Statistical Inference of the Performance of Portuguese Male and Female Students in Math

Jinanwa Chika

Introduction

This report aims to discover if there is a significant difference between the performance of Portuguese female and male students in math using sample data from two Portuguese schools. The dataset includes students' grades, demographics and school related features. We are interested in knowing if there is convincing evidence of a difference between the performance of male and female Portuguese students in math.

We estimate the average math scores for male and females in Portugal by constructing confidence intervals for each. To determine statistical significance, we will conduct hypothesis tests for the difference of mean math scores for males and females. We further test for practical significance in the average math scores of males and females to determine if the difference in mean scores is large enough to matter. A significant difference in the performance of both genders might be a sign of gender inequality in the Portuguese educational system.

Data Set

This dataset is retrieved from Paulo Cortez from the University of Minho, Portugal (Cortez, 2008). It is a multivariate dataset of 33 attributes which include student grades, demographics, social and school related features. It was collected by using school reports and questionnaires (UCI, 2014). Two datasets are provided which include Math and Portuguese performance of students. We focused only on the math dataset because we are interested only in the performance of students in math. The research question is "Is there convincing evidence of a difference between the performance of male and female Portuguese students and is this difference large enough to matter?"

The variables of interest are gender and average math scores. Average math scores are quantitative, discrete variables. It is quantitative because it has a numerical value. It is also discrete because it is countable, that is, the number of values it can possibly take on is finite. Its unit is marks because the data has scores in points not as a percentage. The student math grades in different grade periods are represented as G1, G2, and G3 (final period grade). The math scores are computed for females and males each by taking the average of all three period grades because taking the average across all gives a better representation of the abilities of a student. For example, if only one period is selected, it is possible that a student performed worse than or better than usual by chance alone. The average of the three periods allows for regression to the mean. Because it is quantitative, statistical and practically significant tests can be performed on it.

Gender is a categorical variable because it is divided into distinct groups without any intrinsic ordering. From the data, gender is considered binary; male or female and this analysis considers it as such. There is no specified ordering of male or female. The unit of gender here is Portuguese student(s). Its unit is Portuguese students because other non-human animals have genders and it is important to distinguish the gender and nationality of which animal is considered here. In this case, we are considering the gender of students. In some cultures, there could be a relationship between gender and grades. For example, in a culture of high gender inequality, females are given strenuous house chores. This reduces the number of study hours and this likely results in poorer performance on tests than their male counterparts.

Methods

The data is downloaded into Python with a shareable Google Drive link, then loaded into a Pandas package for analysis. The data is first sorted out into male and female groups with their respective math scores. Table 1 below shows math score summary statistics for the two gender groups (computed in Appendix A).

Table 1. Summary statistics for the math scores (marks) for the two genders: Male and Female Portuguese students

	Male Students	Female Students
Sample size	$n_1 = 561$	$n_2 = 624$
Sample Mean	$\bar{x}_1 = 11.1$	$\bar{x}_2 = 10.3$
Sample Median	$m_1 = 11$	$m_2 = 10$
Sample Standard Deviation	$s_1 = 3.94$	$s_2 = 3.88$
Sample Mode	$M_1 = 10$	$M_2 = 10$

The distribution of the male students' math scores is shown in *Fig. 1* below (Visualized in Appendix B.

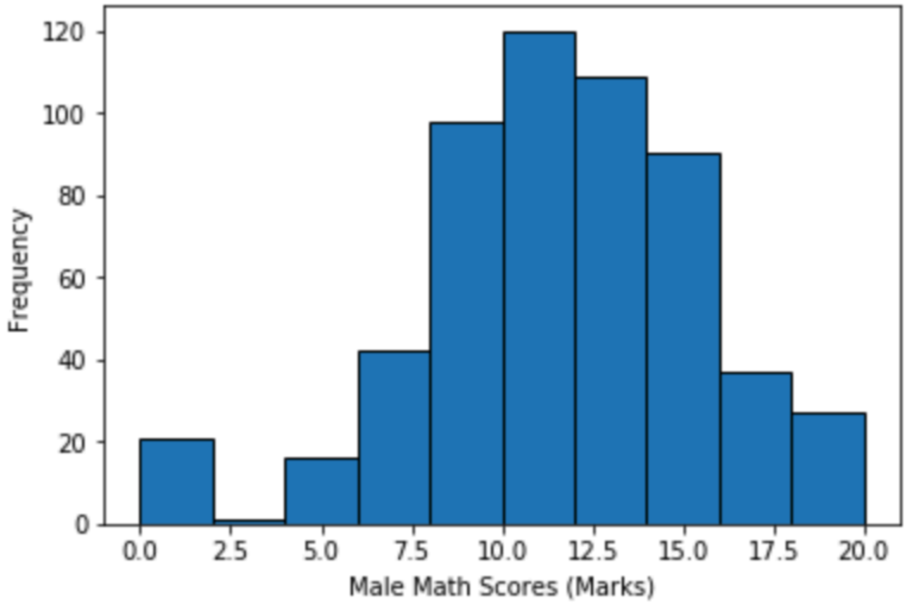


Figure 1. Histogram of the frequency distribution of Male Math scores.

The distribution of the female students' math scores is show in *Fig. 2* below (Visualized in Appendix B).

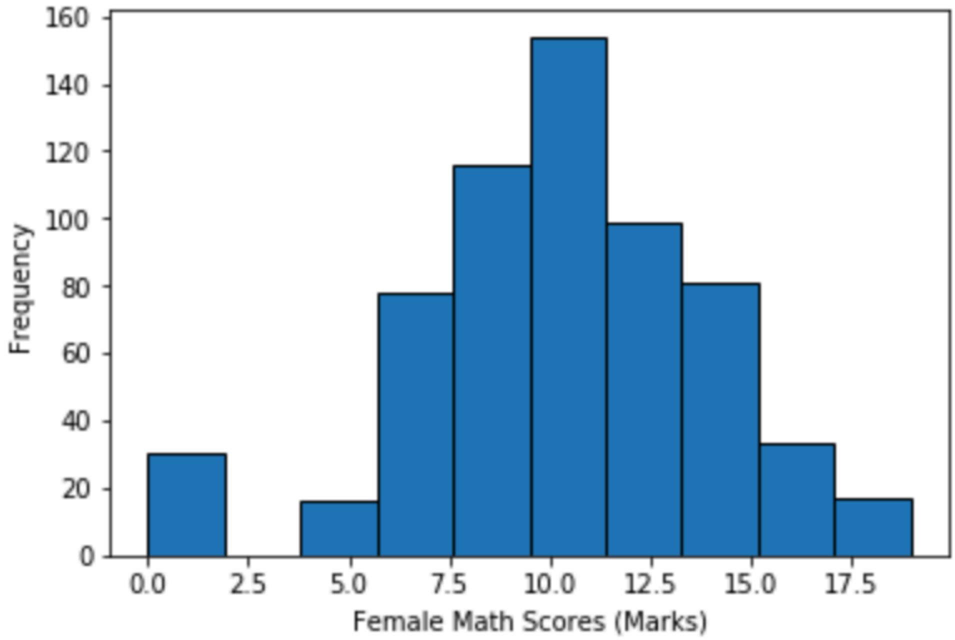


Figure 2. Histogram of the frequency distribution of female math scores

The histogram for male math scores is unimodal and nearly symmetrical with a visible left outlier because the mean and median are nearly equal. The histogram for female math scores is also unimodal and nearly symmetrical with a visible left outlier because the mean and median are almost equal. The male math scores histogram is more spread out than the female because its standard deviation (which represents how spread out values are from the mean) is greater than the females'. The mode (the most frequent marks) for both males and females lie somewhere between 10 and 11 which agrees with the calculated descriptive stat. The mean scores represent the arithmetic average of scores gotten in the math tests.

Confidence Interval for Difference of Means

The difference of the means of math grades of Portuguese male and female students (population parameter) is estimated by constructing a 95% confidence interval which provides a plausible range of values for this parameter. We are using the difference of means because our research question is concerned with the significant difference in the performance of males and females, where performance is measured with the mean math test scores. We intend to use the normal distribution to compute this confidence interval. The conditions to be met are:

1. The individual sample observations are independent of each. I am assuming independence because the sample sizes for males and female students are 561 and 624 respectively and this is very likely to be <10% of the population of Portuguese male and female students. I am also assuming that there was no underlying structure of observations (e.g. consecutive observations) when this data set was collected. These assumptions satisfy this requirement.
2. The individual sample sizes are large, $n \geq 30$ *as a rule of thumb*. Both n_1 (561) *and* n_2 (624) *are* > 30, thus satisfying this condition.
3. The individual population distributions are not strongly skewed. We do not have the population distributions, but the sampling distributions are not strongly skewed as shown in *Fig 1. and Fig2.* and the sample sizes for both distributions are large enough (561 and 624) to accommodate any skew. Thus, this condition is met.

Because these above conditions are met, we can proceed with computing the confidence interval using the normal distribution as guaranteed by the Central Limit Theorem (CLT). This theorem says that the sampling distribution of the sample mean tends to a normal distribution as the sample size gets larger ($n \geq 30$ *as a rule of thumb*). Because my sample sizes are large enough (561 and 624), the CLT ensures the normal models for $\bar{x}_1$ *and* $\bar{x}_2$. Thus, the sample standard deviations s_1 *and* s_2 are good estimates for the standard deviations of the math grades of Portuguese male and female students respectively (population standard deviations).

Since we are considering the difference of means, the point estimate is $\bar{x}_1 - \bar{x}_{2.} =$ 0.8 .The Standard Error (SE) which describes the typical error associated with the point estimate is calculated with the formula:

$SE = \sqrt{\frac{({s_1}^2 \times n_2) + ({s_2}^2 \times n_1)}{n_1 \times n_2}}$ $with\ the\ parameters\ having\ their\ usual\ meanings$. The full calculation is shown in Appendix C. The *SE= 0.23*. The confidence intervals for the Portuguese students are calculated using the formula $point\ estimate\ \pm z^* \times SE$ where $z^* \times SE$ is the margin of error. The full calculations are shown in Appendix D. The 95% confidence interval for the difference between Portuguese male students' math grades and Portuguese female students' math grades is (0.3, 1.19). This means that we are 95% confident that male Portuguese students' math grades are higher than female Portuguese students' math grades by 0.3 to 1.19 marks.

Hypothesis Testing for the Difference of Means

The null hypothesis H_0 represents the case of no difference between the groups. Therefore:

1. H_0: There is no difference between the average math grades of Portuguese male students and those of Portuguese female students. Mathematically, this means $\bar{x}_1 = \bar{x}_2$
2. $H_{A:}$ The alternate hypothesis H_A means that there is some difference between the average math grades of Portuguese male students and those of Portuguese female students. Mathematically, this means $\bar{x}_1 \neq \bar{x}_2$

To apply the normal distribution, we check the 3 conditions, which have been satisfied with justifications above. The chosen significance level α is 0.05, corresponding to the probability of making a type 1 error. A type 1 error is rejecting the null hypothesis when it is true. In this case, it means saying that there is some difference between the math grades of Portuguese male students and those of Portuguese female students when there is actually no difference in real life. A type 2 error is failing to reject the null hypothesis when the alternative is true. In this case, it means saying that there is no difference between the average math grades of Portuguese males and average math grades of Portuguese females when there is indeed a difference in real life. Our research is focused on identifying a difference in performance, so a Type 2 error is costlier, hence the choice for a larger $\alpha = 0.05$. We will perform a 2-tailed hypothesis test since we are only concerned with detecting a statistically significant difference between the two scores which could be in any direction. The Z score for the normal model is computed in Appendix E. The Z score is 3.29. The p -value, which is the probability of observing data that is at least as extreme as the difference of means if the null hypothesis is true, is 0.001. (Calculated in Appendix F). Because $p < \alpha$ $(0.001 < 0.05)$ for the two-tailed test, we reject the null hypothesis. Therefore, we reject that there is no difference between the average math grades of Portuguese male students and those of Portuguese female students. That is there is a statistically significant difference between the two means.

Practical Significance

The effect size is calculated (in Appendix G) using Cohen's d with the formula $Cohen's\ d = \frac{\bar{x}_1 - \bar{x}_{2.}}{\sqrt{\frac{(s_1{}^2 \times n_1 - 1) + (s_2{}^2 \times n_2 - 1)}{n_1 + n_2 - 2}}}$. Cohen's effect size value (d=0.19) suggested a low practical significance.

Results and Conclusion

The 95% confidence interval for the difference between Portuguese male students' math grades and Portuguese female students' math grades is (0.3, 1.19). This means that we are 95% confident that male Portuguese students' math grades are higher than female Portuguese students' math grades by 0.3 to 1.19 marks. The observed p value, $p < \alpha$ (that is $0.001 < 0.05$) for the two-tailed test, suggests that there is a statistically significant difference (strong evidence to reject the null hypothesis) between the average math grades of Portuguese males and that of Portuguese females, therefore we reject the null hypothesis. Thus, we reject that there is no difference between the average math grades of Portuguese male students and those of Portuguese female students. Cohen's effect size with d = 0.19 suggest little practical significance. Thus, even though there is a statistically significant difference between the performance of Portuguese males in math and Portuguese females in math, this difference is not large enough to matter in real life.

Inference

Of a sizeable sample of the math grades of males and females in two Portuguese schools, boys performed better than girls with 0.8 marks. Therefore, on average, we are 95% confident that male Portuguese students perform slightly better than female Portuguese students with higher grades between 0.3 to 1.19 marks. However, this difference is not large enough to matter. This is a type of induction called statistical generalization. It is statistical generalization because we made use of statistical inference tools (such as point estimates, confidence intervals etc.) on a sample (Portuguese students in two schools) to make generalizations about a population (all Portuguese students). This induction is strong because the 95% confidence interval means a 95% confidence of capturing the population difference with the confidence interval. This argument can be made stronger by using a 99% confidence interval which will entail a range of (0.16,1.33) (calculated in Appendix H) which is a wider range.

References

Cortez, P. (2008, June 11). USING DATA MINING TO PREDICT SECONDARY SCHOOL STUDENT PERFORMANCE. Retrieved from http://www3.dsi.uminho.pt/pcortez/student.pdf

UCI. (2014, October 10). *Student Performance Data Set*. Retrieved from Machine Learning Repository website: https://archive.ics.uci.edu/ml/datasets/Student+Performance#

Appendix

Appendix A: Import and Analyze Data

In [1]:
```
import pandas as pd
url_math = 'https://docs.google.com/spreadsheets/d/1MmUvgHBSpbZZ-ytuDrnDZdfV2tZo6dpo/export?format=csv'
#download student math file to program
math_data = pd.read_csv(url_math) #load file into a pandas dataframe
math_data #show table
```

Out[1]:

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	famrel	freetime	goout	Dalc	Walc	health	absences	G1	G2	G3
0	GP	F	18	U	GT3	A	4	4	at_home	teacher	...	4	3	4	1	1	3	6	5	6	6
1	GP	F	17	U	GT3	T	1	1	at_home	other	...	5	3	3	1	1	3	4	5	5	6
2	GP	F	15	U	LE3	T	1	1	at_home	other	...	4	3	2	2	3	3	10	7	8	10
3	GP	F	15	U	GT3	T	4	2	health	services	...	3	2	2	1	1	5	2	15	14	15
4	GP	F	16	U	GT3	T	3	3	other	other	...	4	3	2	1	2	5	4	6	10	10
5	GP	M	16	U	LE3	T	4	3	services	other	...	5	4	2	1	2	5	10	15	15	15
6	GP	M	16	U	LE3	T	2	2	other	other	...	4	4	4	1	1	3	0	12	12	11
7	GP	F	17	U	GT3	A	4	4	other	teacher	...	4	1	4	1	1	1	6	6	5	6
8	GP	M	15	U	LE3	A	3	2	services	other	...	4	2	2	1	1	1	0	16	18	19
9	GP	M	15	U	GT3	T	3	4	other	other	...	5	5	1	1	1	5	0	14	15	15
10	GP	F	15	U	GT3	T	4	4	teacher	health	...	3	3	3	1	2	2	0	10	8	9
11	GP	F	15	U	GT3	T	2	1	services	other	...	5	2	2	1	1	4	4	10	12	12
12	GP	M	15	U	LE3	T	4	4	health	services	...	4	3	3	1	3	5	2	14	14	14
13	GP	M	15	U	GT3	T	4	3	teacher	other	...	5	4	3	1	2	3	2	10	10	11
14	GP	M	15	U	GT3	A	2	2	other	other	...	4	5	2	1	1	3	0	14	16	16
15	GP	F	16	U	GT3	T	4	4	health	other	...	4	4	4	1	2	2	4	14	14	14
16	GP	F	16	U	GT3	T	4	4	services	services	...	3	2	3	1	2	2	6	13	14	14
17	GP	F	16	U	GT3	T	3	3	other	other	...	5	3	2	1	1	4	4	8	10	10
18	GP	M	17	U	GT3	T	3	2	services	services	...	5	5	5	2	4	5	16	6	5	5
19	GP	M	16	U	LE3	T	4	3	health	other	...	3	1	3	1	3	5	4	8	10	10
20	GP	M	15	U	GT3	T	4	3	teacher	other	...	4	4	1	1	1	1	0	13	14	15
21	GP	M	15	U	GT3	T	4	4	health	health	...	5	4	2	1	1	5	0	12	15	15

22	GP	M	16	U	LE3	T	4	2	teacher	other	...	4	5	1	1	3	5	2	15	15	16
23	GP	M	16	U	LE3	T	2	2	other	other	...	5	4	4	2	4	5	0	13	13	12
24	GP	F	15	R	GT3	T	2	4	services	health	...	4	3	2	1	1	5	2	10	9	8
25	GP	F	16	U	GT3	T	2	2	services	services	...	1	2	2	1	3	5	14	6	9	8
26	GP	M	15	U	GT3	T	2	2	other	other	...	4	2	2	1	2	5	2	12	12	11
27	GP	M	15	U	GT3	T	4	2	health	services	...	2	2	4	2	4	1	4	15	16	15
28	GP	M	16	U	LE3	A	3	4	services	other	...	5	3	3	1	1	5	4	11	11	11
29	GP	M	16	U	GT3	T	4	4	teacher	teacher	...	4	4	5	5	5	5	16	10	12	11
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
365	MS	M	18	R	GT3	T	1	3	at_home	other	...	3	3	4	2	4	3	4	10	10	10
366	MS	M	18	U	LE3	T	4	4	teacher	services	...	4	2	2	2	2	5	0	13	13	13
367	MS	F	17	R	GT3	T	1	1	other	services	...	5	2	1	1	2	1	0	7	6	0
368	MS	F	18	U	GT3	T	2	3	at_home	services	...	5	2	3	1	2	4	0	11	10	10
369	MS	F	18	R	GT3	T	4	4	other	teacher	...	3	2	2	4	2	5	10	14	12	11
370	MS	F	19	U	LE3	T	3	2	services	services	...	3	2	2	1	1	3	4	7	7	9
371	MS	M	18	R	LE3	T	1	2	at_home	services	...	4	3	3	2	3	3	3	14	12	12
372	MS	F	17	U	GT3	T	2	2	other	at_home	...	3	4	3	1	1	3	8	13	11	11
373	MS	F	17	R	GT3	T	1	2	other	other	...	3	5	5	1	3	1	14	6	5	5
374	MS	F	18	R	LE3	T	4	4	other	other	...	5	4	4	1	1	1	0	19	18	19
375	MS	F	18	R	GT3	T	1	1	other	other	...	4	3	2	1	2	4	2	8	8	10
376	MS	F	20	U	GT3	T	4	2	health	other	...	5	4	3	1	1	3	4	15	14	15
377	MS	F	18	R	LE3	T	4	4	teacher	services	...	5	4	3	3	4	2	4	8	9	10
378	MS	F	18	U	GT3	T	3	3	other	other	...	4	1	3	1	2	1	0	15	15	15
379	MS	F	17	R	GT3	T	3	1	at_home	other	...	4	5	4	2	3	1	17	10	10	10
380	MS	M	18	U	GT3	T	4	4	teacher	teacher	...	3	2	4	1	4	2	4	15	14	14
381	MS	M	18	R	GT3	T	2	1	other	other	...	4	4	3	1	3	5	5	7	6	7
382	MS	M	17	U	GT3	T	2	3	other	services	...	4	4	3	1	1	3	2	11	11	10
383	MS	M	19	R	GT3	T	1	1	other	services	...	4	3	2	1	3	5	0	6	5	0
384	MS	M	18	R	GT3	T	4	2	other	other	...	5	4	3	4	3	3	14	6	5	5
385	MS	F	18	R	GT3	T	2	2	at_home	other	...	5	3	3	1	3	4	2	10	9	10
386	MS	F	18	R	GT3	T	4	4	teacher	at_home	...	4	4	3	2	2	5	7	6	5	6
387	MS	F	19	R	GT3	T	2	3	services	other	...	5	4	2	1	2	5	0	7	5	0
388	MS	F	18	U	LE3	T	3	1	teacher	services	...	4	3	4	1	1	1	0	7	9	8
389	MS	F	18	U	GT3	T	1	1	other	other	...	1	1	1	1	1	5	0	6	5	0
390	MS	M	20	U	LE3	A	2	2	services	services	...	5	5	4	4	5	4	11	9	9	9
391	MS	M	17	U	LE3	T	3	1	services	services	...	2	4	5	3	4	2	3	14	16	16
392	MS	M	21	R	GT3	T	1	1	other	other	...	5	5	3	3	3	3	3	10	8	7
393	MS	M	18	R	LE3	T	3	2	services	other	...	4	4	1	3	4	5	0	11	12	10
394	MS	M	19	U	LE3	T	1	1	other	at_home	...	3	2	3	3	3	5	5	8	9	9

395 rows × 33 columns

[2]:
```python
female_m_g1 = list(math_data["G1"][math_data["sex"] == "F"]) #extract only female G1 math scores into list
female_m_g2 = list(math_data["G2"][math_data["sex"] == "F"]) #extract only female G2 math scores into list
female_m_g3 = list(math_data["G3"][math_data["sex"] == "F"]) #extract only female G3 math scores into list
fe_composite_math = [] #initialize all female math scores list
for a in female_m_g1: #loop through each element in female_m_gl
    fe_composite_math+=[a] #append element to fe_composite_math
for b in female_m_g2:
    fe_composite_math+=[b]
for c in female_m_g3:
    fe_composite_math+=[c]
print(fe_composite_math,'\n','The total number of female math scores is',len(fe_composite_math))
```

```
[5, 5, 7, 15, 6, 6, 10, 10, 14, 13, 8, 10, 6, 8, 12, 14, 7, 10, 8, 11, 7, 12, 11, 8, 10, 8, 14, 15, 10, 10, 8, 10, 1
0, 16, 7, 8, 16, 8, 11, 11, 5, 7, 9, 7, 8, 13, 7, 16, 7, 11, 7, 8, 11, 7, 7, 10, 7, 14, 7, 10, 16, 13, 8, 7, 7, 12,
8, 10, 12, 11, 4, 16, 9, 14, 8, 6, 10, 10, 11, 9, 14, 6, 14, 8, 10, 13, 16, 9, 12, 8, 11, 8, 14, 18, 9, 16, 8, 9, 7,
10, 10, 7, 11, 9, 7, 8, 12, 8, 14, 6, 7, 9, 6, 6, 16, 13, 9, 16, 12, 13, 13, 13, 7, 14, 10, 17, 10, 9, 12, 6, 9, 15,
11, 10, 12, 10, 9, 12, 8, 10, 18, 13, 15, 12, 18, 10, 10, 14, 12, 15, 17, 14, 12, 9, 14, 13, 15, 13, 8, 9, 11, 11, 1
3, 11, 11, 12, 16, 10, 14, 12, 7, 8, 10, 16, 14, 7, 16, 9, 11, 9, 11, 13, 13, 10, 12, 12, 18, 13, 11, 16, 12, 7, 11,
14, 7, 13, 6, 19, 8, 15, 8, 15, 10, 10, 6, 7, 7, 6, 6, 5, 8, 14, 10, 5, 8, 12, 14, 14, 10, 9, 9, 7, 12, 13, 10, 10,
8, 12, 7, 13, 13, 10, 13, 9, 15, 16, 11, 8, 10, 9, 10, 15, 7, 9, 16, 6, 12, 11, 5, 6, 10, 9, 7, 14, 7, 17, 6, 10, 10,
9, 14, 9, 6, 11, 8, 15, 10, 13, 15, 13, 7, 10, 8, 0, 0, 13, 11, 0, 0, 16, 11, 14, 11, 7, 11, 10, 11, 8, 15, 7, 14, 7,
```

```
11, 13, 17, 9, 13, 7, 12, 8, 15, 18, 9, 16, 10, 9, 6, 10, 9, 7, 12, 9, 7, 8, 13, 10, 15, 6, 7, 10, 6, 5, 16, 13, 8, 1
5, 10, 13, 12, 11, 0, 12, 9, 18, 9, 10, 10, 0, 9, 14, 11, 10, 12, 9, 8, 12, 9, 9, 18, 12, 15, 12, 18, 9, 8, 13, 10, 1
2, 17, 12, 10, 9, 12, 10, 13, 11, 8, 10, 11, 11, 13, 9, 11, 14, 15, 9, 14, 14, 0, 8, 9, 15, 13, 8, 15, 10, 12, 8, 10,
13, 15, 9, 13, 12, 16, 13, 11, 15, 11, 6, 10, 12, 7, 11, 5, 18, 8, 14, 9, 15, 10, 9, 5, 5, 9, 5, 6, 6, 10, 15, 10, 6,
9, 12, 14, 14, 10, 8, 8, 6, 11, 13, 11, 9, 6, 11, 7, 13, 13, 11, 13, 10, 15, 16, 11, 11, 9, 9, 10, 15, 6, 8, 16, 5, 1
1, 11, 5, 6, 10, 8, 6, 14, 8, 18, 6, 10, 10, 10, 14, 8, 6, 11, 8, 16, 10, 13, 15, 13, 8, 11, 9, 0, 0, 12, 11, 0, 0, 1
5, 11, 13, 11, 0, 11, 10, 12, 10, 16, 0, 14, 0, 9, 11, 17, 8, 12, 9, 13, 9, 15, 18, 10, 16, 10, 10, 6, 11, 9, 7, 13,
10, 7, 8, 14, 10, 15, 4, 8, 10, 6, 0, 17, 14, 7, 15, 12, 14, 12, 11, 0, 13, 0, 18, 9, 0, 11, 0, 9, 14, 11, 10, 12, 9,
8, 12, 10, 11, 19, 12, 15, 13, 18, 0, 8, 14, 11, 14, 18, 12, 10, 0, 13, 11, 13, 11, 0, 9, 10, 11, 13, 9, 11, 15, 15,
9, 14, 14, 0, 0, 0, 15, 13, 0, 17, 10, 11, 0, 10, 14, 15, 9, 13, 11, 16, 13, 10, 15, 12, 0, 10, 11, 9, 11, 5, 19, 10,
15, 10, 15, 10, 10, 6, 0, 8, 0]
 The total number of female math scores is 624
```

```
male_m_g1 = list(math_data["G1"][math_data["sex"] == "M"]) #extract only male G1 math scores into list
male_m_g2 = list(math_data["G2"][math_data["sex"] == "M"]) #extract only male G2 math scores into list
male_m_g3 = list(math_data["G3"][math_data["sex"] == "M"]) #extract only male G3 math scores into list
m_composite_math = [] #initialize all male math scores list
for g in male_m_g1:#loop through each element in male_m_g1
    m_composite_math+=[g] #append element to m_composite_math
for h in male_m_g2:
    m_composite_math+=[h]
for i in male_m_g3:
    m_composite_math+=[i]
print(m_composite_math,'\n','The total number of male math scores is',len(m_composite_math))
```

```
[15, 12, 16, 14, 14, 10, 14, 6, 8, 13, 12, 15, 13, 12, 15, 11, 10, 9, 17, 17, 8, 12, 15, 15, 12, 19, 8, 19, 15, 11, 1
4, 9, 13, 13, 10, 12, 9, 11, 8, 10, 11, 15, 11, 8, 11, 11, 7, 16, 10, 16, 16, 10, 18, 18, 9, 15, 11, 13, 9, 14, 16, 1
4, 13, 7, 18, 9, 10, 14, 7, 9, 5, 7, 8, 6, 12, 5, 11, 16, 17, 10, 7, 5, 7, 10, 5, 12, 10, 6, 13, 13, 10, 6, 10, 10,
9, 12, 12, 11, 15, 8, 7, 8, 13, 17, 9, 12, 6, 6, 12, 12, 10, 11, 11, 14, 9, 11, 14, 7, 12, 10, 6, 13, 18, 12, 6, 3, 1
3, 6, 7, 6, 8, 8, 7, 11, 15, 8, 13, 17, 9, 10, 15, 9, 10, 8, 11, 12, 15, 15, 12, 14, 14, 16, 11, 15, 17, 8, 15, 13,
9, 14, 11, 9, 10, 16, 16, 10, 11, 8, 13, 8, 8, 13, 10, 13, 10, 13, 14, 15, 7, 11, 6, 6, 9, 14, 10, 11, 8, 15, 12, 18,
15, 14, 10, 16, 5, 10, 14, 15, 15, 13, 12, 16, 11, 12, 11, 16, 16, 10, 14, 16, 16, 12, 18, 8, 19, 15, 11, 15, 10, 13,
15, 10, 12, 9, 11, 8, 12, 10, 15, 10, 6, 13, 15, 7, 17, 13, 18, 18, 13, 19, 19, 9, 15, 13, 14, 7, 13, 14, 11, 13, 4,
18, 0, 0, 12, 9, 9, 0, 6, 9, 5, 13, 0, 8, 12, 15, 12, 6, 9, 0, 10, 8, 11, 10, 5, 15, 11, 9, 5, 8, 10, 8, 13, 12, 12,
15, 9, 8, 9, 14, 15, 9, 12, 7, 6, 13, 11, 8, 11, 9, 13, 7, 9, 13, 7, 12, 11, 0, 12, 18, 12, 8, 5, 15, 8, 10, 9, 9, 1
2, 9, 11, 14, 8, 12, 17, 9, 9, 14, 9, 11, 8, 9, 10, 14, 13, 11, 13, 12, 15, 11, 14, 18, 9, 12, 11, 12, 15, 10, 8, 10,
15, 15, 10, 13, 7, 13, 7, 8, 11, 10, 12, 10, 13, 12, 14, 6, 11, 5, 5, 9, 16, 8, 12, 9, 15, 11, 19, 15, 14, 11, 16, 5,
10, 15, 15, 16, 12, 11, 15, 11, 11, 12, 17, 16, 12, 15, 18, 15, 12, 18, 11, 20, 14, 10, 15, 9, 12, 15, 10, 14, 10, 1
0, 10, 12, 11, 15, 10, 7, 14, 15, 5, 17, 14, 18, 18, 13, 19, 19, 9, 16, 14, 13, 8, 13, 15, 13, 12, 0, 18, 0, 0, 12,
0, 9, 0, 0, 10, 0, 14, 0, 8, 13, 15, 12, 0, 7, 0, 10, 7, 12, 10, 0, 16, 10, 9, 6, 9, 11, 8, 12, 11, 11, 15, 10, 8, 1
0, 14, 16, 10, 13, 8, 8, 13, 12, 9, 11, 9, 13, 6, 10, 13, 0, 12, 12, 0, 12, 18, 13, 8, 5, 15, 8, 10, 8, 8, 12, 8, 11,
14, 8, 12, 17, 10, 10, 14, 9, 10, 8, 10, 11, 14, 15, 11, 14, 11, 16, 10, 13, 18, 8, 12, 11, 11, 16, 10, 8, 0, 15, 16,
9, 13, 8, 13, 8, 8, 11, 10, 12, 10, 13, 12, 14, 7, 10, 0, 5, 9, 16, 7, 10, 9]
 The total number of male math scores is 561
```

In [4]:

```
import numpy as np
import statistics as stat
def median(array):
    """Function that returns the median of an array object"""
    n=len(array) #gets the number of entries in the list
    array.sort() #sorts the elements in the list in ascending order
    if n % 2 == 0: #there is an even number of elements in list
        middle1 = array[n//2] #first pair of middle number
        middle2 = array[n//2 - 1] #second pair of middle number
        median = (middle1 + middle2)/2 #average the middle pairs of elements
    else: #there is an odd number of elements in list
        median = array[n//2] #element in the middle is median
    if median == stat.median(array):
    #compares median computed by function to the value obtained by statistics module
        return median


def mode(array):
    """Function that returns the mode of a list/array using a counting dictionary"""
    counters={}# Initialize counting dictionary
    list_of_frequency=[] #initializes list of frequencies of repectives data
    for elem in array: #iterates through each element in list
        if elem not in counters: #element not in dictionary key
            counters[elem]=1 # First occurrence, add new number to key and count it once
        else:
            counters[elem] +=1 # Increment existing counter
    for frequency in counters.values(): #iterates through all values in dictionary
        list_of_frequency += [frequency] #add this number to a list
    list_of_frequency.sort() #Sorts the elements of the list in ascending order.
    highest_frequency = list_of_frequency[-1] #gets last element in list;largest element in list
    for keys in counters:
        if counters[keys]==highest_frequency: #if value associated with key is the highest
            mode = keys #assign the mode to this key
    if mode == stat.mode(array): #compare with the mode generated by statistics module
        return mode


def mean(array):
    """Function that returns the mean of list/array object"""
    length = len(array) #total number of elements in list
    total=0 #initializes the sum of elements in list
    for elem in array: #iterates through each element
        total += elem #add element to total
```

```
    mean=total/length #compute the average
    if mean == np.mean(array):  #compare with mean generated by numpy module
        return mean


def standard_deviation(array):
    """Function that returns the standard deviation of a list/array object"""
    total = 0 #initializes sum of elements in the list
    total_square_deviation = 0 #initializes the sum of the squares of the deviations
    deviation_list = [] #initializes list which will contain squares of the deviations
    n = len(array) #number of elements in list
    for elem in array: #loops through each elem in array
        total += elem #adds the elem to total
    average = total/len(array) #computes mean
    for num in array: #loops through each element in array
        deviation_square = (num-average)**2 #square difference between number and mean
        deviation_list += [deviation_square] #add the square of deviations of each element to the list
    for num in deviation_list: #loops through each element in list
        total_square_deviation += num #sums the square of the deviations
    variance = total_square_deviation/(n-1) #computes variance with Bessel's correction for sample data
    std_dev = variance **0.5 #square root of variance

    return std_dev
```

```
print('The average female math score is',mean(fe_composite_math))
print('The average male math score is',mean(m_composite_math))
```

```
The average female math score is 10.325320512820513
The average male math score is 11.073083778966131
```

```
print('The median female math score is',median(fe_composite_math))
print('The median male math score is',median(m_composite_math))
```

```
The median female math score is 10.0
The median male math score is 11
```

```
print('The modal female math score is',mode(fe_composite_math))
print('The modal male math score is',mode(m_composite_math))
```

```
The modal female math score is 10
The modal male math score is 10
```

```
]: print('The standard deviation for female math score is',standard_deviation(fe_composite_math))
   print('The standard deviation for male math score is',standard_deviation(m_composite_math))
```

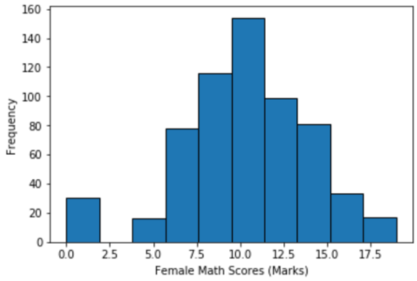
```
The standard deviation for female math score is 3.879600502322402
The standard deviation for male math score is 3.939725236309178
```


```
]: print('The standard deviation for female math score is',standard_deviation(fe_composite_math))
   print('The standard deviation for male math score is',standard_deviation(m_composite_math))
```

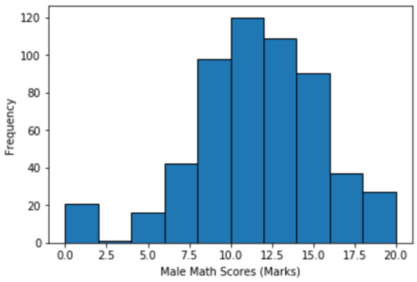
```
The standard deviation for female math score is 3.879600502322402
The standard deviation for male math score is 3.939725236309178
```


Appendix B: Data Visualizations

In [9]:
```
%matplotlib inline
import math
#displays inline image
import matplotlib.pyplot as plt  #imports matplotlib
def bin_number(array):
    """Function provides a recommended number of bins using Sturges formula floor(1+log2(n)), where n is the
    number of data points. It assumes the data follow a normal distribution."""
    n = len(array)
    bin_number = math.floor(1+math.log2(n)) #computes no of bins with sturges formula
    return bin_number

plt.hist(fe_composite_math, bins=bin_number(fe_composite_math), edgecolor='black') #plots histogram
plt.xlabel('Female Math Scores (Marks)') #label x axis
plt.ylabel('Frequency') #label y-axis

plt.show()
```



In [10]:
```
plt.hist(m_composite_math, bins=bin_number(m_composite_math), edgecolor='black')
#plots histogram, using sturges formula because test scores are known to follow normal distributions
plt.xlabel('Male Math Scores (Marks)') #label x axis
plt.ylabel('Frequency') #label y-axis

plt.show()
```



Appendix C: Standard Error of Difference of Sample Means

In [16]:
```
def standard_error_difference_of_means (array1,array2):
    """Function accepts two arrays and computes the standard error of the difference of two means of the arrays
    using the formula (s1**2/n1 + s2**2/n2)**0.5 where n is the sample size and s is the standard deviation of
    sample """
    sample_std1 = standard_deviation(array1) #standard std of sample with bessel's correction
    sample_var1 = sample_std1**2 #computes the sample variance
    sample_std2 = standard_deviation(array2)
    sample_var2 = sample_std2**2 #computes the variance
    sample_size1=len(array1)
    sample_size2 = len(array2)#length of array
    standard_error = ((sample_var1/sample_size1) + (sample_var2/sample_size2))**0.5
    #computes standard error using the formula (s1**2/n1 + s2**2/n2)**0.5
    return standard_error
print('The standard error for the difference of average female math score and male math score means is'\
      ,standard_error_difference_of_means(m_composite_math,fe_composite_math))
```

```
The standard error for the difference of average female math score and male math score means is 0.22757002246734911
```

Appendix D: Confidence Interval for the difference of Sample Means

```
def confidence_interval_difference_of_means(array1,array2,confidence_level):
    """Function accepts two arrays and constructs a confidence interval for difference of two means using the
    formula point estimate +- t* * SE (standard error) where point estimate is xbar1-xbar2"""
    difference = mean(array1)-mean(array2) #difference of sample means
    standard_error = standard_error_difference_of_means (array1,array2) #calculate standard error
    df = min(len(array1),len(array2))
    margin_of_error = standard_error * (st.norm.ppf(confidence_level)) #calculate margin of error with t* * SE
    #st.norm.ppf gets the z score for a 2tailed percentile
    lower_bound = difference - margin_of_error #constructs confidence interval's lower bound
    upper_bound = difference + margin_of_error #constructs confidence interval's upper bound
    confidence_interval = [lower_bound,upper_bound] #creates confidence interval
    #array_name = input('Enter the name of the variable')
    return confidence_interval
print('The 95% confidence interval for the difference of means is'\
      ,confidence_interval_difference_of_means(m_composite_math,fe_composite_math,0.95))
```

Appendix E: Calculating the Z score for the Hypothesis Test

```
def Z_score_hypothesis_test(array1,array2, null_value):
    """Function accepts 2 arrays and a null value and calculates the Z score for the difference of two means using
    the formula Z= difference of means - null value/SE"""
    difference = mean(array1)-mean(array2) #difference of sample means
    standard_error = standard_error_difference_of_means (array1,array2) #calculate standard error
    z_score = (difference-null_value)/standard_error
    #calculates the z score with formula Z= difference of means - null value/SE
    return z_score
print('The Z-score is',Z_score_hypothesis_test(m_composite_math,fe_composite_math,0))
```

```
The Z-score is 3.285860141148005
```

Appendix F: Calculating the P-value for the Hypothesis Test

```
def p_value(array1,array2,null_value,tail_number):
    """Function accepts 2 arrays and a null value and calculates the p value for the hypothesis test of difference
    of two means using the formula 1- percentile(Zscore)"""
    z_score = Z_score_hypothesis_test(array1,array2, null_value) #computes Z score
    z_percentile = (st.norm.cdf(z_score)) #converts z score to a 1-tailed percentile
    p_value = (1-z_percentile)*tail_number #calculates the p_value using the formula 1-percentile(Zscore)
    return p_value
print('The p value is',p_value(m_composite_math,fe_composite_math,0,2))
```

```
The p value is 0.0010167147820114852
```

Appendix G: Calculating Cohen's d

```
def cohen_d(array1,array2):
    """Function accepts 2 arrays and computes the effect size using cohen's d formula given by
    d=mean1-mean2/(((n1-1)*(s1**2)) + ((n2-1)*(s2**2)))/n1+n2-2)**0.5"""
    mean_difference = mean(array1)-mean(array2) #difference of sample means
    len1 = len(array1)
    len2 = len(array2)
    var1 = (standard_deviation(array1))**2
    var2 = (standard_deviation(array2))**2
    denom = len1 +len2-2
    denominator = ((((len1-1)*var1) + ((len2-1)*var2))/denom)**0.5
    cohen_d = mean_difference/denominator
    return cohen_d
print ("Cohen's effect size is,",cohen_d(m_composite_math,fe_composite_math))
```

```
Cohen's effect size is, 0.191332998929746
```

Appendix H: Constructing 99% confidence interval

```
print('The 99% confidence interval for the difference of means is'\
      ,confidence_interval_difference_of_means(m_composite_math,fe_composite_math,0.995))
```

```
The 99% confidence interval for the difference of means is [0.1615817336649389, 1.3339447986262978]
```