TMA4267 Linear statistical models

13. march 2025

Thea Bjørnland

TMA4267 Linear statistical models

About the course

Examination arrangement

Examination arrangement: School exam

Grade: Letter grades

Evaluation	Weighting	Duration	Examination aids
School exam	100/100	4 hours	C

Course content

~~Random vectors. Multivariate normal distribution. Multiple linear regression.~~ Analysis of variance. ~~Multiple hypothesis testing~~. Design of experiments.

Two-factor anova with interactions

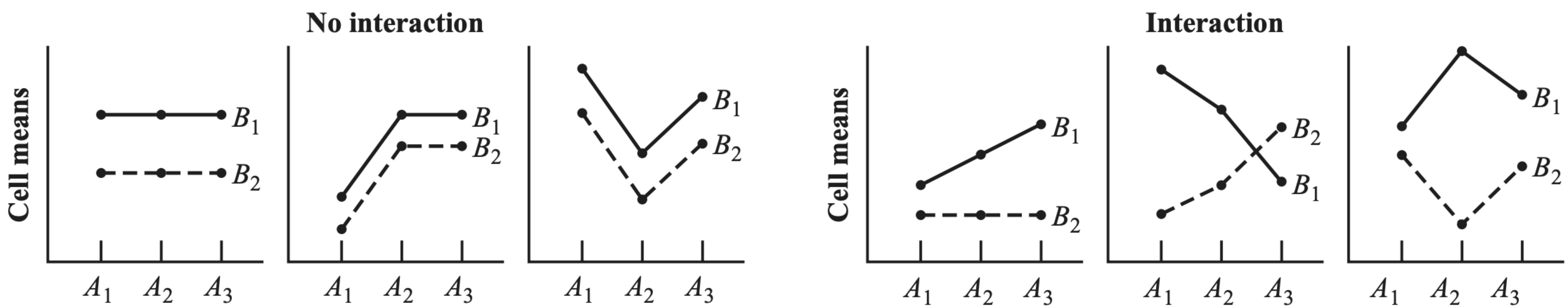


Figure 13.2 Illustration of possible noninteractions and interactions

From the textbook *Statistical methods for psychology* by Howell (2010)

Two-factor anova

Age and memory data

From the textbook *Statistical methods for psychology* by Howell (2010)

11.1 An Example

Many features of the analysis of variance can be best illustrated by a simple example, so we will begin with a study by M. W. Eysenck (1974) on recall of verbal material as a function of the level of processing. The data we will use have the same group means and standard deviations as those reported by Eysenck, but the individual observations are fictional.

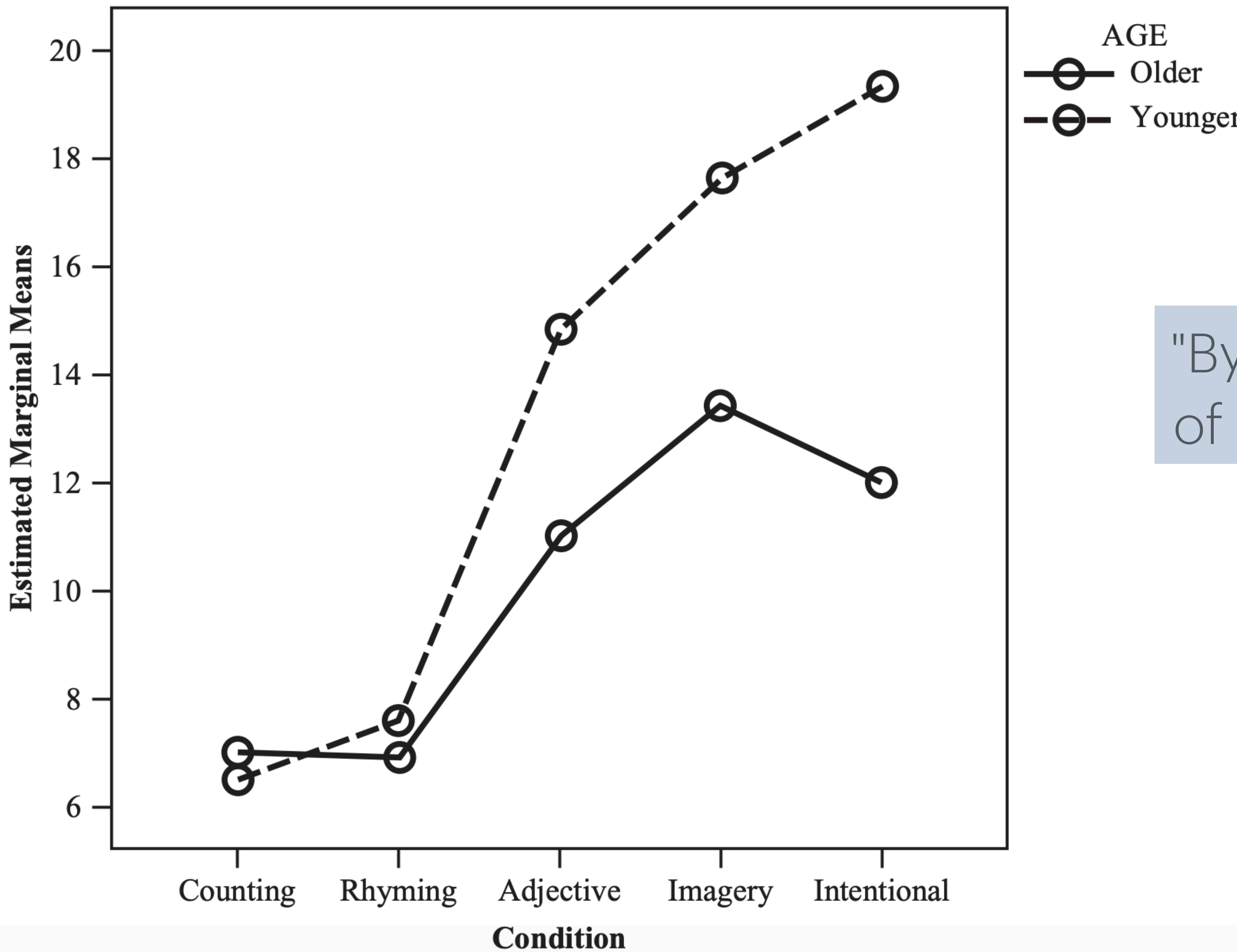
Two-factor anova

Age and memory data

- Eysenck study (1974)
  - 50 older (55-65 years) participants randomly split into 5 groups (10 per group)
  - 50 younger (18-30 years) participants randomly split into 5 groups (10 per group)
  - All groups: read list of 27 words
    - Group 1: Count letters in each word
    - Group 2: Find words that rhyme
    - Group 3: Add adjectives to each word
    - Group 4: Imagery
    - Group 5: Intentionally aim to recall words
  - Afterwards: how many words from the list could the participants recall?

Two-factor anova

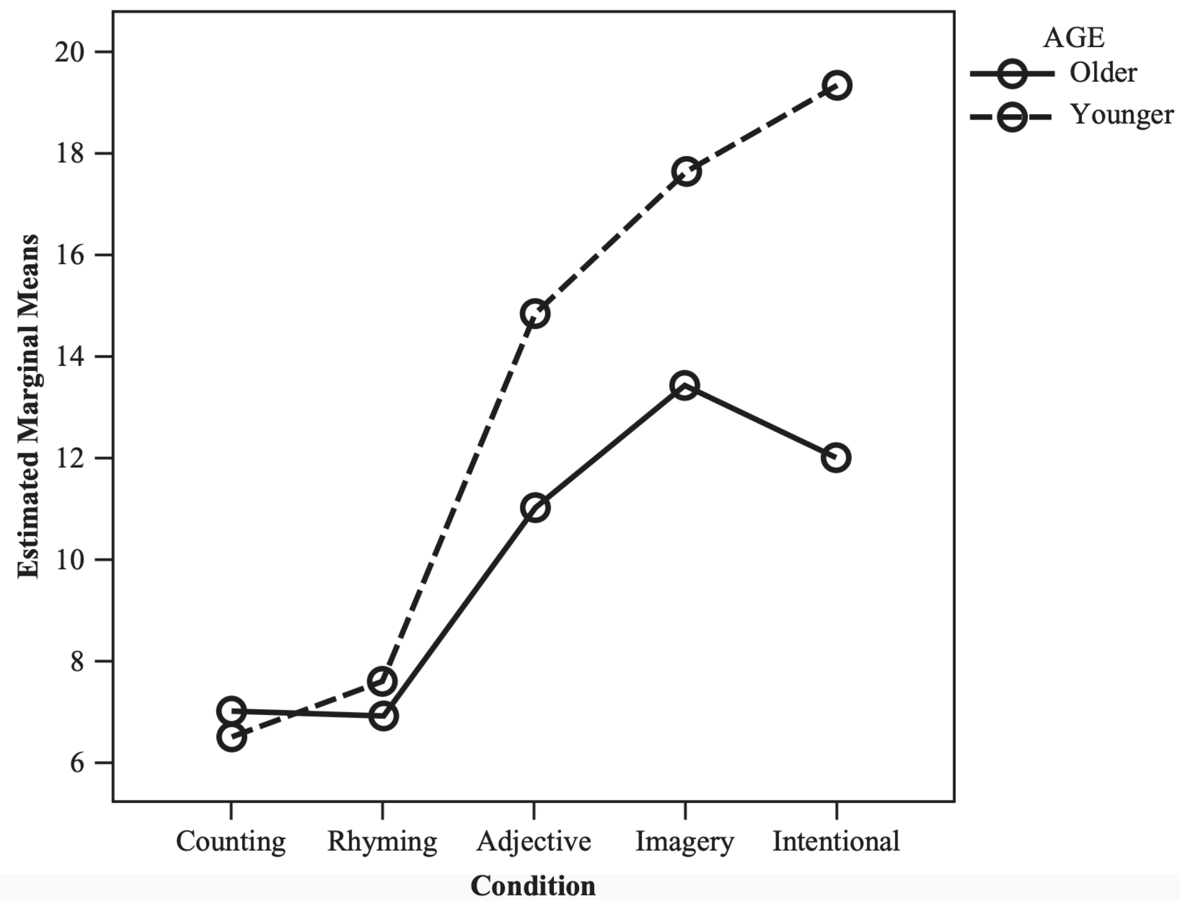
Age and memory data



"By chance" or evidence of an interaction effect?

Two-factor anova

Age and memory data




```
Call:
lm(formula = Words ~ Age * Process, data = ds)

Residuals:
   Min     1Q Median     3Q    Max 
  -7.0   -1.6   -0.5    2.0    9.6 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)    11.6100     0.2833  40.982  < 2e-16 ***
Age1           -1.5500     0.2833  -5.471 3.98e-07 ***
Process1       -4.8600     0.5666  -8.578 2.60e-13 ***
Process2       -4.3600     0.5666  -7.695 1.72e-11 ***
Process3        1.2900     0.5666   2.277  0.02517 *  
Process4        3.8900     0.5666   6.866 8.24e-10 ***
Age1:Process1   1.8000     0.5666   3.177  0.00204 ** 
Age1:Process2   1.2000     0.5666   2.118  0.03694 *  
Age1:Process3  -0.3500     0.5666  -0.618  0.53831    
Age1:Process4  -0.5500     0.5666  -0.971  0.33429    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.833 on 90 degrees of freedom
Multiple R-squared:  0.7293,	Adjusted R-squared:  0.7022 
F-statistic: 26.93 on 9 and 90 DF,  p-value: < 2.2e-16
```

Problem 3

A response variable Y_{ij} was measured, using 15 repetitions for each of four levels of a factor. A regression model of the form $Y_{kj} = \beta_j + \epsilon_{kj}$ was assumed, where $k = 1, 2, \ldots, 15$, $j = 1, 2, 3, 4$, and the ϵ_{kj} were independent $N(0, \sigma^2)$.

Another way to formulate the model is $Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \epsilon_i$, $i = 1, 2, \ldots, 60$, with $x_{ij} = 1$ if the factor was at level j in experiment i and $x_{ij} = 0$ otherwise.

a) Assuming that the factor was at level 1 for $i = 1, \ldots, 15$, at level 2 for $i = 16, \ldots, 30$, at level 3 for $i = 31, \ldots, 45$, and at level 4 for $i = 46, \ldots, 60$, explain how the design matrix X looks (including its dimensions). Show that $(X^{\mathrm{T}}X)^{-1} = \frac{1}{15}I$, with I a 4×4 identity matrix.

Problem 3

A response variable Y_{ij} was measured, using 15 repetitions for each of four levels of a factor. A regression model of the form $Y_{kj} = \beta_j + \epsilon_{kj}$ was assumed, where $k = 1, 2, \ldots, 15$, $j = 1, 2, 3, 4$, and the ϵ_{kj} were independent $N(0, \sigma^2)$.

Another way to formulate the model is $Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \epsilon_i$, $i = 1, 2, \ldots, 60$, with $x_{ij} = 1$ if the factor was at level j in experiment i and $x_{ij} = 0$ otherwise.

a) Assuming that the factor was at level 1 for $i = 1, \ldots, 15$, at level 2 for $i = 16, \ldots, 30$, at level 3 for $i = 31, \ldots, 45$, and at level 4 for $i = 46, \ldots, 60$, explain how the design matrix X looks (including its dimensions). Show that $(X^{\mathrm{T}}X)^{-1} = \frac{1}{15}I$, with I a 4×4 identity matrix.

The least-squares estimates of β_3 and of β_4 were 1.0902858 and 0.1752633, respectively, and the error sum of squares was SSE = 43.04524.

b) Perform a test in which the null hypothesis is $H_0\colon \beta_3 = \beta_4$ and the alternative hypothesis is $H_1\colon \beta_3 \neq \beta_4$. Use significance level 0.05. You should calculate a test statistic and use its distribution under H_0 to arrive at your conclusion.

Kritiske verdier i Fisherfordelingen

$$P(F > f_{0.05,\nu_1,\nu_2}) = 0.05$$

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9	10
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

A corresponding test was performed for all pairs of coefficients. The p-values are given in the following table.

H_0	$\beta_1 = \beta_2$	$\beta_1 = \beta_3$	$\beta_1 = \beta_4$	$\beta_2 = \beta_3$	$\beta_2 = \beta_4$	$\beta_3 = \beta_4$
p-value	0.0251	0.3698	0.0557	0.0022	0.7297	0.0060

c) What is *family-wise error rate* (FWER)? Suggest a method that keeps the familywise error rate below 0.05 when performing the tests. Which null hypotheses are rejected?

Problem 2, exam 2010 (modified)

A study was conducted to determine whether lifestyle change could replace medication in reducing blood pressure among hypertensives (people with very high blood pressure).

The levels of treatment considered were
H = Healthy diet with an exercise program,
M = Medication,
N = No intervention.

12 people with very high blood pressure participated in the experiment. Four of these (randomly selected) were put on a healthy diet and followed an exercise program, four were given medication and the rest got no treatment.

Aim: Compare H and M to N. Find a suitable design matrix

Aim: Compare H to M. How?

Note: Many tests! Use Bonferroni, FWER = 0.05