



Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1911, STATISTIK FÖR BIOTEKNIK

Måndag den åttonde januari 2018 14:00-19:00.

Examinator: Timo Koski, 70 – 237 00 47.

Kursledare: Timo Koski, – 790 71 34 .

Tillåtna hjälpmedel: Formel- och tabellsamling för SF911, Mathematics Handbook (Beta), hjälpreda för miniräknare, miniräknare.

Införda beteckningar skall förklaras och definieras. Resonemang och uträkningar skall vara så utförliga och väl motiverade att de är lätta att följa. Numeriska svar skall anges med minst två siffrors noggrannhet. Tentamen består av 8 uppgifter. Varje korrekt lösning ger 4 poäng. Gränsen för godkänt är preliminärt 16 poäng (ev. bonuspoäng inräknade). Möjlighet att komplettera ges för tentander med, preliminärt, 14–15 poäng. Tid och plats för komplettering kommer att anges på kursens hemsida. Det ankommer på dig själv att ta reda på om du har rätt att komplettera.

Tentamen kommer att vara rättad inom tre arbetsveckor från skrivningstillfället och kommer att finnas tillgänglig på studentexpeditionen minst sju veckor efter skrivningstillfället.

Uppgift 1

a) Data på livslängderna y för proteinerna i proteinfamiljen A visar att livslängderna y är två gånger livslängderna x för proteinerna i proteinfamiljen B. r_{xy} är korrelationskoefficienten. Vilken av följande utsagor **i)-v)** är sann? Motivera kortfattat ditt svar. (2 p)

i) $r_{xy} = 0.5$.

ii) $r_{xy} = -0.5$.

iii) $r_{xy} = 1$.

iv) $r_{xy} = -1$.

v) Denna fråga kan inte besvaras utan att kolla data.

b) I ett kursprojekt mäter man vikten i g och svanslängden i mm hos en grupp vita möss. Korrelationskoefficienten r mellan vikt och svanslängd blev $r = 0.7$. Ett spridningsdiagram (scatterplot) visar en utliggare, som tydligt placerar sig under de övriga punkterna. Det visar sig att utliggaren svarar mot en sjuk mus, och därför elimineras utliggaren från projektet. En av följande utsagor **i)-iv)** är sann, vilken? Motivera kortfattat ditt svar. (2 p)

i) Värdet på r är oförändrat efter att utliggaren har avlägsnats.

ii) Värdet på r växer efter att utliggaren har avlägsnats.

- iii) Värdet på r minskar efter att utliggaren har avlägsnats.
- iv) Denna fråga kan inte besvaras utan att ha tillgång till data.

Uppgift 2

Vi jämför två DNA-sekvenser $\mathbf{x}$ och $\mathbf{y}$ med anpassningen

$$\begin{array}{r} \mathbf{x} = G \downarrow A \downarrow T \downarrow A \downarrow A \downarrow G \downarrow C \downarrow C \downarrow C \downarrow C \downarrow T \downarrow G \downarrow T \downarrow C \downarrow T \\ \mathbf{y} = C A A A A T C C C C A G T C T \end{array}$$

Vi säger att vi har en **match**, om de två nukleotiderna i en position är desamma i de båda sekvenserna. Här är dylika match indikerade av $\downarrow$.

Vi antar att nukleotiderna i $\mathbf{x}$ och $\mathbf{y}$ är slumpässigt valda och oberoende av varandra så att sannolikheten för en match är $1/4$ på var och en av de 15 positionerna. Vi antar att längderna ($=15$) på $\mathbf{x}$ och $\mathbf{y}$ är givna på förhand.

- a) Låt X_1 = antalet match i passning av två sekvenser av längd 15. En av följande utsagor **i)-v)** är sann, vilken? (2 p)

- i) $X_1 \sim \mathcal{Poi}(1/4)$.
- ii) $X_1 \sim \mathcal{Ge}(1/4)$.
- iii) $X_1 \sim \mathcal{Bin}(15, 1/4)$.
- iv) $X_1 \sim \mathcal{NB}(15, 1/4)$.
- v) $X_1 \sim \mathcal{U}(1, \dots, 15)$.

- b) Anta att vi gjorde passningen av två DNA-sekvenser med förlängning av sekvenserna under i övrigt oförändrade antaganden tills vi får 15 match. Låt X_2 = antalet positioner **utan match**, när vi fått 15 match. En av följande utsagor **i)-v)** är sann, vilken? (2 p)

- i) $X_2 \sim \mathcal{Poi}(1/4)$.
- ii) $X_2 \sim \mathcal{Ge}(1/4)$.
- iii) $X_2 \sim \mathcal{Bin}(15, 1/4)$.
- iv) $X_2 \sim \mathcal{NB}(15, 1/4)$.
- v) $X_2 \sim \mathcal{U}(1, \dots, 15)$.

Uppgift 3

Hardy-Weinbergs lag eller Hardy-Weinbergs jämvikt (HWE) för allelfrekvenser i en population kräver ett antal förutsättningar för att gälla. En av följande **i)-vi)** ingår INTE bland förutsättningarna för HWE, vilken?

- i) en mycket stor population.
- ii) inga mutationer av alleler.
- iii) ingen migration in i eller ut ur populationen.

- iv) genetisk drift, d.v.s. ändringar i de relativa allelfrekvenserna inträffar slumpmässigt.
- v) ingen selektion bland genotyperna (d.v.s. ingen skillnad i reproduktionsförmåga mellan individerna).
- vi) parningen är slumpmässig (random mating). (4 p)

Uppgift 4

År 2012 godkände läkemedelsmyndigheten FDA i USA ett första test kit för HIV, OraQuick In-Home HIV Test, för privat bruk. OraQuick accepterades tack vare en undersökning, där 4410 vuxna, som tillhörde en riskgrupp, testades med OraQuick. Dessa individer kände inte till, huruvida de var infekterade eller inte. Det sanna tillståndet på HIV-infektion ($D+$ = infektion) eller inte ($D-$) kontrollerades sedan för var och en av dessa individer med flera andra medicinska undersökningar.

Nedan har vi resultaten från denna undersökning. $T+$ betyder att OraQuick gav resultatet att HIV-infektion föreligger, och $T-$ att HIV-infektion inte föreligger enligt OraQuick.

	$D+$	$D-$
$T+$	TP = 105	FP=1
$T-$	FN = 8	TN = 4296

- a) En av följande utsagor **i)-iv)** är sann, vilken? (1 p)

- i) OraQuicks Sensitivity = 0.9875.
- ii) OraQuicks Sensitivity = 0.9998.
- iii) OraQuicks Sensitivity = 0.9292.
- iv) OraQuicks Sensitivity = 0.9502.

- b) En av följande utsagor **i)-iv)** är sann, vilken? (1 p)

- i) OraQuicks Specificity = 0.9875.
- ii) OraQuicks Specificity = 0.9998.
- iii) OraQuicks Specificity = 0.9292.
- iv) OraQuicks Specificity = 0.9502.

- c) På basis av tabellen ovan räknar vi att den relativa frekvensen för HIV, $P(D+)$, i denna högriskgrupp är = 0.025. Du får ta detta som givet. Vi har att PPV = positive predictive value = $P(D+|T+)$. En av följande utsagor **i)-iv)** är sann, vilken? (2 p)

- i) OraQuicks PPV = 0.9500.
- ii) OraQuicks PPV = 0.9906.
- iii) OraQuicks PPV = 0.9650.

iv) OraQuicks PPV = 0.9700

Uppgift 5

I kvalitetskontroll (QC) vid kundanpassad (eller just-in-time) produktion behandlas ett mätvärde X av en producerad enhet så att man först subtraherar produktionsprocessens medelvärde från X och dividerar skillnaden med standardavvikelsen. Resultatet kallas i QC ett Z -värde¹.

Vi antar att $Z \in \mathcal{N}(0, 1)$. Vi säger att en producerad enhet är ”systemiskt avvikande”, om $|Z| > 2.5758$. De mot olika producerade enheter svarande Z -värdena antas vara oberoende av varandra.

a) Beräkna sannolikheten

$$p_c = P(|Z| > 2.5758). \quad (1 \text{ p})$$

b) Låt nu produktionen löpa och sätt Y = antalet kvalitetsgranskade enheter, när den första systemiskt avvikande enheten uppträder, den första avvikande enheten medräknad. M.a.o. antar Y värdena $1, 2, 3, \dots$

Vilken fördelning har Y ? Lägg märke till att p_c ur a) skall ingå som en statistisk parameter i denna fördelning. (2 p)

c) Beräkna sannolikheten $P(Y \geq 3)$. Ifall Du inte löst del a) av denna uppgift får Du använda (det felaktiga) värdet $p_c = 0.05$. (1 p)

Uppgift 6

En bakteriofag är ett virus som infekterar bakterier. En mikrobiolog studerar en bakterieart och frågar sig hur antalet infektioner med en viss typ av bakteriofager fördelar sig statistiskt. Hen odlar ett antal bakterieodlingar och räknar antalet infekterade bakterier per odling. På basis av sina data presenterar hen följande empiriska sannolikhetsfunktion för X = antalet infektioner per odling.

Antal infektioner per odling	0	1	2	3	4	5
Sannolikhet	0.4	0	0	0.2	0.3	0.1

a) Beräkna koefficienten $VMR = \frac{Var(X)}{E(X)}$ utifrån denna sannolikhetsfunktion. (1 p)

b) Om bakteriofagerna angriper bakterierna slumpmässigt, så är en rimlig modell för X en Poissonfördelning, m.a.o., $X \sim \mathcal{Poi}(\lambda)$. Jämför resultatet i a) med VMR för $\mathcal{Poi}(\lambda)$.

Verkar bakteriofagerna agera slumpmässigt enligt Din bedömning? Vilken nollhypotes och mothypotes testas här? Motivering av svaret förväntas. Varför görs det här inga konfidensintervall för VMR? Vilken statistisk metod skulle kunna ge ett konfidensintervall för VMR? (3 p)

¹j.f.r. Lars Sörqvist: Ständiga förbättringar. En bok om resultatorienterat förbättringsarbete, verksamhetsutveckling och Sex Sigma. Studentlitteratur 2004

Uppgift 7

Man vill undersöka halten av kvartsdamm (mg/m^3) i luften på en arbetsplats. Man antar att mätningarna är fördelade enligt $N(\mu, \sigma^2)$. Vi gör fem oberoende mätningar och får

$$x_1 = 0.086, x_2 = 0.112, x_3 = 0.079, x_4 = 0.092, x_5 = 0.085.$$

- a) Gör ett tvåsidigt 95% konfidensintervall för μ . (2 p)
- b) Ett svenskt hygieniskt gränsvärde för respirabelt kvartsdamm är $0.1 \text{ mg}/\text{m}^3$. Vi vill avgöra om det finns anledning till oro på basis av data ovan. Vi tar nollhypotesen

$$H_o : \mu = 0.1$$

Mothypotesen är

$$H_a : \mu < 0.1$$

Avgör på signifikansnivån 1 % om halten av kvartsdamm på denna arbetsplats i genomsnitt ligger under gränsvärdet. Om konfidensintervallet i **a)** användes här, vilken slutsats skulle man dra? . (2 p)

Uppgift 8

Ett team av elektrokemister gör mätningar av pH-värdet för en kommersiell läskdryck med två olika pH -mätare (egentligen pH-array sensorer)². Ingen av dessa pH-mätare har ett systematiskt fel.

DATA: pH-mätare 1 producerar mätdata $x_1, \dots, x_{n_1}$ med mätresultat har som standardavvikelse σ_1 och pH-mätare 2 ger data $y_1, \dots, y_{n_2}$ med standardavvikelse σ_2 . För skattning av pH-värdet för drycken bildas

$$\bar{x} = \frac{1}{n_1} (x_1 + \dots + x_{n_1})$$

och

$$\bar{y} = \frac{1}{n_2} (y_1 + \dots + y_{n_2})$$

och dessa kombineras (*data fusion* på engelska) genom

$$\hat{\mu}_{obs} = c_1 \bar{x} + c_2 \bar{y},$$

där c_1 och c_2 är konstanter.

MODELL: Låt $x_1, \dots, x_{n_1}$ vara respektive utfall av oberoende X_i med $E(X_i) = \mu$ och $D(X_i) = \sigma_1$ för $i = 1, \dots, n_1$, samt att $y_1, \dots, y_{n_2}$ är respektive utfall av oberoende Y_i med $E(Y_i) = \mu$ och $D(Y_i) = \sigma_2$. Då tar vi

$$\hat{\mu} = c_1 \bar{X} + c_2 \bar{Y},$$

där

$$\bar{X} = \frac{1}{n_1} (X_1 + \dots + X_{n_1})$$

$$\bar{Y} = \frac{1}{n_2} (Y_1 + \dots + Y_{n_2}).$$

Då är $\hat{\mu}_{obs}$ ett utfall av $\hat{\mu}$.

²Yi-Hung Liao & Jung-Chuan Cho: Comparison of pH Data Measured with a pH Sensor Array Using Different Data Fusion Methods. *Sensors* (Basel). 2012; 12(9): 12098–12109.

a) Beräkna

$$E(\hat{\mu}) = E(c_1\bar{X} + c_2\bar{Y}).$$

Hur bör konstanterna c_1 och c_2 väljas så att $E(\hat{\mu}) = \mu$? (2 p)

b) Beräkna

$$Var(\hat{\mu}) = Var(c_1\bar{X} + (1 - c_1)\bar{Y}).$$

Hur bör konstanten c_1 väljas så att $Var(\hat{\mu})$ minimeras? (2 p)

Lycka till!



Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG
TENTAMEN I SF1911 STATISTIK FÖR BIOTEKNIK.
Måndag den åttonde januari 2018 14:00-19:00.

Uppgift 1

- a) iii) $r_{xy} = 1$, om sambandet är en exakt rät linje $y = 2x$. $r_{xy} = 1$ p.g.a. att $2 > 0$. (Ett komplett svar ingår i Problem 1.5.6.).
- b) ii) Värdet på r växer efter att utliggaren har avlägsnats, ty utliggaren placerade sig under de övriga punkterna och r är positiv.

Uppgift 2

- a) iii) $X_1 \sim \text{Bin}(15, 1/4)$.
- b) iv) $X_2 \sim \mathcal{NB}(15, 1/4)$.

Uppgift 3

- iv) genetisk drift, d.v.s. ändringar i de relativa allelfrekvenserna inträffar slumpmässigt.

Uppgift 4

	$D+$	$D-$
$T+$	TP = 105	FP=1
$T-$	FN = 8	TN = 4296

- a) iii) OraQuicks Sensitivity = 0.9292, ty Sensitivity är $= P(T+ | D+) = \frac{TP}{TP+FN}$.
- b) ii) OraQuicks Specificity = 0.9998, ty Specificity $= P(T- | D-) = \frac{TN}{TN+FP}$
- c) i) OraQuicks PPV = 0.9500.
- ii) OraQuicks PPV = 0.9906.
- iii) OraQuicks PPV = 0.9650.
- iv) OraQuicks PPV = 0.9700

PPV= Positivt prediktivt värde

$$= P(D+|T+) = \frac{P(T+|D+)P(D+)}{P(T+)} = 0.9292 \cdot 0.025 / (P(T+)).$$

Här är $P(T+) = P(T+|D+)P(D+) + P(T+|D-)P(D-)$. Vi har $P(T+|D-) = \frac{FP}{FP+TN} = 1/(1+4296) = 0.0002$. Detta ger

$$P(T+) = P(T+|D+)P(D+) + P(T+|D-)P(D-) = 0.9292 \cdot 0.025 + 0.0002 \cdot 0.9975 = 0.0234.$$

Således är

$$PPV = \frac{0.9292 \cdot 0.025}{0.0234} = 0.9927,$$

vilket är **ii)**, bortsett från tryckfel i tentan.

Uppgift 5

a) Vi har att

$$p_c = P(|Z| > 2.5758) = P(Z \leq -2.5758) + P(Z \geq 2.5758)$$

och p.g.a. symmetrin hos fördelningen för Z fås

$$= 2 \cdot P(Z \geq 2.5758) = 2 \cdot 0.005 = 0.01.$$

ty $2.5758 = \lambda_{0.005}$ enligt Tabell 2 i Formel- och tabellsamlingen.

b)

Y = antalet kvalitetsgranskade enheter t.o.m. den första systemiskt avvikande enheten uppträder, är en $Geom(p_c)$ -fördelad stokastisk variabel, där $p_c = 0.01$ enligt del b).

c) Den betraktade händelsen $Y \geq 3$ är i själva verket lika med händelsen att få först två icke-systemiskt avvikande Z -värden efter varandra:

$$P(Y \geq 3) = (1 - p_c)^2 = 0.99^2 \approx 0.98.$$

Om vi vill checka detta mer omständligt med utnyttjande av sannolikhetsfunktionen för ffg(p), kan vi observera att

$$P(Y \geq 3) = 1 - P(Y \leq 2).$$

Eftersom $p_Y(k) = (1 - p_c)^{k-1}p_c$, med $k = 1, 2, \dots$, har vi att

$$\begin{aligned} P(Y \leq 3) &= \sum_{k=1}^2 p_Y(k) = \sum_{k=1}^2 (1 - p_c)^{k-1}p_c \\ &= p_c + (1 - p_c)p_c. \end{aligned}$$

Således

$$\begin{aligned} P(Y \geq 3) &= 1 - P(Y \leq 2) = 1 - p_c - (1 - p_c)p_c \\ &= 1 - 2p_c + p_c^2 = (1 - p_c)^2. \end{aligned}$$

$$\underline{P(Y \geq 3) = 0.98.}$$

Uppgift 6

a) Från tabellen

$$E(X) = 0 \cdot 0.5 + 1 \cdot 0 + 2 \cdot 0 + 3 \cdot 0.2 + 4 \cdot 0.3 + 5 \cdot 0.1 = 2.3.$$

$$E(X^2) = 0^2 \cdot 0.5 + 1^2 \cdot 0 + 2^2 \cdot 0 + 3^2 \cdot 0.2 + 4^2 \cdot 0.3 + 5^2 \cdot 0.1 = 9.1.$$

Således är

$$\text{Var}(X) = E(X^2) - E(X)^2 = 9.1 - 2.3^2 = 3.81.$$

Detta ger VMR(=Variance -to - Mean -Ratio) som

$$\underline{\text{VMR} = 3.81/2.3 \approx 1.66}$$

b) Om $X \sim \mathcal{Poi}(\lambda)$, så fås enligt formelsamlingen $\text{VMR} = \frac{\lambda}{\lambda} = 1$. Jämför resultatet i a) med VMR för $\mathcal{Poi}(\lambda)$. Vi väljer nollhypotesen

$$H_o : \text{VMR}=1$$

och mothypotesen

$$H_a : \text{VMR} \neq 1,$$

ty nollhypotesen svarar mot Poissonfördelningen. Vi har teststatistikan

$$\frac{\text{Var}(X)}{E(X)}.$$

Observera att denna kvot är en statistika, ty dess numeriska värde bestäms utifrån data via den empiriska sannolikhetsfunktionen.

Det finns inga konfidensintervall här, ty ingen levande eller bortgången statistiker har framställt någon exakt eller ens en analytiskt härledd approximativ fördelning för teststatistikan $\frac{\text{Var}(X)}{E(X)}$. Det observerade värdet = 1.66 på VMR är rätt långt ifrån värdet = 1, således verkar det enligt en rimlig bedömning som att bakteriofagerna inte agerar slumpmässigt.

Bootstrap-metoden skulle kunna ge ett approximativt konfidensintervall för VMR och en hypotesprövning.

Uppgift 7

a) För ett tvåsidigt 95% konfidensintervall för μ behöver vi

$$\bar{x} = 0.0908, s_x = 0.0127.$$

Vi har från tabellsamlingen $t_{0.025}(4) = 2.78$. Dessa ger

$$I_\mu = \bar{x} \pm t_{0.025}(4)s_x/\sqrt{5} = 0.0908 \pm 2.78 \cdot 0.0127/\sqrt{5} = [0.07507, 0.1066]$$

b) För

$$H_o : \mu = 0.1$$

och mothypotesen

$$H_a : \mu < 0.1$$

räknar vi ett ensidigt konfidensintervall för μ . Enligt kursens formelsamling bildar vi den t-fördelade teststatistikan

$$t = \frac{\bar{x} - 0.1}{s_x/\sqrt{5}} \sim t(5 - 1) = t(4).$$

Vi har från uppgiften att $\bar{x} = 0.0908$ och $s_x = 0.0128$.

$$t = \frac{\bar{x} - 0.1}{s_x/\sqrt{5}} = \frac{0.0908 - 0.1}{0.0127/\sqrt{5}} = -1.6178.$$

Det ensidiga kritiska området ges av $-t_{0.01}(4) = -3.75$. Vi ser att $t = -1.6178 > -3.65$, och **nollhypotesen förkastas inte på signifikansnivån 0.01**. Det ensidiga konfidensintervallet är

$$I_\mu =] -\infty, 0.1121],$$

och 0.1 ligger i intervallet. Med dessa data har man inte visat att μ ligger under gränsvärdet.

Om konfidensintervallet i **a)** användes här, ser vi att 0.1 ligger i I_μ och nollhypotesen kan ej förkastas på signifikansnivån 0.05 i ett tvåsidigt test.

Uppgift 8

a)

$$E(\hat{\mu}) = E(c_1\bar{X} + c_2\bar{Y}) = c_1E(\bar{X}) + c_2E(\bar{Y}),$$

ty c_1 och c_2 är inte stokastiska variabler. Enligt räknereglerna fås

$$E(\bar{X}) = \frac{1}{n_1} (E(X_1) + \dots + E(X_{n_1})) = \frac{1}{n_1} \cdot n_1\mu = \mu.$$

På samma sätt fås

$$E(\bar{Y}) = \mu.$$

Detta innebär att

$$E(\hat{\mu}) = c_1\mu + c_2\mu = (c_1 + c_2)\mu.$$

Om vi alltså tar $c_1 + c_2 = 1$, så gäller $E(\hat{\mu}) = \mu$.

$$\underline{c_1 + c_2 = 1.}$$

b) Utifrån **a)** ser vi att vi kan välja att minimera

$$Var(\hat{\mu}) = Var(c_1\bar{X} + (1 - c_1)\bar{Y})$$

med bibehållet krav $E(\hat{\mu}) = \mu$.

Oberoendet ger

$$\text{Var}(c_1\bar{X} + (1 - c_1)\bar{Y}) = c_1^2\text{Var}(\bar{X}) + (1 - c_1)^2\text{Var}(\bar{Y}).$$

Men oberoendet ger

$$\begin{aligned}\text{Var}(\bar{X}) &= \frac{1}{n_1^2} (\text{Var}(X_1) + \dots + \text{Var}(X_{n_1})) \\ &= \frac{1}{n_1^2} (\sigma_1^2 + \dots + \sigma_1^2) = \frac{1}{n_1^2} \cdot n_1 \sigma_1^2 = \frac{1}{n_1} \sigma_1^2.\end{aligned}$$

På samma sätt erhåller vi

$$\text{Var}(\bar{Y}) = \frac{1}{n_2} \sigma_2^2.$$

Detta innebär att

$$\text{Var}(\hat{\mu}) = c_1^2 \frac{1}{n_1} \sigma_1^2 + (1 - c_1)^2 \frac{1}{n_2} \sigma_2^2 = c_1^2 \frac{1}{n_1} \sigma_1^2 + (1 - 2c_1 + c_1^2) \frac{1}{n_2} \sigma_2^2.$$

Om vi deriverar detta uttryck m.a.p. c_1 fås

$$\frac{d}{dc_1} \text{Var}(\hat{\mu}) = 2c_1 \frac{1}{n_1} \sigma_1^2 - 2 \frac{1}{n_2} \sigma_2^2 + 2c_1 \frac{1}{n_2} \sigma_2^2.$$

Vi sätter

$$\frac{d}{dc_1} \text{Var}(\hat{\mu}) = 0,$$

och löser m.a.p. c_1 . Vi får

$$2c_1 \frac{1}{n_1} \sigma_1^2 - 2 \frac{1}{n_2} \sigma_2^2 + 2c_1 \frac{1}{n_2} \sigma_2^2 = 0$$

och vi kan dividera bort faktorn 2, så att

$$c_1 \left(\frac{1}{n_1} \sigma_1^2 + \frac{1}{n_2} \sigma_2^2 \right) = \frac{1}{n_2} \sigma_2^2.$$

Detta ger

$$c_1 = \frac{\frac{1}{n_2} \sigma_2^2}{\frac{1}{n_1} \sigma_1^2 + \frac{1}{n_2} \sigma_2^2}.$$

Ifall $n_1 = n_2$ och $\sigma_1 = \sigma_2$ får vi

$$c_1 = \frac{1}{2}.$$

vilket känns naturligt.

$$c_1 = \frac{\frac{1}{n_2} \sigma_2^2}{\frac{1}{n_1} \sigma_1^2 + \frac{1}{n_2} \sigma_2^2}.$$