

Internt seminar RKBU 21 mars 2014:

Multiple hypoteser eller multiple sammenlikninger

Stian Lydersen

 Revidert 24 mars 2014
<http://folk.ntnu.no/slyderse/medstat/Multcomp21mars2014.pdf>
www.ntnu.no

Om du ønsker, kan du sette inn navn, tittel på foredraget, o.l. her.

2

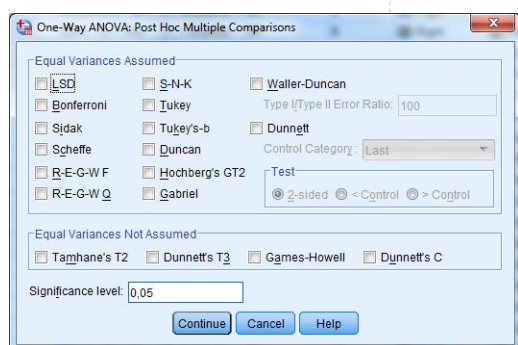
Multiple hypoteser eller multiple sammenlikninger.

Hva bør du gjøre hvis du skal undersøke flere hypoteser samtidig? Dette er tilfelle hvis du har mer enn én avhengig variabel, dvs flere endepunkt, eller hvis mer enn to grupper (feks eksponeringsgrupper eller behandlingsgrupper) skal sammenliknes. Dersom hver av hypotesene testes ved signifikansnivå α , vil sannsynligheten for å gjøre minst en type II feil (FWER, familywise error rate) bli betydelig større enn α . En måte å håndtere dette på, er å bruke en korreksjon (feks Bonferroni) som kontrollerer for at FWER ikke overskrider α . En spesiell situasjon er når man har svært mange hypoteser, feks 1000 hypoteser i en genstudie. Da blir det for konservativt å bevare FWER. Man aksepterer da en viss andel falske positive funn (false discovery rate, FDR) blant de genene som påstås å ha effekt. En annen spesiell situasjon er når man skal sammenlikne bare tre grupper. Her er det faktisk ikke nødvendig å justere (ved Bonferroni e.l.) for å bevare FWER, såfremt man gjør parvise sammenlikninger bare hvis omnibus-hypotesen (alle tre gruppene er like) forkastes.

Jeg vil diskutere følgende spørsmål:
 Når bør vi justere for multiple hypoteser?
 Skal vi kontrollere FWER eller FDR?
 Hvilken metode bør brukes?

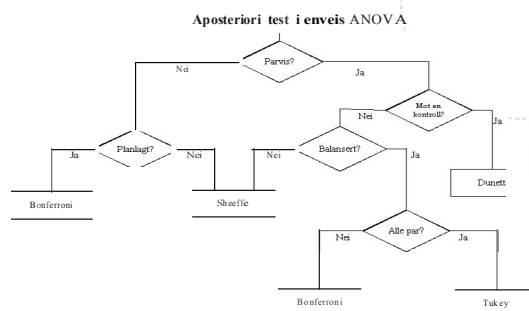
www.ntnu.no

3


www.ntnu.no

4

Notat fra Harald Johnsen 2006:


www.ntnu.no

5

Eksempel
 Kaasbøll, J., Lydersen, S., & Indredavik, M. S. 2012,
 "Psychological symptoms in children of parents with chronic pain-the HUNT study", Pain, vol. 153, no. 5, pp. 1054-1062.

4 grupper ungdommer: Foreldre med kroniske smerter:

- Ingen,
- bare mor (M),
- bare far (F),
- både mor og far (MF)

2 avhengige variable:

- Angst/depresjon (SCL-5)
- Atferdsproblemer

Separate analyser for jenter og gutter

www.ntnu.no

6

Kaasbøll J, Lydersen S, Indredavik M: (Pain, 2012)
 "Psychological symptoms in children of parents with chronic pain – the HUNT study"

Results adjusted for age – boys:

Parents with chronic pain	Number of children	Risk for conduct problems: Odds ratio (OR)		
		estimate	Conf. int.	P-value
None	801	1 (ref.)		
Only mother	289	1.30	1.02 to 1.67	0.036
Only father	216	0.99	0.74 to 1.31	0.93
Both parents	117	1.36	0.96 to 1.93	0.087

www.ntnu.no

Tre(?) problemstillinger:

- Flere avhengige variable
- Sammenlikning mellom flere enn 2 grupper
- Flere subgrupper

Flere avhengige variable
("Uavhengige" eller "urelaterte" hypoteser)

- Rothman, K. J. 1990, "No adjustments are needed for multiple comparisons", *Epidemiology*, vol. 1, no. 1, pp. 43-46.
 - Justere når flere hypoteser i samme artikkel?
 - Splitte opp i flere artikler?
 - Justere for alle hypoteser jeg tester i min karriere?
 - Justere for alle hypoteser menneskeheten tester?
- Ett primært endepunkt, legge "mindre vekt" på signifikante funne på sekundære endepunkt
- Pragmatisk forslag (Peter Fayers): Two-sided p-values < 0.05 are taken to indicate statistical significance. Due to multiple hypotheses, p-values between 0.01 and 0.05 should be interpreted with caution.

Eksempel:
Senn, S. 2007, *Statistical issues in drug development*, 2nd ed, John Wiley. Chapter 10: Multiplicity
RCT, medikamenter for lungefunksjon hos astmatikere:

Mulige avhengige variable:

- FEV
- PEV
- RAW
- ...
- Bivirkninger

"Multiplicity continues to be a field of much research and controversy."

	Akseptert	Forkastet	Totalt
Sanne nullhypoteser	U	V	m_0
Usanne nullhypoteser	T	S	$m - m_0$
	$m - R$	R	m

Per comparison error rate PCER: $E[V / m]$

Hvis hver nullhypotese testes med nivå α :

Hvis alle nullhypotesene er sanne: $E[V / m] = \alpha$

Hvis noen nullhypoteser er usanne: $E[V / m] < \alpha$

Family-wise error rate FWER: $P(V \geq 1)$

(Also called experimentwise error rate)

False discovery rate FDR: $E[R / m]$

er forventet andel sanne nullhypoteser blant de som forkastes.

Weak control of FWER:
FWER $\leq \alpha$ given that all null hypotheses are true
($m_0 = m$)

Strong control of FWER:
FWER $\leq \alpha$ for any combination of true and false null hypotheses

Strong control of FWER is mandated by regulators in the USA in all confirmatory clinical trials (Phase III) trials.

In the rest of this presentation, control of FWER means in the strong sense.

Problem:

Ved å bruke signifikansnivå α på hver av hypotesetestene, blir sannsynligheten for minst ett falskt positivt funn, dvs FWER, (betydelig) større enn α .

Alternative løsninger:

- Bare en hypotese per studie (ett primært endepunkt)
- Kombinere de avhengige variablene til ett globalt endepunkt (for eksempel summere delskårene til en totalskåre i et spørreskjema)
- Bare et signifikant resultat for et forhåndsdefinert primært endepunkt tillater analyse av sekundære endepunkt ("gate-keeper")
- Når flere grupper (for eksempel behandlinger) sammenliknes, bruk en global test først.
- Korreksjon for multiple tester
- Kontroller FDR, ikke FWER

Control of FDR is relevant with a large number of null hypotheses, such as in genomics.

Goeman, J.J. & Solari, A. 2014. Tutorial in biostatistics: Multiple hypothesis testing in genomics. Statistics in Medicine

Distributional assumptions:

- P-value based procedures: Based on univariate p-values. No assumptions about joint distribution of test statistics. F.ex. Bonferroni, Holm
- Semiparametric p-value based procedures. F.ex. Šidák, Hochberg, Hommel assuming independent or positively correlated p-values.
- Parametric procedures assuming a specific multivariate distribution, f.ex. Dunnett, Dunn, Tukey, Scheffé in one-way ANOVA (without additional covariates) assuming normal distribution.
- Resampling-based procedures (bootstrap or permutation methods). Control FWER only when sample size approaches infinity (Dmitrienko and D'Agostino, 2013, page 5181)

P-value based procedures tend to perform poorly compared to parametric and resampling based procedures, when there are many hypotheses or strongly correlated test statistics. (Dmitrienko et al, 2010, page 49 - 50)

m hypoteser, p-verdier $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m)}$.

Bonferronikorreksjon:

Forkast bare hypoteser med $mp_{(j)} \leq \alpha$ (ekvivalent: $p_{(j)} \leq \tilde{\alpha} = \alpha / m$)

Kontrollerer alltid $FWER \leq \alpha$

Bonferroniulikheten gjelder alltid:

$$P(F_1 \cup F_2 \cup \dots \cup F_m) \leq P(F_1) + P(F_2) + \dots + P(F_m) = m\tilde{\alpha}$$

Šidák korreksjon:

Forkast bare hypoteser med $p_{(j)} \leq \tilde{\alpha} = 1 - (1 - \alpha)^{1/m}$

Kontrollerer $FWER \leq \alpha$ ved uavhengige p-verdier, eller uavhengige testobservatorer, eller multinormalfordelte testobservatorer.

Šidák-ulikheten gjelder med = for uavhengige og < for positivt korrelerte p-verdier:

$$P(F_1 \cup F_2 \cup \dots \cup F_m)$$

$$\leq 1 - [1 - P(F_1)][1 - P(F_2)] \dots [1 - P(F_m)] = 1 - (1 - \tilde{\alpha})^m$$

Eksempler:

	α	m	$\tilde{\alpha}$
Bonferroni	0.05	3	0.01667
Šidák	0.05	3	0.01695
Bonferroni	0.05	10	0.00500
Šidák	0.05	10	0.00512

«The Šidák procedure provides a fairly small improvement over the Bonferroni procedure ... For this reason, the Šidák procedure and related multiple testing procedures have not found many applications in a clinical setting ...» (Dmitrienko and D'Agostino 2013, page 5188)

Kirk (2013, Tabell 5.1-1 side 169) anbefaler Šidák-korreksjon fremfor Bonferronikorreksjon for Dunn's test.

19

Holm step-down correction (også kalt Bonferroni step-down correction):
 Hvis $p_{(1)} \leq \alpha_{(1)} = \alpha / m$ forkast hypotese (1)
 Hvis hypotese (1) ble forkastet:
 Hvis $p_{(2)} \leq \alpha_{(2)} = \alpha / (m-1)$ forkast hypotese (2)
 Hvis hypotese (2) ble forkastet:
 Hvis $p_{(3)} \leq \alpha_{(3)} = \alpha / (m-2)$ forkast hypotese (3)
 ...
 Hvis hypotese (m-1) ble forkastet:
 Hvis $p_{(m)} \leq \alpha_{(m)} = \alpha / 1 = \alpha$ forkast hypotese (m)

Dvs:
 I steg j , bruk Bonferronikorreksjon på de $m - j + 1$ siste p-verdiene.

Kontrollerer alltid $FWER \leq \alpha$.
 Uniformt mindre konservativ (høyere styrke) enn Bonferroni

NTNU
Det skapende universitet

www.ntnu.no

20

Hochberg step-up correction

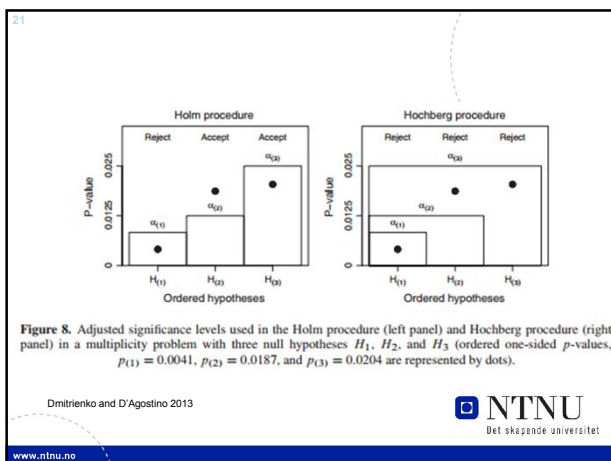
Hvis $p_{(m)} \leq \alpha_{(m)} = \alpha$ forkast alle m hypotesene.
 Hvis $p_{(m-1)} \leq \alpha_{(m-1)} = \alpha / 2$ forkast hypotese (m-1), (m-2), ..., 1
 Hvis $p_{(m-2)} \leq \alpha_{(m-2)} = \alpha / 3$ forkast hypotese (m-2), ..., 1
 ...
 Hvis $p_{(1)} \leq \alpha_{(1)} = \alpha / m$ forkast hypotese (1)

Kontrollerer også $FWER \leq \alpha$ under forholdsvis generelle betingelser (multivariat normalfordelte testobservatorer med ikke-negative korrelasjonskoeffisienter)
 Mindre konservativ (høyere styrke) enn Holm (og Bonferroni)

NB! Hochberg som kontrollerer FWER må ikke forveksles med Benjamini-Hochberg som bare kontrollerer FDR!

NTNU
Det skapende universitet

www.ntnu.no



22

Hommel's (1988) procedure is more powerful than Hochberg's but is more difficult to understand and apply. Let j be the largest integer for which $p_{(m-j+k)} < k\alpha/j$ for all $k = 1, \dots, j$. If no such j exists, reject all hypotheses; otherwise, reject all H_i with $p_i \leq \alpha/j$. Both i and j , by the way, go from 1 to m .

(Shaffer, 1995, page 571, adapted to our notation with m instead of n)

Hommel's procedure is explained in a slightly different way in the Appendix of Dmitrienko and d'Agostino (2013)

NTNU
Det skapende universitet

www.ntnu.no

23

Benjamini-Hochberg

Hvis $p_{(m)} \leq \alpha_{(m)} = m\alpha / m = \alpha$ forkast alle m hypotesene.
 Hvis $p_{(m-1)} \leq \alpha_{(m-1)} = (m-1)\alpha / m$ forkast hypotese (m-1), (m-2), ..., 1
 Hvis $p_{(m-2)} \leq \alpha_{(m-2)} = (m-2)\alpha / m$ forkast hypotese (m-2), ..., 1
 ...
 Hvis $p_{(1)} \leq \alpha_{(1)} = \alpha / m$ forkast hypotese (1)

Kontrollerer ikke $FWER \leq \alpha$
 "The Benjamini & Hochberg procedure has valid FDR control if the test statistics underlying the p-values are positively correlated for one-sided tests, under a different condition that allows some negative correlations for two-sided tests." (Goeman and Solari 2014)

NTNU
Det skapende universitet

www.ntnu.no

24

m hypoteser, p-verdier $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m)}$.

p-value unadjusted	Bonferroni	Holm step-down	Hochberg step-up	Benjamini-Hochberg step-up
$p_{(1)}$	$mp_{(1)}$	$mp_{(1)}$	$\min\left(mp_{(1)}, \dots, \frac{mp_{(m)}}{m}\right)$	$\min\left(\frac{mp_{(1)}}{1}, \dots, \frac{mp_{(1)}}{m}\right)$
$p_{(2)}$	$mp_{(2)}$	$\max\left(mp_{(1)}, (m-1)p_{(2)}\right)$	$\min\left((m-1)p_{(2)}, \dots, \frac{mp_{(m)}}{m}\right)$	$\min\left(\frac{mp_{(2)}}{2}, \dots, \frac{mp_{(1)}}{m}\right)$
$p_{(m-1)}$	$mp_{(m-1)}$	$\max\left(mp_{(1)}, \dots, 2p_{(m-1)}\right)$	$\min\left(2p_{(m-1)}, \dots, \frac{mp_{(m)}}{m}\right)$	$\min\left(\frac{mp_{(m-1)}}{m-1}, \dots, \frac{mp_{(1)}}{m}\right)$
$p_{(m)}$	$mp_{(m)}$	$\max\left(mp_{(1)}, \dots, p_{(m)}\right)$	$\frac{mp_{(m)}}{m}$	$\frac{mp_{(m)}}{m}$

NTNU
Det skapende universitet

www.ntnu.no

www.ntnu.no

www.ntnu.no

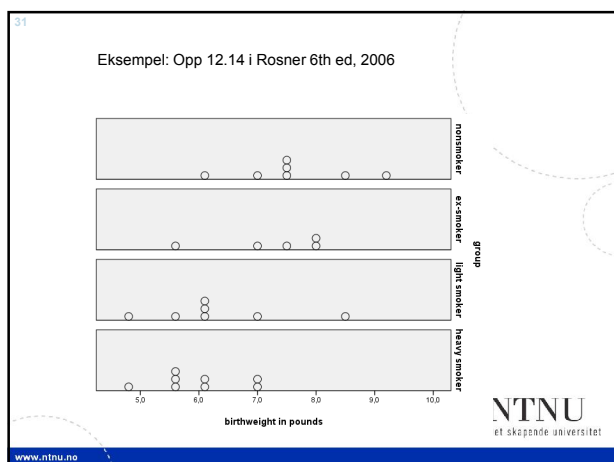


28

www.ntnu.no

www.ntnu.no

www.ntnu.no



32

Descriptives

birthweight in pounds

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
nonsmoker	7	7,586	,9616	,3635	6,896	8,475
ex-smoker	5	7,240	,9127	,4082	6,107	8,373
light smoker	7	6,329	1,1398	,4308	5,274	7,383
heavy smoker	8	6,013	,7200	,2546	5,411	6,614
Total	27	6,730	1,1090	,2134	6,291	7,168

ANOVA

birthweight in pounds

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	11,673	3	3,891	4,408	,014
Within Groups	20,304	23	,883		
Total	31,976	26			

www.ntnu.no

33

SPSS Compare means eller GLM

One-Way ANOVA: Post Hoc Multiple Comparisons

Equal Variances Assumed

☒ LSD ☐ S-N-K ☐ Waller-Duncan

☐ Bonferroni ☐ Tukey Type I/Type II Error Ratio: 100

☐ Sidak ☐ Tukey's-b ☐ Dunnett

☐ Scheffe ☐ Duncan Control Category: Last

☐ R-E-G-W F ☐ Hochberg's GT2

☐ R-E-G-W Q ☐ Gabriel

Equal Variances Not Assumed

☐ Tamhane's T2 ☐ Dunnett's T3 ☐ Games-Howell ☐ Dunnett's C

Significance level: 0,05

Continue Cancel Help

Hochberg's GT2 er ikke Hochberg step-up!

NTNU
Det skapende universitet

www.ntnu.no

34

SPSS GLM med minst en kovariat

Univariate: Options

Estimated Marginal Means

Factor(s) and Factor Interactions: Diagnose1

Display Means for: Diagnose1

☒ Compare main effects

Confidence interval adjustment: LSD(none)

☐ LSD(none) ☐ Bonferroni ☐ Sidak

Display

☐ Descriptive statistics ☐ Spread vs. level plot

☐ Estimates of effect size ☐ Residual plot

☐ Observed power ☐ Lack of fit

☐ Parameter estimates ☐ General estimable function

☐ Contrast coefficient matrix

Significance level: 05 Confidence intervals are 95,0%

Continue Cancel Help

NTNU
Det skapende universitet

www.ntnu.no

35

SPSS Complex samples GLM

Complex Samples General Linear Model: Hypothesis Tests

Test Statistic

☒ F ☐ Adjusted F ☐ Chi-square ☐ Adjusted Chi-square

Sampling Degrees of Freedom

☒ Based on sample design ☐ Fixed Value:

Adjustment for Multiple Comparisons

☒ Least significant difference ☐ Sequential Sidak ☐ Sequential Bonferroni ☐ Sidak ☐ Bonferroni

«Sequential» Betyr Holm step-down anvendt på hhv Sidak og Bonferroni

Continue Cancel Help

NTNU
Det skapende universitet

www.ntnu.no

36

Eksempel: 4 grupper.

Global hypotese:

$$\mu_1 = \mu_2 = \mu_3 = \mu_4$$

Totalt $\binom{4}{2} = \frac{4!}{(4-2)!2!} = 6$ parvise sammenligninger:

$$\mu_1 = \mu_2$$

$$\mu_1 = \mu_3$$

$$\mu_1 = \mu_4$$

$$\mu_2 = \mu_3$$

$$\mu_2 = \mu_4$$

$$\mu_3 = \mu_4$$

NTNU
Det skapende universitet

www.ntnu.no

37

Bare teste mot en kontroll:

$$\mu_1 = \mu_2$$

$$\mu_1 = \mu_3$$

$$\mu_1 = \mu_4$$

NTNU
Det skapende universitet

www.ntnu.no

38

Lineære kontraster :

$$\sum_{i=1}^k c_i \mu_i = 0 \quad \text{hvor} \quad \sum_{i=1}^k c_i = 0$$

For eksempel:

$$\mu_1 - \mu_2 = 0$$

$$\mu_3 - \mu_4 = 0$$

$$\frac{1}{2}(\mu_1 + \mu_2) - \frac{1}{2}(\mu_3 + \mu_4) = 0$$

NTNU
Det skapende universitet

www.ntnu.no

39

Lineære kontraster er ortogonale hvis

$$\sum_{j=1}^k c_{ij} c_{i'j} = 0$$

De tre lineære kontrastene i eksempelet ovenfor er ortogonale. Parvise kontraster som inneholder samme forventningsverdi, som for eksempel alle parvise kontraster eller teste mot en kontroll, er ikke ortogonale.

NTNU
Det skapende universitet

www.ntnu.no

40

De fleste prosedyrer for multiple sammenlikninger bruker testobservatorer (test statistics) av typen

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{MS_{error} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

som vanligvis er fordelt som Student's t eller Studentized range.

NTNU
Det skapende universitet

www.ntnu.no

41

Kirk (2013) «Experimental design. Procedures for the behavioural sciences» 4th ed, Chapter 5 «Multiple Comparison tests» (side 154-208). (Første utgave kom i 1969)

Fra omtalen på bokomslaget:
«Up-to-date coverage of multiple comparison procedures».

Table 5.1_1:
http://folk.ntnu.no/slyderse/medstat/Kirk_table5.1_1.pdf

NTNU
Det skapende universitet

www.ntnu.no

42

Anbefalte prosedyrer ved p (≥4) grupper, enveis ANOVA:

Test hva?	Likt antall (balansert) eller lik varians i gruppene?	Anbefalt test
p-1 grupper mot en kontrollgruppe	Balansert og lik varians	Dunnett
	Ubalansert eller ulik varians	Dunnett med modifikasjoner
C planlagte kontraster	Balansert og lik varians	Dunn(-Bonferroni) Dunn- Sidak (Dunn-) Holm Dunn-Hochberg?
	Ulik varians	Som over med Welch frihetsgrader
Alle parvise kontraster	Balansert og lik varians	Tukey Fisher-Hayter REGW [*] F, FQ og Q
	Ubalansert	Tukey-Kramer Fisher-Hayter
	Ulik varians	Brown-Forsythe
Planlagte og uplanlagte kontraster	Balansert eller ubalansert	Scheffé
	Ulik varians	Dunnett's T3 Dunnett's C Games-Howell

I stor grad basert på Kirk (2013, Table 5.1-1).
Der flere alternativer er listet, er det i økende statistisk styrke
* REGW står for Ryan, Einot, Gabriel, Welch

NTNU
Det skapende universitet

www.ntnu.no

43

Notat fra Harald Johnsen 2006:

Aposteriori test i enveis ANOVA

Dette er fremdeles enkle og gode råd. Evt bruke Holm eller Hochberg istedenfor Bonferroni.

NTNU
Det skapende universitet

www.ntnu.no

44

Sammenlikne grupper justert for minst en kovariat (for eksempel alder eller kjønn). («ANCOVA»)

Ved 4 eller flere grupper:
Bonferroni < Holm < Hochberg < Hommel
Šidák er såvidt bedre enn Bonferroni

Ved 3 grupper:
Kombiner global test med lokal test ved parvise sammenlikninger

NTNU
Det skapende universitet

www.ntnu.no

45

Closed testing procedure for m hypotheses $H_1, H_2, \dots, H_m$:

Define the closed family of hypotheses. For each non-empty set of indices $I \subseteq \{1, 2, \dots, m\}$, consider the intersection hypothesis $\bigcap_{i \in I} H_i$.

Define local (that is, unadjusted) α -level tests for the intersection hypotheses.

Reject H_i at level α if and only if all intersection hypotheses containing H_i are rejected by local α -level tests.

(Marcus et al. 1976) showed that this closed testing procedure for the hypotheses $H_1, H_2, \dots, H_m$ controls the FWER in the strong sense at the α -level.

NTNU
Det skapende universitet

www.ntnu.no

46

Spesialtilfelle:
Parvise sammenlikninger av 3 parametre, feks forventningsverdi i 3 grupper:

$$H_1: \mu_2 = \mu_3, H_2: \mu_1 = \mu_3, H_3: \mu_1 = \mu_2$$

Global hypotese:
 $H_{Global}: \mu_1 = \mu_2 = \mu_3$

NTNU
Det skapende universitet

www.ntnu.no

47

I følge «The closed testing procedure» forkaster vi H_1 på nivå α hvis vi forkaster alle disse på lokalt nivå α :

$$H_1$$

$$H_1 \cap H_2$$

$$H_1 \cap H_3$$

$$H_1 \cap H_2 \cap H_3 (= H_{Global})$$

Men

$$H_1 \cap H_2 = \{(\mu_2 = \mu_3) \cap (\mu_1 = \mu_3)\} = \{\mu_1 = \mu_2 = \mu_3\} = H_{Global}$$

og

$$H_1 \cap H_3 = \{(\mu_2 = \mu_3) \cap (\mu_1 = \mu_2)\} = \{\mu_1 = \mu_2 = \mu_3\} = H_{Global}$$

så vi forkaster H_1 på nivå α hvis vi forkaster H_1 på lokalt nivå α og forkaster H_{Global} på nivå α .

NTNU
Det skapende universitet

www.ntnu.no

48

Hvis vi forkaster $\mu_i = \mu_j$ bare når både lokal p-verdi og p-verdi for global test er under α så bevarer vi FWER! Dvs vi utfører testen for $\mu_i = \mu_j$ (ujustert) bare hvis vi forkaster H_{Global} på nivå α .

Ekvivalent:
 p -verdien for den parvise testen settes lik den største av den lokale og globale p -verdien

Dette er lite kjent. Feks Kirk (2013, side 164 – 165) påpeker dette for enveis ANOVA. Men det gjelder generelt.

NB!
Dette gjelder bare for 3 grupper. Det gjelder ikke for eksempel for 4 grupper, selv når du bare har 3 hypoteser (som 3 alternativer mot en kontroll).

NTNU
Det skapende universitet

www.ntnu.no

49

Eksempel fra Weider et al (2014), Tabell 3:

Descriptives

Skåre på WAIS Informasjon- Skalert skåre

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min
					Lower Bound	Upper Bound	
AN	41	10,51	3,264	,510	9,48	11,54	
BN	40	10,00	2,418	,382	9,23	10,77	
Kontroll	40	11,85	2,833	,448	10,94	12,76	
Total	121	10,79	2,945	,268	10,26	11,32	

ANOVA

Skåre på WAIS Informasjon- Skalert skåre

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	73,069	2	36,535	4,457	,014
Within Groups	967,344	118	8,198		
Total	1040,413	120			

NTNU
Det skapende universitet

www.ntnu.no

50

Global p = 0.0136

Par	p-verdi				
	Šidák	Tukey HSD	Dunnett	LSD (ujustert)	Max (ujustert, global)
AN vs BN	0.807	0.701	-	0.422	0.422
AN vs HC	0.109	0.094	0.069	0.038	0.038
BN vs HC	0.0137	0.013	0.009	0.005	0.0136

NTNU
Det skapende universitet

www.ntnu.no

51

Testing, P-verdier og konfidensintervall:

- Metodene tar utgangspunkt i testing (aksepter eller forkast ved gitt signifikansnivå)
- Forholdsvis rett frem å beregne p-verdier (minste signifikansnivå som ville medført forkastning)
- Konfidensintervall kan beregnes for noen metoder, men ikke for flerstegs metoder.

NTNU
Det skapende universitet

www.ntnu.no

52

References

Dmitrienko, A., D'Agostino, R.B., & Huque, M.F. 2013. Key multiplicity issues in clinical drug development. *Statistics in Medicine*, 32, (7) 1079-1111.

Dmitrienko, A. & D'Agostino, R. 2013. Tutorial in Biostatistics: Traditional multiplicity adjustment methods in clinical trials. *Statistics in Medicine*, 32, 5172-5218

Dmitrienko, A., Tamhane, A.C., & Bretz, F. 2010. Multiple testing problems in pharmaceutical statistics Boca Raton, FL, Chapman & Hall/CRC.

Goeman, J.J. & Solari, A. 2014. Tutorial in biostatistics: Multiple hypothesis testing in genomics. *Statistics in Medicine*

Kaasbol, J., Lydersen, S., & Indredavik, M.S. 2012. Psychological symptoms in children of parents with chronic pain-the HUNT study. *Pain*, 153, (5) 1054-1062

Kirk, R.E. 2013. Experimental design. Procedures for the behavioral sciences, 4th ed. Thousand Oaks, Sage Publications.

NTNU
Det skapende universitet

www.ntnu.no

53

References (continued)

Marcus, R., Peritz, E., & Gabriel, K.R. 1976. Closed Testing Procedures with Special Reference to Ordered Analysis of Variance. *Biometrika*, 63, (3) 655-660

Rosner, B. 2006. Fundamentals of biostatistics, 6th ed ed. Belmont, CA, Thomson-Brooks/Cole.

Rothman, K.J. 1990. No adjustments are needed for multiple comparisons. *Epidemiology*, 1, (1) 43-46

Senn, S. 2007. Statistical issues in drug development, 2nd ed. Chichester, England, John Wiley & Sons.

Shaffer, J.P. 1995. Multiple Hypothesis Testing. *Annual Reviews Psychology*, 46, 561-584

Weider, S., Indredavik, M.S., Lydersen, S., & Hestad, K. 2014. Intellectual Function in Patients with Anorexia Nervosa and Bulimia Nervosa. *Eur.Eat Disord.Rev.*, 22, (1) 15-22

NTNU
Det skapende universitet

www.ntnu.no