NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET
INSTITUTT FOR MATEMATISKE FAG

Fagleg kontakt under eksamen:
Institutt for matematiske fag, Gløshaugen
Oddgeir Samset, 99 54 27 79

EKSAMEN I EMNE TMA4260 INDUSTRIELL STATISTIKK

Onsdag 3. desember 2003
Tid: 09.00 – 14.00

Hjelpemiddel: B - Alle trykte og handskrivne hjelpemiddel tillatne. Bestemt, enkel kalkulator tillate.

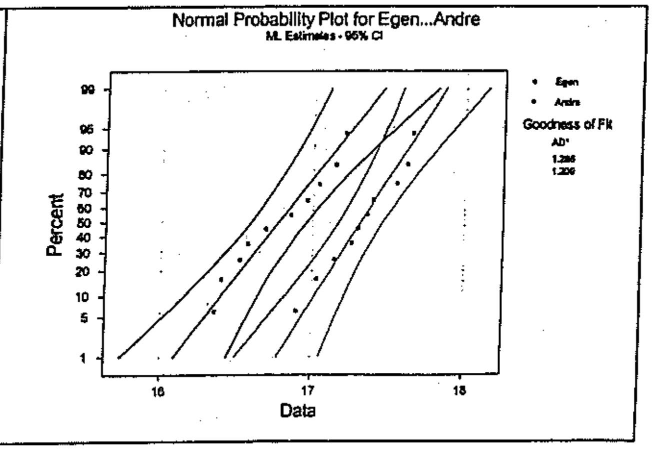
Sensuren fell 6. januar 2004

Oppgåve 1

Ei bedrift som produserer og leverer kulelager utfører ein benchmark av kvaliteten på tilsvarande kulelager frå ein annan leverandør. Det er spesielt kulene i kulelageret som er av interesse. Desse er laga i herda stål og bør vere så harde som mogleg. I tillegg ynskjer ein sjølvsagt at variansen i hardleiken til kulene er så liten som mogleg.

Hardleiken vert testa for 10 kuler frå eigen produksjon og 10 kuler frå produksjonen til den andre leverandøren. Data frå undersøkinga og nødvendige utskrifter frå MINITAB er gitt under. Eit innleiande normalplott av data syner at det ikkje er grunnlag for å påstå at data kjem frå ei anna fordeling enn normalfordelinga.

Row	Egen	Andre
1	16.85	16.90
2	16.40	17.03
3	17.21	17.65
4	16.35	17.40
5	16.52	17.26
6	17.04	17.15
7	16.96	17.62
8	17.15	17.30
9	16.69	17.36
10	16.57	17.55



a)

Er variansen i kulenes hardleik frå den andre leverandøren sin produksjon forskjellig frå variansen i kulenes hardleik frå bedrifta sin eigen produksjon? Formuler dette spørsmålet som eit hypotesetestingsproblem. Skriv ned og gjer greie for dei føresetnader som du eventuelt må gjere, set opp uttrykket for testobservator, utfør testen og dra konklusjon. Bruk 5% signifikansnivå.

b)

Er det grunn til å påstå at hardleiken til den andre leverandøren sine kuler er større enn for bedrifta sine eigne kuler? Formuler dette spørsmålet som eit hypotesetestingsproblem. Skriv ned og gjer greie for dei føresetnader som du eventuelt må gjere, set opp uttrykket for testobservator, utfør testen og dra konklusjon. Bruk 5% signifikansnivå. Vis spesielt korleis ein finn p-verdien, og forklar kva den fortel oss.

Finn også eit 95% konfidensintervall for differansen i hardleik.

```
Two-sample T for Andre vs Egen

         N      Mean     StDev   SE Mean
Andre   10    17.322     0.248     0.078
Egen    10    16.774     0.312     0.099

Difference = mu Andre - mu Egen
Estimate for difference:  0.548
95% lower bound for difference: 0.330
T-Test of difference = 0 (vs >): T-Value = 4.35  P-Value = 0.000  DF = 18
Both use Pooled StDev = 0.282
```

c)

Korleis vil du utføre testen under spørsmål **b)** dersom det innleiande normalplottet i staden hadde vist at det var grunn til å påstå at data kjem frå ei anna fordeling enn normalfordelinga? Skriv ned og gjer greie for dei føresetnader som du eventuelt må gjere, set opp uttrykket for testobservator, utfør testen og dra konklusjon. Bruk 5% signifikansnivå. Gi kommentar til resultatet.

Ei sortert liste over alle 20 observasjonane er gitt under med indeks frå eigen eller annan produksjon.

Row	Obs	Indeks	Row	Obs	Indeks
1	16.35	E	11	17.15	E
2	16.40	E	12	17.15	A
3	16.52	E	13	17.21	E
4	16.57	E	14	17.26	A
5	16.69	E	15	17.30	A
6	16.85	E	16	17.36	A
7	16.90	A	17	17.40	A
8	16.96	E	18	17.55	A
9	17.03	A	19	17.62	A
10	17.04	E	20	17.65	A

Oppgåve 2

På grunnlag av benchmark-resultata bestemte bedrifta seg for å studere herdeprosessen sin nærare, og følgjande fire faktorar ble valt ut: A – Tilsetjing av karbon, B – Herdetemperatur, C – Herdetid og D – Avkjølingstemperatur. Valt design og resultatet frå forsøket er gitt under.

Row	StdOrder	A	B	C	D	Hardhet
1	1	-1	-1	-1	1	15.32
2	2	1	-1	-1	-1	18.24
3	3	-1	1	-1	-1	17.18
4	4	1	1	-1	1	16.90
5	5	-1	-1	1	-1	15.95
6	6	1	-1	1	1	17.52
7	7	-1	1	1	1	14.26
8	8	1	1	1	-1	18.59

a)

Kva er generator og definerande relasjon til designet, og kva resolusjon har designet? Skriv opp alias-strukturen.

Finn estimatorene for hovudeffekten A og samspelet AC.

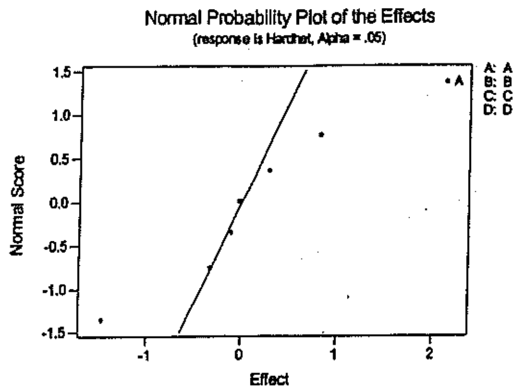
b)

Kva er variansen til estimatorene for hovudeffekten A og samspelet AC?

Gå ut frå at variansen i hardleik for kulene til bedrifta er den same nå som tidlegare. Bruk dei 10 observasjonane frå eigenproduserte kuler i Oppgåve 1 til å undersøkje om samspelet AC er signifikant forskjellig frå 0. Bruk 5% signifikansnivå. Kva vert konklusjonen på forsøket så langt?

Estimated Effects and Coefficients for Hardhet (coded units)

Term	Effect	Coef
Constant		16.7450
A	2.1350	1.0675
B	-0.0250	-0.0125
C	-0.3300	-0.1650
D	-1.4900	-0.7450
A*B	-0.1100	-0.0550
A*C	0.8150	0.4075
A*D	0.2850	0.1425



Analysis of Variance for Hardhet (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	4	13.776	13.776	3.4439	*	*
2-Way Interactions	3	1.515	1.515	0.5050	*	*
Residual Error	0	0.000	0.000	0.0000		
Total	7	15.291				

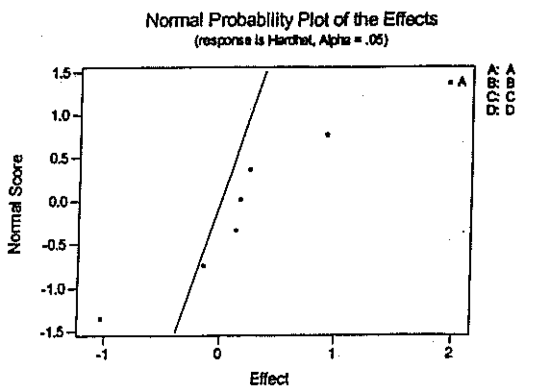
c)

Bedrifta er godt nøgd med resultatet frå undersøkinga så langt, og det vert bestemt at ein også skal utføre den andre halvfraksjonen. Den andre halvfraksjonen og resultatet frå forsøket er gitt under.

Row	StdOrder	A	B	C	D	Hardhet
1	1	-1	-1	-1	-1	16.57
2	2	1	-1	-1	1	16.72
3	3	-1	1	-1	1	15.76
4	4	1	1	-1	-1	17.69
5	5	-1	-1	1	1	14.59
6	6	1	-1	1	-1	18.63
7	7	-1	1	1	-1	16.18
8	8	1	1	1	1	17.86

Estimated Effects and Coefficients for Hardhet (coded units)

Term	Effect	Coef
Constant		16.7500
A	1.9500	0.9750
B	0.2450	0.1225
C	0.1300	0.0650
D	-1.0350	-0.5175
A*B	-0.1450	-0.0725
A*C	0.9100	0.4550
A*D	0.1650	0.0825



Analysis of Variance for Hardhet (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	4	9.901	9.901	2.4753	*	*
2-Way Interactions	3	1.753	1.753	0.5842	*	*
Residual Error	0	0.000	0.000	0.0000		
Total	7	11.654				

Bruk dette til å finne ukonfunderte estimat for hovudeffektane og to-faktor samspela.

Gå ut frå at ein vil estimere variansen til effektane ut frå dei høgare ordens samspela. Forklar korleis ein kan gjere dette, og finn estimatet. Er det fornuftig å ta med fire-faktor samspelet i denne utrekninga? Forklar.

I ettertid spurte ein av operatørane som deltok i forsøket om ein kunne utført den første halvfraksjonen i **a)** i to blokker. Dette ville i så fall letta utføringa av forsøket vesentleg. Kva ville du svart operatøren ?

Oppgåve 3

Eit mogleg samspel mellom faktor A – Tilsetjing av karbon og faktor C – Herdetid verkar interessant. Dersom ein kan redusera herdetida så kan ein auke produksjonen.

Ein av medarbeidarane i bedrifta har lest vidare i boka til Box, Hunter og Hunter og fant ut at ein kanskje kan nytte responsflate-teknikkar for å studere prosessen vidare og mogleg optimalisere den. Eit nytt forsøksdesign vert planlagt rundt forsøk nr 2 i første halvfraksjonen i Oppgåve 2 då dette forsøket gav stor hardleik samstundes som herdetida var lav.

La Y vere hardleiken, og la x_1 og x_2 vere dei koda nivåa for hhv karbon (A) og herdetid (C). Følgjande første-ordens design blei utført, kor verdien 0 indikerer eit senterpunkt i designet:

y	y_{c1}	y_{c2}	y_{c3}	y_{c4}	y_1	y_2	y_3	y_4
x_1	0	0	0	0	-1	1	-1	1
x_2	0	0	0	0	-1	-1	1	1

a)

Gjer greie for dei nødvendige føresetnadene og finn estimatorer for β_0, β_1 og β_2 uttrykt med $y_{c1},\ldots,y_4$ for ein første-ordens regresjonsmodell med forventning

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 .$$

Forklar korleis du kan testa om $\beta_1 \neq 0$ i denne modellen.

b)

Frå Oppgåve 2 veit vi at det er indikasjon på samspel mellom karbon og herdetid, og vi mistenkjer difor at ein første-ordens modell ikkje er nok. Vi ynskjer difor å sjekke om vi har ei krumma flate slik at vi i staden burde utvida designet vårt til eit *sentralt samansett design* for estimere ein andre-ordens modell

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 .$$

Teikn opp første-ordens designet frå **a)**. La $\bar{y}_f$ vere gjennomsnittet av dei fire punkta i 2^2 designet, og la $\bar{y}_c$ vere gjennomsnittet av dei fire senterpunkta.

Tenk deg så at du sit i sentrum av designet og ser utover responsflata. Forklar kvifor $\bar{y}_f - \bar{y}_c$ verkar som eit fornuftig mål på krumheita i responsflata, og vis at dersom andre-ordens modellen over er den riktige modellen så vil $E(\bar{y}_f - \bar{y}_c) = \beta_{11} + \beta_{22}$. Bruk dette til å forklara korleis ein kan lage ein hypotesetest om krumheit.

TMA4260 Industriell statistikk
Løsningsforslag Eksamen 3 desember 2003

Oppgave 1

a) Antagelser:

Egen $X_1, X_2, \ldots, X_{10}$ uavh identisk $N(\mu_x, \sigma_x^2)$
Andre $Y_1, Y_2, \ldots, Y_{10}$ uavh identisk $N(\mu_Y, \sigma_Y^2)$
X_i uavh Y_j $\forall ij$

Oppgaven gir ingen forsikring om at dataene er uavh. Vi burde derfor sjekke opp hvordan forsøket er randomisert.

Test: $H_0: \sigma_x^2 = \sigma_y^2$ mot $H_1: \sigma_x^2 \neq \sigma_y^2$ nivå $\alpha = 5\%$

Under H_0 vil $\frac{S_x^2}{S_y^2} \sim F_{9,9}$ Innsatt $\frac{S_x^2}{S_y^2} = \frac{0.312^2}{0.248^2} = \underline{1.5}$

Forkast H_0 dersom $F_{observert} > F_{\alpha/2,9,9} = 4.03$

$\Rightarrow$ Data gir <u>ikke</u> grunnlag for å forkaste H_0

b) Antagelser som over, og i tillegg at $\sigma_x^2 = \sigma_y^2 = \sigma^2$

Test: $H_0: \mu_Y \leq \mu_X$ mot $H_1: \mu_Y > \mu_X$ nivå $\alpha = 5\%$

Naturlig estimator: $\bar{Y} - \bar{X} \sim N(\mu_Y - \mu_X, \frac{2}{n}\sigma^2)$

Under H_0 vil $\frac{\bar{Y}-\bar{X}}{\sqrt{\frac{2}{10}S_{pool}^2}} \sim T_{18}$ $S_{pool}^2 = \frac{9S_x^2 + 9S_y^2}{18}$

p-verdi $= P(T_{18} \geq T_{observert} / H_0) = P(T_{18} \geq 4.35 / H_0) = \underline{0.00}$

p-verdi beskriver sannsynligheten for å få det vi har observert eller noe enda mer ekstremt gitt H_0 riktig

$\Rightarrow$ Data gir grunnlag for å forkaste H_0 og påstå at $\mu_Y > \mu_X$

95% konfidensintervall

$$P\left(t_{\alpha/2,18} < \frac{\bar{Y}-\bar{X}-(\mu_Y-\mu_X)}{\sqrt{\frac{2}{n} S_{pool}^2}} \le t_{\alpha/2,18} \right) = 1-\alpha$$

$$\bar{Y}-\bar{X} \pm t_{\alpha/2,18} \cdot \sqrt{\tfrac{2}{10} S_{pool}^2}$$

$$0.548 \pm 2.101 \cdot \sqrt{\tfrac{2}{10} \cdot 0.282^2} \qquad \underline{(0.283,\ 0.813)}$$

c) Kan benytte Wilcoxon's to-utvalg test

Antar:

$X_1, \dots, X_{10}$ uavh. identisk og kontinuerlig $F_X(x)$

$Y_1, \dots, Y_{10}$ uavh. identisk og kontinuerlig $F_Y(y)$

X_i uavh Y_j $\forall ij$

$F_X(x)$ og $F_Y(y)$ har samme form men er forskjøvet ift hv

Test: $H_0: \mu_Y \le \mu_X$ mot $H_1: \mu_Y > \mu_X$ nivå $\alpha = 5\%$

Forkast H_0 dersom rang-sum til Y blir stor eller rang-sum til X blir liten

$$W_X = \underline{63.5} \qquad W_Y = \frac{20 \cdot 21}{2} - W_X = 210 - 63.5 = 146.5$$

$$\begin{aligned} p\text{-verdi} &= P(W_X \le 63.5 \mid H_0) \\ &= P(W_X - \min W_X \le 63.5 - 55 \mid H_0) \\ &= P(U \le 8.5 \mid H_0) < \underline{0.005} \qquad (n_1 = 8,\ n_2 = 8) \end{aligned}$$

$\Rightarrow$ Konklusjonen blir den samme som under b)

Ekstrapolasjon i tabellen ($n_1 = 10$ og $n_2 = 10$) gir p-verdi ≈ 0.0015 (0.001 fra Minitab). En noe høyere p-verdi enn for testen i b) er som forventet.

Oppgave 2

a) Generator $D = -ABC$ eller $I = -ABCD$

Definerende relasjon $I = -ABCD$ $\qquad R = IV$

Alias-struktur:

$I \to I - ABCD$

$A \to A - BCD$ $\qquad AB \to AB - CD$

$B \to B - ACD$ $\qquad AC \to AC - BD$

$C \to C - ABD$ $\qquad AD \to AD - BC$

$D \to D - ABC$

Estimator for A

$$l_A = \hat{A} = \frac{(Y_2 + Y_4 + Y_6 + Y_8) - (Y_1 + Y_3 + Y_5 + Y_7)}{4}$$

$$l_{AC} = \hat{AC} = \frac{(Y_1 + Y_3 + Y_6 + Y_8) - (Y_2 + Y_4 + Y_5 + Y_7)}{4}$$

b) Anta at $Y_1, \ldots, Y_8$ uavh. normalfordelt med konstant varians σ_Y^2

$$Var(\hat{A}) = \frac{8}{16}\sigma_Y^2 = \underline{\frac{1}{2}\sigma_Y^2}$$

$$Var(\hat{AC}) = \underline{\frac{1}{2}\sigma_Y^2}$$

95% konfidensintervall for AC

$$\hat{AC} \pm t_{\alpha/2,\,9}\sqrt{\frac{1}{2}\sigma_Y^2}$$

$$0.815 \pm 2.262\sqrt{\frac{1}{2}0.312^2}$$

$$0.815 \pm 0.499$$

Konfidensintervallet inneholder <u>ikke</u> verdien 0

$\Rightarrow$ Data gir grunnlag for å påstå at AC er forskjellig fra 0.

Konklusjon på forsøket så langt :

Aktive effekter (kontraster) : A, D og AC
Inerte effekter (kontraster) : B, C, AB og AD

Dersom vi kan anta at tre-faktor samspill kan neglisjeres så er A og D aktive hovedeffekter. Dessverre klarer designet ikke å skille AC og -BD som er fullstendig konfundert.

c) $\hat{A} = (l_A + l_A')/2 = (2.135 + 1.950)/2 = \underline{2.043}$

$\hat{BCD} = (l_A' - l_A)/2 = \underline{-0.093}$

$\hat{B} = 0.110$
$\hat{C} = -0.100$
$\hat{D} = -1.263$
$\hat{AB} = -0.128$
$\hat{AC} = 0.863$
$\hat{AD} = 0.225$
$\hat{BC} = -0.060$

$\hat{BD} = 0.048$
$\hat{CD} = -0.018$
$\hat{ABC} = 0.228$
$\hat{ABD} = 0.230$
$\hat{ACD} = 0.135$
$\hat{ABCD} = -0.005$ (konf med blokk)

Fire-faktor samspillet er konfundert med med blokk og bør derfor utelates.

La E_i^* angi de $n^*=4$ tre-faktor samspillene

$E_i^* \sim N(0, \sigma^2_{effekt})$

$$\hat{\sigma}^2_{effekt} = \frac{\sum E_i^{*2}}{n^*} = \frac{0.2275^2 + 0.2300^2 + 0.1350^2 + 0.0925}{4}$$

$$= \underline{0.033}$$

eller $\hat{\sigma}^2_{effekt} = \frac{4}{16} \underbrace{MS_{trefaktor}}_{\hat{\sigma}_y^2} = \frac{0.13144}{4} = \underline{0.033}$

$I = -ABCD$

Anta Blokk = ABC ⇒ Blokk konf med hovedeffekt !
Anta Blokk = AB ⇒ Blokk konf med to-faktor samspill !
Uheldig for oss siden vi har aktive to-faktor <u>samspill</u>

Oppgave 3

a) Modell: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$

Antagelser: ε_i uavh $N(0, \sigma_Y^2)$ $\forall i$

$$X = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \\ \hline 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$Y = X\beta + \varepsilon$$

$$\hat{\beta} = (X'X)^{-1} X'Y$$

$$= \begin{bmatrix} 1/8 & & 0 \\ & 1/4 & \\ 0 & & 1/4 \end{bmatrix} X'Y$$

$$\hat{\beta}_0 = \frac{Y_{c1} + Y_{c2} + Y_{c3} + Y_{c4} + Y_1 + Y_2 + Y_3 + Y_4}{8}$$

$$\hat{\beta}_1 = \frac{(Y_2 + Y_4) - (Y_1 + Y_3)}{4} \qquad \hat{\beta}_2 = \frac{(Y_3 + Y_4) - (Y_1 + Y_2)}{4}$$

$$Var(\hat{\beta}_1) = \frac{4}{16}\sigma_Y^2 = \frac{1}{4}\sigma_Y^2$$

σ_Y^2 er ukjent men kan estimeres vha senterpunkt

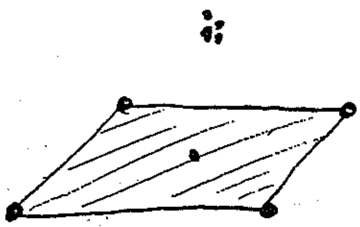
$$\hat{\sigma}_y^2 = \frac{\sum_{i=1}^{4} (y_{ci} - \bar{y}_c)^2}{3}$$

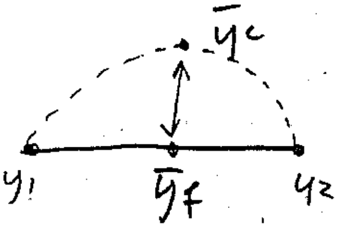
$$H_0: \beta_1 = 0 \quad \text{mot} \quad H_1: \beta_1 \neq 0$$

Under H_0 vil $\dfrac{\hat{\beta}_1}{\sqrt{\frac{1}{4}\hat{\sigma}_y^2}} \sim T_3$

p-verdi $= 2 \cdot P(T_3 \geq |T_{observert}| \mid H_0)$

b)





Dersom $\bar{y}_f - \bar{y}_c$ er forskjellig fra 0 tyder det på at responsene i senterpunktet ligger utenfor planet. Dette tyder på en krum flate

Anta at $E[Y] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \beta_{11} X_1^2 + \beta$:

$$
\begin{aligned}
E[\bar{Y}_f] &= \frac{1}{4} E[Y_1 + Y_2 + Y_3 + Y_4] \\
&= \frac{1}{4} [\beta_0 - \beta_1 - \beta_2 + \beta_{12} + \beta_{11} + \beta_{22} \\
&\quad + \beta_0 + \beta_1 - \beta_2 - \beta_{12} + \beta_{11} + \beta_{22} \\
&\quad + \beta_0 - \beta_1 + \beta_2 - \beta_{12} + \beta_{11} + \beta_{22} \\
&\quad + \beta_0 + \beta_1 + \beta_2 + \beta_{12} + \beta_{11} + \beta_{22}] \\
&= \underline{\beta_0 + \beta_{11} + \beta_{22}}
\end{aligned}
$$

$E[\bar{Y}_c] = \beta_0 \qquad E[\bar{Y}_f - \bar{Y}_c] = \underline{\beta_{11} + \beta_{22}}$

Test for krumhet : $H_0 : \beta_{11} + \beta_{22} = 0 \quad H_1 : \beta_{11} + \beta_{22} \neq 0$

$$\mathrm{Var}[\bar{Y}_f - \bar{Y}_c] = \frac{1}{4}\sigma_y^2 + \frac{1}{4}\sigma_y^2 = \frac{1}{2}\sigma_y^2$$

Under Ho vil $\dfrac{\bar{Y}_f - \bar{Y}_c}{\sqrt{\frac{1}{2}\hat{\sigma}_y^2}} \sim T_3$

p-verdi $= 2 \cdot P(T_3 \geq |T_{observert}| \,|\, H_0)$