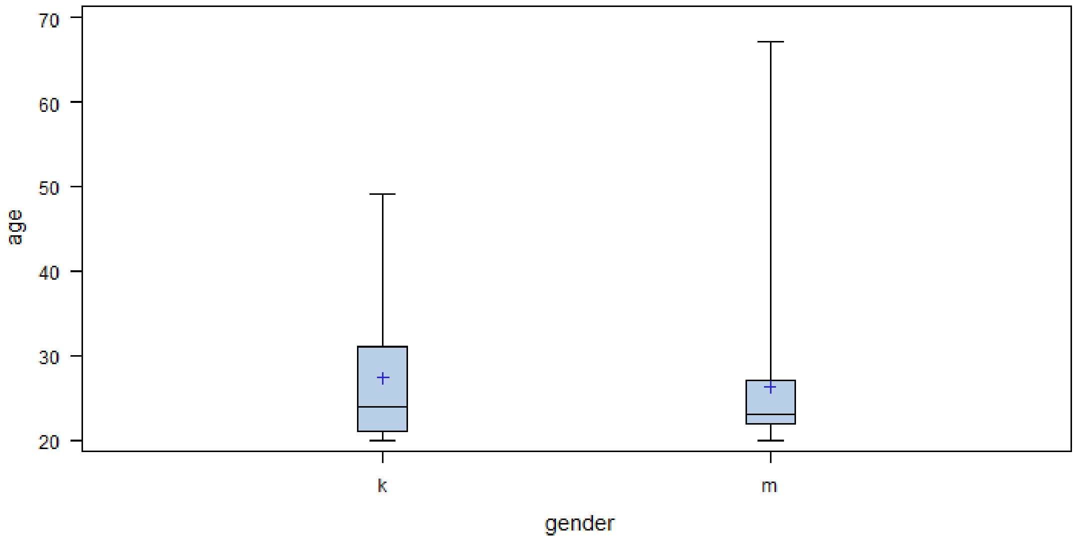


Figure A4 Box plot of Age versus gender

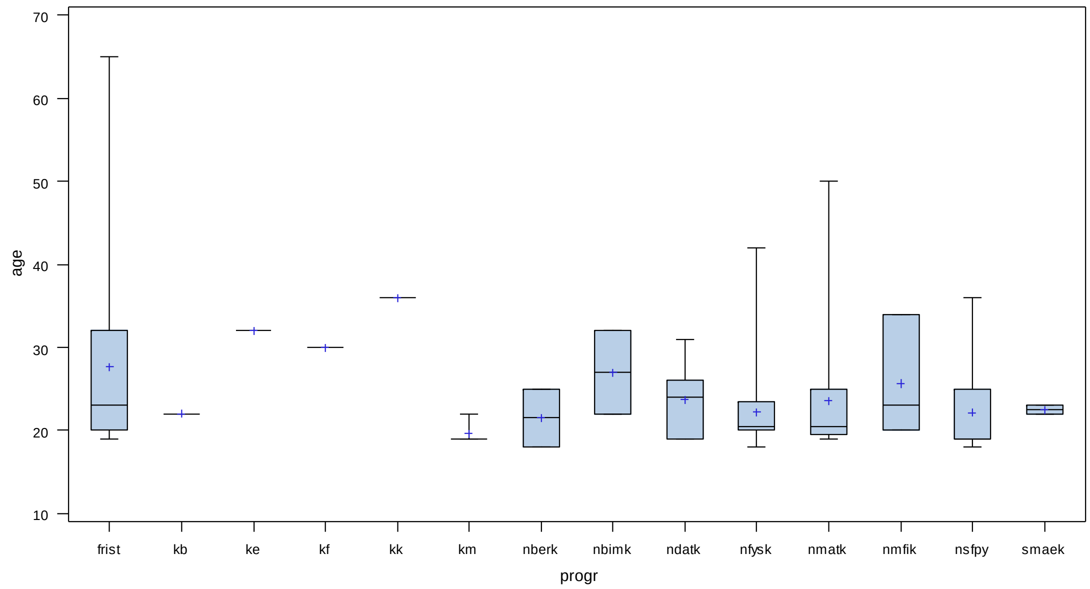


Figure A5 Box plot of age versus program

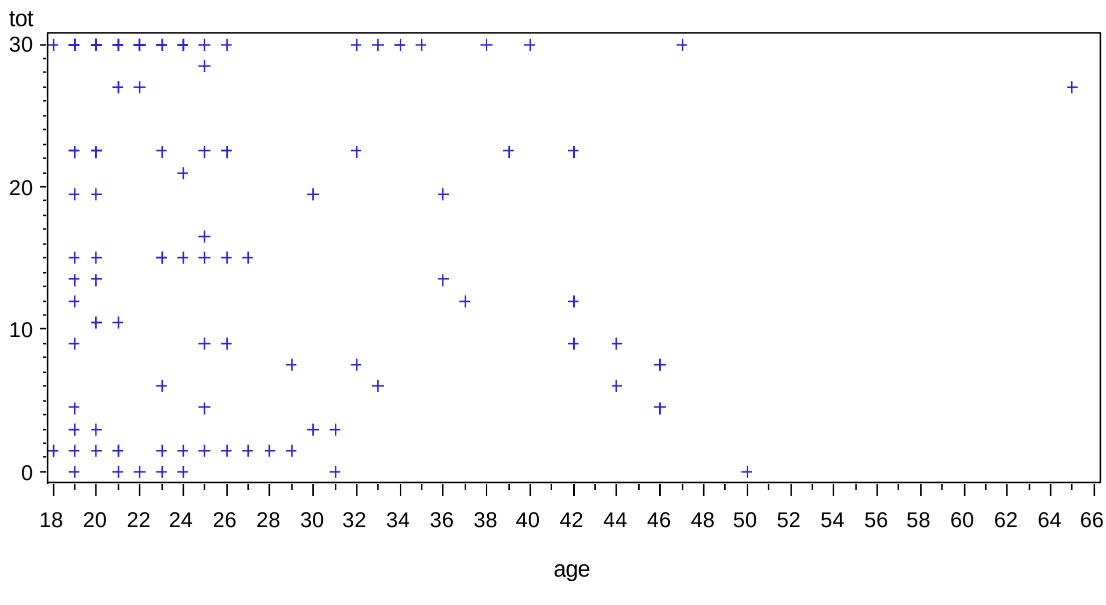


Figure A6 Scatter graph of total credit versus age

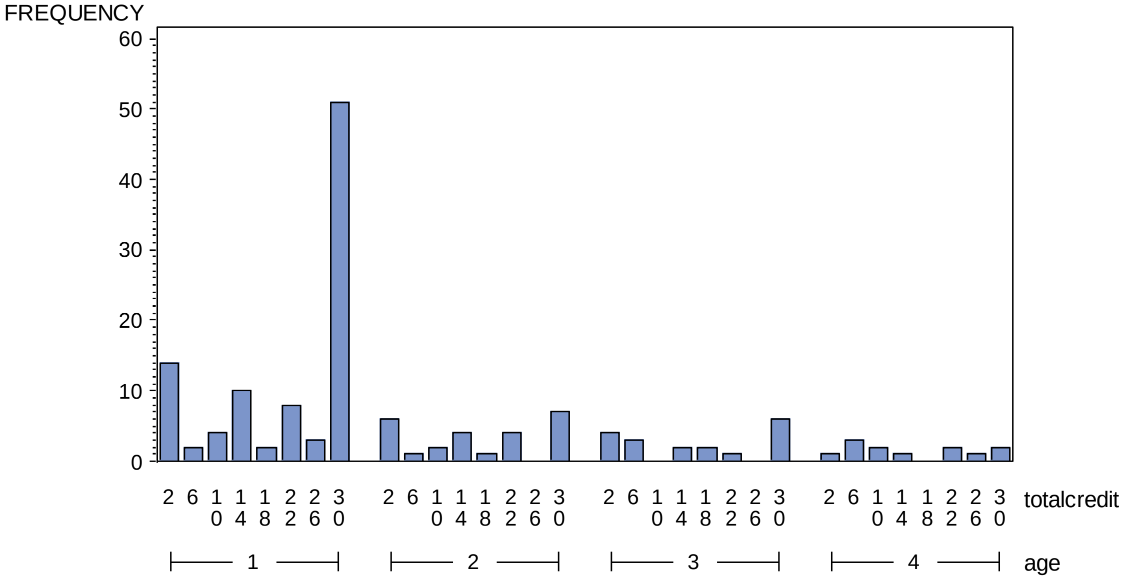


Figure A7 Frequency distribution of total credit versus classified age. “1” stands for the age category 18-23, “2” stands for 24-28, “3” stands for age 29-38 and “4” stands for 39-65.

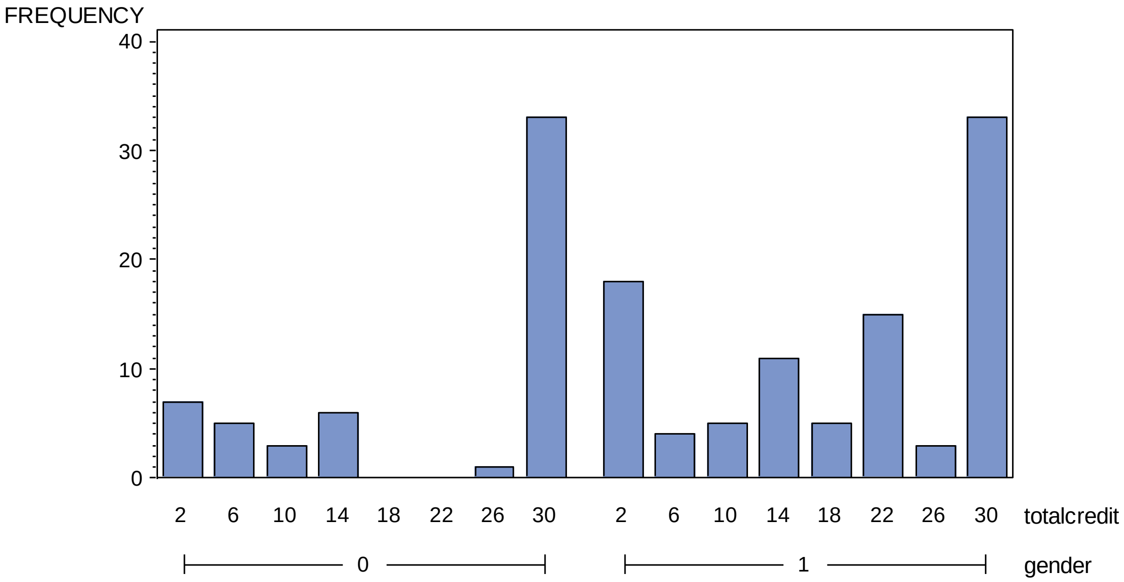


Figure A8 Frequency distribution of total credit versus gender. “0” stands for female and “1” stand for male.

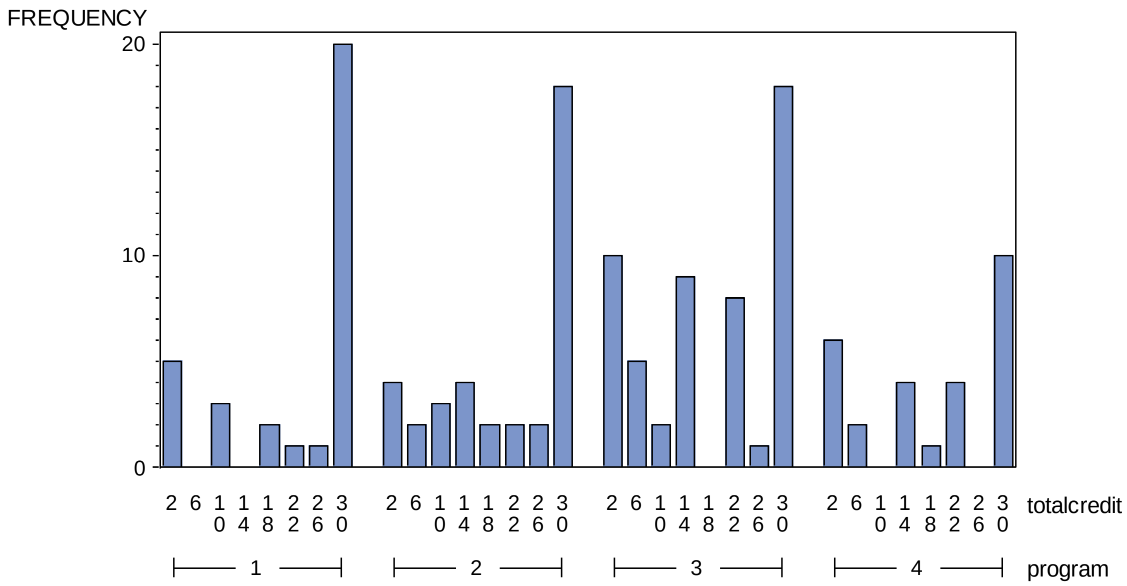


Figure A9 Frequency distribution of total credit versus classified program. "1" stands for the program category Mathematics, "2" stands for Physics, "3" stands for Fristånde and "4" stands for Rest.

Table A1 Frequency table of gender versus program

Program	Female	Male	Total
FRIST	19	34	53
KB	0	1	1
KE	1	0	1
KF	0	1	1
KK	0	1	1
KM	3	2	5
NBERK	0	2	2
NBIMK	2	0	2
NDATK	0	7	7
NFYSK	13	23	36
NMATK	6	14	20
NMFIK	2	1	3
NSFPY	8	7	15
SMAEK	1	1	2
Total	55	94	149

Table A2 Frequency table of classified age versus grade

grade	(18-23)	(24-28)	(29-38)	(39-65)	total
A	13	2	0	1	16
B	17	1	1	0	19
C	7	0	2	0	9
D	11	3	2	0	16
E	3	0	1	1	5
Total	51	6	6	2	65

Table A3 Frequency table of gender versus grade

grades	Frequency for female	Frequency for male	Total
A	8	8	16
B	8	11	19
C	5	4	9
D	9	7	16
E	3	2	5
Total	33	32	65

Table A4 Frequency table of classified program versus grade

grade	Mathematics	Physics	Fristånde	Rest	total
A	1	6	7	2	16
B	6	4	5	4	19
C	2	5	1	1	9
D	8	2	4	2	16
E	3	1	1	0	5
Total	20	18	18	9	65