Variance

Simple analysis of variance

Confidence Interval of the Mean

t values

Do two means differ

F test

Simple analysis of variance

So far, we've assumed that all observed variance comes from a single, random source.

- not likely
- there can be many sources of variance

We'll now introduce a way to analyze the variance in sample sets.

Analysis of variance

$$s^2_{total} = s^2_1 + s^2_2 + ... + s^2_k$$

In general, when sources of variance are linearly related (independent and uncorrelated), the variances are additive.

We often need to do experiments to evaluate the magnitude and sources of variance.

Determining sources of variance

Let's start with a simple example where there should only be two potential sources of variance.

Sample	Replicates	Mean
1	15.9, 16.1, 16.3	16.1
2	14.9, 15.1, 15.3	15.1
3	15.8, 15.8, 15.8	15.8
4	16.2, 16.0, 15.9	16.0

In this example, a series of four samples are obtained and each is analyzed in triplicate.

Determining sources of variance

Since there is error in any measurement, it's not surprising that the means are different.

We want to know if the difference is due to variance in the method or real sample differences.

Simple two level model

$s^2_{between}$ = variance of sample material

s^2_{within} = variance of analytical method

s^2_{total} = $s^2_{between}$ + s^2_{within}

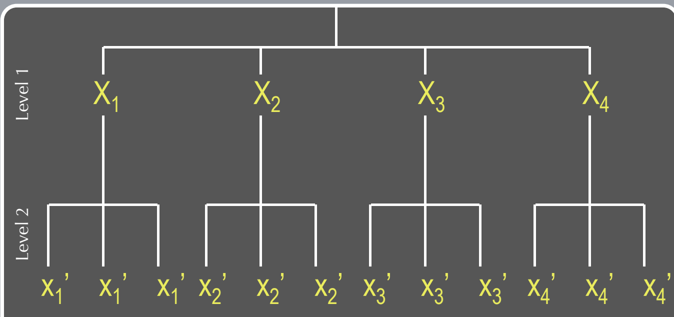
Simple two level model

We have two potential sources of variance.

- Samples may actually be different.
- Run to run errors.

A simple set of calculations can be used to sort out the sources of variance.

The F test can then be used to determine if the variance values are significant.



Level 1 gives us an idea as to sample variability

Level 2 tells use about the method variability

Simple two level model

Simple analysis of variance

First, calculate sum of squares for all values in your sample :

$$SS_T = \sum\left(x_i - \bar{x}_T\right)^2$$

x_T = Grand Mean

mean of all the points

The variation of the total mean can the be calculated as:

$$MS_T = \frac{\sum\left(x_i - \bar{x}_T\right)^2}{df_T}$$

df_T = total - 1

Next, calculate the between sample variance

$$SS_{between} = \sum n_r\left(\bar{x}_s - \bar{x}_T\right)^2$$

Then the mean square for the samples

$$MS_{between} = \frac{SS_{between}}{df_s}$$

n_r = # replicates per sample

$\bar{x}_s$ = mean of each sample

df_s = # samples − 1

Since you know SS_T and $SS_{between}$, you can find the within sample variance by:

$$SS_T = SS_{between} + SS_{within}$$

The mean sum of squares for our replicates is then:

$$MS_{within} = \frac{\left(SS_T - SS_{between}\right)}{df_T - \#\ samples}$$

Back to the example.

#	Replicates		
1	15.9	16.1	16.3
2	14.9	15.2	15.3
3	14.8	15.8	15.8
4	16.2	16.0	15.9

Source	df	SS	MS
Total SS_T	11	2.107	0.192
Sample $SS_{between}$	3	1.893	0.631
Replicate SS_{within}	8	0.213	0.024

Simple analysis of variance

OK. We've done several calculations. Now what?

We can now use the F test to determine if there is a significant difference between the two sources of variance.

$$F = \frac{s^2_{big}}{s^2_{small}}$$

F is then compared to F_C to see if the difference is significant. This will be covered in a bit.

F test

Simple analysis of variance

Now we can use the F test to determine if the samples are really different.

$$F = \frac{MS^2_{big}}{MS^2_{small}} = \frac{MS^2_{sample}}{MS^2_{replicate}}$$

$$= \frac{0.628}{0.026} = 24.2$$

F_c for 95% confidence and $df_{larger} = 3$ and $df_{smaller} = 8$ is 4.07, so our samples are different.

Using Excel

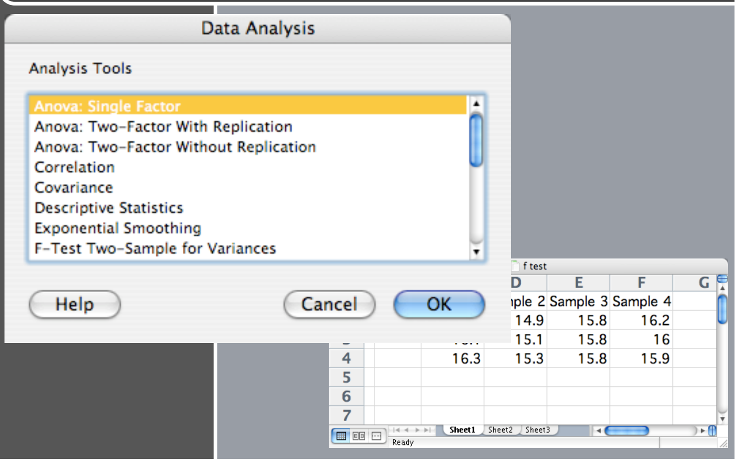
f test

	A	B	C	D	E	F	G
1			Sample 1	Sample 2	Sample 3	Sample 4	
2			15.9	14.9	15.8	16.2	
3			16.1	15.1	15.8	16	
4			16.3	15.3	15.8	15.9	
5							
6							
7							

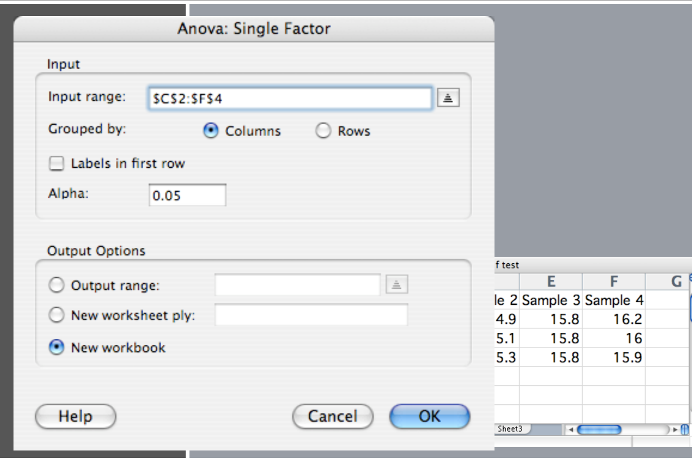
Sheet1 Sheet2 Sheet3

Ready

Using Excel



Using Excel



Using Excel

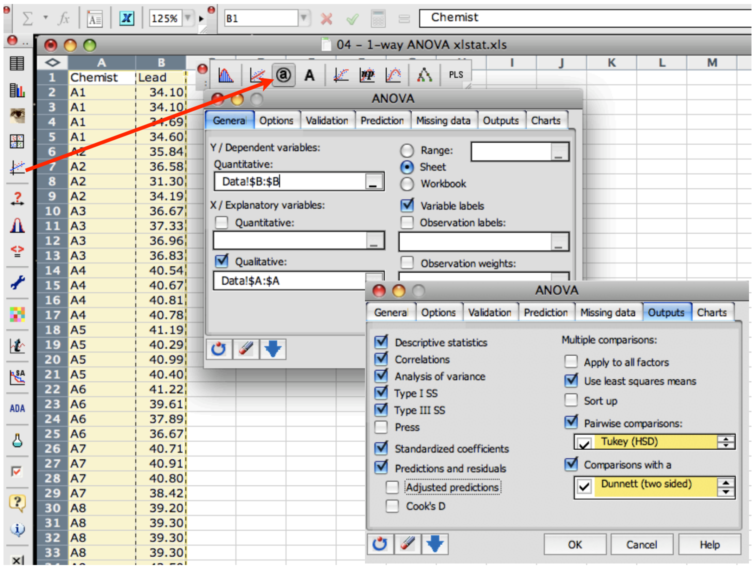
1		Sample 1	Sample 2	Sample 3	Sample 4		
2		15.9	14.9	15.8	16.2		
3		16.1	15.1	15.8	16		
4		16.3	15.3	15.8	15.9		
5							
6	Anova: Single Factor						
7	SUMMARY						
8	*Groups*	*Count*	*Sum*	*Average*	*Variance*		
9	Sample 1	3	48.3	16.1	0.04		
10	Sample 2	3	45.3	15.1	0.04		
11	Sample 3	3	47.4	15.8	0		
12	Sample 4	3	48.1	16.033	0.0233		
13							
14	ANOVA						
15	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
16	Between Samples	1.8825	3	0.6275	24.29	0.0002	4.0662
17	Within samples	0.2067	8	0.0258			
18							
19	Total	2.0892	11				

Replicate	A1	A2	A3	A4	A5	A6
1	34.10	35.84	36.67	40.54	41.19	41.22
2	34.10	36.58	37.33	40.67	40.29	39.61
3	34.69	31.30	36.96	40.81	40.99	37.89
4	34.60	34.19	36.83	40.78	40.40	36.67
	A7	A8	A9	A10	A11	A12
1	40.71	39.20	42.50	39.75	36.04	44.36
2	40.91	39.30	42.30	39.69	37.03	45.73
3	40.80	39.30	42.50	39.23	36.84	45.25
4	38.42	39.30	42.50	39.73	36.24	45.34

Twelve chemists assayed a sample for Pb to see if they got the same results. Each used the same furnace AA and spiked authentic serum sample. Results are in ug Pb/L

Current Federal limit is 100 ug Pb/L in blood.

Using XLStat



Using XLStat

- Note: XLstat does not report Fc values - just the P value - the probability that your values are NOT different.
- Data must be ordered in a single column.
- Two methods are available for calculating sum of squares for your groups - Type I and III. These are only useful for more complex multivariable ANOVA
- Might as well review them at this point.

Sum of squares analysis.

Type I (Sequential)
The Sums of Squares obtained by fitting effects in the order specified in the model. Type I SS for each effect will change if the order of the effects in the model is changed.

Type III (Marginal)
The Sums of Squares obtained by fitting each effect after all the other terms in the model. The Type III SS do not depend upon the order in which effects are specified in the model.

Sum of squares analysis.

Type I SS - Useful to explore unbalanced experimental data - where some effects are measured more than others. Can also show flaws in an experimental design (next chapter)

Type III Sums of Squares are preferable in most cases since they correspond to the variation attributable to an effect after correcting for any other effects in the model. They are unaffected by the frequency of observations.

With a balanced experiment (all combinations measured with equal frequency), Type I and III give the same results.

Analysis of variance:

Source	DF	Sum of squares	Mean squares	F	Pr > F
Model	11	438.943	39.904	40.264	< 0.0001
Error	36	35.678	0.991		
Corrected Total	47	474.621		F_{crit} = 2.07	

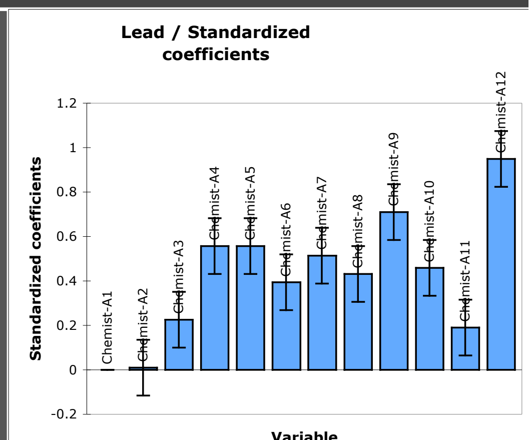
There is a difference. Can we tell what it is?

In this example, there is <0.01% chance of there NOT being a difference.

XLStat results.

This plot shows how each chemist performed.

While results have been normalized, you'd get the same basic plot with the raw data.

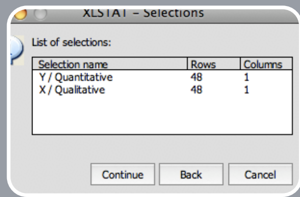


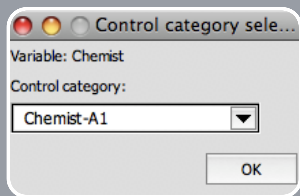
Using XLStat

The Dunnett test is used to compare samples (your chemists) to a control.

There actually is no control but the test provides a useful way of comparing results.

In this case, choose Chemist A1 because his/her results were the lowest, causing the results to be positive for the others.





Dunnett test

- Compares group means.
- Each is pitted against one control or reference group.
- Calculate a t test values for each group comparison.
- Test typically can only be used when all groups are of equal size.

Dunnett test

Category	Difference	Standardized difference	Critical value	Critical difference	Pr > Diff	Significant
A1 vs A12	-10.798	-15.339	2.890	2.034	0.000	Yes
A1 vs A9	-8.078	-11.475	2.890	2.034	0.000	Yes
A1 vs A5	-6.345	-9.014	2.890	2.034	0.000	Yes
A1 vs A4	-6.328	-8.989	2.890	2.034	0.000	Yes
A1 vs A7	-5.838	-8.293	2.890	2.034	0.000	Yes
A1 vs A10	-5.227	-7.426	2.890	2.034	0.000	Yes
A1 vs A8	-4.903	-6.964	2.890	2.034	0.000	Yes
A1 vs A6	-4.475	-6.357	2.890	2.034	0.000	Yes
A1 vs A3	-2.575	-3.658	2.890	2.034	0.007	Yes
A1 vs A11	-2.165	-3.076	2.890	2.034	0.032	Yes
A1 vs A2	-0.105	-0.149	2.890	2.034	1.000	No

Tukey Test

$$n_h = \frac{\bar{x}}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

Harmonized mean

- "Honestly Significantly Different (HSD) test.
- Based on pairwise comparison among means.

$$t_s = \frac{M_i - M_j}{\sqrt{\frac{MSE}{n_h}}}$$

M_i - M_j = difference between pair means
MSE = mean square error
n_h = the harmonized mean

Harmonized mean is the weighted arithmetic mean, with each value's weight being the reciprocal of the value.

Tukey test

Compares each chemist's results to see if there is a significant difference.

Contrast	Difference	Standardized difference	Critical value	Pr > Diff	Significant
A12 vs A1	10.798	15.339	3.490	< 0.0001	Yes
A12 vs A2	10.693	15.190	3.490	< 0.0001	Yes
A12 vs A11	8.633	12.263	3.490	< 0.0001	Yes
A12 vs A3	8.223	11.681	3.490	< 0.0001	Yes
A12 vs A6	6.323	8.982	3.490	< 0.0001	Yes
A12 vs A8	5.895	8.374	3.490	< 0.0001	Yes
A12 vs A10	5.570	7.913	3.490	< 0.0001	Yes
A12 vs A7	4.960	7.046	3.490	< 0.0001	Yes

Tukey test

Provides grouping of chemists with statistically similar results (95% confidence.)

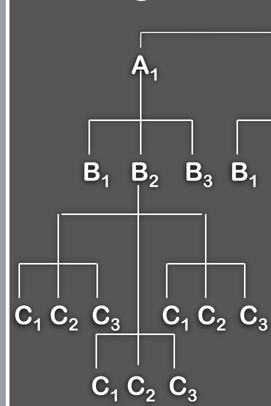
Chemist	Means	Groups					
A12	45.170	A					
A9	42.450		B				
A5	40.718		B	C			
A4	40.700		B	C			
A7	40.210		B	C			
A10	39.600			C			
A8	39.275			C	D		
A6	38.848			C	D	E	
A3	36.948				D	E	
A11	36.538					E	F
A2	34.478						F
A1	34.373						F

One-Factor ANOVA Table

Analysis of variance

Source	SS	df	MS	F
Total	SST	(N-1)		
Between (samples)	$SS_{between}$	(k-1)	$\frac{SS_{between}}{(k-1)}$	
Within (replicates)	SS_{within}	(N-1)-(k-1)	$\frac{SS_{within}}{(N-1)-(k-1)}$	$\frac{MS_{between}}{MS_{within}}$

Adding more levels

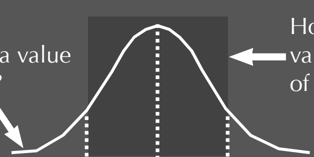


When adding more levels, things rapidly become more complex.

A factorial design is required. Covered in unit 5.

Confidence interval of the mean

How likely is a value to occur here?



How often will a value fall outside of this range?

Confidence interval

This will tell you where most of your data should occur.

A common calculation to report variability of data.

A quick way of identifying outlying values.

$$C.I. = \mu \pm \frac{Z\sigma}{\sqrt{N}}$$

$Z = probability\ factor$

Confidence interval of the mean

C.I. (%)	2 sided Z
90	1.645
95	1.960
99	2.575
99.99995	5.000

This is for large (N>100) data sets.

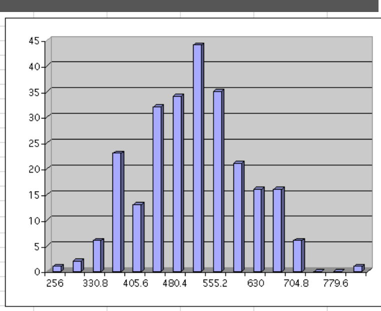
Z comes from infinity row of the t table.

Example - failure time of a lightbulb

lifetime
466
443
401
488
631
437
524
517
627
548
362
331
633
577
591
489
424
407
546
359
573
379
496
424
542
485
338
274

life time	
Mean	491.57
Median	489.5
Mode	548
Standard Deviatior	95.64
Sample Variance	9147.59
Kurtosis	-0.15
Skewness	0.06
Range	561
Minimum	256
Maximum	817
Sum	122893
Count	250

Bin	*Frequency*
256	1
293.4	2
330.8	6
368.2	23
405.6	13
443	32
480.4	34
517.8	44
555.2	35
592.6	21
630	16
667.4	16
704.8	6



Confidence intervals - Z		Mean value of 492
0.90	482	502
0.95	480	504
0.99	476	508
0.9999995	462	522

Confidence interval of the mean

$$C.I. = \bar{x} \pm t\left(\frac{s_x}{\sqrt{N}}\right)$$

Standard deviation of the mean.

Degrees of freedom	Confidence level 90% $t_{.95}$	95% $t_{.975}$	99% $t_{.995}$
1	6.31	12.7	63.7
2	2.92	4.30	9.92
3	2.35	3.18	5.84
4	2.13	2.78	4.60
5	2.02	2.57	4.03
6	1.94	2.45	3.71
7	1.90	2.36	3.50
8	1.86	2.31	3.36
9	1.83	2.26	3.25
10	1.81	2.23	3.17

We seldom collect a near infinite data set. It much more common to work with smaller sets.

We can then rely on the t test.

t values

t values account for error introduced based on sample size, degrees of freedom and potential sample skew. Actually use χ^2 distribution - chi squared.

$$\chi^2_{n-1} = (n-1)\, s^2 \;/\; \sigma^2$$

This allows us to estimate population variance from sample variance. All of this is tied together into the t values.

Degrees of freedom	90% $t_{.95}$	95% $t_{.975}$	99% $t_{.995}$
	Confidence level		
1	6.31	12.7	63.7
2	2.92	4.30	9.92
3	2.35	3.18	5.84
4	2.13	2.78	4.60
5	2.02	2.57	4.03
6	1.94	2.45	3.71
7	1.90	2.36	3.50
8	1.86	2.31	3.36
9	1.83	2.26	3.25
10	1.81	2.23	3.17

t values

Example

Data: 1.01, 1.02, 1.10, 0.95, 1.00

mean = 1.016

s_x = 0.0541

$s_{\bar{x}}$ = 0.0242

t values for 4 degrees of freedom

90% confidence = 2.13

95% confidence = 2.78

Example

$$CI_{90\%} = 1.016 \pm \frac{2.13 \times 0.0541}{5^{1/2}}$$

$$= 1.02 \pm 0.05 \quad (\pm 5\%)$$

$$CI_{95\%} = 1.016 \pm \frac{2.78 \times 0.0541}{5^{1/2}}$$

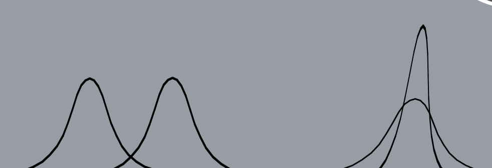
$$= 1.02 \pm 0.07 \quad (\pm 7\%)$$

A	B	C	D	E	F
K in water sample, ppm					
Replicate analysis					
14.7		Mean	14.97		average(a3..a9)
15.1		Std. Dev.	0.189		stdev(a3..a9)
15					
14.9		df	6		count(a3..a9)-1
15.3		t value, 95%	2.447		tinv(0.05,6)
14.9		CI - +/-	0.175		t * stdev / sqrt(7)
14.9					
TINV(probability, degrees of freedon)					
probability - use 1 - target confidence level					
Example - for 95%, use 0.05					

t test example

You can have samples that are considered significantly different and still have the same mean.

Beyond the mean



In both examples, the populations would be considered to be different - even though the means, medians and modes are identical in example on the right.

The F test

This test can be used to tell if two populations are different based on changes in variance.

Examples

- Has the measurement precision changed?
- Has the method been altered?
- Were there any significant changes due to the lab or analyst?

The F test

Calculation of F

$$F = \frac{s^2_{larger}}{s^2_{smaller}}$$

F is always 1 or greater and depends on the confidence level and degrees of freedom for both data sets

You can look up the F_C value for the desired levels or use a spreadsheet.

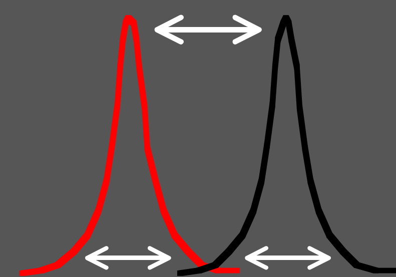
Example

- A - mean = 50 mg/l, s = 2.0 mg/l, n = 5, df = 4
- B - mean = 45 mg/l, s = 1.5 mg/l, n = 6, df = 5
- $F = 2^2 \ / \ 1.5^2 \ = \ 1.78$

F_C is 5.19 at 95% confidence

The variance values are essentially the same so the means must really differ.

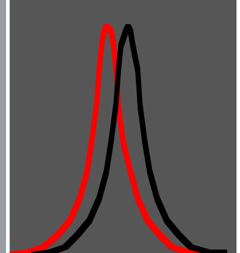
For this example, F would not exceed F_C but the means are significantly different.



You need to be concerned with differences in both the mean and sample variance.

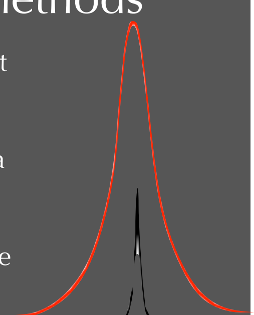
Comparison of the methods

The difference in the means is smaller than the sample variance.



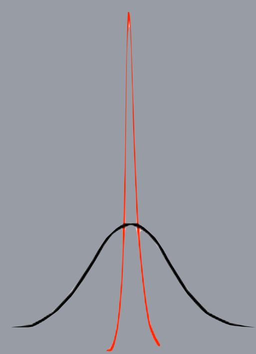
Comparison of the methods

- Here, the means are identical but the distributions look different.
- However, the lower curve is for a much smaller data set.
- The F test would show them to be the same.
- It accounts for the variations in sample size - using df.



Comparison of the methods

- In this case, the two groups are of similar size but with a significant difference in distribution.
- Again, the case type test would not work because the means are so close.
- The F test would indicate that they were different.



Comparison of the methods

No test can be relied on to provide all of the answers.

You must always look at your data, considering the mean, variance and sample size.

If need be, do multiple tests.

Excel example analysis

Was one exam 'harder' than another?

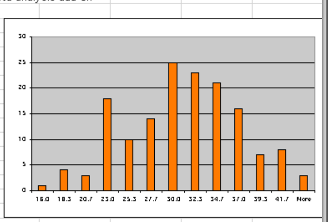
- A group of 153 students took two different multiple choice examinations.
- As a group, did they perform differently on the first and second exam?

Exam 1	Exam 2	Exam 1		Exam 2	
16	27				
17	27	Mean	30.16	Mean	30.75
17	26	Median	31	Median	31
18	26	Mode	32	Mode	29
18	26	Standard Deviation	6.10	Standard Deviation	5.72
19	25	Sample Variance	37.26	Sample Variance	32.70
19	16	Kurtosis	-0.48	Kurtosis	-0.63
20	31	Skewness	-0.11	Skewness	-0.21
21	29	Range	28	Range	26
21	27	Minimum	16	Minimum	16
21	22	Maximum	44	Maximum	42
21	25	Sum	4614	Sum	4705
21	29	Count	153	Count	153
21	29				
21	21				
21	23				

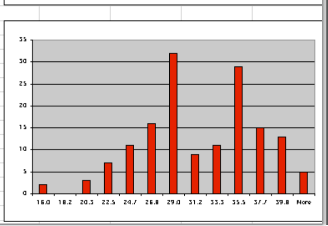
Grades

Histogram function built into the 'data analysis add-on'

Exam 1 score	Frequency
16.0	1
18.3	4
20.7	3
23.0	18
25.3	10
27.7	14
30.0	25
32.3	23
34.7	21
37.0	16
39.3	7
41.7	8
More	3



Exam 2 score	Frequency
16.0	2
18.2	0
20.3	3
22.5	7
24.7	11
26.8	16
29.0	32
31.2	9
33.3	11
35.5	29
37.7	15
39.8	13
More	5



Grades

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Exam 1	153	4614	30.157	37.265
Exam 2	153	4705	30.752	32.701

ANOVA

ce of Varia	SS	df	MS	F	P-value	F crit
Between (	27.0621	1	27.062	0.7736	0.3798	3.8722
Within Gro	10634.8	304	34.983			
Total	10661.9	305				