Analysis of Variance
Factorial experiments

Informally: The basic ANOVA situation

Two variables: 1 Categorical, 1 Quantitative

Main Question: Do the (means of) the quantitative variables depend on which group (given by categorical variable) the individual is in?

If categorical variable has only 2 values: can use standard 2-sample t-test

ANOVA allows for 3 or more groups

Example:

Question: Does maximum daily temperature depend on the season?

- **Response: Max daily temp; Group/categorical variable/factor: season**
- **Categories/factor levels: Spring, Summer, Fall, and Winter.**

ANOVA will answer the question if the mean maximum daily temps differ by season.

Informally: An example ANOVA situation

Subjects: 25 patients with blisters.

Treatments: Treatment A, Treatment B, Placebo
Measurement: # of days until blisters heal

Data	[and means]:
• A: 5,6,6,7,7,8,9,10	[7.25]
• B: 7,7,8,9,9,10,10,11	[8.875]
• P: 7,9,9,10,10,10,11,12,13	[10.11]

Are the differences in the mean healing time significant?

Informal Investigation

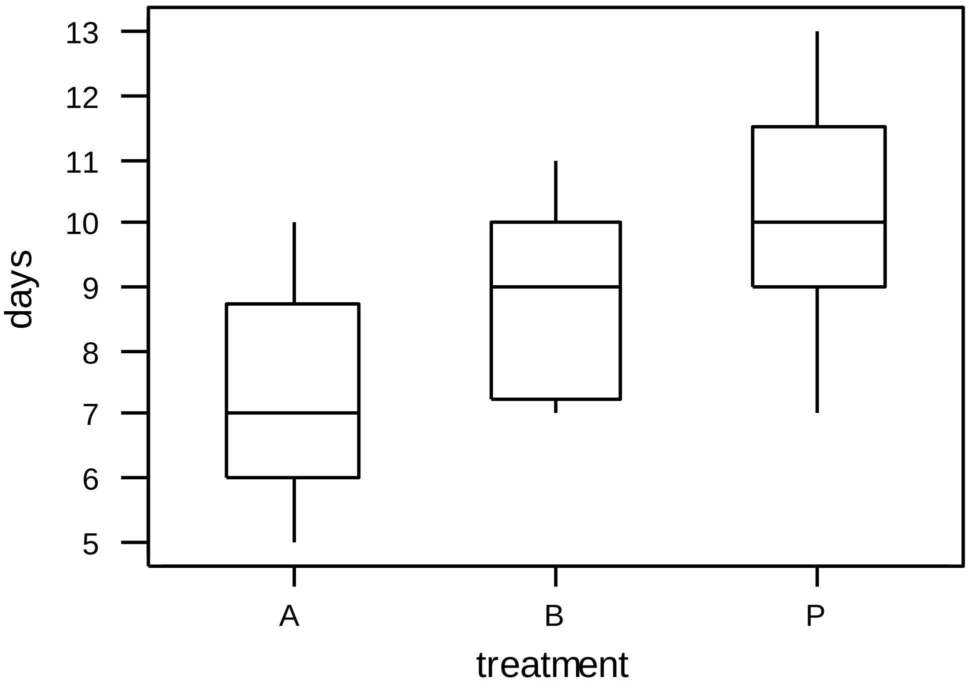
Graphical investigation:

- **side-by-side box plots**
- **multiple histograms**

Whether the differences between the groups are significant depends on

- **the difference in the means**
- **the standard deviations of each group**
- **the sample sizes**

Informally: Side by Side Boxplots



Informally: What does ANOVA do?

At its simplest (there are extensions) ANOVA tests the following hypotheses:

H_0: The means of all the groups/factor levels are equal.

H_a: Not all the means are equal

- **doesn't say how or which ones differ.**
- **Can follow up with "multiple comparisons"**

One-Factor Experiments-vocabulary

- In general, a factorial experiment involves several variables.
- One variable is the response variable, which is sometimes called the outcome variable or the dependent variable.
- The other variables are called factors.
- The question addressed by a factorial experiment is whether varying the levels of the factors produces a difference in the mean of the response variable.
- If there is just a single factor, then we say that it is a one-factor experiment.
- The different values of the factor are called the levels of the factor and can also be called treatments.
- The objects upon which measurements are made are called experimental units.
- The units assigned to a given treatment are called replicates.

Example: patients with blisters

- **Subjects: 25 patients with blisters.**

Treatments: Treatment A, Treatment B, Placebo

Measurement: # of days until blisters heal

- **Response: healing time in days**
- **Factor = treatment**
- **This is a one-factor experiment.**
- **Levels of the factor: A, B, P**
- **Patients with blisters are experimental units.**
- **Patients assigned to the same treatment are called replicates.**

Completely Randomized Experiments

- **Definition: A factorial experiment in which experimental units are assigned to treatments at random, with all possible assignments being equally likely, is called a completely randomized experiment.**

- **In many situations, the results of an experiment can be affected by the order in which the observations are taken.**

- **The ideal procedure is to take the observations in random order.**

- **In a completely randomized experiment, it is appropriate to think of each treatment as representing a population, and the responses observed for the units assigned to that treatment as a simple random sample from that population.**

Treatment Means

- The data from the experiment consists of several random samples, each from a different population.

- The population means are called **treatment means**.

- The **questions of interest** concern the treatment means :
  - whether they are all equal, and if not,
  - which ones are different,
  - how big the differences are, and so on.

- To make a formal determination as to whether the treatment means differ, a **hypothesis test is needed**.

One-Way Analysis of Variance

- **We have *I* samples, each from a different treatment.**
- **The treatment means are denoted $\mu_1, \ldots, \mu_I$.**
- **The sample sizes are denoted $J_1, \ldots, J_I$.**
- **The total number in all the samples combined is denoted by N, $N = J_1 + \ldots + J_I$**
- **The hypothesis that we wish to test is**

H_0: $\mu_1 = \ldots = \mu_I$ versus H_1: two or more of the μ_i are different

- **To test this hypothesis, we use a method known as one-way analysis of variance (ANOVA).**

Notation Needed

- **Since there are several samples, we use a double subscript to denote the observations.**
- **Specifically, we let X_{ij} denote the *j*th observation in the *i*th sample.**
- **The sample mean of the *i*th sample (mean response for treatment i):**

$$\overline{X}_{i.} = \frac{\sum_{j=1}^{J_i} X_{ij}}{J_i}$$

- **The sample grand mean:**

$$\overline{X}_{..} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J_i} X_{ij}}{N} = \frac{\sum_{i=1}^{I} J_i \overline{X}_{i.}}{N}$$

Example

An experiment was performed on welding fluxes. The objective was to determine which flux produces the strongest weld. There were 4 fluxes differing in the chemical composition. Several welds were made using each flux. The results of weld hardness are in the table.

Question: For the data in the table below, find $I, J_1, \ldots, J_I, N, X_{23}, \overline{X}_{3.}$ **, and** $\overline{X}_{..}$

TABLE: hardess measurements

Flux	Sample Values	Mean	SD
A	250, 264, 256, 260, 239	253.8	9.7570
B	263, 254, 267, 265, 267	263.2	5.4037
C	257, 279, 269, 273, 277	271.0	8.7178
D	253, 258, 262, 264, 273	262.0	7.4498

Example 1 cont.

Answer:

- **There are four samples, so I = 4.**
- **Each sample contains five observations, so $J_1 = J_2 = J_3 = J_4 = 5$.**
- **The total number of observations is N = 20.**
- **The quantity X_{23} is the third observation in the second sample, which is 267. The quantity $\overline{X}_{3.}$ is the sample mean of the third sample. This value is presented in the table and is 271.0.**
- **We can use the equation on a previous slide:**

$$\overline{X}_{..} = \frac{5(253.8) + 5(263.2) + 5(271.0) + 5(262.0)}{20} = 262.5$$

Treatment Sum of Squares

- **The variation of the sample means around the sample grand mean is measured by a quantity called the treatment sum of squares (SSTr), which is given by**

$$\mathrm{SSTr} = \sum_{i=1}^{I} J_i \left(\bar{X}_i - \bar{X}_{..}\right)^2 = \sum_{i=1}^{I} J_i \bar{X}_{i.}^{2} - N\bar{X}_{..}^{2}$$

- **Note that each squared distance is multiplied by the sample size corresponding to its sample mean, so that the means for the larger samples count more.**

- **SSTr provides an indication of how different the treatment means are from each other.**

- **If SSTr is large, then the sample means are spread widely, and it is reasonable to conclude that the treatment means differ and to reject H_0.**

- **If SSTr is small, then the sample means are all close to the sample grand mean and therefore to each other, so it is plausible that the treatment means are equal.**

Error Sum of Squares

- In order to determine, whether SSTr is large enough to reject H_0, we compare it to another sum of squares, called the error sum of squares (SSE).
- SSE measures the variation in the individual sample points around their respective sample means.
- This variation is measured by summing the squares of the distances from each point to its own sample mean, that is summing the squared residuals.
- SSE is given by

$$\mathrm{SSE} = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (X_{ij} - \bar{X}_{i.})^2$$

- SSE depends only on the distances of the sample points from their own means and is not affected by the location of the treatment means relative to one another.

- So, SSE measures only the underlying random variation in the process being studied.

Example 1 cont.

For the weld data, compute SSTr and SSE. MINITAB

One-way ANOVA: A, B, C, D Method

Null hypothesis All means are equal
Alternative hypothesis At least one mean is different
Significance level $\alpha = 0.05$

Equal variances were assumed for the analysis.

Factor Information

Factor	Levels	Values
Factor	4	A, B, C, D

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	3	743.4	247.80	3.87	0.029
Error	16	1023.6	63.97		
Total	19	1767.0			

Model Summary

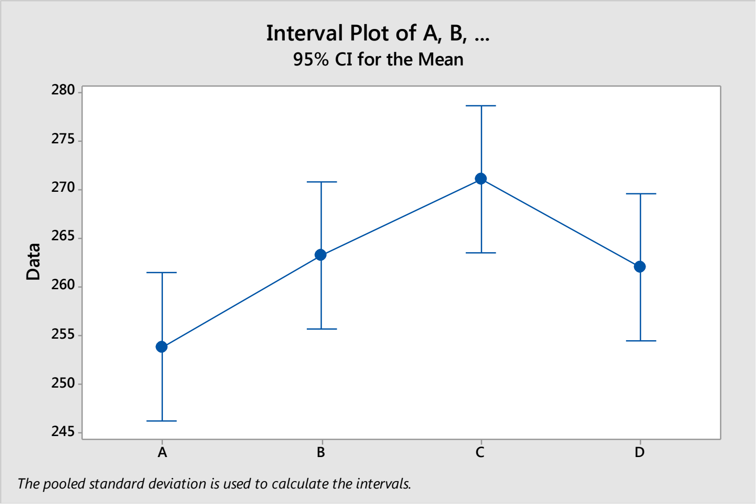
S	R-sq	R-sq(adj)	R-sq(pred)
7.99844	42.07%	31.21%	9.49%

Means

Factor	N	Mean	StDev	95% CI
A	5	253.80	9.76	(246.22, 261.38)
B	5	263.20	5.40	(255.62, 270.78)
C	5	271.00	8.72	(263.42, 278.58)
D	5	262.00	7.45	(254.42, 269.58)

Pooled StDev = 7.99844

Interval Plot of A, B, ...



Assumptions for the One-Way ANOVA

The standard one-way ANOVA hypothesis test are valid under the following conditions:

1. **The treatment populations must be normal.**
2. **The treatment populations must all have the same variance, which we will denote by σ^2.**

To check:

1. **Look at a normal probability plot for each sample and see if the assumption of normality is violated.**
2. **The assumption of equal variances is difficult to check with only a few observations in each sample. If it seems reasonable, then it can be concluded that this assumption is satisfied. This can be checked with the residual plot and examining how spread out each sample is.**

The Rule of Thumb: Standard deviations of each group are approximately equal if the ratio of the largest to the smallest sample st. dev. is less than 2:1.

Standard Deviation Check

Blister example.

Variable	treatment	N	Mean	Median	StDev
days	A	8	7.250	7.000	1.669
	B	8	8.875	9.000	1.458
	P	9	10.111	10.000	1.764

Compare largest and smallest standard deviations:

- **largest: 1.764**
- **smallest: 1.458**
- **1.764/1.458 = 1.2099 < 2. So, we can assume equal variances.**

Note: variance ratio of 4:1 is equivalent.

When Assumptions Are Met

- We can compute the means of SSE and SSTr.

- The mean of SSTr depends on whether H_0 is true, because SSTr tends to be smaller when H_0 is true and larger when H_0 is false.

- The mean of SSTr satisfies the condition

  $\mu_{SSTr} = (I - 1)\sigma^2$, when H_0 is true

  $\mu_{SSTr} > (I - 1)\sigma^2$, when H_0 is false.

- The likely size of SSE, and its mean, does not depend on whether H_0 is true. The mean of SSE is given by $\mu_{SSE} = (N - I)\sigma^2$.

The *F* test for One-Way ANOVA

To test H_0: $\mu_1 = \ldots = \mu_I$ versus H_1: two or more of the μ_i are different

1. **Compute SSTr.**
2. **Compute SSE.**
3. **Compute MSTr = SSTr/(I – 1) and MSE = SSE/(N – I).**
4. **Compute the test statistic: F^* = MSTr / MSE.**
5. **Find the *P*-value by finding $P(F_{I-1,\ N-1} > F^*)$.**

Note: The total sum of squares, SST = SSTr + SSE.

We use MINITAB for the computation of the test statistic and the p-value. The results are reported in the ANOVA table

Example 1 cont.

For the weld data, compute MSTr, MSE, and *F*. Find the *P*-value for testing the null hypothesis that all the means are equal. What do you conclude?

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	3	743.4	247.80	3.87	0.029
Error	16	1023.6	63.97		
Total	19	1767.0			

Confidence Intervals for the Treatment Means

A level 100(1 – α)% confidence interval for μ_i is given by

$$\bar{X}_{i.} \pm t_{N-1,\alpha/2}\sqrt{\frac{\text{MSE}}{J_i}}$$

Weld Example cont.

Find a 95% confidence interval for the mean hardness of welds produced with flux A.

$$\bar{X}_{i.} \pm t_{N-1,\alpha/2}\sqrt{\frac{\text{MSE}}{J_i}}$$

Means

Factor	N	Mean	StDev	95% CI
A	5	253.80	9.76	(246.22, 261.38)
B	5	263.20	5.40	(255.62, 270.78)
C	5	271.00	8.72	(263.42, 278.58)
D	5	262.00	7.45	(254.42, 269.58)

$X_{1.}$ = 253.80, MSE = 63.97, N =20, I=4, α=0.05, so $t_{16,\,0.025}$ = 2.12.

The 95% CI for μA is:

253.2 ± (2.12)($253.2 \pm (2.12) * \sqrt{63.97/5}) = 253.8 \pm 7.6 =$ (246.22, 261.38)

Blister example: Minitab ANOVA Output

Analysis of Variance for days

Source	DF	SS	MS	F	P
treatment	2	34.74	17.37	6.45	0.006
Error	22	59.26	2.69		
Total	24	94.00			

\# of data values - # of levels of the factor/treatment

1 less than # of levels of the factor/treatment

1 less than # of individuals/replicates

Blister example: Minitab ANOVA Output

```
Analysis of Variance for days
Source      DF        SS        MS        F        P
treatment    2     34.74     17.37     6.45    0.006
Error       22     59.26      2.69
Total       24     94.00
```

$$\sum_{obs}(x_{ij}-\bar{x}_i)^2$$

$$\sum_{obs}(x_{ij}-\bar{\bar{x}})^2$$

$$\sum_{obs}(\bar{x}_i-\bar{\bar{x}})^2$$

SS stands for sum of squares, ANOVA splits this into 3 parts

Blister example: Minitab ANOVA Output

Analysis of Variance for days

Source	DF	SS	MS	F	P
treatment	2	34.74	17.37	6.45	0.006
Error	22	59.26	2.69		
Total	24	94.00			

MSTr = SSTr / DFTr
MSE = SSE / DFE

F = MSTr / MSE

P-value comes from F(DFTr,DFE)

Test H_0: $\mu_1 = \ldots = \mu 3$ versus *Ha*: two or more of the μ_i are different
Take significance level =0.05

Not all treatment means are the same.

The ANOVA Table

```
One-way ANOVA: A, B, C, D

Source     DF        SS        MS     F      P
Factor      3    743.40   247.800  3.87  0.029
Error      16   1023.60    63.975
Total      19   1767.00

S = 7.998   R-Sq = 42.07%   R-Sq(adj) = 31.21%


                           Individual 95% CIs For Mean Based on
                           Pooled StDev
Level  N    Mean  StDev   ---+---------+---------+---------+---
A      5  253.80   9.76   (-------*-----)
B      5  263.20   5.40           (-----*------)
C      5  271.00   8.72                  (------*------)
D      5  262.00   7.45          (------*------)
                           ---+---------+---------+---------+---
                           250       260       270       280

Pooled StDev = 8.00
```

Blister example: Where's the Difference?

Once ANOVA indicates that the groups do not all have the same means, what do we do?

```
Analysis of Variance for days
Source      DF        SS        MS        F        P
treatmen     2     34.74     17.37     6.45    0.006
Error       22     59.26      2.69
Total       24     94.00
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level        N      Mean     StDev ----------+---------+---------+------
A            8     7.250     1.669 (-------*-------)
B            8     8.875     1.458           (-------*-------)
P            9    10.111     1.764                    (------*-------)
                                   ----------+---------+---------+------
Pooled StDev =     1.641                    7.5       9.0      10.5
```

Clearest difference: P is worse than A (CI's don't overlap)