Kapitel 12
Variansanalyse

Peter Tibert Stoltze
stat@peterstoltze.dk

Elementær statistik
F2011

Version 7. april 2011

Indledning

- Sammenligning af middelværdien i to grupper indenfor en stikprøve kan foretages med t-test i en passende variant
- Hvis der er tre eller flere grupper kan man lave en række parvise t-tests, men ved mange tests skal man justere p-værdien (signifikanssandsynligheden) for disse multiple sammenligninger
- Bedre er det at lave variansanalyse (ANOVA), hvor vi sammenligner variansen indenfor grupper med variansen mellem grupper for at vurdere, om grupperne er forskellige
- Grupperne er karakteriset ved en eller flere faktorer på et bestemt niveau — faktorerne er typisk målt på nominal eller ordinal skala

Eksempel

- Et eksempel kan være score i en læsetest, hvor der for hver elev på et bestemt klassetrin er noteret resultat, køn og familiebaggrund
- Køn er en faktor med to niveauer:
  - Pige
  - Dreng
- Familiebaggrund er en faktor med her tre niveauer:
  - Bor hos begge forældre
  - Bor hos én af sine forældre
  - Bor ikke hos forældre

Eksempel *(fortsat)*

- I alt haves $2 \times 3 = 6$ grupper, og variansanalysen giver os et mål for faktorernes indflydelse på responset (scoren)
- Betragter vi kun køn kan vi klare os med et t-test
- Betragter vi kun familiebaggrund skal vi lave ensidet variansanalyse
- Betragter vi både køn og familiebaggrund (og eventuelt vekselvirkninger mellem disse) skal vi lave tosidet variansanalyse

Balancerede data

- Er der tale om et forsøg, så kan man planlægge at have lige mange observationer i hver gruppe — man siger da, at forsøget er balanceret
- Er der tale om en ikke-eksperimentel undersøgelse er data oftest ubalancerede — dette behøver ikke at være et problem
- I balancerede forsøg kan variansanalysen beregnes i hånden, da kvadratsummerne kan beregnes på en let måde
- Der er lidt mere arbejde i ubalancerede forsøg...

Ikke-parametrisk alternativ

- Ved sammenligning mellem to stikprøver kan t-testet erstattes af Mann-Whitney testet, hvis forudsætninger for t-testet ikke er opfyldt (fx hvis data ikke kan antages at stamme fra normalfordelte populationer)
- Det ikke-parametriske alternativ til ensidet ANOVO hedder Kruskal-Wallis testet, der er en udvidelse af Mann-Whitney testet til at håndtere tre eller flere grupper
- Vi gennemgår ikke dette test, men nu ved I at det findes...

Notation og model

- Antag at vi har a uafhængige stikprøver på normalfordelte populationer med *samme* varians, og at hver stikprøve består af n observationer
- Lad nu X_{ij} betegne den j'te observation i den i'te stikprøve
- Vi har nu an uafhængige stokastiske variable X_{ij}, $i = 1, \ldots, a$ og $j = 1, \ldots, n$, hvor $X_{ij} \sim \mathrm{N}(\mu_i, \sigma^2)$

Notation og model *(fortsat)*

- Vores model er nu
$$X_{ij} = \mu_i + \varepsilon_{ij}$$
hvor $\{\varepsilon_{ij}\}$ er uafhængige og $\mathrm{N}(0, \sigma^2)$
- Modellen kan også skrives
$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$
hvor
$$\mu = \frac{1}{a}\sum_{i=1}^{a} \mu_i \quad \text{og} \quad \alpha_i = \mu_i - \mu$$
så α_i er effekten af α-faktorens $i$te niveau

Alternativ formulering

- En alternativ formulering af modellen er

$$\begin{aligned} x_{ij} &= x_{ij} + \bar{x}_{i\cdot} - \bar{x}_{i\cdot} + \bar{x}_{\cdot\cdot} - \bar{x}_{\cdot\cdot} \\ &= \bar{x}_{\cdot\cdot} + \underbrace{(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})}_{\text{gruppesnit ift. fællessnit}} + \underbrace{(x_{ij} - \bar{x}_{i\cdot})}_{\text{individ ift. gruppesnit}} \end{aligned}$$

- En observation kan altså dekomponeres i et generelt niveau, gruppens afvigelse fra det generelle niveau og endelig en individuel afvigelse fra gruppens niveau.

Hypotese og test

- Hypotesen om, at alle stikprøverne har ens gennemsnit (faktoren α har ingen betydning) kan formuleres som

$$\mathrm{H}_0 : \quad \mu_1 = \mu_2 = \ldots = \mu_a$$

eller ækvivalent hermed

$$\mathrm{H}_0 : \quad \alpha_i = 0, \quad i = 1, \ldots, a$$

- Alternativhypotesen er at $\alpha_i \neq 0$ for i hvert fald ét i

$$\mathrm{H}_1 : \quad \exists i \in \{1, \ldots, a\}, \alpha_i \neq 0$$

Hypotese og test

Variation	Kvadratsum SAK	Frihedsgrader DF	Middelsum MS	F-værdi F_{obs}
Mellem grupper	SAK_a	$a-1$	$\text{MS}_a = \dfrac{\text{SAK}_a}{a-1}$	$F_{\text{obs}} = \dfrac{\text{MS}_a}{\text{MS}_e}$
Indenfor grupper	SAK_e	$a(n-1)$	$\text{MS}_e = \dfrac{\text{SAK}_e}{a(n-1)}$	
Total	SAK_{tot}	$an-1$		

Hypotese og test

- Beregning af kvadratsummer:

$$\text{SAK}_{\text{tot}} = \sum_{i=1}^{a}\sum_{j=1}^{n}(X_{ij} - \bar{X}_{\cdot\cdot})^2 = \sum_{i=1}^{a}\sum_{j=1}^{n} X_{ij}^2 - \frac{1}{an}X_{\cdot\cdot}^2$$

$$\text{SAK}_a = n\sum_{i=1}^{a}(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2 = \frac{1}{n}\sum_{i=1}^{a} X_{i\cdot}^2 - \frac{1}{an}X_{\cdot\cdot}^2$$

$$\text{SAK}_e = \sum_{i=1}^{a}\sum_{j=1}^{n}(X_{ij} - \bar{X}_{i\cdot})^2 = \text{SAK}_{\text{tot}} - \text{SAK}_e$$

Hypotese og test

- Signifikanssandsynligheden p er sandsynligheden for at få en lige så stor eller større teststørrelse F_{obs} givet at nulhypotesen er sand
- Teststørrelsen beregnes ud fra en F-fordeling med $a-1$ frihedsgrader i tæller og $a(n-1)$ frihedsgrader i nævner:

$$p = \mathrm{P}(F_{a-1,a(n-1)} \leq F_{obs})$$

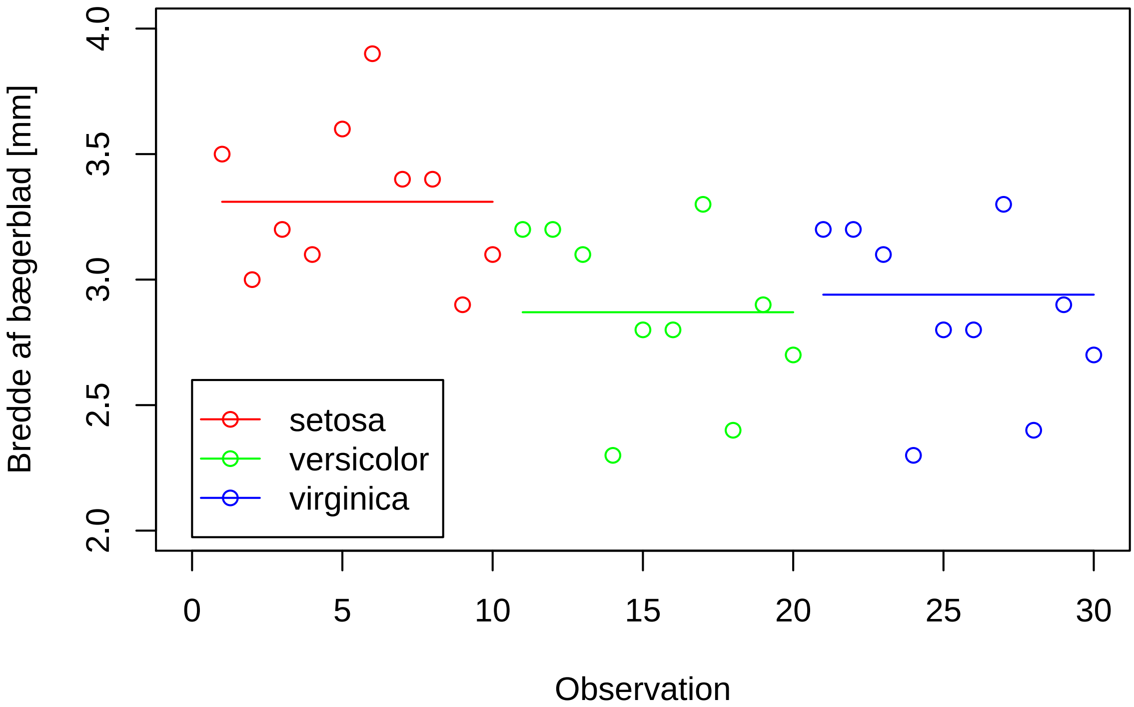
- Hvis p er meget lille følger det, at nulhypotesen (alle gruppper er ens) ikke er ret sandsynlig, så vi antager i stedet alternativhypotesen (mindst to grupper er parvist forskellige)

Eksempel: Bægerblade på *Iris sp.*

- Data indsamlet på Gaspe halvøen i østlige Canada
- Længde og bredde af bæger- og kronblade for 50 individer af tre forskellige arter (i alt 150 observationer af 5 variable)
- Analysen beskrevet i 1936 af R.A. Fisher, der er en af grundlæggerne af den moderne statistik

setosa	*versicolor*	*virginica*
3,5	3,2	3,3
3,0	3,2	2,7
3,2	3,1	3,0
3,1	2,3	2,9
3,6	2,8	3,0
3,9	2,8	3,0
3,4	3,3	2,5
3,4	2,4	2,9
2,9	2,9	2,5
3,1	2,7	3,6

Eksempel: Bægerblade på *Iris sp.*

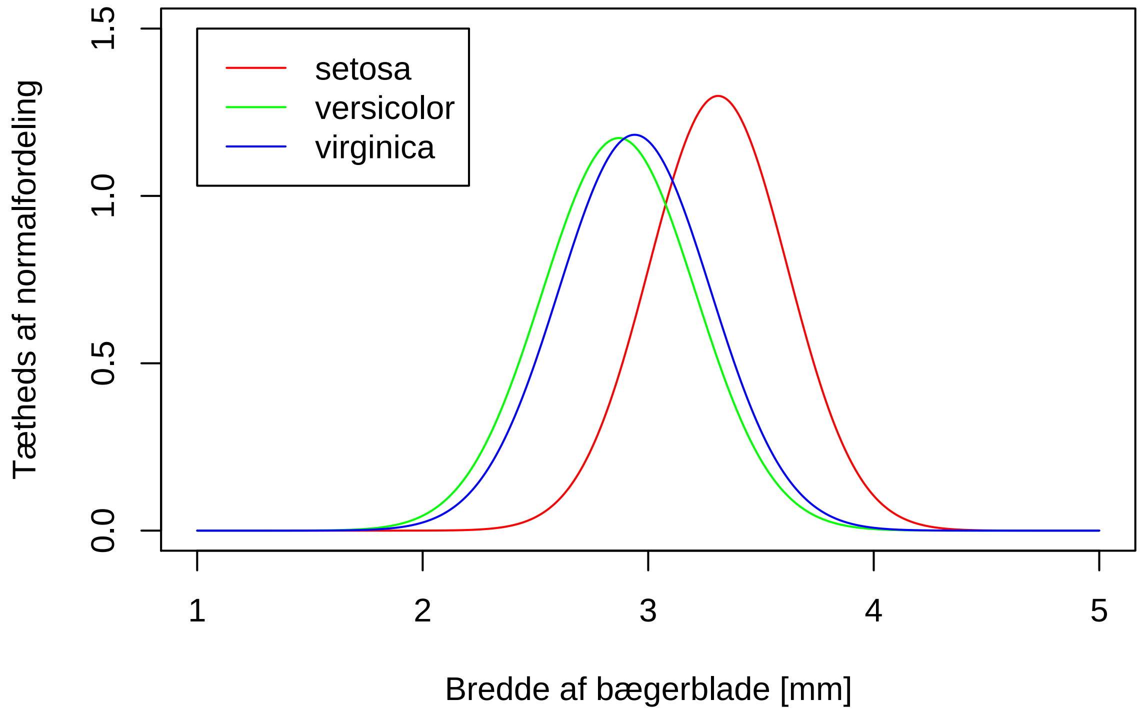


Eksempel: Bægerblade på *Iris sp.*

i	art	n_i	$\bar{X}_{i\cdot}$	S_i
1	*setosa*	10	3,31	0,307
2	*versicolor*	10	2,87	0,340
3	*virginica*	10	2,94	0,337

- Arten *setosa* skiller sig ud fra de to andre med et højere gennemsnit og en lidt lavere spredning
- Vi tester ikke eksplicit for, om spredningerne kan antages at være ens (det burde vi faktisk gøre, fx med Bartletts test eller Levenes test)

Eksempel: Bægerblade på *Iris sp.*



Eksempel: Bægerblade på *Iris sp.*

i	art	n_i	$X_{i\cdot}$
1	*setosa*	10	33,1
2	*versicolor*	10	28,7
3	*virginica*	10	29,4
		30	91,2

$$\sum_{i=1}^{3}\sum_{j=1}^{10} X_{ij}^2 = 281,28$$

$$\text{SAK}_{\text{tot}} = 281,28 - \frac{1}{3\cdot 10} 91,2^2 = 4,0320$$

$$\text{SAK}_a = \frac{1}{10}(33,1^2 + 28,7^2 + 29,4^2) - \frac{1}{3\cdot 10} 91,2^2 = 1,1180$$

$$\text{SAK}_{\text{res}} = 4,0320 - 1,1180 = 2,9140$$

Eksempel: Bægerblade på *Iris sp.*

i	x_{ij}	$\bar{x}_{i\cdot}$	$\bar{x}_{\cdot\cdot}$	$x_{ij} - \bar{x}_{\cdot\cdot}$	$x_{ij} - \bar{x}_{i\cdot}$	$\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot}$	Total $(x_{ij} - \bar{x}_{\cdot\cdot})^2$	Within $(x_{ij} - \bar{x}_{i\cdot})^2$	Between $(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2$
1	3,5	3,31	3,04	0,46	0,19	0,27	0,2116	0,0361	0,0729
1	3,0	3,31	3,04	-0,04	-0,31	0,27	0,0016	0,0961	0,0729
1	3,2	3,31	3,04	0,16	-0,11	0,27	0,0256	0,0121	0,0729
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	3,1	3,31	3,04	0,06	-0,21	0,27	0,0036	0,0441	0,0729
2	3,2	2,87	3,04	0,16	0,33	-0,17	0,0256	0,1089	0,0289
2	3,2	2,87	3,04	0,16	0,33	-0,17	0,0256	0,1089	0,0289
2	3,1	2,87	3,04	0,06	0,23	-0,17	0,0036	0,0529	0,0289
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2	2,7	2,87	3,04	-0,34	-0,17	-0,17	0,1156	0,0289	0,0289
3	3,3	2,94	3,04	0,26	0,36	-0,10	0,0676	0,1296	0,0100
3	2,7	2,94	3,04	-0,34	-0,24	-0,10	0,1156	0,0576	0,0100
3	3,0	2,94	3,04	-0,04	0,06	-0,10	0,0016	0,0036	0,0100
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
3	3,6	2,94	3,04	0,56	0,66	-0,10	0,3136	0,4356	0,0100
				0,00	0,00	0,00	4,0320	2,9140	1,1180

Eksempel: Bægerblade på *Iris sp.*

Variation	Kvadratsum SAK	Frihedsgrader DF	Middelsum MS	F-værdi F_{obs}
Mellem grupper	1,1180	2	0,5590	5,1795
Indenfor grupper	2,9140	27	0,1079	
Total	4,0320	29		

Eksempel: Bægerblade på *Iris sp.*

- Et interval for signifikanssandsynligheden p bestemmes ved at sammenligne med kritiske værdier:

α	0,05	0,01	0,001
$F_{1-\alpha,2,27}$	3,35	5,49	9,02

- Vi kan se at $0,01 < p < 0,05$
- Nøjagtig beregning giver at p = 0,013 så vi kan forkaste nulhypotesen (alle arter ens) med ret stor sikkerhed

Eksempel: Bægerblade på *Iris sp.*

- Da F-testet er signifikant kan vi være interesserede i at vide hvordan de tre stikprøver er forskellige: Er alle tre forskellige eller er to indbyrdes ens
  - Sammenligning af to grupper kan ske med et t-test, hvor signifikanssandsynligheden korrigeres for det antal sammenligninger, man planlægger at lave (Bonferroni)
  - En mere præcis metode er at benytte Scheffes metode, der omfatter et F-test, der i opbygning minder om t-testet
  - Med mange grupper er den letteste metode dog Fishers LSD

Eksempel: Bægerblade på *Iris sp.*

- Hvis F-testet for ANOVA-tabellen er signifikant kan man starte med at lave parvise sammenligninger. Med Sheffes metode sammenlignes to grupper (i og j) med følgende test:

$$F = \frac{(\bar{X}_{i\cdot} - \bar{X}_{j\cdot})^2}{S_e^2(\frac{1}{n_i} + \frac{1}{n_j})}$$

- Teststørrelses evalueres mod en F-test, der har frihedsgrader fra de to første rækker i ANOVA-tabellen. Desuden ganges de kritiske værdier med $k-1$, hvor k er antallet af grupper.
- I vores tilfælde benyttes fraktiler fra en F fordeling med $(2, 27)$ frihedsgrader, der ganges med $3-1$:

	0.05	0.01
$F_{2,27}$	3.35	5.49
$F^\star$	6.70	10.98

Eksempel: Bægerblade på *Iris sp.*

- Vi starter med at teste *versicolor* mod *virginica*

$$F = \frac{(2.87 - 2.94)^2}{0.1079(\frac{1}{10} + \frac{1}{10})} = 0.2271$$

hvilket ses ikke at være signifikant. Vi kan altså (i statistisk forstand) ikke se forskel på de to arter.
- Vi sammenligner derfor *setosa* mod de to andre arter under et

$$F = \frac{(3.31 - \frac{28.7+29.4}{10+10})^2}{0.1079(\frac{1}{10+10} + \frac{1}{10})} = 10.1343$$

og denne sammenligning ses at være signifikant med $0.01 < p < 0.05$

Eksempel: Bægerblade på *Iris sp.*

- Er antallet af grupper højt kan vi benytte Fishers *LSD* (Least SignificantDifference) til hurtigt at få et overblik over hvilke grupper der kan siges at være forskellige. Størrelsen beregnes på niveauet α som

$$LSD_\alpha = t_{1-\frac{\alpha}{2},DF_w}\sqrt{MS_w(\frac{1}{n_i}+\frac{1}{n_j})}$$

- Da alle stikprøver har samme størrelse behøver vi kun at beregne en enkelt LSD-værdi

$$LSD_{0.05} = 2.052\sqrt{0.1079(\frac{1}{10}+\frac{1}{10})} = 0.3014$$

- Grupper der har en absolut forskel større end 0.3014 kan siges at være forskellige (på 5 pct niveau)

Eksempel: Bægerblade på *Iris sp.*

$\lvert X_{i\cdot} - X_{j\cdot}\rvert$	*setosa*	*versicolor*	*virginica*
setosa	0.00	0.44	0.37
versicolor	0.44	0.00	0.07
virginica	0.37	0.07	0.00

Eksempel: Bægerblade på *Iris sp.*

- For det fulde Iris-datasæt med 50 observationer af hver art er der bestemt følgende summer og kvadratsummer

i	Art	n_i	$X_{i\cdot}$	$X_{i\cdot}^2$
1	*setosa*	50	171,4	594,60
2	*versicolor*	50	138,5	388,47
3	*virginica*	50	148,7	447,33

Variation	Kvadratsum SAK	Frihedsgrader DF	Middelsum MS	F-værdi F_{obs}
Mellem grupper	11,34	2	5,67	$49{,}13^{**}$
Indenfor grupper	16,97	147	0,1154	
Total	28,31	149		

Eksempel: Bægerblade på *Iris sp.*

- Ny beregning af LSD giver

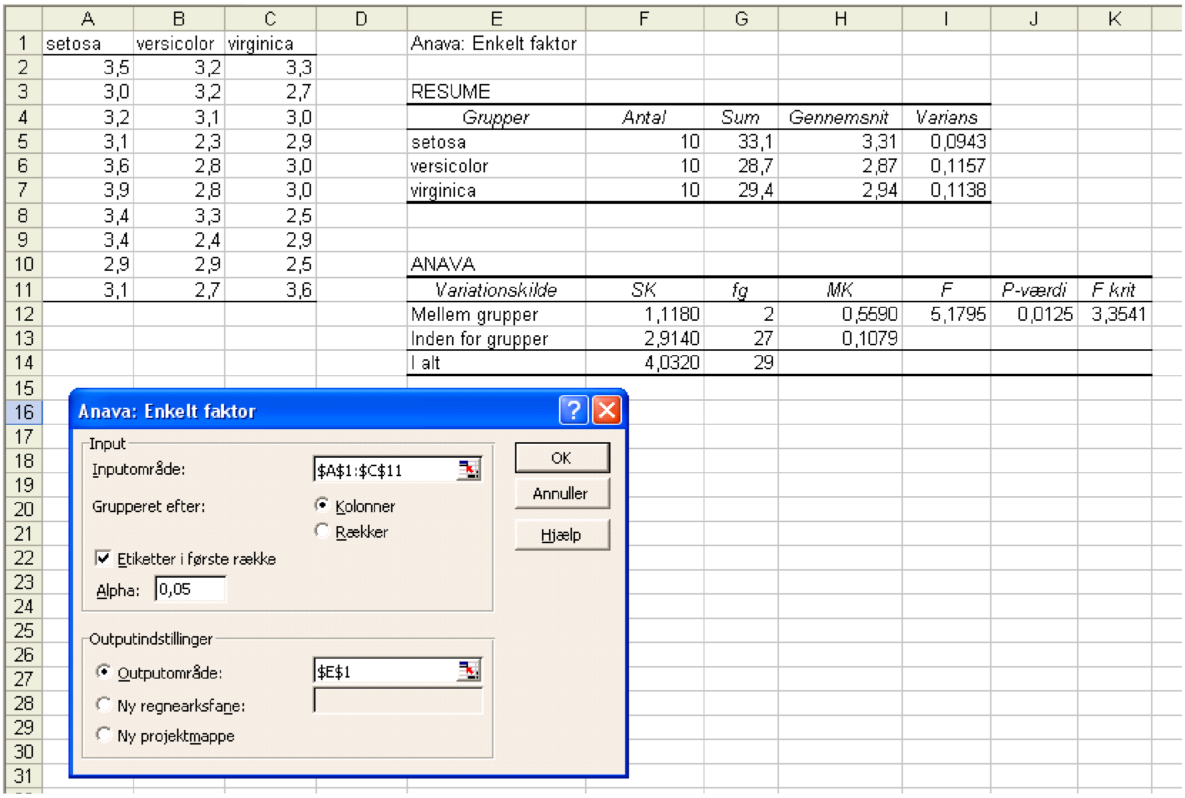
$$LSD_{0.05} = 1.980\sqrt{0.1154(\frac{1}{50} + \frac{1}{50})} = 0.1345$$

- Med 10 observationer i hver gruppe var tallet mere end dobbelt så stort, 0.3014
- Forskellen er

$$\bar{x}_{versicolor} - \bar{x}_{virginica} = 2,7700 - 2,9740 = 0,2040$$

er nu signifikant!

Eksempel: Beregning med Excel



setosa	versicolor	virginica
3,5	3,2	3,3
3,0	3,2	2,7
3,2	3,1	3,0
3,1	2,3	2,9
3,6	2,8	3,0
3,9	2,8	3,0
3,4	3,3	2,5
3,4	2,4	2,9
2,9	2,9	2,5
3,1	2,7	3,6

Anava: Enkelt faktor

RESUME

Grupper	Antal	Sum	Gennemsnit	Varians
setosa	10	33,1	3,31	0,0943
versicolor	10	28,7	2,87	0,1157
virginica	10	29,4	2,94	0,1138

ANAVA

Variationskilde	SK	fg	MK	F	P-værdi	F krit
Mellem grupper	1,1180	2	0,5590	5,1795	0,0125	3,3541
Inden for grupper	2,9140	27	0,1079			
I alt	4,0320	29				

Tosidet variansanalyse uden vekselvirkning: Blokforsøg

- Ensidet ANOVA blev brugt til at analysere forsøg med én faktor, α, men nu udvider nu med endnu en faktor, β
- Hvis vi har én observation for hver kombination af α og β kan vi undersøge hovedvirkningerne af de to faktorer med en model for blokforsøg
- (Hvis vi har gentagelser indenfor kombinationer af α og β kan vi også undersøge for vekselvirkninger — dette kommer senere. . .)
- Man kan tænke på blokforsøg som en udvidelse af den ensidede variansanalyse, hvor gentagelserne er organiseret på tværs efter et eller andet fællestræk. . .

Eksempel: Design af forsøg

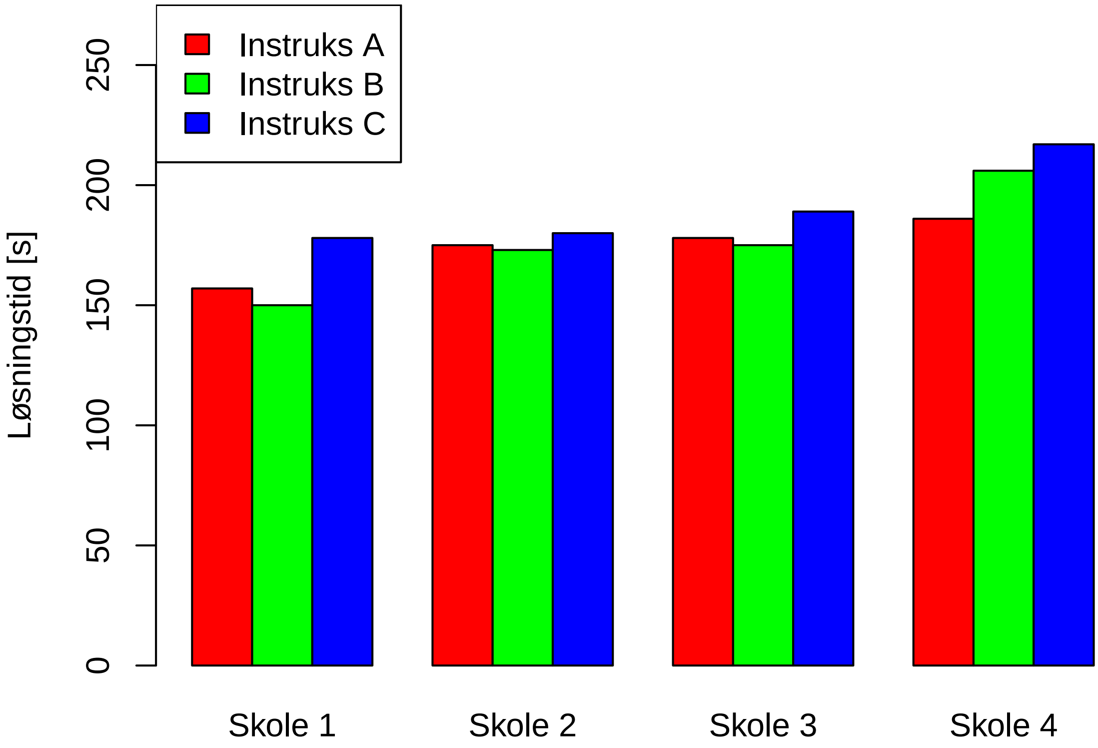
- På 4 skoler, hvor der på hver skole er 3 klasser på 6. klassetrin, har man lavet en samarbejdsprøve
- Den enkelte klasse skal løse den samme opgave på tid, men der er udarbejdet tre forskellige instruktioner til opgaven
- På hver skole fordeles de tre instrukser ved lodtrækning tilfældigt på klasserne
- Man registrerer hvor mange sekunder klassen er om at løse opgaven
- Instruksen er behandlingen og skolen er blokken

Eksempel: Data

Tid [sek]	Skole 1	Skole 2	Skole 3	Skole 4	Snit
Instruks A	157	175	178	186	174,0
Instruks B	150	173	175	206	176,0
Instruks C	178	180	189	217	191,0
Snit	161,7	176,0	180,7	203,0	180,0

- Er der en signifikant behandlingseffekt, dvs. giver instruksen en sikker påvirkning af løsningstiden?
- Er der en signifikant blokeffekt, dvs. er der forskellige niveauer for skolerne?
- Kunne blokeffekten være den interessante?

Eksempel: Illustration af data



Teori: Statistisk model

- Modellen for den ensidede variansanalyse,

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

  udvider vi med en blokeffekt β_j, $j = 1, \ldots, b$ så modellen nu kan skrives som

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

- De tilfældige fejl $\{\varepsilon_{ij}\}$ antages uafhængige og fra en normalfordeling med middelværdi 0 og varians σ^2
- Der gælder at $\sum_i \alpha_i = \sum_j \beta_j = 0$

Teori: ANOVA-tabel

Variation	Frihedsgrader DF	Middelsum MS	F-værdi F_{obs}
Rækker (behandling)	$a-1$	$\text{MS}_a = \dfrac{\text{SAK}_a}{a-1}$	$F_a = \dfrac{\text{MS}_a}{\text{MS}_e}$
Søjler (blokke)	$b-1$	$\text{MS}_b = \dfrac{\text{SAK}_b}{b-1}$	$F_b = \dfrac{\text{MS}_b}{\text{MS}_e}$
Residual	$(a-1)(b-1)$	$\text{MS}_e = \dfrac{\text{SAK}_e}{(a-1)(b-1)}$	
Total	$ab-1$		

- Der gælder at $\text{SAK}_{\text{tot}} = \text{SAK}_a + \text{SAK}_b + \text{SAK}_e$
- Detaljer i separat note...

Eksempel: Beregning af SAKer

i \ j	1	2	3	4	$\sum$
1	157	175	178	186	696
2	150	173	175	206	704
3	178	180	189	217	764
$\sum$	485	528	542	609	2164

$$\sum_{i=1}^{3}\sum_{j=1}^{4} X_{ij}^2 = 393938$$

$$\text{SAK}_{\text{tot}} = 393938 - \frac{1}{3\cdot 4}2164^2 = 3696,67$$

$$\text{SAK}_a = \frac{1}{4}(696^2 + 704^2 + 764^2) - \frac{1}{3\cdot 4}2164^2 = 690,67$$

$$\text{SAK}_b = \frac{1}{3}(485^2 + 528^2 + 542^2 + 609^2) - \frac{1}{3\cdot 4}2164^2 = 2643,33$$

$$\text{SAK}_{\text{res}} = 3696,67 - 690,67 - 2643,33 = 362,67$$

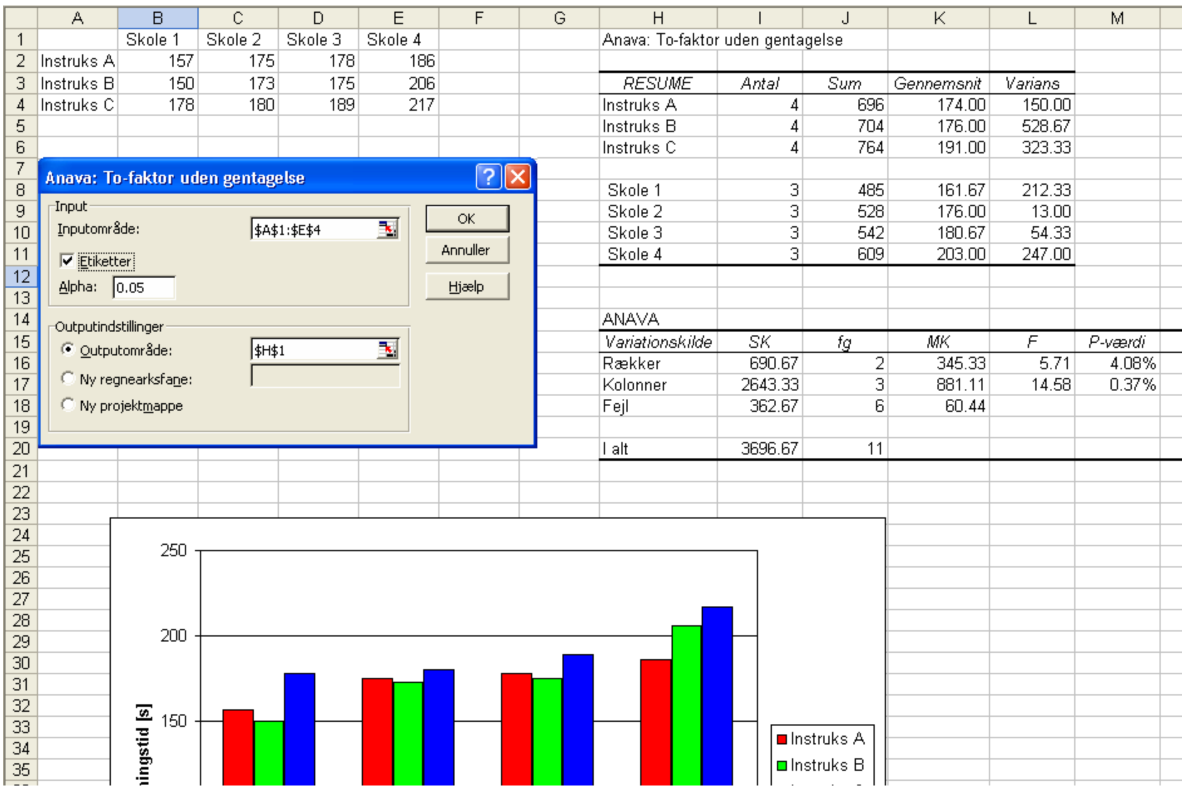
Eksempel: ANOVA-tabel

Variation	SAK	Frihedsgrader	Middelsum	F-værdi
Behandling (instruks)	690,67	2	345,33	5,71*
Blok (skole)	2643,33	3	881,11	14,58**
Residual	362,67	6	60,44	
Total	3696,67	11		

- Kritiske værdier:

α	0,05	0,01	0,001
$F_{2,6}$	5,14	10,92	27,00
$F_{3,6}$	4,76	9,78	23,70

Blokforsøg: Beregning med Excel



Tosidet variansanalyse med vekselvirkning

- Har vi gentagelser indenfor hver kombination af de to faktorer, så kan vi estimere en vekselvirkning
- Vekselvirkning betyder, at effekten af en faktor også afhænger af den anden faktor — modellen er ikke længere rent additiv
- Den statistiske model er nu udvidet til

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

  hvor henholdsvis X_{ijk} er den $k$te gentagelse og γ er vekselvirkningen for kombination ij af de to faktorer
- De tilfældige fejl $\{\varepsilon_{ijk}\}$ antages uafhængige og fra en normalfordeling med middelværdi 0 og varians σ^2
- Der gælder at $\sum \alpha_i = \sum \beta_j = \sum \gamma_{ij} = 0$

Vekselvirkning: ANOVA-tabel

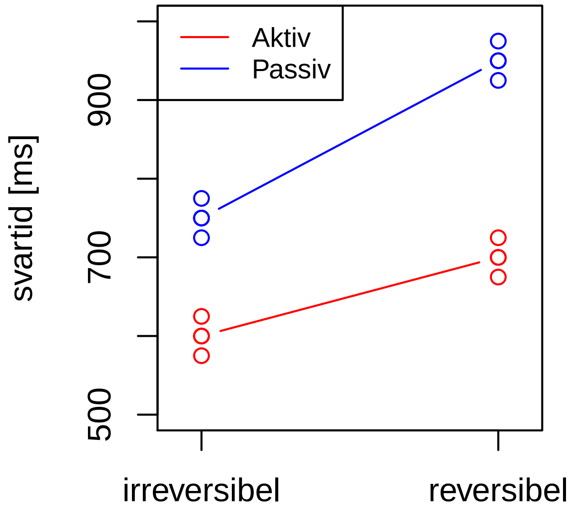
Variation	Frihedsgrader DF	Middelsum MS	F-værdi F_{obs}
Rækker (faktor A)	$a-1$	$\text{MS}_A = \dfrac{\text{SAK}_A}{a-1}$	$F_A = \dfrac{\text{MS}_A}{\text{MS}_e}$
Søjler (faktor B)	$b-1$	$\text{MS}_B = \dfrac{\text{SAK}_B}{b-1}$	$F_B = \dfrac{\text{MS}_B}{\text{MS}_e}$
Vekselvirkning ($A \times B$)	$(a-1)(b-1)$	$\text{MS}_{AB} = \dfrac{\text{SAK}_{AB}}{(a-1)(b-1)}$	$F_{AB} = \dfrac{\text{MS}_{AB}}{\text{MS}_e}$
Residual	$ab(n-1)$	$\text{MS}_e = \dfrac{\text{SAK}_e}{ab(n-1)}$	
Total	$abn-1$		

- Detaljer i separat note...

Vekselvirkning: Eksempel

- Svartider på sætninger der varierer mht reversibilitet og aktiv–passiv

	Aktiv	Passiv
Irreversibel	575	725
	600	750
	625	775
Reversibel	675	925
	700	950
	725	975



- Det ses at kombinationen aktiv–irreversibel giver meget korte svartider, mens kombinationen passiv–reversibel giver meget lange svartider

Vekselvirkning: Beregning

Tabel 12.4 + 12.5

$i \sim$ rækker (1 = irreversibel, 2 = reversibel)
$j \sim$ kolonner (1 = aktiv, 2 = passiv)
$k \sim$ gentagelser ($k = 1, 2, 3$)

$$X_{ij\cdot} = \sum_{k=1}^{3} X_{ijk} \quad ,$$

$i \backslash j$	1	2	$\cdot$
1	1800	2250	4050
2	2100	2850	4950
$\cdot$	3900	5100	9000

$$\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{3} X_{ijk}^2 = 6.950.000$$

$$SAK_a = \frac{1}{2\cdot 3}(4050^2 + 4950^2) - \frac{1}{2\cdot 2\cdot 3} 9000^2 = 67.500$$

$$SAK_b = \frac{1}{2\cdot 3}(3900^2 + 5100^2) - \frac{1}{2\cdot 2\cdot 3} 9000^2 = 120.000$$

$$SAK_e = 6.950.000 - \frac{1}{3}(1800^2 + 2100^2 + 2750^2 + 2850^2) = 5.000$$

$$SAK_{tot} = 6.950.000 - \frac{1}{2\cdot 2\cdot 3} 9000^2 = 200.000$$

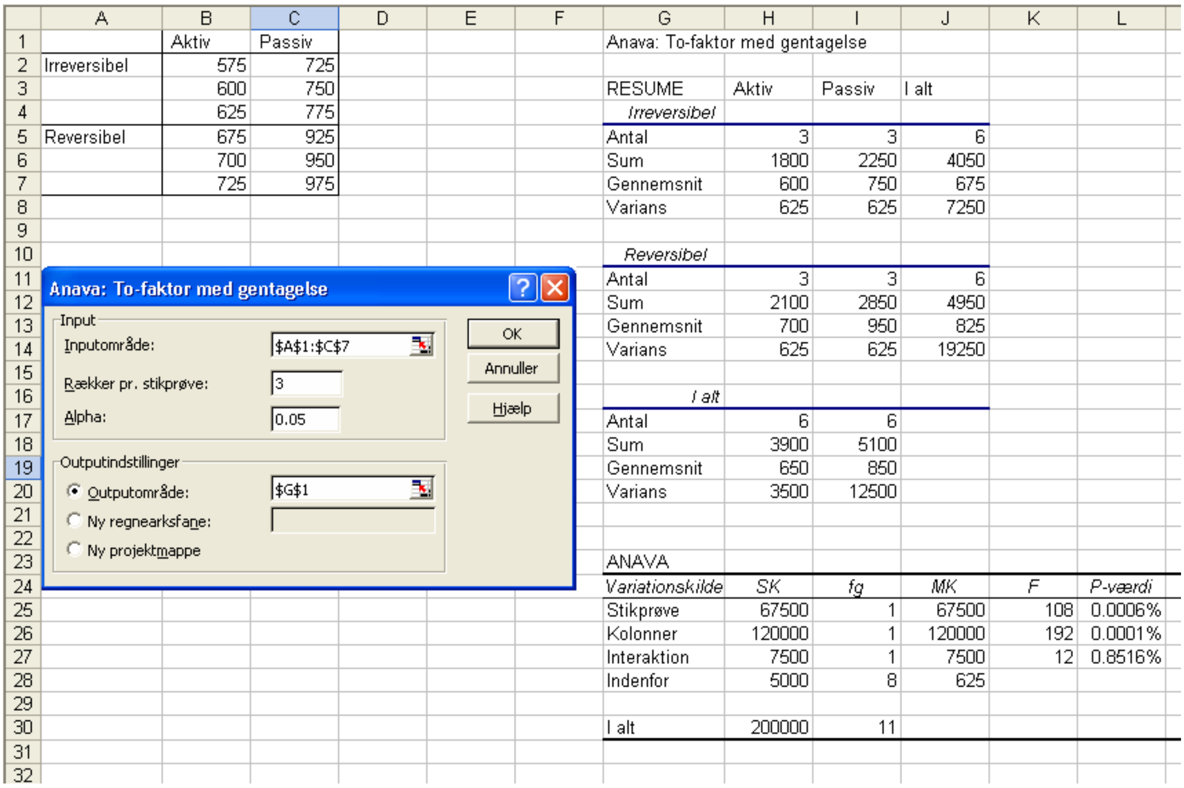
$$SAK_{ab} = 200.000 - 67500 - 120.000 - 5000 = 7500$$

Variation	SAK	DF	MS	F
Rækker	67.500	1	67.500	108 (***)
Søjler	120.000	1	120.000	192 (***)
a × b	7.500	1	7.500	12 **
Rest	5.000	8	625	
Total	200.000	11		

Kritiske værdier:

p	5%	1%	0,1%
$F_{1,8}$	5,32	11,26	25,41

Vekselvirkning: Beregning med Excel



	Aktiv	Passiv
Irreversibel	575	725
	600	750
	625	775
Reversibel	675	925
	700	950
	725	975

Anava: To-faktor med gentagelse

RESUME	Aktiv	Passiv	I alt
Irreversibel			
Antal	3	3	6
Sum	1800	2250	4050
Gennemsnit	600	750	675
Varians	625	625	7250
Reversibel			
Antal	3	3	6
Sum	2100	2850	4950
Gennemsnit	700	950	825
Varians	625	625	19250
I alt			
Antal	6	6	
Sum	3900	5100	
Gennemsnit	650	850	
Varians	3500	12500	

ANAVA

Variationskilde	*SK*	*fg*	*MK*	*F*	*P-værdi*
Stikprøve	67500	1	67500	108	0.0006%
Kolonner	120000	1	120000	192	0.0001%
Interaktion	7500	1	7500	12	0.8516%
Indenfor	5000	8	625		
I alt	200000	11			