Föreläsning 6: Hypotestester (forts.)

Johan Thim (johan.thim@liu.se)

8 januari 2019

Vi fortsätter nu exkursionen i hypotesernas förlovade land. Fokus kommer vara på den vanligaste typen av hypotestester, nämligen när datan antas vara normalfördelad. Vi kommer nu åter stifta bekantskap med t- och χ^2-fördelningar.
Låt oss börja med ett enklare exempel.

Exempel

I en fabrik med mångårig erfarenhet tillverkar man material av en viss tjocklek. Man mäter med jämna mellanrum tjockleken på 9 nytillverkade material och testar om medelvärdet uppfyller $|\overline{x} - 5.0| > 0.05$. Om så är fallet stoppas tillverkningen och tekniker får gå igenom maskineriet. Ansvarig för metodutvecklingen vet av erfarenhet att $\sigma = 0.1$. Om vi antar normalfördelning, vad är bästa signifikansnivån för testet om $H_0 : \mu = 5$ testas mot $H_1 : \mu \neq 5$?

Lösning. Vi antar att $X_i \sim N(\mu, \sigma^2) = N(\mu, 0.1^2)$, $i = 1, 2, \ldots, 9$, är oberoende. För att testa H_0 ställer vi upp teststorheten

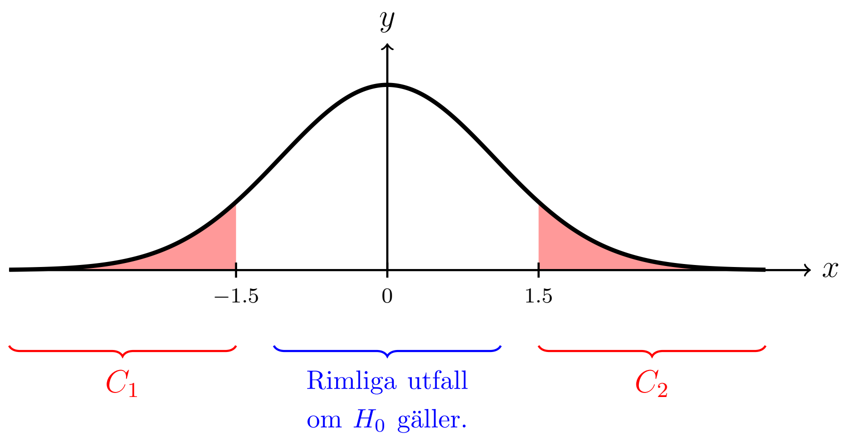
$$Z = \frac{\overline{X} - 5.0}{0.1/\sqrt{9}}.$$

Om H_0 är sann så är $Z \sim N(0, 1)$. Om vi jämför med fabrikens test så ser vi att

$$\left| \frac{0.1}{3} Z \right| > 0.05 \quad \Leftrightarrow \quad |Z| > 1.5$$

ger det kritiska området

$$C = \{z \in \mathbf{R} : |z| > 1.5\}.$$



Så signifikansnivån α kan om fördelningen ser symmetrisk ut enligt ovan beräknas enligt

$$\begin{aligned} p &= P(Z \leq -1.5) + P(Z \geq 1.5) = 2P(Z \leq -1.5) \\ &= 2\Phi(-1.5) = 2(1 - \Phi(1.5)) = 0.1336. \end{aligned}$$

Den bästa signifikansnivån vi kan välja är alltså $\alpha = 0.1336$.

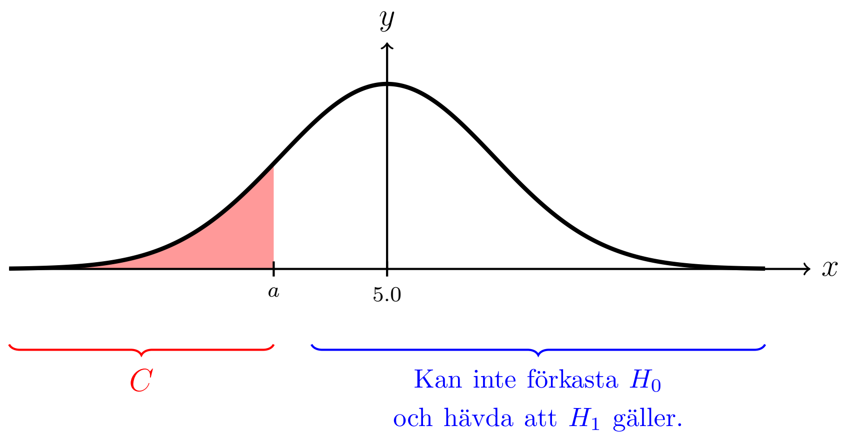
Exempel

Fabriken har fått en ny beställare som inte har något problem om materialet blir tjockare. Ansvarig tänker lite snabbt och ställer upp ett test med samma signifikansnivå för att endast testa att materialet inte blir för tunt. Hur ser testet ut nu och varför är detta antagligen inte vad man vill göra?

Lösning. Vi har fortfarande $H_0 : \mu = 5$ men mothypotesen ges nu av $H_1 : \mu < 5$. Vi kan använda samma teststorhet och om H_0 är sann så är

$$Z = \frac{\overline{X} - 5.0}{0.1/\sqrt{9}} \sim N(0, 1).$$

Vi söker en gräns a så att $\overline{X} < a$ med sannolikheten $\alpha = 0.1336$ om H_0 är sann.



Det är tydligt att

$$\overline{X} < a \quad \Leftrightarrow \quad Z = \frac{\overline{X} - 5.0}{0.1/3} < \frac{a - 5.0}{0.1/3},$$

så

$$\begin{aligned} 0.1336 = P\left(Z < \frac{a - 5.0}{0.1/3}\right) \quad &\Leftrightarrow \quad \frac{a - 5.0}{0.1/3} = \Phi^{-1}(0.1336) = -1.1095 \\ &\Leftrightarrow \quad a - 5.0 = -0.0370. \end{aligned}$$

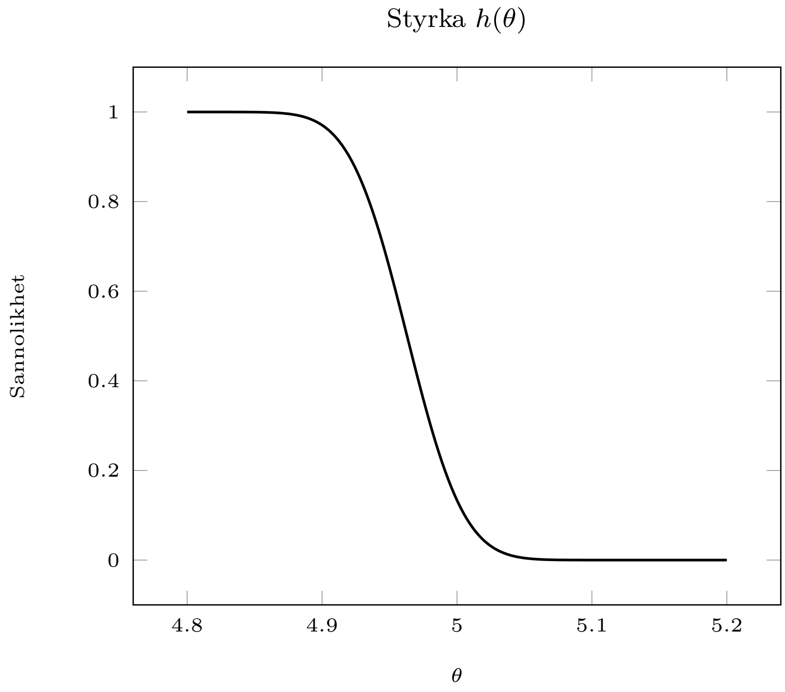
Det sökta värdet blir alltså $a = 4.9630$. Detta test blir alltså mer känsligt för att materialet är för tunnt än det föregående. Om den nya beställaren har samma tolerans för fel som de tidigare är det kanske mer strategiskt att istället sänka signifikansnivån till hälften.

Exempel

Föregående hypotestest har en lite udda signifikansnivå. Hur ser styrkefunktionen ut?

Lösning. Styrkefunktionen definieras enligt

$$\begin{aligned} h(\theta) &= P(H_0 \text{ förkastas} \mid \mu = \theta) = P\left(\overline{X} < 4.9630 \;\middle|\; \overline{X} \sim N\left(\theta, \frac{0.1^2}{9}\right)\right) \\ &= \Phi\left(\frac{4.9630 - \theta}{0.1/3}\right) = \Phi(148.89 - 30\theta). \end{aligned}$$



Exempel

En nyanställd i fabriken (med en kurs i statistisk inferens i bagaget) påtalar att det kanske är olämpligt att anta att variansen är känd och att man borde skatta den från mätningen. Vid en mätning fick man stickprovsvariansen 0.0144, vad ger testet $|\overline{x} - 5.0| > 0.05$ för signifikansnivå i denna situation?

Lösning. Vi testar således $H_0 : \mu = 5.0$ mot $H_1 : \mu \neq 5.0$ och som testvariabel blir

$$T = \frac{\overline{X} - 5.0}{S/\sqrt{9}} \sim t(8)$$

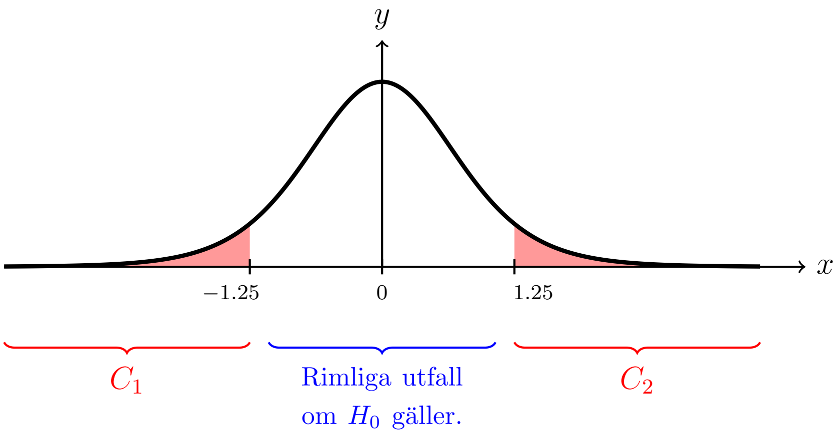
om H_0 är sann. Analogt med första exemplet måste då

$$\left|\frac{s}{3}\,t\right| > 0.05 \quad \Leftrightarrow \quad |t| > \frac{0.15}{s},$$

vilket ger det kritiska området

$$C = \left\{t \in \mathbf{R} : |t| > \frac{0.15}{s}\right\} = \{t \in \mathbf{R} : |t| > 1.25\}$$

i vårt fall. Fördelningen för T är symmetrisk lik normalfördelningen, så situationen är snarlik.



Så p-värdet kan om fördelningen ser symmetrisk ut enligt ovan beräknas enligt

$$\begin{aligned} p &= P(T \leq -1.25) + P(T \geq 1.25) = 2P(T \leq -1.25) \\ &= 2F_T(-1.25) = 2 \cdot 0.1233 = 0.2466. \end{aligned}$$

Här använde vi `tcdf`(-1.25,8) i MATLAB (vi har inga tabeller i formelsamlingen för att slå på t-fördelningar i ”den riktningen”).
Den bästa signifikansnivån vi kan välja är alltså i princip $\alpha = 0.25$. Mindre lyckat! Testet kanske behöver ändras.

2 Väntevärde för ett stickprov

I föregående samling exempel såg vi vad som hände när vi kände till variansen exakt och vad som hände när vi behövde uppskatta den från mätdatan. Osäkerheten ökar vid varje skattning, men känner vi inte exakta värden är skattningarna nödvändiga.
Låt oss undersöka den generella situationen. Vi har ett stickprov $X_1, X_2, \ldots, X_n$ från en normalfördelning $N(\mu, \sigma^2)$. Vi kan därför tänka oss att

$$X_i = \mu + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2),\ i = 1, 2, \ldots, n,$$

där ϵ_i är obereoende. Notera att samtliga variabler har samma varians. Som föregående avsnitt visade är det skillnad på när vi känner variansen exakt och när den behöver skattas.
Vi börjar med att testa

$$H_0 : \mu = \mu_0 \qquad \text{mot} \qquad H_1 : \mu \neq \mu_0.$$

Givetvis kan man vilja testa mot $H_1' : \mu > \mu_0$ eller $H_1'' : \mu < \mu_0$ också, vi kommer ta upp något sådant exempel också.

2.1 Känd varians

Eftersom

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

kan vi när σ är känd direkt använda $\overline{X}$ som teststorhet. Men för att göra det hela systematiskt och analogt med fallet då σ inte är känd skapar vi en testvariabel

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

under förutsättning att H_0 är sann. Vad vi egentligen gör är att vi utnyttjar att $\overline{X}$ är en skattning (konsistent och väntevärdesriktig) av (det okända) väntevärdet μ. Testet går ut på att se om det uppmätta värdet på skattningen sticker ut så mycket från vad som är förväntat att det gör H_0 orimlig.
Det kritiska området C ges av

$$P(Z \in C \mid H_0) = \alpha$$

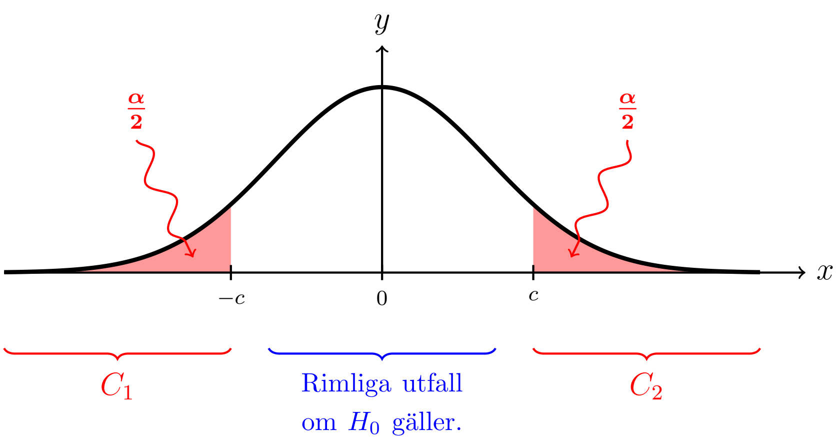
där vi av symmetriskäl (eftersom $Z \sim N(0,1)$) kan – för något $c > 0$ – uttrycka C enligt

$$C = \{z \in \mathbf{R} : |z| > c\} = \{z \in \mathbf{R} : z > c \text{ eller } z < -c\}$$

Vi noterar att C består av två delar C_1 och C_2 där talen i C_1 är negativa och talen i C_2 är positiva. Återigen, av symmetriskäl måste

$$P(Z \in C_1) = P(Z \in C_2) = \frac{\alpha}{2}.$$

Gränsen hittar vi i tabell genom att leta reda på ett tal $c = \Phi^{-1}(1 - \alpha/2)$ (sitter du med MATLAB kan du använda `c` = -norminv(alpha/2)).



2.1.1 Approximativt test via CGS

Som vi såg på förra föreläsningen kan man använda approximationer för att utföra hypotestest. Om vi i vårt fall inte vet att X_i är normalfördelad kan vi ändå via centrala gränsvärdessatsen säga att

$$\overline{X} \overset{\text{appr.}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

om $n \geq 30$ (lite beroende på hur skev fördelningen för X_i är). Som teststorhet använder vi sedan

$$\frac{\overline{X} - \mu}{s/\sqrt{n}} \overset{\text{appr.}}{\sim} N(0,1).$$

Notera att vi ersätter σ med s utan att förändra fördelningen (eftersom vi redan håller på med approximationer vet vi inte om det blir bättre med t-fördelningen). Faktum är att vi kan göra detta även om variablerna är lite beroende. Det finns flera varianter av CGS som kan hantera lite olika situationer.

2.2 Okänd varians

Om vi inte känner till σ så kan vi inte direkt använda $\overline{X}$ som teststorhet och inte heller Z från föregående stycka fungera bra (vad ska vi göra med den okända storheten σ?). Vad vi brukar göra är att ersätta σ^2 med stickprovsvariansen s^2, vilket vi tidigare visat leder till t-fördelningen. Så, då gäller att

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

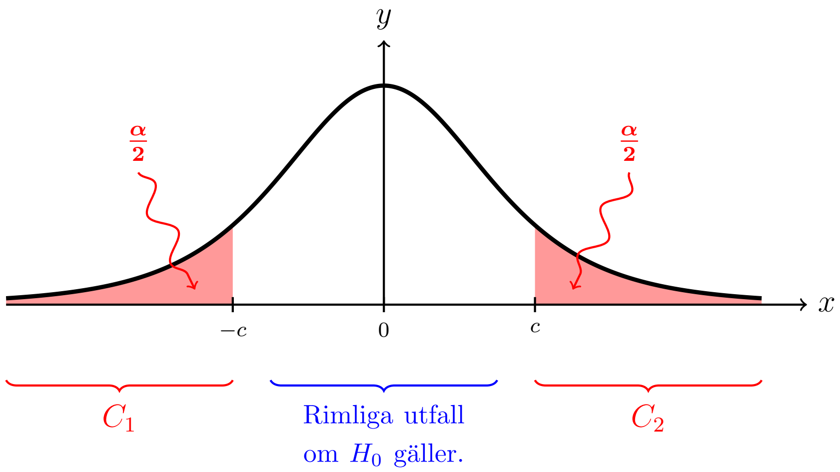
om H_0 är sann. Helt analogt med föregående situation erhåller vi nu

$$C = \{t \in \mathbf{R} : t > c \text{ eller } t < -c\}$$

där

$$P(T > c) = P(T < -c) = \frac{\alpha}{2}.$$

Gränsen hittar vi i tabell genom att leta reda på ett tal $c = F_T^{-1}(1 - \alpha/2)$ (sitter du med MATLAB kan du använda c = -tinv(alpha/2)).



2.3 Varianstest

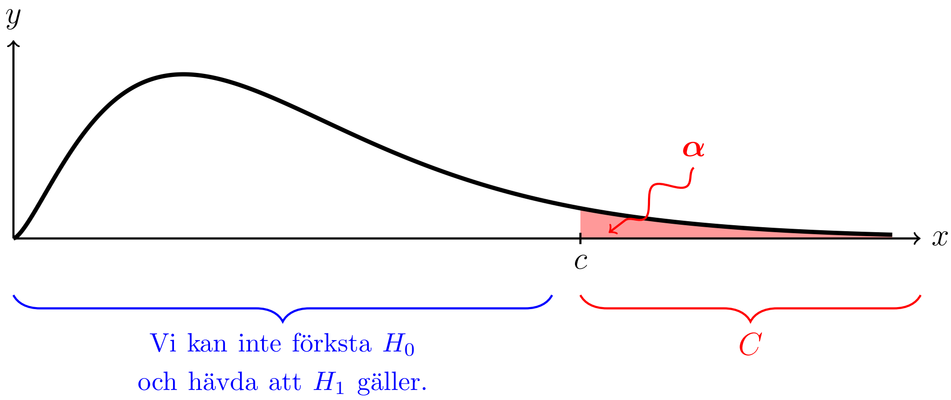
Om någon ställer upp en hypotesen $H_0 : \sigma^2 = \sigma_0^2$, kan vi utföra ett hypotestest? Givetvis kan vi det. Vi vet att

$$V = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1)$$

om H_0 är sann. Så detta är en lämplig teststorhet. Alternativt kan man även använda S^2 som teststorhet. Hur ska mothypotesen se ut? Det vanligaste brukar vara $H_1 : \sigma^2 > \sigma_0^2$ för att vi oftast endast är intresserade av att se till att variansen inte blir för stor. Men givetvis kan man även testa mot $H_1 : \sigma^2 \neq \sigma_0^2$ (eller $\sigma^2 < \sigma_0^2$ för den delen). Vi söker nu en gräns c så att

$$\alpha = P(V > c) = P\left(S^2 > \frac{c\sigma_0^2}{n-1}\right)$$

och definierar det kritiska området som $C =]c, \infty[$. Vi använder tabell för att hitta c (eller i MATLAB funktionen chi2inv(alpha,n-1)).



Exempel
Tillbaka till fabriken. Efter kritiken så vill man testa om $s = 0.12$ indikerar att det tidigare antagandet $\sigma = 0.1$ är för lågt. Utför testet med signifikansnivå 5%.

Lösning. Låt nollhypotesen vara $H_0 : \sigma = 0.1$ och mothypotesen $H_1 : \sigma > 0.1$. Vi har $n = 9$ och därmed blir

$$V = \frac{8\,S^2}{0.1^2} \sim \chi^2(8).$$

Vi hittar $c = 15.5073$ (chi2inv(0.95,8) eller ur tabell). Vi ser nu att

$$v = \frac{8 \cdot 0.12^2}{0.1^2} = 11.52 < c.$$

Vi kan alltså *inte* förkasta H_0. Betyder det att man hade rätt när fabriken sade att $\sigma = 0.1$?

3 Hypotestester och konfidensintervall

Den observante läsaren har nog redan reflekterat över att det vi ägnat oss åt är ganska snarlikt de föregående föreläsningarna om konfidensintervall. Vi ställer upp liknande storheter (ja, identiska för det mesta) men istället för att stänga in något okänt i ett intervall så testar vi skattningen mot ett kritiskt område.
Ett annat sätt att testa hypoteserna på är att ställa upp konfidensintervall och sedan testa om intervallet täcker nollhypotesen eller ej.

Exempel
I ett husvagnsdrivet laboratorie kokar Janne ihop den suspekta kemikalen $C_{10}H_{15}N$. Varje vecka startar han med samma mängd utgångsmaterial och följer samma procedur. Janne har fått en ny köpare och har hävdat att han kan producera 500 gram i veckan. För att inte riskera problem med hälsan vill Janne testa hypotesen $H_0 : \mu = 500$ mot $H_1 : \mu > 500$ på signifikansnivån 1%. Under 16 veckor producerar han i snitt 525 gram med stickprovsstandardavvikelsen 30 gram.

Lösning. Vi antar normalfördelning och ställer upp ett enkelsidigt konfidensintervall för väntevärdet μ. Låt

$$T = \frac{\overline{X} - \mu}{S/\sqrt{16}} \sim t(15).$$

Då gäller att

$$P(T < t) = 0.99$$

om $t = 2.6025$ (ur tabell). Eftersom

$$T < t \quad \Leftrightarrow \quad \frac{\overline{X} - \mu}{S/4} < t \quad \Leftrightarrow \quad \frac{S\,t}{4} > \overline{X} - \mu \quad \Leftrightarrow \quad \mu > \overline{X} - \frac{S\,t}{4}.$$

så erhåller vi konfidensintervallet

$$I_\mu = \left(\overline{x} - \frac{s\,t}{4},\ \infty\right) = (505.48,\ \infty)\,.$$

Om H_0 är sann så kommer $\mu = 500 \in I_\mu$ med sannolikheten 99%. Eftersom detta inte är sant kan vi förkasta H_0. Ska Janne sitta lugnt i båten att han inte lovat för mycket? Mäter Jannes test rätt sak?

4 Generellt om hypotestester

Innan vi avslutar med en diskussion och tester vid flera stickprov tar vi och summerar lite att tänka på.

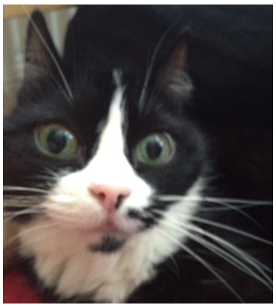
(i) Formulera hypoteser *innan* du tittar på datan. Att du vill göra ett enkelsidigt test ska inte bero på hur datan ser ut. Av den anledningen är den vanligaste typen av tester två-sidiga.

(ii) Men vissa situationer är alltid enkelsida på grund av konstruktion. Vi kommer se det i samband med regressionsanalysen där vi till exempel vet att varians minskar med fler förklaringsvariabler.

(iii) Kom ihåg när hypotestestet ställs upp att det är *mothypotesen* vi vill styrka.

(iv) Var *mycket* försiktig med tolkning av resultaten.

Att man inte förkastar H_0 betyder inte att H_0 gäller. Ett klassiskt exempel handlar om fyrbenta djur: låt H_0 : djuret har fyra ben och H_1 : djuret har inte fyra ben. Vi vill undersöka om en observation är en häst och testar H_0 mot H_1. Bara för att vi inte kan förkasta H_0 när djuret är en katt betyder det inte att det är en häst, eller hur? Kanske ett urartat exempel, det kan vara betydligt mer diffust att läsa av resultaten rätt i andra fall.



(v) Signifikansnivån kan också vara missvisande. Vid stora stickprov kan man ofta se en skillnad och förkasta H_0 även om skillnaden kanske inte spelar någon större roll i praktiska fall.

5 Flera stickprov

Vi kan givetvis betrakta flera stickprov samtidigt. Ofta är man intresserade av att testa om de har samma väntevärde och/eller samma varians. Men vi kan ställa upp test för linjärkombinationer av väntevärdena direkt.
Låt $X_1, X_2, \ldots, X_m$ och $Y_1, Y_2, \ldots, Y_n$ vara stickprov från $N(\mu_X, \sigma_X^2)$ respektive $N(\mu_Y, \sigma_Y^2)$. Vi kan ställa upp hypotestester för linjärkombinationen $c_1\mu_X + c_2\mu_Y$. Om varianserna är kända kan vi direkt använda att

$$Z = \frac{c_1\overline{X} + c_2\overline{Y} - (c_1\mu_X + c_2\mu_Y)}{\sqrt{c_1^2\sigma_X^2/m + c_2^2\sigma_Y^2/n}} \sim N(0,1),$$

alltså precis samma variabel vi såg när vi tog fram konfidensintervall för $c_1\mu_X + c_2\mu_Y$.

5.1 $\boldsymbol{\sigma_X = \sigma_Y = \sigma}$ okänd

Helt analogt med motsvarande situation när vi tog fram konfidensintervall använder vi att

$$T = \frac{c_1\overline{X} + c_2\overline{Y} - (c_1\mu_X + c_2\mu_Y)}{S\sqrt{c_1^2/m + c_2^2/n}} \sim t(m+n-2),$$

där S^2 är den sammanvägda variansskattningen. Vi betraktar ett exempel.

Exempel

Janne har fått konkurens av den före detta lärlingen Rossana som använder samma metod. Under 9 veckor producerar hon i snitt 600 gram med en stickprovsstandardavvikelse på 50 gram. Testa på signifikansnivån 1% hypotesen $H_0 : \mu_1 = \mu_2$ mot $H_1 : \mu_1 < \mu_2$ med antagandet att variansen är densamma, där μ_1 är Jannes förväntade värde och μ_2 är Rossanas. Borde köparen byta leverantör?

Lösning. Vi formulerar om enligt $H_0 : \mu_2 - \mu_1 = 0$ mot $H_1 : \mu_2 - \mu_1 > 0$. Om H_0 är sann så gäller att

$$T = \frac{\overline{Y} - \overline{X}}{S\sqrt{1/9 + 1/16}} = \frac{\overline{Y} - \overline{X}}{0.4167\,S} \sim t(23),$$

och det kritiska området blir

$$C = \{t \in \mathbf{R} : t > 2.4999\}$$

eftersom $P(T < 2.4999) = 0.99$. Med uppmätta siffrorna blir

$$t = \frac{600 - 525}{0.4167 \cdot s_p} = \frac{75}{0.4167 \cdot 38.16} = 4.7161,$$

där

$$s_p^2 = \frac{15\,s_1^2 + 8\,s_2^2}{23} = 38.16^2$$

är den sammanvägda variansskattningen. Eftersom $4.7161 \in C$ förkastar vi H_0. Blir det samma resultat om vi testar mot $H_1 : \mu_1 \neq \mu_2$? (svar: ja, förkasta H_0. Vad ändras?)

Vi kan givetvis ta fram ett konfidensintervall $I_{\mu_2-\mu_1}$ och testa nollhypotesen genom att undersöka om $0 \in I_{\mu_2-\mu_1}$ också (förkasta H_0 om $0 \notin I_{\mu_2-\mu_1}$).

5.2 Test för variansskillnad

Så med föregående avsnitt i tankarna är en rimlig fråga om vi kanske kan testa huruvida varianserna är lika eller inte. Vi gör detta med ett så kallat F-test. Vi testar hypotesen

$$H_0 : \sigma_X^2 = \sigma_Y^2 = \sigma^2$$

mot

$$H_1 : \sigma_X^2 \neq \sigma_Y^2.$$

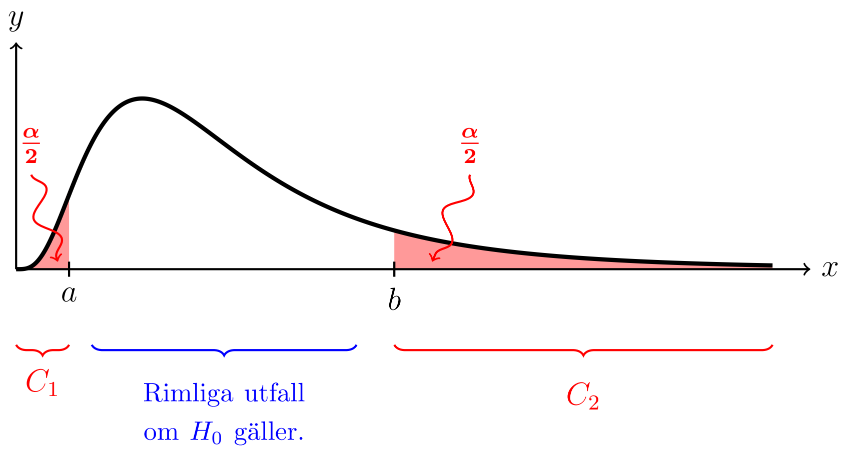
Om H_0 är sann, så gäller att $(m-1)S_X^2/\sigma^2 \sim \chi^2(m-1)$ och $(n-1)S_Y^2/\sigma^2 \sim \chi^2(n-1)$. Enligt tidigare resultat vet vi att följande då gäller:

$$V = \frac{\dfrac{(m-1)S_X^2}{\sigma^2}/(m-1)}{\dfrac{(n-1)S_Y^2}{\sigma^2}/(n-1)} = \frac{S_X^2}{S_Y^2} \sim F(m-1, n-1).$$

Vi söker nu ett kritiskt område C så att

$$\alpha = P(V \in C \mid H_0)$$

och som vi anser styrker H_1 om vi får utfall där. En figur kan vara lämplig för att se hur fördelningen ser ut.



Vi hittar gränser a och b ur tabell (eller med `finv`(p, n-1, m-1) i MATLAB) så att

$$P(V < a) = P(V > b) = \frac{\alpha}{2}.$$

Exempel

Var det rimligt att anta att Jannes och Rossanas tillvägagångssätt hade samma varians? Testa med signifikansnivån 5%.

Lösning. Med $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2$ och $H_1 : \sigma_1^2 \neq \sigma_2^2$. Med

$$V = \frac{S_1^2}{S_2^2} \sim F(15, 8)$$

hittar vi det kritiska området

$$C = \{v \in [0, \infty[: v < a \text{ eller } v > b\}$$

med $a = 0.3126$ (`finv`(0.025, 15, 8)) och $b = 4.1012$ (`finv`(0.975, 15, 8)). Eftersom

$$v = \frac{30^2}{50^2} = 0.36 \notin C$$

kan vi inte förkasta H_0. Varianserna kan vara lika (men är de det?). Vad händer på signifikansnivån 1%? (samma sak, kan man säga det utan att räkna?)