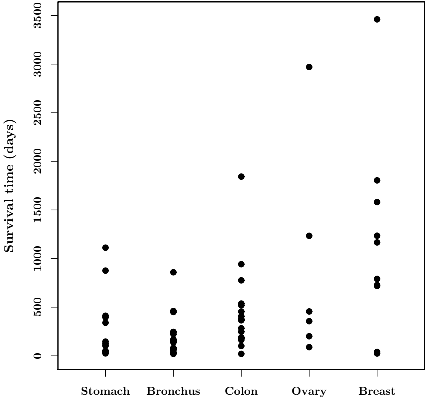
Today: One- and two-way analysis of variance, multivariate analysis of variance

- Practical remarks
- Example: Survival times in terminal human cancer
- One-way analysis of variance
- Multiple comparisons
- Example: The effect of work site and health program on weight loss
- Two-way analysis of variance
- Example: BMI and diastolic blood pressure
- Multivariate analysis of variance (MANOVA)

Example: Survival times in terminal human cancer

Data: Survival times (days) for terminal cancer patients.
Cameron & Pauling (1978).

Stomach	Bronchus	Colon	Ovary	Breast
124	81	248	1234	1235
42	461	377	89	24
25	20	189	201	1581
45	450	1843	356	1166
412	246	180	2970	40
51	166	537	456	727
1112	63	519		3808
46	64	455		791
103	155	406		1804
876	859	365		3460
146	151	942		719
340	166	776		
396	37	372		
	223	163		
	138	101		
	72	20		
	245	283		



Question: Is there any difference between the survival times corresponding to the different types of cancer? If so, how do they differ?

Survival times in terminal human cancer: Means and standard deviations

We have six groups of very different sizes.

The sample sizes, averages, and standard deviations are given by

Group	Type	n_i	$\bar{x}_i$	s_i
1	Stomach	13	286	346
2	Bronchus	17	212	210
3	Colon	17	457	427
4	Ovary	6	884	1099
5	Breast	11	1396	1239

We want to compare the **expected survival times** in the six groups.

Assumptions

The survival times are **independent** and **normally distributed** with

$$\text{Group } i: \qquad \text{Mean } \mu_i, \qquad \text{sd } \sigma_i, \qquad i = 1, \ldots, 5$$

The hypothesis of interest is that the mean survival time does not depend on the type of cancer

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

In order to test this hypothesis (using an exact F-test) we need to assume **constant sd**.

One-way analysis of variance:

$$\text{Group } i: \qquad \text{Mean } \mu_i, \qquad \text{sd } \sigma, \qquad i = 1, \ldots, 5$$

Bartlett's test for the hypothesis of equal sd's (H_0: $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5$):

$$B = 48.1 \sim \chi^2(4), \qquad p < 0.0001$$

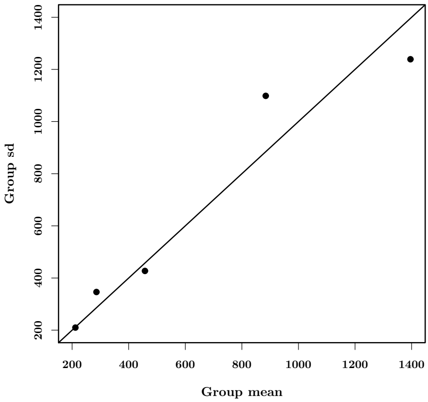
Survival times in terminal human cancer: Transformation

Often a transformation helps.

If the **standard deviation is proportional to the mean** a **log-transformation** is appropriate.

log-survival times:

Group	Type	n_i	$\bar{x}_i$	s_i
1	Stomach	13	4.97	1.25
2	Bronchus	17	4.95	0.95
3	Colon	17	5.75	1.00
4	Ovary	6	6.15	1.26
5	Breast	11	6.56	1.65



Bartlett's test for equal standard deviations now gives a test statistic of $B = 4.81$ which compared to a $\chi^2(4)$-distribution results in a p-value of 0.31.

Survival times in terminal human cancer: One-way ANOVA for the transformed data

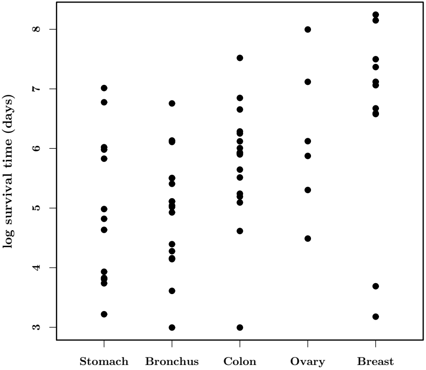
One-way analysis of variance:

log-survival times **independent** and **normally distributed**

Group i : Mean μ_i, sd σ, $i = 1, \ldots, 5$

The hypothesis of interest is still that the mean (log) survival time does not depend on the type of cancer

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$



One-way analysis of variance: The F-test for equal means

This hypothesis of equal means is tested using the F-test statistic

$$F = \frac{s_{\mathrm{B}}^2}{s^2}$$

Here s_{B}^2 measures the variation **between** group means.

Similarly, s^2 is a measure of the variation **within** groups.

If the hypothesis is true then F follows an F-distribution with $(5-1, 64-5)$ degrees of freedom.

Survival times in terminal cancer: The F-test for equal means

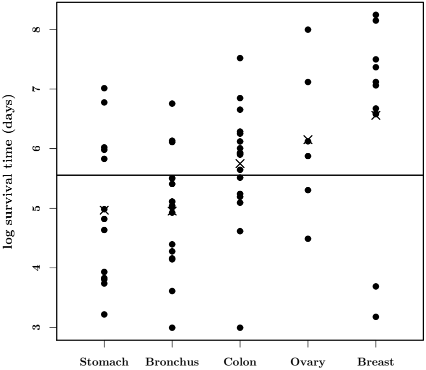
In this case we have

$s_B^2 = 6.122$ and $s^2 = 1.428$

The F-test statistic is given by

$$F = \frac{6.122}{1.428} = 4.29$$

This is to be compared to an $F(4, 59)$-distribution, and we get a p-value of 0.004.



Conclusion: There is clear evidence against the hypothesis that the mean (log) survival times are the same for the different types of cancer.

Survival times in terminal cancer: Confidence intervals

The estimated mean (log) survival times are just the group averages

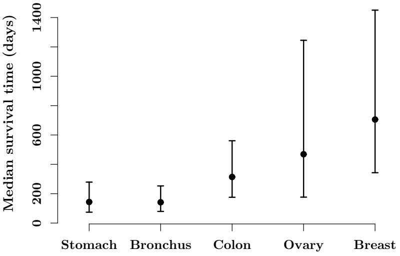
$$\hat{\mu}_1 = 4.97, \quad \hat{\mu}_2 = 4.95, \quad \hat{\mu}_3 = 5.75, \quad \hat{\mu}_4 = 6.15, \quad \hat{\mu}_5 = 6.56$$

The confidence interval for the expected log-survival time in the i'th group is given by

$$\text{Group } i: \quad \bar{x}_i \pm t_{0.975}(59) \cdot s/\sqrt{n_i}$$

Estimated **median** survival times (with corresponding confidence intervals) are obtained by transforming the estimates and confidence limits by the exponential function.

Group	Estimate	95%−CI
1	143.7	[74.0, 279.0]
2	141.6	[79.3, 252.9]
3	313.9	[175.8, 560.7]
4	469.0	[176.7, 1245.0]
5	705.3	[342.9, 1450.5]



Survival times in terminal cancer: CIs for differences

Question: Are any two groups significantly different?

A way to examine this is to compute *p*-values and 95%-confidence intervals for all pairwise differences. For group 1 and 2 we get (on **log-scale**):

$$\mu_2 - \mu_1 : \quad \bar{x}_2 - \bar{x}_1 \pm t_{0.975}(59) \cdot s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

The back-transformed differences are

Difference	Estimate	95%-confidence interval	*p*-value
2 − 1	0.99	[0.41 , 2.38]	0.973
3 − 1	2.18	[0.90 , 5.27]	0.081
4 − 1	3.26	[1.00 , 10.62]	0.050
5 − 1	4.91	[1.84 , 13.07]	0.002
3 − 2	2.22	[0.98 , 5.03]	0.057
4 − 2	3.31	[1.06 , 10.31]	0.039
5 − 2	4.98	[1.97 , 12.56]	0.001
4 − 3	1.49	[0.48 , 4.65]	0.482
5 − 3	2.25	[0.89 , 5.67]	0.085
5 − 4	1.50	[0.45 , 5.06]	0.504

Survival times in terminal cancer: Multiple comparisons

With so many (10) pairwise comparisons you run the risk of rejecting that the survival time in two groups are the same (when in fact they are) by pure chance. For each test this **type I error** is 5%, so the overall chance of falsely rejecting a true hypothesis can become much larger than 5% when many tests are made.

A (conservative) way of adjusting for multiple comparisons is the so-called **Bonferroni** adjustment where the significance level is divided by the number of comparisons. In this case we would only say that two groups are significantly different if the p-value was less than $0.05/10 = 0.005$. **But much better to report the actual p-values**.

Conclusion: The survival times corresponding to Stomach and Bronchus cancer are significantly lower (a factor 5, 95%-CI: 2-12) than for Breast cancer.

None of the other pairwise comparisons yield significant group differences. Can we conclude that all the other survival times are the same (for example Stomach - Colon and Colon - Breast)?

Survival times in terminal cancer: Residual Q-Q plot

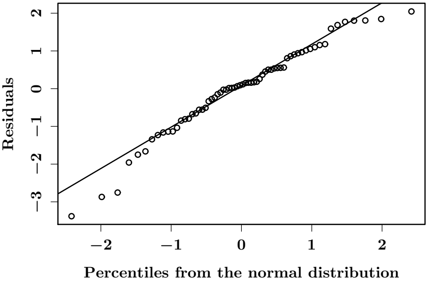
The assumptions behind the one-way analysis of variance include

1. **Independence** between groups. In **each group**: Independent observations from the same population.
2. The distribution in each population can be described by a **normal distribution**
3. The populations have a **common standard deviation**

Recall that the **residuals** are defined as

$$\text{res} = \text{observed} - \text{expected}$$

If the one-way analysis of variance model describes the data well, we have that the distribution of the residuals is normal. Whether or not this is reasonable is best investigated by means of a Q-Q plot.



Survival times in human cancer: The analyzes in Stata I

In **Stata** the one-way analysis of variance is done using `oneway` and `anova`:

```
oneway logsurv type

                   Analysis of Variance
    Source              SS         df       MS            F     Prob > F
------------------------------------------------------------------------
Between groups     24.4865571      4   6.12163929   4.29     0.0041
Within  groups     84.269591      59   1.42829815
------------------------------------------------------------------------
    Total         108.756148      63   1.72628807

Bartlett's test for equal variances: chi2(4) = 4.8090
                                       Prob>chi2 = 0.307
```

Survival times in human cancer: The analyzes in Stata II

```
anova logsurv type
margins, over(type) expression(exp(predict(xb)))
```

```
         |            Delta-method
         |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------+----------------------------------------------------------------
type     |
 Stomach |   143.7276   47.64068     3.02   0.003     50.35358    237.1016
Bronchus |   141.6232   41.05059     3.45   0.001     61.16554    222.0809
   Colon |    313.915   90.99071     3.45   0.001     135.5765    492.2535
   Ovary |   469.0064   228.8298     2.05   0.040     20.50823    917.5046
  Breast |    705.286   254.1431     2.78   0.006     207.1746    1203.397
--------------------------------------------------------------------------
```

Survival times in human cancer: The analyzes in Stata III

```
pwcompare type, eff eform
```

```
                     |   exp(b)  Std. Err.    t  P>|t|  [95% Conf. Interval]
---------------------+--------------------------------------------------------
type                 |
Bronchus vs Stomach | .9853586  .4338786 -0.03 0.973  .4082649     2.37819
   Colon vs Stomach | 2.184097  .9617138  1.77 0.081  .9049397    5.271379
   Ovary vs Stomach | 3.263162  1.924764  2.01 0.050   1.00242    10.62252
  Breast vs Stomach | 4.907102  2.402548  3.25 0.002  1.842242    13.07084
  Colon vs Bronchus |  2.21655  .9086101  1.94 0.057  .9759953    5.033934
  Ovary vs Bronchus | 3.311649  1.879393  2.11 0.039  1.063816    10.30914
 Breast vs Bronchus | 4.980017  2.303024  3.47 0.001   1.97401    12.56355
     Ovary vs Colon | 1.494055  .8478909  0.71 0.482   .479942    4.650982
    Breast vs Colon | 2.246742  1.039013  1.75 0.085  .8905774    5.668064
    Breast vs Ovary | 1.503788  .9121126  0.67 0.504  .4467734    5.061575
------------------------------------------------------------------------------
```

Example: The effect of work site and health program on weight loss

Data: Difference between pre- and post-intervention weights (pounds) after 6 months of participation by intervention program at two sites. Five moderately overweight women for each combination. Data from **Forthofer & Lee(1995)**.

Program	Office site					Factory site				
Diet clinic	6	2	10	−1	8	3	15	4	8	6
Exercise club	3	4	−2	6	−2	−4	6	8	−2	3
Both Programs	8	12	7	10	5	15	8	10	16	3

Questions:

Q1. Is there an effect of the programs?

Q2. Is it the same effect?

Q3. Does the difference between programs depend on work site?

Weight loss, work site, and health program: Mean and sd

The means and standard deviations in the six groups are

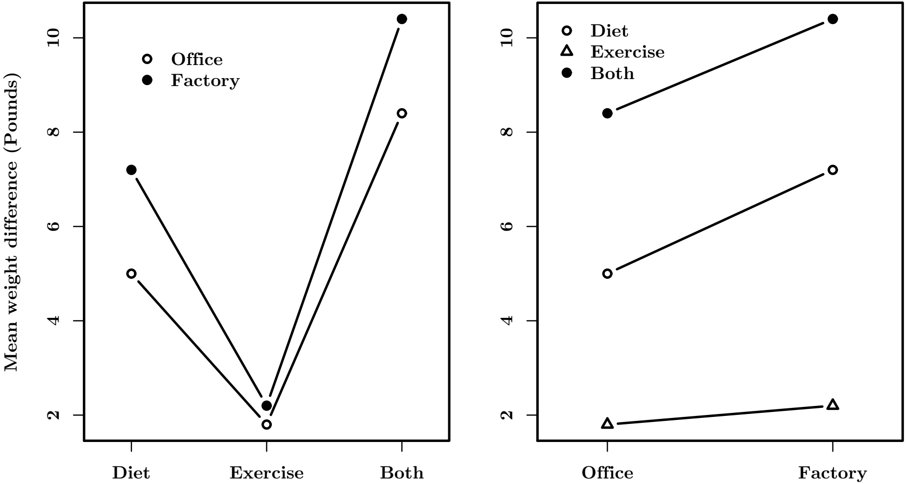
Group	Program	Site	Mean	SD
1	Diet	Office	5.0	4.47
2	Diet	Factory	7.2	4.76
3	Exercise	Office	1.8	3.63
4	Exercise	Factory	2.2	5.12
5	Both	Office	8.4	2.70
6	Both	Factory	10.4	5.32

Even though the variation appears to be different in the six groups then the **Bartlett's test** for equal standard deviations gives a p-value of 0.84.

The F-test for the hypothesis of no mean difference between the six groups (using the **one-way analysis of variance model**) gives a p-value of 0.03.

Weight loss, work site, and health program: Interaction plots

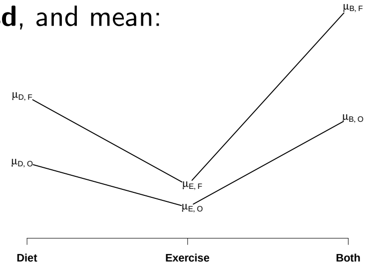
A useful way of examining the data is via the so-called **interaction plots**.



Assumptions

The statistical analysis that allows us to answer all of our questions is the **two-way analysis of variance**:
The weight differences in the six groups are described by **independent normal distributions** with a **common sd**, and mean:

$\mu_{D,O}$	$\mu_{D,F}$
$\mu_{E,O}$	$\mu_{E,F}$
$\mu_{B,O}$	$\mu_{B,F}$



This just means that there is a **level of weight loss for each combination** of intervention program and work site:

Possibly non-parallel interaction plots

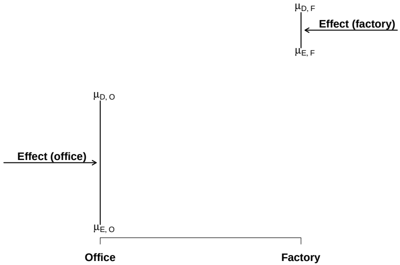
Weight loss and programs: Difference between programs

The **effect of diet compared to the effect of exercise:**

$$\mu_D - \mu_E$$

(expected weight loss when on diet minus the expected weight loss when doing exercise).

Hang on: This difference may not be the same in both work environments!



Office: $\mu_{D,O} - \mu_{E,O}$

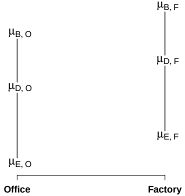
Factory: $\mu_{D,F} - \mu_{E,F}$

In order to talk about **the effect** of diet compared to the effect of exercise, we must test that it is not affected by work site (if it is, we must report the effect for each site).

Weight loss, work site, and health program: Interaction?

The hypothesis of no interaction between program and work site:

$$H_0: \quad \mu_{D,O} - \mu_{E,O} = \mu_{D,F} - \mu_{E,F} \quad \text{and}$$
$$\mu_{D,O} - \mu_{B,O} = \mu_{D,F} - \mu_{B,F}$$



(and thereby $\mu_{B,O} - \mu_{E,O} = \mu_{B,F} - \mu_{E,F}$).

If the hypothesis is accepted, then we conclude that:

- There is no interaction between program and work site.
- The effect of diet compared to the effect of exercise is <u>not</u> modified by (does not depend on) work site (no effect modification).

Note: This is the relevant hypothesis when we want to answer **Q3**.

Two-way analysis of variance: Testing for no interaction

Note that if program does not interact with site then site does not interact with program (!), so the hypothesis of no interaction between program and work site can also be written as

$$H_0: \mu_{D,O} - \mu_{D,F} = \mu_{E,O} - \mu_{E,F} = \mu_{B,O} - \mu_{B,F}$$

The F-test statistic corresponding to the hypothesis of no interaction is given by

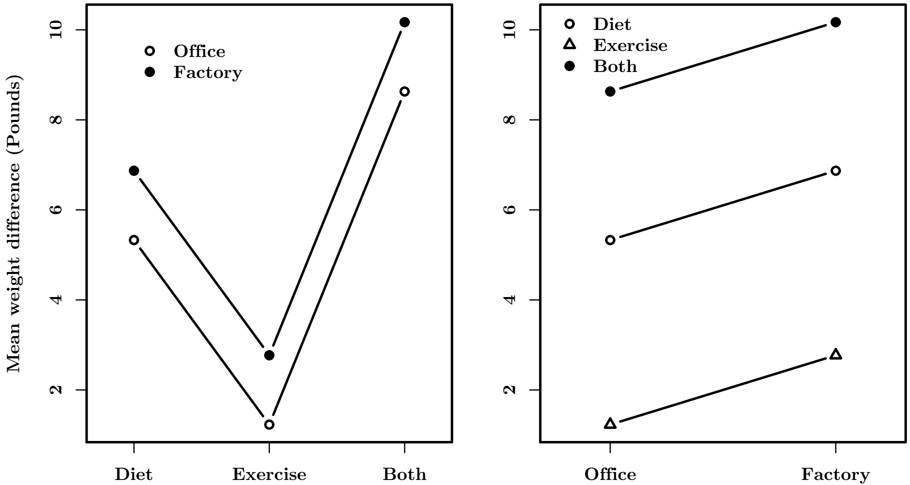
$$F = 0.12$$

which compared to an $F(2, 24)$-distribution results in a p-value of 0.88. We conclude there is no evidence in the data against the hypothesis of no interaction between program and work site. This corresponds to accepting that the curves in the **interaction plots** are **parallel**.

If we consider two randomly chosen women where one follows the **Diet** program and the other the **Exercise** program then we would expect the same difference between the two women's weight loss no matter if they both have office work or both factory work.

Weight loss, work site, and program: Parallel interaction plots

The model that results in parallel interaction plots is also called the **additive model**.



Two-way analysis of variance: Testing for no main effect

Having accepted that there is no interaction the next natural question is whether there is any difference between the intervention programs (and equivalently between the work sites). The corresponding hypothesis is:

$$H_0: \mu_{D,O} = \mu_{E,O} = \mu_{B,O} \qquad (H_0: \mu_{D,O} = \mu_{D,F})$$

and thereby $\mu_{D,F} = \mu_{E,F} = \mu_{B,F}$ ($\mu_{E,O} = \mu_{E,F}$, $\mu_{B,O} = \mu_{B,F}$).
In this case we get

Program: $F = 7.51 \sim F(2, 26)$, $p = 0.0027$ **Q2**
Work site: $F = 0.96 \sim F(1, 26)$, $p = 0.34$

The conclusion is that there is no real evidence against the hypothesis of no difference between the two work sites, whereas there is clear evidence against the hypothesis of no difference between the programs.

Weight loss, work site, and health program: Estimated effects

Even though the average weight loss is greater in the Factory site compared to the Office site for all three programs, then this difference is not significant.

One could continue and analyze the data using the **one-way analysis of variance model** (with program as the factor of interest) but that is typically not of interest.

The estimated effects are

Program:

Effect	Estimate	95%-confidence interval	p-value
Exercise - Diet	-4.10	$[-8.03, -0.17]$	0.042
Both - Diet	3.30	$[-0.63, 7.63]$	0.096
Both - Exercise	7.40	$[3.47, 11.33]$	0.001

Site:

$$\text{Factory - Office:} \quad 1.53, \quad [-1.68, 4.74], \quad p = 0.335$$

Weight loss, work site, and health program: Fitted values

The fitted values in the six groups are

Group	Program	Site	Mean	95%-confidence interval
1	Diet	Office	5.33	[2.12, 8.54]
2	Diet	Factory	6.87	[3.66, 10.08]
3	Exercise	Office	1.23	[−1.98, 4.44]
4	Exercise	Factory	2.77	[−0.44, 5.98]
5	Both	Office	8.63	[5.42, 11.84]
6	Both	Factory	10.17	[6.96, 13.38]

We see that we cannot rule out that there is in fact no significant weight loss when enrolled in the exercise program, whereas there is evidence against no weight loss for the other two programs. **Q1**

Weight loss, work site, and program: Q-Q plot for residuals

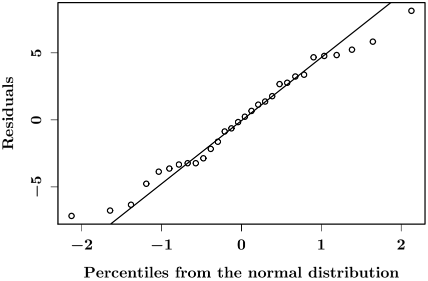
As in the one-way ANOVA the assumptions behind the two-way analysis of variance include

1. **Independence** between groups. In **each group**: Independent observations from the same population.
2. The distribution in each population can be described by a **normal distribution**
3. The populations have a **common standard deviation**

Again we consider the **residuals**:

$$\text{res} = \text{observed} - \text{expected}$$

If the two-way analysis of variance model describes the data well, we have that the distribution of the residuals is normal. Whether or not this is reasonable is best investigated by inspecting a Q-Q plot.



Weight loss, work site, and program: The analyzes in Stata I

In **Stata** the two-way analysis of variance is done using `anova`:

```
anova wloss program##site
       Number of obs =       30     R-squared     =  0.3871
       Root MSE      = 4.42907     Adj R-squared =  0.2594

      Source |  Partial SS    df       MS           F     Prob > F
-------------+----------------------------------------------------
       Model |  297.366667     5  59.4733333     3.03     0.0293
             |
     program |  274.866667     2  137.433333     7.01     0.0040
        site |  17.6333333     1  17.6333333     0.90     0.3525
program#site |  4.86666667     2  2.43333333     0.12     0.8839
             |
    Residual |       470.8    24  19.6166667
-------------+----------------------------------------------------
       Total |  768.166667    29  26.4885057
```

Weight loss, work site, and program: The analyzes in Stata II

The data is analyzed using the additive model in the following way:

```
anova wloss program site

        Number of obs =      30     R-squared     =  0.3808
        Root MSE      = 4.27725     Adj R-squared =  0.3093

    Source |  Partial SS    df       MS           F     Prob > F
-----------+----------------------------------------------------
     Model |       292.5     3        97.5     5.33     0.0054
           |
   program |  274.866667     2  137.433333     7.51     0.0027
      site |  17.6333333     1  17.6333333     0.96     0.3353
           |
  Residual |  475.666667    26  18.2948718
-----------+----------------------------------------------------
     Total |  768.166667    29  26.4885057
```

Weight loss, work site, and program: The analyzes in Stata III

Fitted values can be obtained using `margins` as in the one-way ANOVA:

```
margins program#site

-----------------------------------------------------------------------
                |            Delta-method
                |     Margin   Std. Err.      t    P>|t| [95% Conf. Interval]
----------------+------------------------------------------------------
program#site    |
    Diet#Office| 5.333333    1.561831    3.41   0.002    2.122944  8.543723
   Diet#Factory| 6.866667    1.561831    4.40   0.000    3.656277  10.07706
Exercise#Office| 1.233333    1.561831    0.79   0.437   -1.977056  4.443723
Exercise#Factory| 2.766667   1.561831    1.77   0.088    -.443723  5.977056
    Both#Office| 8.633333    1.561831    5.53   0.000    5.422944  11.84372
   Both#Factory| 10.16667    1.561831    6.51   0.000    6.956277  13.37706
-----------------------------------------------------------------------
```

Weight loss, work site, and program: The analyzes in Stata IV

Estimated effects can be retrieved in this way:

```
pwcompare program site, eff

-------------------------------------------------------------------------
                  |                            Unadjusted      Unadjusted
                  | Contrast  Std. Err.   t      P>|t|  [95% Conf. Interval]
------------------+------------------------------------------------------
program           |
 Exercise vs Diet |     -4.1  1.912845 -2.14   0.042 -8.031908   -.1680917
     Both vs Diet |      3.3  1.912845  1.73   0.096 -.6319083    7.231908
 Both vs Exercise |      7.4  1.912845  3.87   0.001  3.468092    11.33191
                  |
site              |
Factory vs Office | 1.533333  1.561831  0.98   0.335 -1.677056    4.743723
-------------------------------------------------------------------------
```

Example: BMI and diastolic blood pressure for three groups

Data: BMI (kg/m^2) and diastolic blood pressure (mmHg) for three groups of subjects: Non-smoking controls, non-smoking type 2 diabetics, and smoking controls.

Controls (non-smokers)				Diabetics (non-smokers)				Controls (smokers)			
BMI	DBP	BMI	DBP	BMI	DBP	BMI	DBP	BMI	DBP	BMI	DBP
24.3	72	21.6	56	23.0	69	22.9	69	24.4	81	23.9	72
25.4	81	26.7	90	23.6	65	25.7	89	27.6	108	24.2	82
24.2	75	26.4	84	25.4	88	26.0	81	21.0	73	23.3	78
23.2	77	22.9	70	24.9	79	22.8	69	26.6	100	24.0	78
23.9	58	24.7	74	27.8	96	25.3	78	21.9	71	22.9	73
24.1	80	24.8	75	25.8	78	22.1	75	25.0	82	20.5	63
25.5	82	24.1	78	26.8	86	23.7	76	22.9	73	23.7	75
23.9	70	21.9	66	28.8	104	25.0	83	26.7	92	25.1	80
25.0	77	25.1	78	27.9	96	26.1	88	22.4	73	25.4	91
22.3	73	27.9	98	23.0	67	26.8	87	27.0	94	27.3	103

BMI and diastolic blood pressure: Questions of interest

Purpose: We want to compare smokers and non-smokers, and diabetics and non-diabetics with respect to some common public health characteristics namely BMI and diastolic blood pressure.

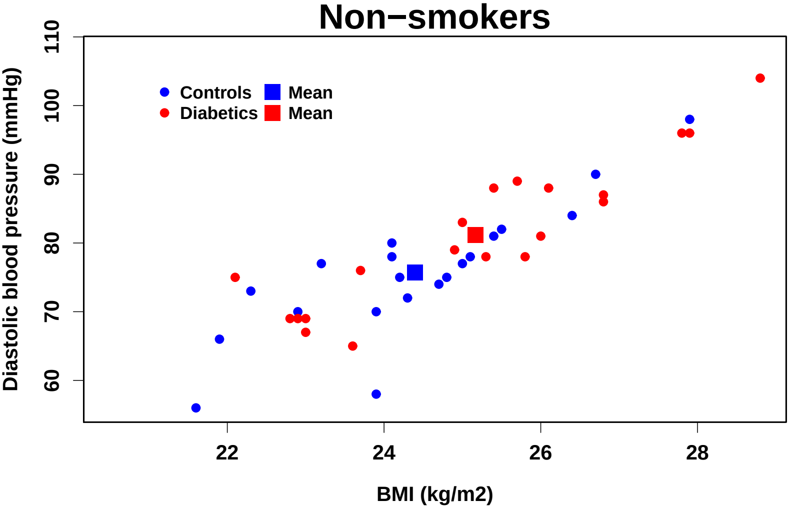
We could compare the groups based on each characteristic separately, but

- We would rather include all information in **one analysis** instead of splitting the analysis up in several analyses each based on a small chunk of the data.
- We might expect that the BMI and diastolic blood pressure measurements on the same individual are **correlated**.
- We might expect an **effect** of smoking and/or diabetes **on both characteristics**.

One way to use all the available information and investigate these expectations is to use **MANOVA**.

We will focus on comparing non-smoking diabetics and non-diabetics (groups 1 and 2), and comparing smoking and non-smoking controls (groups 1 and 3).

BMI and diastolic BP: Effect of diabetes - scatterplot (Group 1 and 2)



BMI and diastolic BP: Effect of diabetes - individual t-tests

We can compare diabetic and non-diabetic non-smokers separately for each characteristic:

BMI:

Group	n	Mean	sd	95%−CI
Controls	20	24.4	1.586	23.6 − 25.2
Diabetics	20	25.2	1.913	24.4 − 26.0

$$t = -1.395 \sim t(38), \quad p = 0.17$$

Diastolic blood pressure:

Group	n	Mean	sd	95%−CI
Controls	20	76	9.614	71 − 80
Diabetics	20	81	10.693	77 − 86

$$t = -1.695 \sim t(38), \quad p = 0.10$$

Conclusion: No significant effect of diabetes for either characteristic.

Assumptions

For each subject:

- Several measurements (for example BMI and diastolic blood pressure).
- The measurements can either be different characteristics or the same quantity measured several times (repeated measurements which we will return to on day 2).

When comparing two (or more) groups:

Assumptions:

1. Each measured variable is described by a **normal distribution**.
2. Measurements corresponding to different subjects are **independent**. Measurements corresponding to the same subject are correlated.
3. Standard deviations and correlations between the different measurements on the same subject are the **same** in the different groups.

Comment: The assumptions are similar to those behind ANOVA.

Comparing two groups: Hotellings T^2

Comparing **two groups** with respect to several variables (for example BMI and diastolic blood pressure) by testing the hypothesis

$$H_0: \quad \mu_{\text{Control, BMI}} = \mu_{\text{Diabetic, BMI}} \quad \text{and} \quad \mu_{\text{Control, DBP}} = \mu_{\text{Diabetic, DBP}}$$

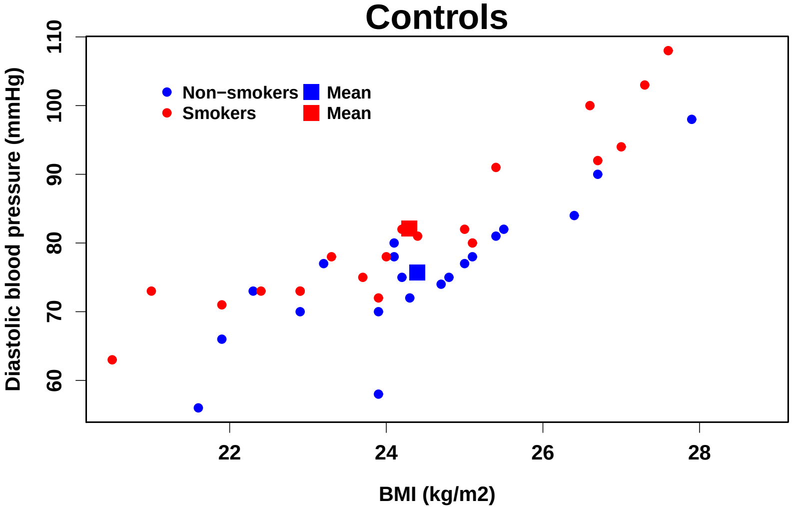
Hotellings T^2 is the multivariate equivalent of the *t*-test.
In the example we get:

$$T^2 = 2.918, \quad F = 1.421 \sim F(2, 37), \quad p = 0.25$$

Conclusion: No evidence against the hypothesis of equal means in the two groups.

So neither the individual *t*-tests nor the multivariate test lead us to conclude that there is a significant effect of diabetes on the chosen public health characteristics.

BMI and diastolic BP: Effect of smoking - scatterplot (Group 1 and 3)



Effect of smoking: Individual t-tests and multivariate test

Comparing group 1 (non-smoking controls) and 3 (smoking controls) we get:

Individual t-test:

Characteristic	Group	Mean	sd	*p*
BMI	Non-smoker	24.4	1.586	
	Smoker	24.3	2.062	0.86
DBP	Non-smoker	76	9.614	
	Smoker	82	12.048	0.07

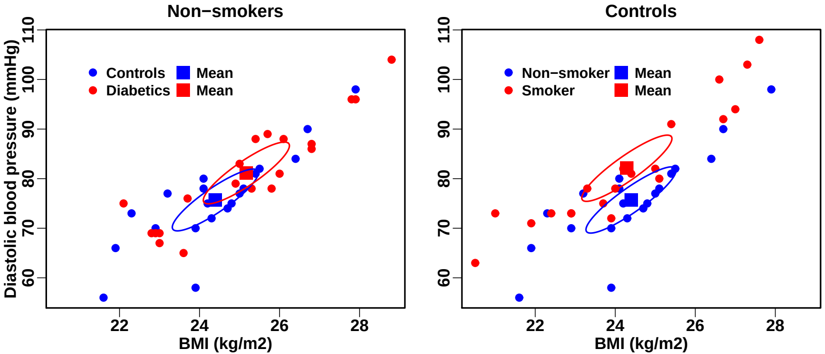
Multivariate test:

$$\mathrm{T}^2 = 20.662, \quad F = 10.059 \sim F(2, 37), \quad p = 0.0003$$

Conclusion: Based on the individual t-tests there is no significant effect of smoking on the public health characteristics under study here. However, when combining the two characteristics we find a **clearly significant effect of smoking** by the multivariate test!

BMI and diastolic BP: Confidence regions for the means

The different results (a higher p-value when comparing controls and diabetics and a lower p-value when comparing smokers and non-smokers) can best be understood by looking at the confidence regions for the means.



If the difference is in the direction dictated by the association between BMI and diastolic blood pressure, it takes a much larger difference to get a significant difference between the 2 groups.

BMI and diastolic BP: Checking the assumptions behind MANOVA

How should we go about validating the assumptions behind the MANOVA?

- For each of the measured variables we can make validations as in ANOVA.
- The tests are **not affected** by moderate departures from normality, especially for large number of observations in each group.
- The F-test is more **sensitive** to departures from the assumption of equal standard deviations and correlations, **unless**
- The **group sample sizes are almost equal**.

Standard deviations and correlations in the three groups:

Control, non-smoker

$$\begin{array}{c} \\ \text{BMI} \\ \text{DBP} \end{array} \begin{array}{c} \begin{array}{cc} \text{BMI} & \text{DBP} \end{array} \\ \begin{pmatrix} 1.586 & \\ 0.84 & 9.614 \end{pmatrix} \end{array}$$

Diabetic, non-smoker

$$\begin{array}{c} \\ \text{BMI} \\ \text{DBP} \end{array} \begin{array}{c} \begin{array}{cc} \text{BMI} & \text{DBP} \end{array} \\ \begin{pmatrix} 1.913 & \\ 0.91 & 10.693 \end{pmatrix} \end{array}$$

Control, smoker

$$\begin{array}{c} \\ \text{BMI} \\ \text{DBP} \end{array} \begin{array}{c} \begin{array}{cc} \text{BMI} & \text{DBP} \end{array} \\ \begin{pmatrix} 2.062 & \\ 0.93 & 12.048 \end{pmatrix} \end{array}$$

BMI and diastolic BP: Testing equality of standard deviations and correlations

It is possible to test the hypothesis of equal standard deviations and correlations in group 1 and 2 (non-smoking controls and diabetics):

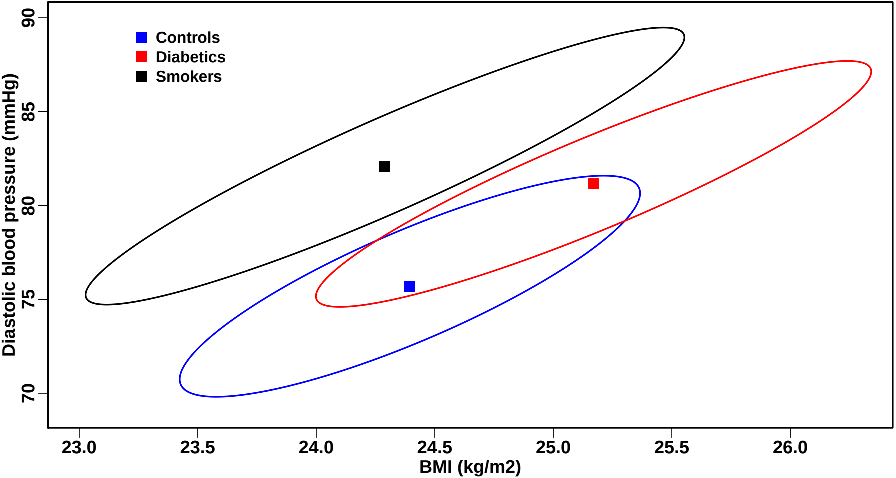
$$M = 1.20 \sim \chi^2(3), \quad p = 0.75$$

Conclusion: There is no evidence against the hypothesis of equal standard deviations and correlations in the two groups.
Similarly, we can test the hypothesis of equal standard deviations and correlations in group 1 and 3 (smoking and non-smoking controls):

$$M = 1.79 \sim \chi^2(3), \quad p = 0.62$$

It is possible to test the hypothesis of equal group means **without assuming equal standard deviations and correlations**, but the price is a lower power.

BMI and diastolic BP: Confidence regions with unequal standard deviations and correlations



MANOVA: Comments

- MANOVA constitutes a way to **compare groups** based on **several correlated measurements** on each subject securing an overall significance level (something which is not easily done using a t-test for each variable).
- Sometimes we will achieve an overall significant group difference if several non-significant differences are combined in a MANOVA (but that is not always the case).
- This also means that there may be no obvious way to present the results of the analysis as the **groups may not differ significantly for any single variable**.
- The results depend on the correlation structure, but the analyses (for example which and how many variables to include in the analysis) should be **planned in advance** before looking at the data.

BMI and diastolic blood pressure: The analyses in Stata

In **Stata** Hotellings T^2 can be calculated using `hotelling`:

```
hotelling BMI diastol if group==1 | group==2, by(group)

2-group Hotelling's T-squared = 2.9180715
F test statistic: ((40-2-1)/(40-2)(2)) x 2.9180715 = 1.4206401
H0: Vectors of means are equal for the two groups
              F(2,37) =    1.4206
       Prob > F(2,37) =    0.2544
```

In **Stata** `mvtest` can be used to test equal sd's and correlations:

```
mvtest covariances BMI diastol if group==1 | group==2, by(group)

Test of equality of covariance matrices across 2 samples
      Modified LR chi2 =  1.273624
       Box F(3,259920) =         .      Prob > F =        .
           Box chi2(3) =      1.20   Prob > chi2 =  0.7528
```

BMI and diastolic BP: Comparing group means using `mvtest`

In **Stata** group means can also be compared using `mvtest`:

```
mvtest mean BMI diastol if group==1 | group==2, by(group)
```

However, it is possible to compare groups means without assuming that the standard deviations and correlations are equal across groups. In **Stata** this is also done using `mvtest`:

```
mvtest mean BMI diastol if group==1 | group==2, het by(group)


Test for equality of 2 group means, allowing for heterogeneity


         MNV F(2,36.6) =      1.42
              Prob > F =    0.2547
```

The conclusion is unchanged as the standard deviations and correlations are very similar in the two groups.