



# Statistiske beregninger på enkeltforsøg

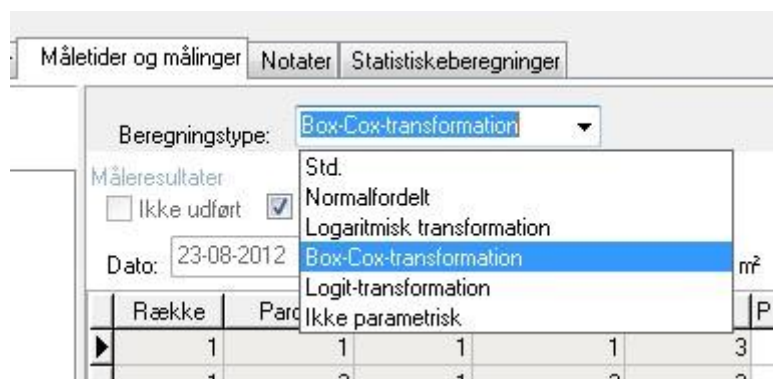
Hanne Justesen Bach<sup>a</sup>

<sup>a</sup> Teknologisk Institut

Statistiske beregninger i NFTS kan regnes på alle måleparametre, der måles på parcelniveau. Beregningsprocedurerne håndteres via PC-Markforsøg af en administrator i samarbejde med fagkonsulenten fra SEGES eller forsøgsejeren og inkluderer statistik for normalfordelte, lognormal-, logit-normal- og boxcox-normal-fordelte data, samt non-parametriske rank tests. Output fra de statistiske analyser er beregnede middelværdier (LSmeans), 95 pct.-konfidensintervaller, signifikansgrupper jf. en bogstavstest, og - såfremt data er normalfordelte – en LSD værdi. Modelantagelser tjekkes visuelt vha. såkaldte residual- og qq-plots. Dokumentet her beskriver, hvordan de statistiske resultater fremkommer jf. beregningsproceduren og hvordan resultater tolkes og modelantagelser tjekkes.

## Indstilling af beregningstype

Der er pt. mulighed for at bruge 6 forskellige beregningstyper i Danmark (opsætning kun muligt for administratorer).



## Statistiske beregninger – hvordan?

Beregninger sættes op og redigeres i PC-markforsøg af administratorer. Det vil sige at for Landsforsøgene er det fagkonsulenterne fra SEGES og forsøgsplanlæggerne i Teknologisk Institut, der i fællesskab opsætter, tester og validerer beregningsopsætning. Siden sommeren 2014 opsættes der automatisk en normalfordelt beregning af alle parcelludbytter i danske forsøg. Forsøgslederne kan selv regne statistik på "egne planer".

# Hvad betyder de forskellige indstillinger for beregnings-type?

## Std.

Indstillingen "Std." betyder at der ikke beregnes statistik for alle måleparametre (undtagen PARCEL-UDBYTTE). Ingen ekstra beregning er nødvendigt og enkeltforsøgsresultatudskrift er som hidtil, og angiver kun måleparametrenes gennemsnitsværdier eller den normale beregning af udbytter som tidligere. F.eks.:

<b>C02: 21 dage efter sidste behandling</b>
<b>03-08-2012 ST. 81</b>
<b>CIKADER, NYMFER</b>
<b>/10 blade</b>
64
28
14
14
1

## Normalfordelt

Normalfordelt" betyder, at der beregnes statistik (dvs. konfidensintervaller og bogstavstest, samt LSD værdi på den utransformerede måleparameter) og der antages normalfordelte residualer og varians-homogenitet. Enkeltforsøgsresultatudskrift viser for alle led LSmeans værdier og konfidensintervaller, samt en LSD værdi. Bogstavstestens signifikansgrupper vises også såfremt  $p < 0.05$ . Se nedenfor for en nærmere beskrivelse af LSmeans, konfidensintervaller, bogstavstest, p-værdier og LSD.

## Logaritmisk transformation, Box-Cox-transformation og Logit-transformation

For disse tre indstillinger gælder følgende:

1. Først foretages den valgte transformation af måleparameterens værdier,
2. derefter beregnes statistik på den transformerede skala under antagelsen om normalfordelte residualer og varianshomogenitet, og
3. til sidst bliver resultaterne tilbagetransformeret og de tilbagetransformerede resultater afrapporteret.

For at denne procedure er meningsfuld, skal transformationen resultere i modelresidualer, der er normalfordelte med en rimelig approksimation. *Dette tjekkes af administratorer før resultater offentliggøres.* Se nedenfor for en beskrivelse hvordan modelantagelserne om normalfordelte residualer kan tjekkes.

Ved disse tre indstillinger beregnes der statistik (dvs. konfidensintervaller og bogstavstest, men ingen LSD). Enkeltforsøgsresultatudskrift viser konfidensintervaller for middelestimerne og såfremt  $p < 0.05$  også bogstavstestens signifikansgrupper. Behandlinger med forskelligt bogstav er signifikant forskellige. F.eks.:

P05:				
23-08-2012 ST. 91				
	PLANTEFARVE karakter 0-10	Nedre konf.	Øvre konf. int.	Signifikansgruppe
1	3	2,9	3,6	d
2	4	4,0	5,0	c
3	5	4,5	5,5	bc
4	5	4,9	6,1	ab
5	6	5,4	6,6	a

## Ikke-parametrisk

Beregningsindstillingen "Ikke parametrisk" betyder, at der beregnes statistik på rangordensværdien efter at måleparameterens værdier er blevet rangordnet efter deres størrelse (Kruskal-Wallis test). Testresultater afrapporteres kun i internt dokument, se nedenfor, og fremgår IKKE af enkeltforsøgsresultatsiden.

## LSmeans

LSmeans står for Least Squares means og er de fra modellen beregnede middelværdier for hvert led. LSmeans er valide estimater såfremt modelantagelserne er opfyldte.

## Konfidensintervaller

Et konfidensinterval betegner et interval beliggende mellem en nedre og en øvre værdi (Nedre konf. og Øvre konf. int.), som forventes at indeholde den sande parameterværdi. Det kunne f.eks. være middeludbyttet for en sort. Alle konfidensintervaller i NFTS er 95%-konfidensintervaller. Et 95% konfidensinterval angiver med 95% sandsynlighed det interval, som inkluderer den sande parameterværdi. Eller med andre ord, hvis man gentog eksperimentet mange gang, så ville de beregnede 95%-konfidensintervaller indeholde den sande parameterværdi i 95% af tilfældene.

## LSD-værdier

- LSD-værdier eller Fisher's Least Significant Difference værdier udtrykker, hvor stor en forskel, der skal være imellem 2 behandlinger, for at de er signifikant forskellige ved et signifikansniveau på 0,05. Hvis en udbytteforskel mellem to forsøgsled er lig med eller større end LSD-værdien, er forskellen statistisk sikker.  
LSD-værdier er meget nyttige og letforståelige tal, men man skal være varsom med at tolke dem for skarpt, idet der ikke tages højde for multiple testing problematikken, se nedenfor.
- LSD2: LSD-værdien for faktor 2, altså en statistisk sammenligning af behandlingerne i faktor 2. Man skal være forsigtig med at sammenligne de enkelte ledværdier imellem behandlingerne i faktor 2, da der kan være en vekselvirkning imellem faktor 1 og faktor 2. Det vil være mere korrekt at beregne gennemsnittet af behandlingerne i faktor 2 og sammenligne dem v.h.j.a. LSD2.
- LSD12: LSD-værdien for kombinationer af faktor 1 og faktor 2. Man bør ved et tofaktoriel forsøg først betragte LSD12. Er der angivet en værdi, betyder det, at der er vekselvirkning mellem de to faktorer, og LSD1 og LSD2 er dermed uinteressante.

## p-værdier

På resultatudskrifterne af enkeltforsøg er der anført en p-værdi, der er et andet mål for sikkerheden i forsøgsresultaterne. Hvis p-værdien er større end 0,05 er det ikke muligt med 95 pct. sikkerhed at sige, om der er forskel på behandlingerne. Størrelsen af p-værdien udtrykker således graden af sikkerhed - jo mindre p-værdi jo mere sikker. Ved et signifikansniveau på 0,05 betegnes behandlingsforskelle med test p-værdier  $< 0,05$  som statistisk signifikante.

## Signifikansgrupper fremkommet ved bogstavstest

Bogstavstestens signifikansgrupper vist på resultatsiden er fremkommet ved parvise sammenligninger vha. LSD værdien (Fisher's Least Significant Difference test). Behandlinger med forskelligt bogstav er signifikant forskellige. F.eks.:

P06: Ved optagning				
24-10-2012 ST.				
	TØRSTOF % af råvare	Nedre konf.	Øvre konf. int.	Signifikansgruppe
1	26,6	26,2	27,1	b
2	26,4	25,9	26,9	b
3	26,5	26,0	27,0	b
4	27,6	27,1	28,1	a
5	26,9	26,4	27,4	ab

LSD			
Måletid	Dato	Måling af	Resultaternes sikkerhed
P06	24-10-2012	TØRSTOF. % af råvare	lsd1=0,7 p1=0,01743

Det vil sige at:

- Led 1, 2, og 3 er IKKE signifikant forskellige fra hinanden.
- Led 1, 2, og 3 er signifikant forskellige fra led 4, men IKKE fra led 5.
- Led 4 er signifikant forskellig fra led 1, 2 og 3, men IKKE fra led 5.
- Led 5 er ikke signifikant forskellig fra nogle af de andre led.

Vises der ikke signifikansgrupper på den aktuelle måleparameter betyder det, at der ikke er signifikante forskelle mellem behandlingerne (Vises også som ns under LSD værdi).

Signifikansgrupperne vist på resultatsiden tager ikke højde for multiple testing, se nedenfor.

## Modelkontrol og detaljeret information vedr. de statistiske analyser

Under afsnittet Notater kan der downloades to \*.rtf dokumenter, der kan åbnes i Microsoft Word eller tilsvarende tekstbehandlingsprogram.

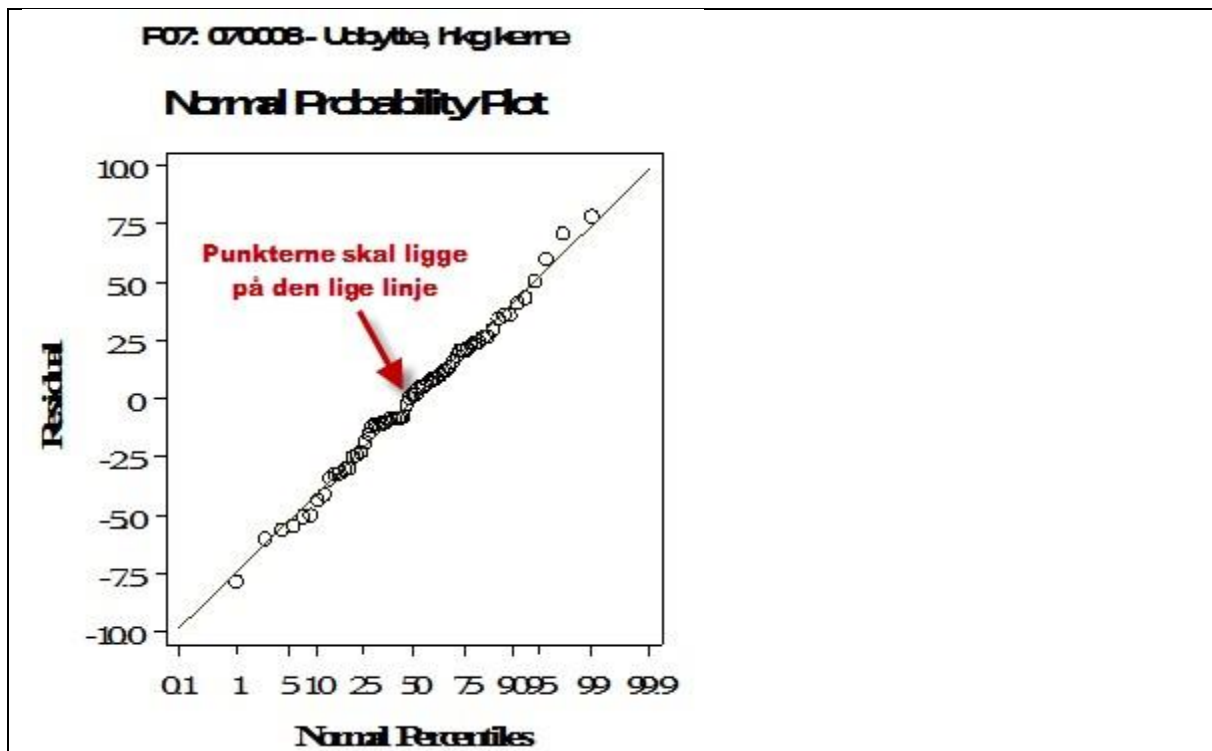
Notater			
Dato	Tid	Vedrørende	Notater
			Beregnete led-estimer (LSmeans) og statistiske analyser <a href="#">Download (rtf)</a>
			Observerede rå-data og modelkontrol <a href="#">Download (rtf)</a>

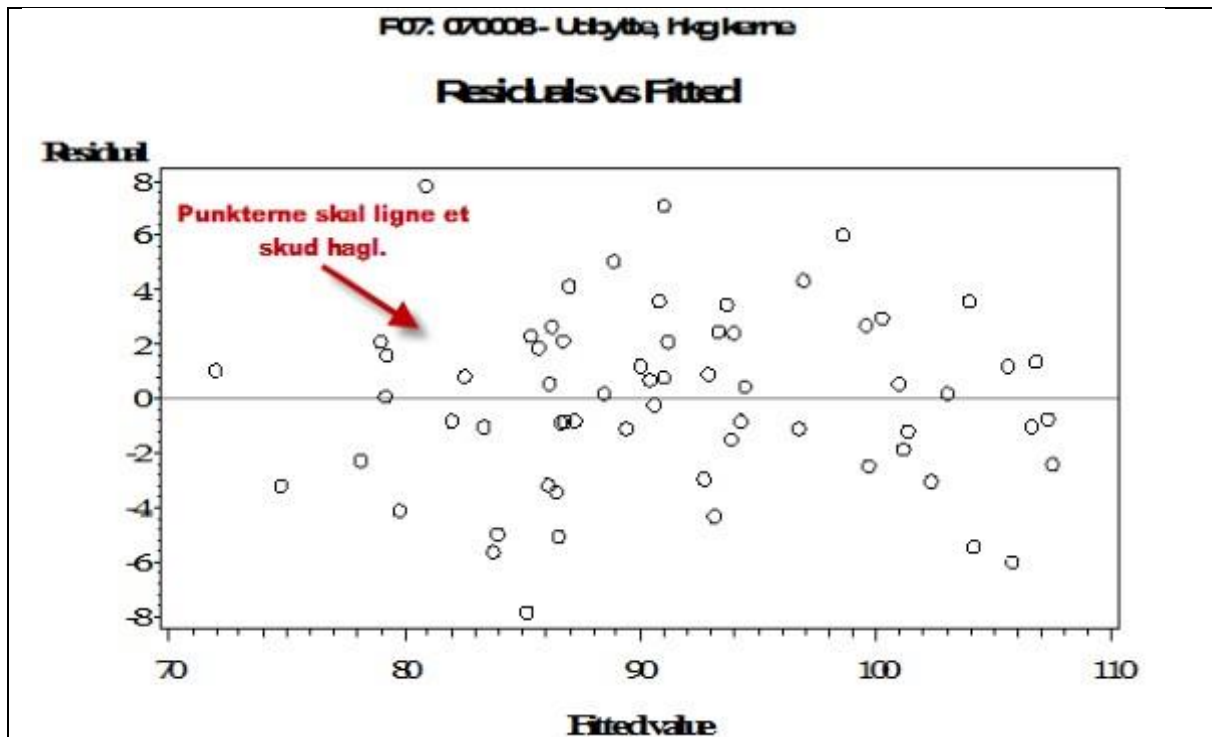
Det første dokument indeholder detaljeret information vedr. de statistiske analyser for alle måleparametre, der er beregnet statistik for. Dette inkluderer information vedr. en eventuel transformation, information vedr. variationens fordeling mellem modellens varianskomponenter (blok, alphablok, residual), modellens ANOVA tabel med F-testenes detaljer, en LSmeans tabel med standard error (SEM), antal frihedsgrader og 95%-konfidensintervaller, en LSmeans tabel med signifikansgrupper fra to bogstavstests (Unadjusted og Tukeys), samt én tabel af alle parvise tests - led mod led - for hver af de to bogstavstests Unadjusted og Tukeys. Se nedenfor for en beskrivelse af Unadjusted og Tukeys.

Det andet dokument indeholder for alle måleparametre og led de observerede rå-data (registreringer), deres median, gennemsnit, standardafvigelse og CV-værdi (Coefficient of variation). Derudover indeholder dokumentet modelkontrol-plots for de måleparametre, der er beregnet statistik for.

## Modelkontrol

For indstillingen *Normalfordelt*, *Logaritmisk transformation*, *Box-Cox-transformation* og *Logit-transformation* skal der foretages modelkontrol. p-værdier, LSD-værdier og bogstavstestens signifikansgrupper er kun valide, såfremt modelkontrollens tjek passerer. Modelkontrollen består i 1) et visuelt tjek, om residualerne er normalfordelte; dertil bruges qq-plottet (se venstre plot nedenfor) og 2) et visuelt tjek, om middelværdi-estimer og residualvariationen er uafhængige af hinanden; dertil bruges residualplottet (se højre plot nedenfor):





## Multiple testing: Unadjusted LSD vs. Tukeys honest difference

Dokumentet "Beregnete led-estimer (LSmeans) og statistiske analyser" indeholder to bogstavstests (Unadjusted og Tukeys).

Den første tabel over parvise sammenligninger baserer sig på Fisher's Least Significant Difference test. Denne test er en parvis sammenligning af leddenes LSmeans vha. LSD værdien. Denne kaldes også for Unadjusted, idet testen ikke tager højde for multiple testing.

Den anden tabel over parvise sammenligninger baserer sig på Tukey's range test, også kaldt Tukey's honest significance test, som er en parvis sammenligning af leddenes LSmeans under hensyntagen af multiple testing. Med et stigende antal parvise tests stiger også sandsynligheden for en falsk positiv (type I fejl). Tukeys metode korrigerer p-værdien for antal gennemførte tests.

LSD baserer sig på antagelsen, at der kan ses bort fra korrektionen for multiple testing og er dermed som udgangspunkt mindre konservativ end Tukeys metode. LSD er dog traditionelt brugt i landbrugsvidenskabelige kredse, og dette er grunden, hvorfor LSD er brugt i NFTS.