Main topics

- Principles of experimental design
- 1-Factor Anova
- Block designs
- Factorial designs
- Fractional factorials
- Split plot designs

Multi-Factor Experiments

1. Twoway anova

2. More than two factors

Hypertension: Effect of biofeedback

Biofeedback + Medication	Biofeedback	Medication	Control
158	188	186	185
163	183	191	190
173	198	196	195
178	178	181	200
168	193	176	180

Main effects

Treatment means:

	Medication		
	no	yes	
Biofeedback			Total
no	190	186	188
yes	188	168	178
Total	189	177	183

Main effect of biofeedback: $188 - 178 = 10$ mmHg

Question: What is the main effect of medication? 12mmHg

Interaction

Effect of biofeedback with medi: 18 mmHg

$\neq$

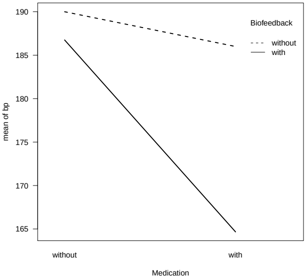
Effect of biofeedback without medi: 2 mmHg

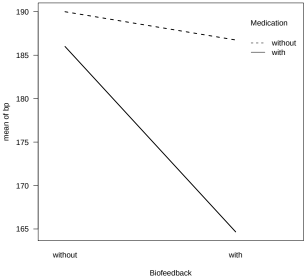
$\longrightarrow$ Interaction

Interaction effect: half this difference=8

Question: What do we get when we compare the effect of medication with and without biofeedback?

Interaction plots





Model for two factors

$$Y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + \epsilon_{ijk}$$

$$i = 1, \ldots, I; j = 1, \ldots, J; k = 1, \ldots, n.$$

$$\sum A_i = 0,\ \sum B_j = 0,\ \sum_i (AB)_{ij} = \sum_j (AB)_{ij} = 0.$$

A_i : ith effect of factor A

B_j : jth effect of factor B

$\mu + A_i + B_j$: overall mean + effect of factor A on level i + effect of factor B on level j

$(AB)_{ij}$: deviation from additive model

Parameter estimation

$\hat{\mu} = y_{...}$, $\hat{A}_i = y_{i..} - y_{...}$ and $\hat{B}_j = y_{.j.} - y_{...}$
$\widehat{AB}_{ij} = y_{ij.} - (\hat{\mu} + \hat{A}_i + \hat{B}_j) = y_{ij.} - y_{i..} - y_{.j.} + y_{...}$

	Medication A		
Biofeedback B	no	yes	Total
no	190	186	$y_{.1.} = 188$
yes	188	168	$y_{.2.} = 178$
Total	$y_{1..} = 189$	$y_{2..} = 177$	$y_{...} = 183$

$\hat{\mu} = 183,\ \hat{A}_1 = -\hat{A}_2 = 6,\ \hat{B}_1 = -\hat{B}_2 = 5$

Predicted values of the additive model

Predictions: $(\hat{\mu} + \hat{A}_i + \hat{B}_j)$

	Medication	
Biofeedback	no	yes
no	194	182
yes	184	172

$$\hat{y}_{11} = \mu + \hat{A}_1 + \hat{B}_1 = 183 + 6 + 5 = 194$$

$$\widehat{AB}_{11} = \widehat{AB}_{22} = -\widehat{AB}_{12} = -\widehat{AB}_{21} = -4.$$

Decomposition of Variability

$$SS_{tot} = SS_A + SS_B + SS_{AB} + SS_{res}$$

$$\begin{aligned}
SS_{tot} &= \sum\sum\sum(y_{ijk} - y_{...})^2 \\
SS_A &= \sum\sum\sum(y_{i..} - y_{...})^2 \\
SS_B &= \sum\sum\sum(y_{.j.} - y_{...})^2 \\
SS_{AB} &= \sum\sum\sum(y_{ij.} - y_{i..} - y_{.j.} + y_{...})^2 \\
SS_{res} &= \sum\sum\sum(y_{ijk} - y_{ij.})^2
\end{aligned}$$

degrees of freedom: main effect with I levels: $I - 1$ df,
interaction between 2 factors with I and J levels: $(I - 1)(J - 1)$ df.

Anova table

Source	SS	df	MS	F	P value
medi	720	1	720	11.52	0.004
bio	500	1	500	8.00	0.012
medi:bio	320	1	320	5.12	0.038
residual	1000	16	62.5		
total	2540	19			

Treatment effects

	effect size	C.I.
medi without bio:	$y_{21.} - y_{11.} = 186 - 190 = -4$	$(-18, 10)$
medi with bio:	$y_{22.} - y_{12.} = 168 - 188 = -20$	$(-34, -6)$
bio without medi:	$y_{12.} - y_{11.} = 188 - 190 = -2$	$(-16, 12)$
bio with medi:	$y_{22.} - y_{21.} = 168 - 186 = -18$	$(-32, -4)$

(standard error: $\sqrt{2 \cdot MS_{res}/5} = 5$)

Question: How are the CI limits calculated?

More than two factors

	bio/medi	bio	medi	control
without diet	158	188	186	185
	163	183	191	190
	173	198	196	195
	178	178	181	200
	168	193	176	180
with diet	162	162	164	205
	158	184	190	199
	153	183	169	171
	182	156	165	161
	190	180	177	179

Treatment means

without diet(C)

	medication (A)		
bio (B)	no	yes	total
no	190	186	188
yes	188	168	178
total	189	177	183

with diet (C)

	medication (A)		
bio (B)	no	yes	total
no	183	173	178
yes	173	169	171
total	178	171	174.5

Main effects and interactions

Main effects A, B, C

Difference of the response on the two levels averaged over all other factor levels.

medication (A):	$183.5 - 174 = 9.5$
biofeedback (B):	$183 - 174.5 = 8.5$
diet(C):	$183 - 174.5 = 8.5$

2-way interactions AB, AC, BC

Average over all but two factors.

	medi (A)		
Bio (B)	no	yes	total
no	186.5	179.5	183
yes	180.5	168.5	174.5
total	183.5	174	178.75

Effect of medi without bio: 7
Effect of medi with bio: 12
Interaction effect: 2.5

Main effects and interactions cont.

3-way interaction ABC

Difference of the 2-way interaction effect between the levels of the third factor.

Interaction effect AB without diet= 8
Interaction effect AB with diet= -3
Half this difference=11/2=5.5

Model and Anova table

$$Y_{ijkl} = \mu + A_i + B_j + C_k + (AB)_{ij} + (AC)_{ik} + (BC)_{jk} + (ABC)_{ijk} + \epsilon_{ijkl}$$

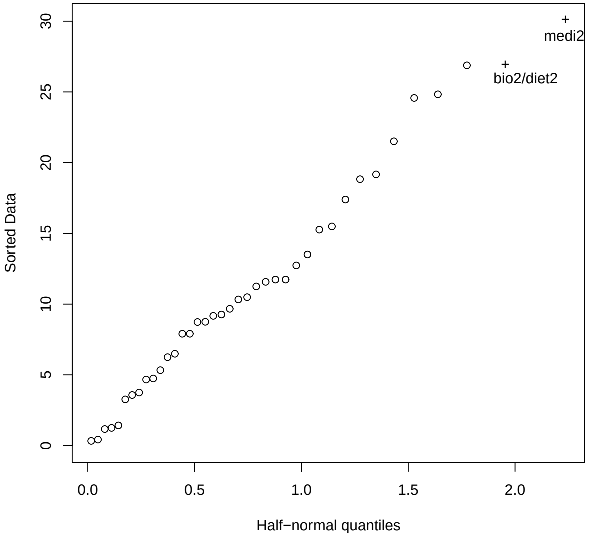
with constraints $\sum A_i = 0, \ldots, \sum_k (ABC)_{ijk} = 0$

Source	df	MS=SS/df	F
A	$I-1$		MS_A/MS_{res}
B	$J-1$		MS_B/MS_{res}
C	$K-1$		MS_C/MS_{res}
AB	$(I-1)(J-1)$		MS_{AB}/MS_{res}
AC	$(I-1)(K-1)$		MS_{AC}/MS_{res}
BC	$(J-1)(K-1)$		MS_{BC}/MS_{res}
ABC	$(I-1)(J-1)(K-1)$		MS_{ABC}/MS_{res}
Residual	«Difference»	$MS_{res} = \hat{\sigma}^2$	
Total	$IJKn-1$		

Anova table

Source	SS	df	MS	F	P value
medi	902.5	1	902.5	6.33	0.017
bio	722.5	1	722.5	5.06	0.031
diet	722.5	1	722.5	5.06	0.031
medi:bio	62.5	1	62.5	0.44	0.51
medi:diet	62.5	1	62.5	0.44	0.51
bio:diet	22.5	1	22.5	0.16	0.69
medi:bio:diet	302.5	1	302.5	2.12	0.15
Residual	4566.0	32	142.7		
Total	7363.5	39			

Half normal plot



Unbalanced Factorials

uncorrelated estimators:

$$SS_{tot} = SS_A + SS_B + SS_{AB} + \underbrace{SS_{res}}_{SS_C + \ldots + SS_{res'}}$$

correlated estimators:

$$SS_{tot} = SS'_A + SS'_B + SS'_{AB} + SS_C + \ldots + SS_{res'}$$

SS Typ I: SS_A ignores all other SS

SS Typ II: SS_A takes into account all other main effects, ignores all interactions

SS Typ III: SS_A takes into account all other effects

Calculation of SS's

by model comparison

For SS Typ I:

model 1:	$Y_{ijk} = \mu + \epsilon_{ijk}$	$SS_{e1} = SS_T$
model 2:	$Y_{ijk} = \mu + A_i + \epsilon_{ijk}$	SS_{e2}
model 3:	$Y_{ijk} = \mu + A_i + B_j + \epsilon_{ijk}$	SS_{e3}
model 4:	$Y_{ijk} = \mu + A_i + B_j + AB_{ij} + \epsilon_{ijk}$	$SS_{e4} = SS_{res}$

Rat genotype

- Litters of rats are separated from their natural mother and given to another female to raise.
- 2 factors: mother's genotype (A, B, I, J) and litter's genotype (A, B, I, J)
- response: average weight gain of the litter.

Full model

```
> summary(aov(y~mother*genotype,data=gen))
                Df  Sum Sq Mean Sq F value   Pr(>F)
mother           3  771.61 257.202  4.7419 0.005869 *
genotype         3   63.63  21.211  0.3911 0.760004
mother:genotype  9  824.07  91.564  1.6881 0.120053
Residuals       45 2440.82  54.240
---
> summary(aov(y~genotype*mother,data=gen))
                Df  Sum Sq Mean Sq F value   Pr(>F)
genotype         3   60.16  20.052  0.3697 0.775221
mother           3  775.08 258.360  4.7632 0.005736 *
genotype:mother  9  824.07  91.564  1.6881 0.120053
Residuals       45 2440.82  54.240
```

SS Typ III in R

```
> drop1(mod1,test="F")
Model: y ~ mother * genotype
            Df  Sum Sq    RSS    AIC  F value  Pr(F)
<none>                 2440.8 257.04
mother:geno  9  824.07 3264.9 256.79  1.6881 0.1201

> drop1(mod1,.~.,test="F")
Model: y ~ mother * genotype
            Df  Sum Sq   RSS    AIC   F value   Pr(F)
<none>                 2440.8 257.04
mother       3  582.25 3023.1 264.09  3.5782 0.02099 *
geno         3  591.69 3032.5 264.28  3.6362 0.01968 *
mother:geno  9  824.07 3264.9 256.79  1.6881 0.12005
```

Offer for a 6-year old car

- Planned experiment to see whether the offered cash for the same medium-priced car depends on gender or age (young, middle, elderly) of the seller.
- 6 factor combinations with 6 replications each.
- Response variable y is offer made by a car dealer (in $ 100)
- Covariable: overall sales volume of the dealer

Analysis of Covariance

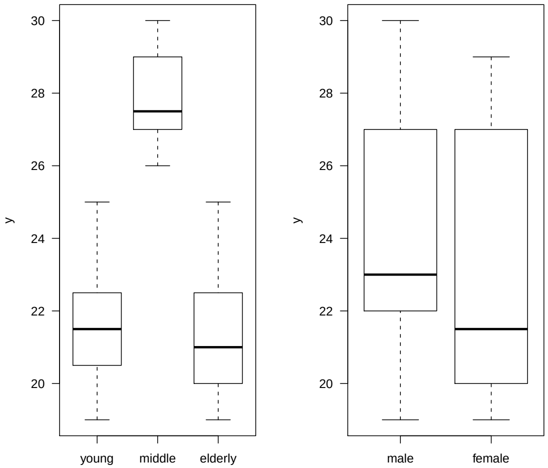
- Covariates can reduce MS_{res}, thereby increasing power for testing.
- Baseline or pretest values are often used as covariates. A covariate can adjust for differences in characteristics of subjects in the treatment groups.
- It should be related only to the response variable and not to the treatment variables (factors).
- We assume that the covariate will be linearly related to the response and that the relationship will be the same for all levels of the factor (no interaction between covariate and factors).

Model for two-way ANCOVA

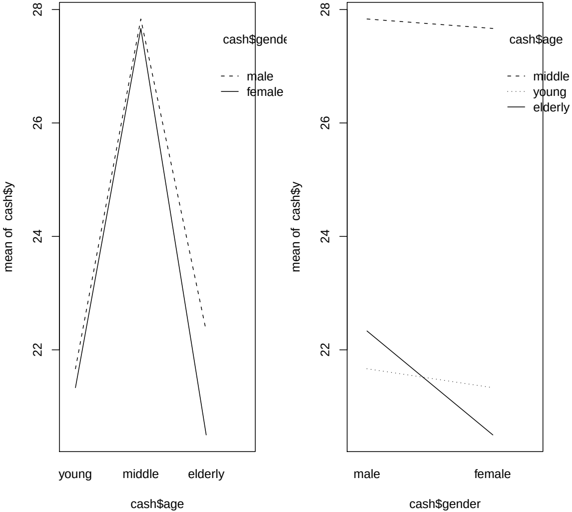
$$Y_{ijk} = \mu + \theta x_{ijk} + A_i + B_j + (AB)_{ij} + \epsilon_{ijk}$$

$$\sum A_i = \sum B_j = \sum (AB)ij = 0, \quad \epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$$

Effect of Age and Gender



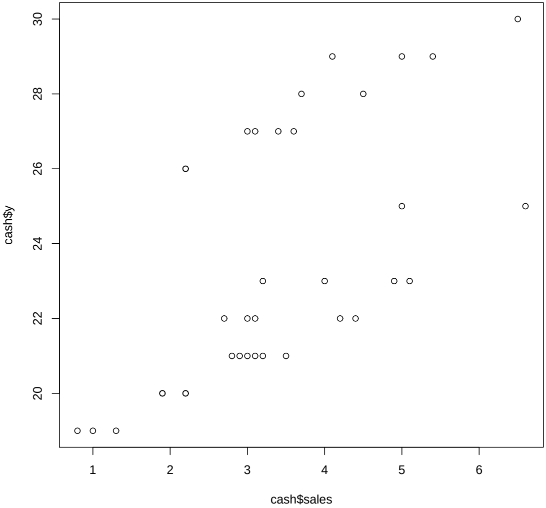
Interaction effect of Age and Gender



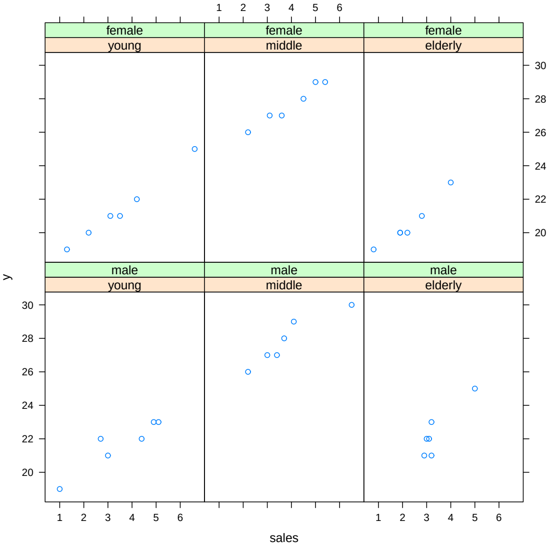
Two-way Anova

```
> mod1=aov(y~age*gender,data=cash)
> summary(mod1)
           Df Sum Sq MeanSq Fvalue  Pr(>F)
age         2 316.72 158.36 66.29  9.79e-12***
gender      1   5.44   5.44  2.28   0.1416
age:gender 2   5.06   2.53  1.06   0.3597
Residuals 30  71.67   2.39
```

Sales and Cash Offer



Sales and Cash Offer by Group



Two-way Ancova

```
> mod2=aov(y~sales+age*gender,data=cash)
> summary(mod2)
          Df  SumSq MeanSq Fvalue  Pr(>F)
sales      1 157.37 157.37 550.22 < 2e-16***
age        2 231.52 115.76 404.75 < 2e-16***
gender     1   1.51   1.51   5.30 0.02874*
age:gender 2   0.19   0.10   0.34 0.71422
Residuals 29   8.29   0.29
```