2 Sample project

This Maths Studies project has been graded by a moderator. As you read through it, you will see comments from the moderator in boxes like this:

Moderator's comment:

At the end of the sample project is a summary of the moderator's grades, showing how the project has been graded against all the criteria A to G. These criteria are explained in detail in chapter 13 of the Mathematical Studies textbook.

Reading projects and the moderator's comments will help you to see where marks are gained and lost, and will give you helpful tips in writing your own project.

Is lung capacity affected by smoking, sport, height or gender

Moderator's comment:
The project has a title

Table of contents

Introduction

Aim of the project:

This project is aimed to figure out if smoking, sport, gender or height influences lung capacity. In order to figure out what influences the lung capacity, data will be collected and analyzed. Comparisons between chosen smokers and non-smokers will be carried out in order to see whether smoking influences the lung capacity. It is expected that general lung capacity is going to be bigger for non-smokers and smaller for smokers, male and female, thus it is the aim of the project and will be investigated through analysis of data.

1. To check if this hypothesis is true, measuring lung capacity procedure will take place.
2. I wanted to test 40 individuals, divided into female and male groups, smokers and non smoker.
   (10 male smokers, 10 male non-smokers, 10 female smokers, 10 female non-smokers). So I picked students at random from the IB diploma programme in my school and asked if they were smokers or non smokers. Once I had 10 of each gender that were smokers and 10 that were non smokers I asked them to complete the questionnaire and then started the test that would measure their lung capacity.
3. Age from 16 till 21 years old because 16 is the legal age for smoking in the Netherlands.
4. Every person gets 3 tries to blow in the lung capacity meter so the average value can be picked.

The type of questions asked:

In order to investigate as it was mentioned before we need to collect and analyze the data. In order to do that, questionaires and forms are going to be composed for the 40 individuals that are tested for the lung capacity.

Questionaires are going to contain these types of questions:

1. Gender? Male Female
2. Age? .....
3. Are you an athlete? yes no

The data taken from individuals that perform the lung capacity test.

1. Height
2. Lung capacity

Hypotheses

1. Smoking decreases lung capacity
2. Lung volume depends on height, gender and also if the individual is an athlete. People who do sports have a higher lung volume which can be independent of height.

3 Larger lung capacity expected to be: **Smaller lung capacity expected to be:**

Males	Females
Non-smokers	Smokers
Athletes	Non-athletes

I will find the mean lung capacity for each of the 40 participants and set up a table with the information collected. Each individual blows into the lung capacity meter three times and I will find the mean of these three blows to use in my analysis. This gives me a more reliable reading than if the participant only blew once into the meter. I will compare the means of the lung capacities for each of the groups in order to find out which group has the largest lung capacity and which one the smallest. I will also find the standard deviation as this may be useful when deciding on the groupings for the chi-squared test to see if lung capacity is independent of gender or of smoking. I will also compare the mean lung capacity of athletes and non-athletes to find out if athletes have larger lung capacities than non-athletes and I will draw a scatter diagram to find out if there is any correlation between height and lung capacity. If it appears that there is a correlation then I will find the correlation coefficient and possibly the equation of the regression line if the correlation coefficient is moderate to strong.

Moderator's comment:
The project has a title, a task and a fairly detailed plan that is followed.

Data

See Appendix for **raw** data.

Moderator's comment:
The raw data is relevant, sufficient in quality but not in quantity and is set up for use.

Lung capacity data was collected with a Spirometer.

Smoker females

Age:	16	16	16	17	17	17	18	19	20	20
Height:	166 cm	170 cm	165 cm	161 cm	171 cm	164 cm	175 cm	165 cm	170 cm	165 cm
Lung capacity:	2500 cc	2500 cc	2200 cc	2600 cc	2800 cc	2200 cc	2500 cc	2600 cc	2800 cc	2800 cc
Athlete:	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Non – smoker females

Age:	17	17	17	17	18	18	18	19	19	19
Height:	170 cm	160 cm	178 cm	156 cm	171 cm	163 cm	164 cm	175 cm	170 cm	163 cm
Lung capacity:	3000 cc	2000 cc	3000 cc	2500 cc	3100 cc	2900 cc	2000 cc	2600 cc	2900 cc	2700 cc
Athlete:	No	No	No	No	No	Yes	No	Yes	No	Yes

Smoker males

Age:	17	17	18	18	18	18	19	19	19	20
Height:	185 cm	173 cm	183 cm	182 cm	175 cm	189 cm	187 cm	186 cm	177 cm	185 cm
Lung capacity:	3300 cc	3300 cc	4000 cc	3900 cc	4000 cc	4000 cc	3500 cc	4600 cc	3500 cc	4100 cc
Athlete:	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No	Yes

Non – smoker males

Age:	16	16	16	16	17	17	18	18	18	18
Height:	179 cm	172 cm	171 cm	175 cm	176 cm	178 cm	175 cm	179 cm	180 cm	176 cm
Lung capacity:	4200 cc	3100 cc	3500 cc	4100 cc	3100 cc	3800 cc	4400 cc	3500 cc	2600 cc	4000 cc
Athlete:	Yes	Yes	No	Yes	No	No	No	Yes	No	Yes

Hypothesis:

Female Non-smokers should have less lung capacity than Male Non-smokers while Female Smokers should also have less lung capacity compared to Male Smokers. Generaly it is expected that lung capacity differs in gender, because females generally have smaller lungs.

Calculating the means

Moderator's comment:
Simple process

Non Smoker Females

Lung capacity: (2000, 2000, 2500, 2600, (2700, 2900), 2900, 3000, 3000, 3100)

Mean: $\frac{(2000+2000+2500+2500+2600+2700+2900+2900+3000+3000+3100)}{10} = \frac{26700}{10} = 2670\,cc$

Smoker females

Lung capacity: (2200, 2200, 2500, 2500, (2500, 2600), 2600, 2800, 2800, 2800)

Mean: $\frac{(2200+2200+2500+2500+2500+2600+2600+2800+2800+2800)}{10} = \frac{25500}{10} = 2550\,cc$

Non- smoker males

Lung capacity: (2600, 3100, 3100, 3500, (3500, 3800), 4000, 4100, 4200, 4400)

Mean: $\frac{(2600+3100+3100+3500+3500+3800+4000+4100+4200+4400)}{10} = \frac{36300}{10} = 3630\,cc$

Smoker males

Lung capacity: (3300, 3300, 3500, 3500, 3900, 4000, 4000, 4000, 4100, 4600)

Mean: $\frac{(3300+3300+3500+3500+3900+4000+4000+4000+4100+4600)}{10} = \frac{38200}{10} = 3820\,cc$

These values confirm that males have larger lung capacity than females and female non-smokers have larger lung capacity than female smokers. However, male smokers have larger lung capacity than male non-smokers. This was an unexpected result but could be explained by the fact that there were more males who played sport and were smokers than non-smokers.

Standard deviation

The standard deviation is going to be calculated to find out how close the data is to the mean in each case. I will take the standard deviation into account when deciding on the groupings for the lung capacity in the chi-squared test.

Process: Find the deviation of each entry from the mean, then square these values. Next find the mean of the squared values and take the square root of this answer.

Moderator's comment:
Simple process

Non – smoker females lung capacity

Mean: 2670

Standard deviation

x_i	x_i – mean	$(x_i - \text{mean})^2$
2000	(–670)	448900
2000	(–670)	448900
2500	(–170)	28900
2600	(–70)	4900
2700	30	900
2900	230	52900
2900	230	52900
3000	330	108900
3000	330	108900
3100	430	184900
		Total: 1441000

$$\text{SD}: \sqrt{\frac{1441000}{10}} = 380$$

Non – smoker males lung capacity:

Mean: 3630 cc

Standard deviation

X_i	X_i – mean	$(X_i - \text{mean})^2$
2600	(–1030)	1060900
3100	(–530)	280900
3100	(–530)	280900
3500	(–130)	16900
3500	(–130)	16900
3800	170	28900
4000	370	136900
4100	470	220900
4200	570	324900
4400	770	592900
		Total: 2961000

$$\textbf{SD}: \sqrt{\frac{2961000}{10}} = 544$$

Smoker females lung capacity:

Mean: 2550 cc

Standard deviation

X_i	X_i – mean	$(X_i$ – mean$)^2$
2200	(–350)	122500
2200	(–350)	122500
2500	(–50)	2500
2500	(–50)	2500
2500	(–50)	2500
2600	50	2500
2600	50	2500
2800	250	62500
2800	250	62500
2800	250	62500
		Total: 445000

SD: $\sqrt{\frac{445000}{10}} = 211$

Smoker male lung capacity:

Mean: 3820 cc

Standard deviation

X_i	X_i – mean	$(X_i$ – mean$)^2$
3300	–520	270400
3300	–520	270400
3500	–320	102400
3500	–320	–102400
3900	80	6400
4000	180	32400
4000	180	32400
4000	180	32400
4100	280	78400
4600	780	608400
		Total: 1503600

SD: $\sqrt{\frac{1503600}{10}} = 392$

	Female Non-smoker	Female Smoker	Male Non-smoker	Male Smoker
Mean lung capacity	2670	2550	3630	3820
Standard deviation of lung capacity	380	211	544	392

The standard deviations shows that male non-smokers have the largest spread of data from the mean and female smoker's lung capacities are the least widespread.

The average lung capacity of athletes and non athletes:

Are you an athlete?
Yes: 20 people

Average of lung capacity of athletes = $\frac{66500}{20} = 3325$
No: 20 people

Average of lung capacity of non-athletes = $\frac{61700}{20} = 3085$

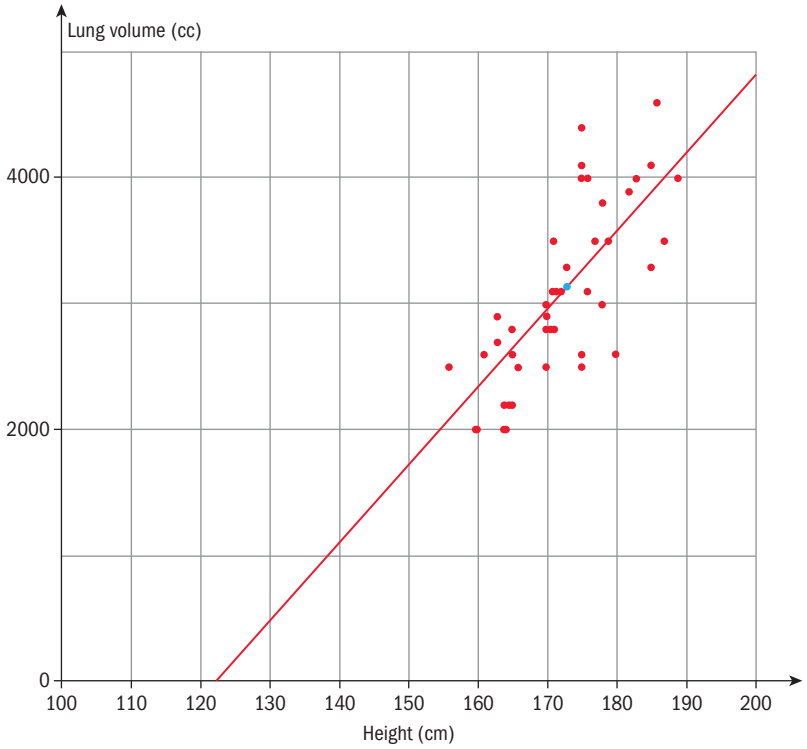
(3325 > 3085)

To conclude these results by looking at athlete lung capacity and non-athlete statistics, it is generally expected that athletes have bigger lungs.

This may not be completely accurate due to some reasons for example; if a person was biking or skating to school every day and they are a non-athlete then their lungs would expand due to continuous inhale exhale motion which means that he/she gained more lung capacity than other non-athlete people who came to school by car etc.

Now I will plot a scatter graph of height and lung capacity, and calculate the correlation coefficient to see if there is a relationship between the two.

Height vs Lung capacity



Method on calculator: L_1– height

L_2– Lung volume

$r = 0.734$

This is moderately strong so I can find the equation of the regression line.

From the GDC the equation of the regression line is:

$y = 63.3x - 7784$

e.g. if $x = 170$ cm then $y = 63.3 \times 170 - 7784 = 2977$

which is a reasonable answer. With this I can conclude that there is a correlation between height and lung capacity.

Moderator's comment: This is also a simple process because everything has been done using technology.

χ^2 Test

I will use the chi-squared test at the 5% level of significance to find out whether or not certain sets of data are independent or not. I will test to see if lung capacity is independent of gender and if lung capacity is independent of smoking. In order for my expected values to be greater than 5 I had to group my lung capacity as shown in the table. I took into consideration the means and standard deviations of the lung capacities of all four groups to help me decide on the range of values to take for the lung capacity groups.

$\mathbf{H_0}$: Lung capacity is independent of gender.

$\mathbf{H_1}$: Lung capacity is dependent of gender.

	< 3000 cc	3000 cc – 3500 cc	> 3500 cc	
Male	1	8	11	20
Female	17	3	0	20
Total	18	11	11	40

Expected values:

$\frac{20\times18}{40} = 9,\ \frac{20\times11}{40} = 5.5,\ \frac{20\times11}{40} = 5.5$

	< 3000 cc	3000 cc – 3500 cc	> 3500 cc	
Male	9	5.5	5.5	20
Female	9	5.5	5.5	20
Total	18	11	11	40

Male

$\frac{(1-9)^2}{9} = \frac{64}{9} = 7.11$

$\frac{(8-5.5)^2}{5.5} = \frac{6.25}{5.5} = 1.14$

$\frac{(11-5.5)^2}{5.5} = \frac{30.25}{5.5} = 5.5$

Female

$\frac{(17-9)^2}{9} = \frac{64}{9} = 7.11$

$\frac{(3-5.5)^2}{5.5} = \frac{6.25}{5.5} = 1.14$

$\frac{(0-5.5)^2}{5.5} = \frac{30.25}{5.5} = 5.5$

Chi squared test statistic = 27.5

Degrees of freedom = $(2-1)\times(3-1)=2$

Critical value = 5.991

At 5% significance level, 27.5 > 5.991, therefore we reject the null hypothesis.

That implies that lung capacity is dependent on gender.

Moderator's comment:
This is a further process

$\mathbf{H_0}$: Lung capacity is independent of smoking.

$\mathbf{H_1}$: Lung capacity is dependent on smoking.

	< 3000 cc	3000cc – 3500 cc	> 3500 cc	
Smoker	10	4	6	20
Non-smoker	8	7	5	20
Total	18	11	11	40

Expected values:

$\frac{20\times18}{40}=9,\ \frac{20\times11}{40}=5.5,\ \frac{20\times11}{40}=5.5$

	< 3000 cc	3000 cc – 3500 cc	> 3500 cc	
Smoker	9	5.5	5.5	20
Non-smoker	9	5.5	5.5	20
Total	18	11	11	0

Smoker

$\frac{(10-9)^2}{2}=\frac{1}{9}=0.11$

$\frac{(8-5.5)^2}{5.5}=\frac{2.25}{5.5}=0.40$

$\frac{(6-5.5)^2}{5.5}=\frac{0.25}{5.5}=0.05$

Non-smoker

$\frac{(8-9)^2}{9}=\frac{1}{9}=0.11$

$\frac{(7-5.5)^2}{5.5}=\frac{2.25}{5.5}=0.40$

$\frac{(5-5.5)^2}{5.5}=\frac{0.25}{5.5}=0.05$

Chi squared test statistic = 1.12
Degrees of freedom = $(2-1)\times(3-1)=2$
Critical value = 5.991

At 5% value 1.12 < 5.991, therefore we accept the null hypothesis that lung capacity is independent of smoking.

Validity:

I relied on the honesty of my school friends regarding their height and whether they smoked or played sports. The results would have been more valid if I had measured the people myself and double checked if they played sports or smoked.

At the start of the data collection every individual was instructed on how to perform the lung capacity task, but some of the individuals did not take it seriously and unsuspected underperformance distorted the data recordings which could have impacted the answer. The end result was positive for the hypothesis that lung capacity is dependent on gender. The negative result for the hypotheses that lung capacity is independent of smoking was unexpected but could have been caused by reasons given above and also the fact that more smokers played sports than non-smokers. As was previously mentioned, students who cycled regularly to school but indicated that they were non-athletes, may have built up a larger lung capacity than those who came to school by car. When using the chi-squared test I made sure that my expected values were more than five otherwise the test would have been invalid. Only when I saw from the scatter graph that there appeared to be a relationship between height and lung capacity did I find the correlation coefficient. Because this was moderately strong then it was relevant for me to find the equation of the regression line. Obviously, if I had tested more students then my results would have been more reliable.

Moderator's comment: Validity has been discussed.

Conclusion

To conclude, comparisons between smokers and non-smokers were carried out in order to see whether smoking influences the lung capacity. It was expected that general lung capacity was going to be bigger for non-smokers and smaller for smokers, both male and female. However, although this was true for the females, the male smokers had a larger lung capacity than the non-smokers. The hypothesis that smoking decreases lung capacity wasn't valid, because calculations that were carried out showed independence between smoking and lung capacity. Gender hypothesis was carried out and the results proved that lung capacity is dependant on gender, which proves a theory that males have a bigger lung capacity than females. Another hypothesis was stated that athletes have bigger lung capacities than non-athletes and this proved to be valid. The correlation coefficient on height v lung capacity proved that there is a relation between the two of them and it is moderately strong. The equation of the regression line was also found and this could be used for prediction purposes.

Bibliography

IB Course Companion: Mathematical Studies; Bedding, Coad, Forrest, Fussey and Waldman; 08/03/2007

Appendix raw data

To find the average lung capacity I added up the three trials and divided by 3

For example: $\frac{(2400 + 2500 + 2600)}{3} = \frac{7500}{3} = 2500$

Female smokers

Age	Height	Lung Capacity 1	Lung Capacity 2	Lung Capacity 3	Average Lung Capacity
16	166	2400	2500	2600	2500
16	170	2550	2500	2450	2500
16	165	2000	2200	2400	2200
17	161	2550	2750	2500	2600
17	171	2800	2850	2750	2800
17	164	2100	2300	2200	2200
18	175	2350	2550	2600	2500
19	165	2700	2500	2600	2600
20	170	2750	2800	2850	2800
20	165	2800	2900	2700	2800

Female non-smokers

Age	Height	Lung Capacity 1	Lung Capacity 2	Lung Capacity 3	Average Lung Capacity
17	170	3050	3100	2850	3000
17	160	2000	2000	2000	2000
17	178	3000	2900	3100	3000
17	156	2550	2500	2450	2500
18	171	3100	3000	3200	3100
18	163	2950	2850	2900	2900
18	164	2000	2050	1950	2000
19	175	2650	2550	2600	2600
19	170	2900	2800	3000	2900
19	163	2650	2700	2750	2700

Male smokers

Age	Height	Lung Capacity 1	Lung Capacity 2	Lung Capacity 3	Average Lung Capacity
17	185	3150	3350	3400	3300
17	173	3200	3250	3450	3300
18	183	3900	4000	4100	4000
18	182	3950	3850	3900	3900
18	175	4050	4100	3850	4000
18	189	3900	4000	4100	4000
19	187	3200	3450	3850	3500
19	186	4400	4650	4750	4600
19	177	3500	3500	3500	3500
20	185	4050	4100	4150	4100

Male non-smokers

Age	Height	Lung Capacity 1	Lung Capacity 2	Lung Capacity 3	Average Lung Capacity
16	179	4000	4200	4400	4200
16	172	3150	3050	3100	3100
16	171	3300	3600	3600	3500
16	175	4000	4000	4300	4100
17	176	3150	3150	3000	3100
17	178	3750	3800	3850	3800
18	175	4250	4450	4500	4400
18	179	3500	3450	3550	3500
18	180	2300	2650	2850	2600
18	176	3950	4150	3900	4000

Summary of moderator's comments

Criterion	Grade	Comment
A	3	The project does have a title, a statement of the task and a description of the plan which is quite detailed. (3 out of 3 marks awarded.)
B	2	Relevant data has been collected. The data is sufficient in quality but not in quantity. However, it has been set up for use in the chi-squared test. The student should have tested more than 10 people in each category. (2 marks awarded, out of a possible 3.)
C	5	All the mathematical processes used are accurate and relevant. (5 out of 5 marks awarded.)
D	2	The interpretations are consistent with the processes used but there is no meaningful discussion. (2 marks awarded, out of a possible 3.)
E	1	There is an attempt made to discuss the validity of the processes used and the data collection process. (1 out of 1 mark awarded.)
F	2	The project is structured but does not always flow well. (2 marks awarded, out of a possible 3.)
G	2	Notation and terminology are correct throughout the project. (2 out of 2 marks awarded.)
Total grade	**16**	