

1 Purpose of the test

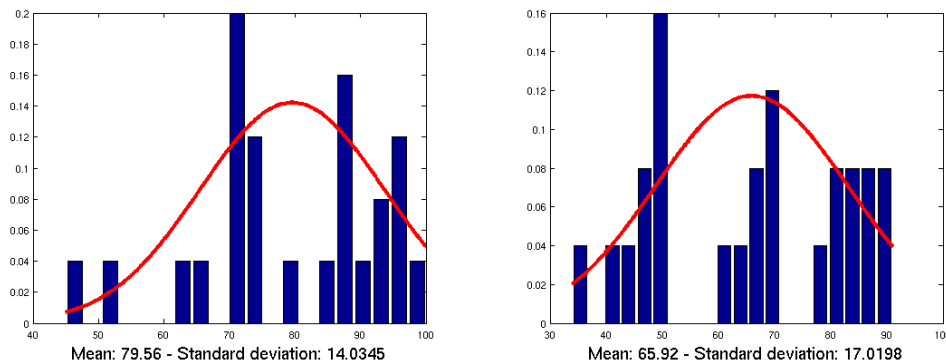
In this test I intend to test the belief of mine that students grades in the exams are not completely random but they reflect to a certain extent the abilities of the students. To this extent I will measure if there is a statistically significant correlation between the scores my students obtained on the first midterm and scores that my students obtained in the second midterm. I don't expect to find a complete correlation because I do believe that grades reflect also the amount of work that a student can put in the course and that can fluctuate over time.

2 The data

The data analyzed is summarized in the following table. For every student, the scores on their two first midterms have been reported. To protect the privacy of my students, the scores have been permuted and slightly modified. The results might therefore not apply to the real situation but they still constitute a good example.

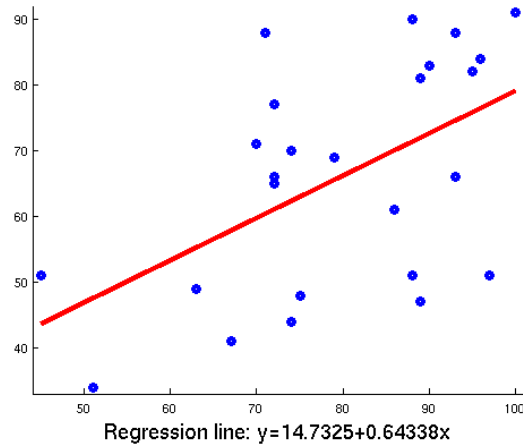
86 71 93 89 74 75 96 88 67 93 63 89 74 72 45 51 100 88 72 95 79 90 70 97 72
61 88 66 47 70 48 84 90 41 88 49 81 44 77 51 34 91 51 65 82 69 83 71 51 66

On Figure 2 a histogram of the scores for the first midterm is presented and on 2 a histogram of the score for the scores of the second midterm is presented.



It can be deduced from the pictures that the spread for the scores is about the same for both tests and the mean was much lower for the second test, most likely this is an indication of the fact that the second test was harder than the first test.

In Figure 2 a scatterplot of the data collected for the two midterms is shown. It can be noticed that a group of about 7 people takes the upper right corner of the picture. These are the students that performed well in both test. The fact that the lower right corner is much more populated than the upper left corner is an indication that there is a fairly large amount of people that performed well in the first test but did not in the second test. Overall the data doesn't tightly cluster around the line of regression, shown in red, so this is an indication that the correlation was not too strong.



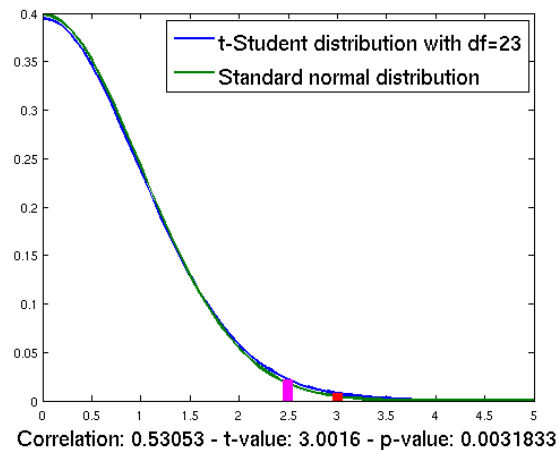
Computing the correlation coefficient using the formula

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}}$$

gave a value of $r = 0.53$, that overall is not a very high value. Although the amount of evidence from the $n = 25$ students still makes it statistically significant. Using the formula

$$t = r \sqrt{\frac{n-2}{1-r^2}}$$

gave a value of $t = 3.00$. This value is quite high and for the t -student distribution with $n - 2 = 23$ degrees of freedom.

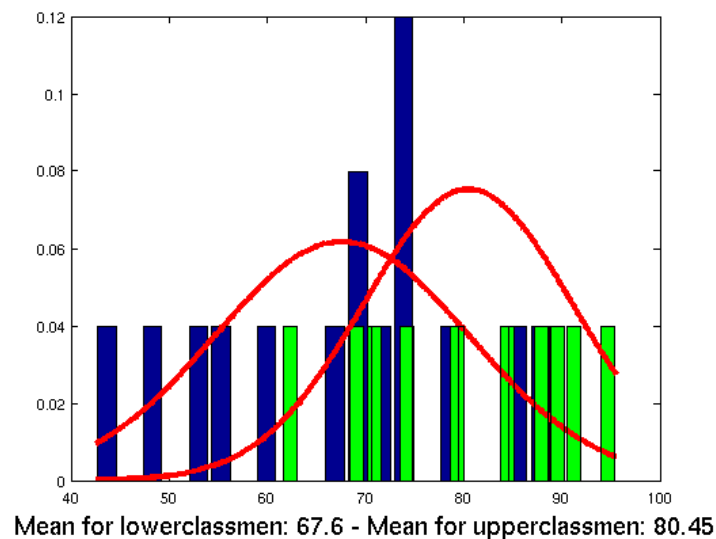


We performed a right-tailed test, as shown in Figure 2, and we obtained a P -value equal to 0.32%, therefore the correlation is statistically significant.

3 Second experiment

Looking at the data for the previous test I noticed that some students performed well over both tests. I wondered if this group of students belonged to the same class year and then I wondered if in general lowerclassmen and upperclassmen performed equally in the tests. It is perfectly reasonable to believe that upperclassmen, by virtue of being more experienced, tend to score higher on written exams. On the other side, freshmen might be more motivated at doing well in the class because it is more likely that they are taking this class to take further courses for which this class is a requirement. I am going to compare the two means for the two sets and measure if the distance between these is statistically significant. My belief is that the two groups perform much differently on average.

In Figure 3 a double histogram is presented where frequencies of average scores for the two midterms are presented, in blue for the lowerclassmen and in green for the upperclassmen. The two normal distribution with same mean and same standard deviation for the two populations have been overimposed to enhance the differences. Overall it does seem that the two groups performed very differently.



To test my last statement, a t -test was performed. It was found that lowerclassmen achieved an average score of 67.6 while upperclassmen achieved an average score of 80.45. Although only 10 people belonged to the upperclassmen population. Using the formula

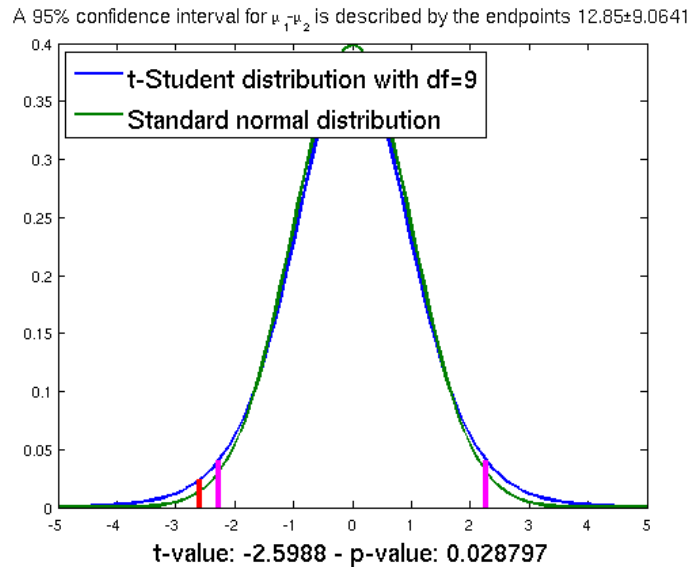
$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{(s_1^2/n_1) + (s_2^2/n_2)}}$$

using $n_1 = 15$ and $n_2 = 10$ a t -value of -2.60 was found for a t -Student distribution with $n_2 - 2 = 8$ degrees of freedom, as shown in Figure 3. For a two-tailed test, this value produced a P -value of 2.9%. Therefore the result is statistically significant at the 5% level. A 95%

confidence interval for the mean difference was computed using the formula

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

and the interval found was $\mu_1 - \mu_2 = 12.85 \pm 9.06$.



4 Conclusion

Looking at the data I confirmed my hypothesis that scores are linearly correlated on different midterms. I also found out, unexpectedly, that class differences can explain a significant part of the difference in the different performance of my students. Even if I don't have an explanation for this finding, I consider this fact an important result to keep in mind and to further study. Different factors could be the cause of this and at this point it would be premature to indicate one rather than another. A more interesting analysis could be made analyzing the final grades of other Dartmouth classes where there is high variability of student class.

5 Appendix

Matlab code used to display the histogram for the first midterm:

```
%% Produce a histogram of the data for the first midterm
[count xout] = hist(m1,20);
bar(xout,count/n);
hold on;
```

```
X = min(m1):.1:max(m1);
plot(X,5*normpdf(X,mean(m1),std(m1,1)),'Color','red','LineWidth',4);
hold off;
xlabel(['Mean: ' num2str(mean(m1)) ' - Standard deviation: '
        num2str(std(m1,1))],'FontSize',16);
```

Matlab code to display a scatterplot for the data together with the line of best fit:

```
x = m1; y = m2;
scatter(x,y,'LineWidth',4);
n = 25;
b1 = (n*sum(x.*y)-sum(x)*sum(y)) / (n*sum(x.^2)-sum(x)^2);
b0 = mean(y) - b1 * mean(x);
line([min(x) max(x)],[b0+b1*min(x) b0+b1*max(x)],'Color','red','LineWidth',4);
axis([min(x)-1 max(x)+1 min(y)-1 max(y)+1]);
xlabel(['Regression line: y=' num2str(b0) '+' num2str(b1) 'x'],'FontSize',16);
```

Matlab code to compute the correlation among the data and the corresponding t -value:

```
r = (n*sum(x.*y)-sum(x)*sum(y)) / ...
    sqrt((n*sum(x.^2)-sum(x)^2) * (n*sum(y.^2)-sum(y)^2));
t = r * sqrt((n-2)/(1-r^2));
df = n-2;
p = 1 - tcdf(t,df);
X = 0:.1:5;
plot(X,[tpdf(X,df); normpdf(X)],'LineWidth',2);
line([t t],[0 tpdf(t,df)],'Color','red','LineWidth',8);
alpha = .01;
talpha = tinverse(1-alpha,df);
line([talpha talpha],[0 tpdf(talpha,df)],'Color','magenta','LineWidth',8);
legend({'t-Student distribution with df=' ...
        num2str(df)],'Standard normal distribution'],'FontSize',16);
xlabel(['Correlation: ' num2str(r) ' - t-value: ' ...
        num2str(t) ' - p-value: ' num2str(p)],'FontSize',16);
```

Matlab code to display a histogram of the scores obtained by lowerclassmen and upperclassmen:

```
old = year == 08 | year == 09;
young = year == 10 | year == 11;
x1 = (m1(young)+m2(young)) / 2;
x2 = (m1(old)+m2(old)) / 2;

[count1 x1out] = hist(x1,20);
[count2 x2out] = hist(x2,20);
```

```

bar(x1out,count1/n);
hold on;
X = min([x1 x2]):.1:max([x1 x2]);
plot(X,2*normpdf(X,mean(x1),std(x1,1)), 'Color','red','LineWidth',4);
bar(x2out,count2/n,'g');
plot(X,2*normpdf(X,mean(x2),std(x2,1)), 'Color','red','LineWidth',4);
hold off;
xlabel(['Mean for lowerclassmen: ' num2str(mean(x1)) ...
        ' - Mean for upperclassmen: ' num2str(mean(x2))], 'FontSize',16);

```

Matlab code to compute a confidence interval for the distance of the mean of the score between lowerclassmen and upperclassmen:

```

xbar1 = mean(x1); n1 = length(x1);
xbar2 = mean(x2); n2 = length(x2);
t = (xbar1 - xbar2) / sqrt(var(x1)/n1 + var(x2)/n2);
df = min(n1-1,n2-1);
p = 2*tcdf(t,df);
X = -5:.1:5;
plot(X,[tpdf(X,df); normpdf(X)], 'LineWidth',2);
line([t t],[0 tpdf(t,df)], 'Color','red','LineWidth',4);
alpha = .05;
talpha = tinverse(alpha/2,df);
line([talpha talpha],[0 tpdf(talpa,df)], 'Color','magenta','LineWidth',4);
line([-talpa -talpa],[0 tpdf(-talpa,df)], 'Color','magenta','LineWidth',4);
legend({'t-Student distribution with df=' ...
        num2str(df)], 'Standard normal distribution'},2, 'FontSize',16);
xlabel(['t-value: ' num2str(t) ' - p-value: ' num2str(p)], 'FontSize',16);
width = tinverse(1-alpha,df) * sqrt(var(x1)/n1 + var(x2)/n2);
title(['A ' num2str(100*(1-alpha)) ...
        '% confidence interval for \mu_1-\mu_2 is described by the endpoints ' ...
        num2str(xbar2-xbar1) '\pm' num2str(width)], 'FontSize',16);

```