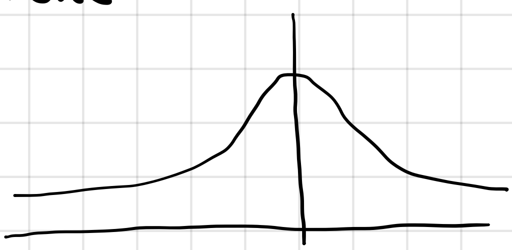
Crash course: Statistik

Sannolikhet vs statistik

- Sannolikhet är en teori och statistik är det som vi får i praktiken

Sannolikhet	Statistik
• Väntevärde μ: = det man får i genomsnitt om man gör oändligt många försök	• Oändligt många försök är inte realistiskt i praktiken • Medelvärde: $\bar{x}$ = det man får i genomsnitt om man gör ändligt många försök miniräknare: (TI-84) stat → edit → L1 → sätt in alla värde (2nd) → list → math → mean( ) → välja L1 (via stat)
• Varians σ^2 - avvikelse 	• Stickprovsvariansen s^2 $s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$ → $n-1$ istället för n för att vi ska få det väntevärdesriktigt ska bli σ^2 i genomsnitt - väljer ut några slumpmässiga stickprov, representerar alla resultat (praktiken)

• Stickprovsvarians på miniräknare:

stat → edit → skapa lista tex L1 → sätt in värde

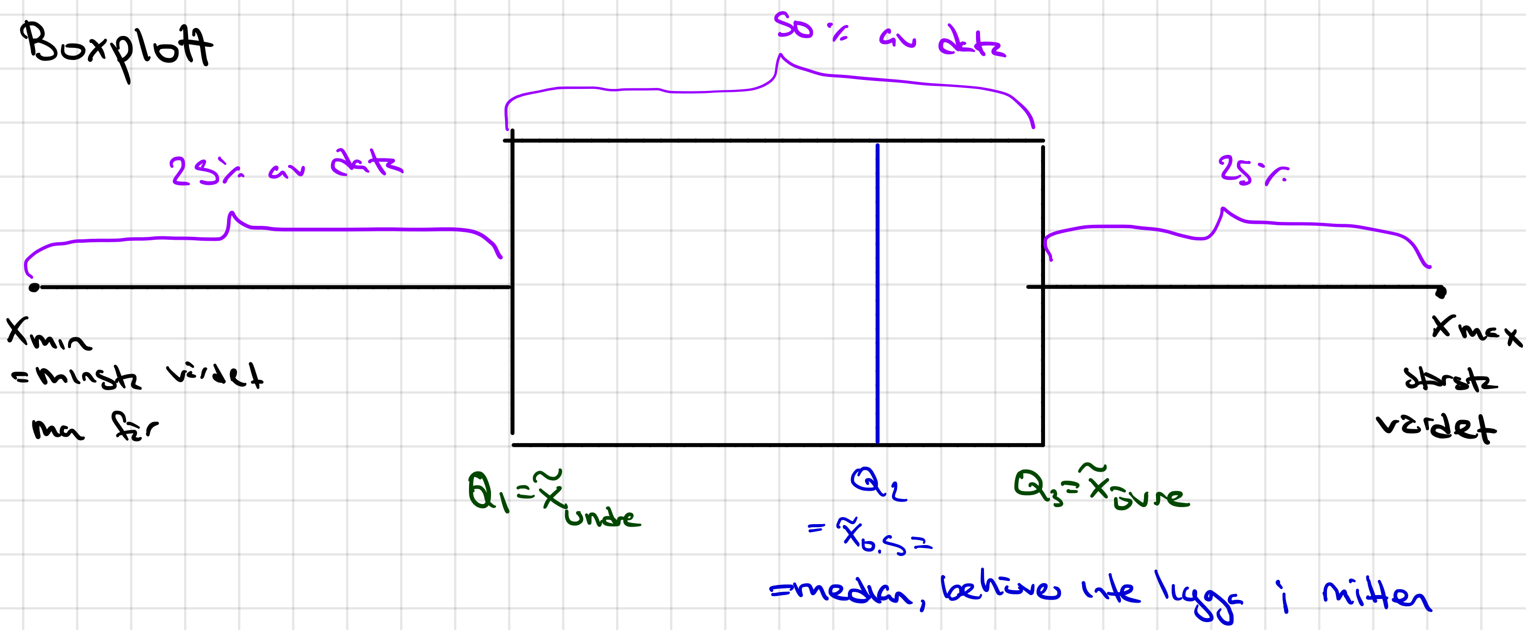
2nd → list - math → varians (för s^2) / std dev (s) } → välj rätt lista

• Statistik: populationsvarians: $= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

- Ingen osäkerhet, tex i ett samhälle då kollar man alla individer

• I alla koncept vi ska kolla på i denna kurs använd s^2

Boxplott



• Q_1 = första kvantilen, kallas även för 25%-percentilen, 25% av all data ligger till vänster

• $Q_2 = \tilde{x}_{0.5}$ = median, andra kvantilen, 50%-percentilen, 50% av data ligger till vänster

• Q_3 = tredje kvantilet, 75%-percentilen, 75% av data ligger till vänster

- (Q_1, Q_3) är kvantilintervall (50 % av data)
- $Q_3 - Q_1$ = kvantilavstånd
- (X_{min}, X_{max}) är variationsintervall

$X_{max} - X_{min}$ = variationsbredd

Hur tar man fram Q_x?

- Anta att vi söker Q_2 dvs medianen

- om vi har n st mätdata borde det vara mätdata nummer k där $\frac{k}{n} = 0{,}50$

väljer k till det heltal som uppfyller olikheten

$$\boxed{0.50\,n \leq k \leq 0.50\,n + 1}$$

- fungerar på alla sätt för alla Q_x

exempel

Anta att $n = 11$ dvs vi har 11 st data. Vad är Q_1 och Q_3?

Q_1: $0.25\,n \leq k \leq 0.25\,n + 1$

$0.25\,n = 0.25 \times 11 = \frac{11}{4} = 2.75$

$0.25\,n + 1 = 3.75$

$2.75 \leq k \leq 3.75 \Rightarrow k = 3$ är det heltal som uppfyller olikheten $\Rightarrow$

$Q_1 = X_3$

Q_1 Q_3 i växande ordning

$X_1 \quad X_2 \quad X_3 \quad X_4 \quad X_5 \quad X_6 \quad X_7 \quad X_8 \quad X_9 \quad X_{10} \quad X_{11}$

Q_3: $0.75n \leq k \leq 0.75n + 1$

$0.75 \times 11 = 8.25$
$0.75 \times 11 + 1 = 9.25$ $\Big\}$ $k = 9$ $\Rightarrow$ $Q_3 = x_9$

Vad händer om vi får flera k som uppfyller olikheten?

$n = 12$, Q_1: $0.25 \times 12 = 3$
$0.25 \times 12 + 1 = 4$ $\Big\}$ $k = 3$ och $k = 4$ båda uppfyller olikheten

x_3 och x_4 $\Rightarrow$ ta medelvärdet

$$Q_1 = \frac{x_3 + x_4}{2}$$

tenta 2020-05-26 # 6

Ange kvantilavstånd för följande mätdata

7 10 10 11 12 12 12 12 13 15 19 20 21

21 21 22 23 26 26 26 27 30 36

Lösning: kvantilavstånd: $Q_3 - Q_1$ (se i FS)

$n = 24$

Q_1: $0.25 \times 24 \leq k \leq 0.25 \times 24 + 1$

$6 \leq k \leq 7$

Q_1: $\frac{x_6 + x_7}{2} = \frac{12 + 12}{2} = 12$

Q_3: $0.75 \times 24 \leq k \leq 0.75 \times 24 + 1$

$18 \leq k \leq 19$ Q_3: $\frac{x_{18} + x_{19}}{2} = \frac{23 + 26}{2} = 24.5$

k.a: $24.5 - 12 = \underline{\underline{12.5}}$

Punkt skattning

- En skattning av θ kallas θ^*_{obs} (observerade skattningen av θ)

- θ^*_{obs} är ett utfall av den stokastiska variabeln θ^*

- θ^* - stickprovs variabel : $\theta^* = \theta(X_1, X_2 \dots X_n)$

Skattningar i vanliga fördelningar

- viktigt för konfidens intervall

- Binomial-fördelning: $X \in Bin(n, p)$: n = antal försök vi gör, känd
  x - antal gånger vi lyckas.
  p - vill skatta

$$p^*_{obs} = \frac{x}{n}$$

- Hypergeometrisk fördelning: $X \in Hyp(N, n, p)$, väldigt lik Bin,

$$p^*_{obs} = \frac{x}{n}$$

- Poisson-fördelning: $X \in Po(\mu)$. $\mu = E(X)$, väntevärde, det som behövs för att kunna räkna ut sannolikheter

  väntevärde är medelvärde i statistik:

$$\mu^*_{obs} = \bar{x}$$

- Exponential-fördelning: $X \in Exp(\lambda)$, λ okänd:

$$E(X_i) = \frac{1}{\lambda} \Rightarrow \lambda = \frac{1}{E(X_i)} \quad \lambda^*_{obs} = \frac{1}{\bar{x}}$$

• ffg-fördelning: $X \in ffg(p)$, vill skatta p. $p = \frac{1}{E(X_i)}$ $\quad p^*_{obs} = \frac{1}{\bar{x}}$

• Normalfördelning: $X \in N(\mu, \sigma)$ $\quad \mu^*_{obs} = \bar{x}$

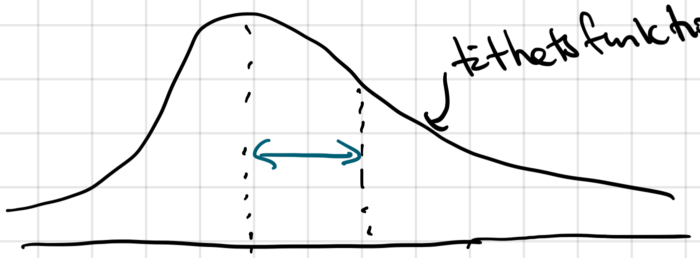
$\sigma^*_{obs} = s$

Väntevärdesriktighet

• Vill att praktiken ska vara lik teorin.

• En punktskattning θ^*_{obs} sägs vara väntevärdesriktig om tillhörande stickprovsvariabel θ^* har väntevärdet θ

$$E(\theta^*) = \theta$$



- detta är tex en icke-väntevärdesriktig skattning då det är ett stort systematiskt fel

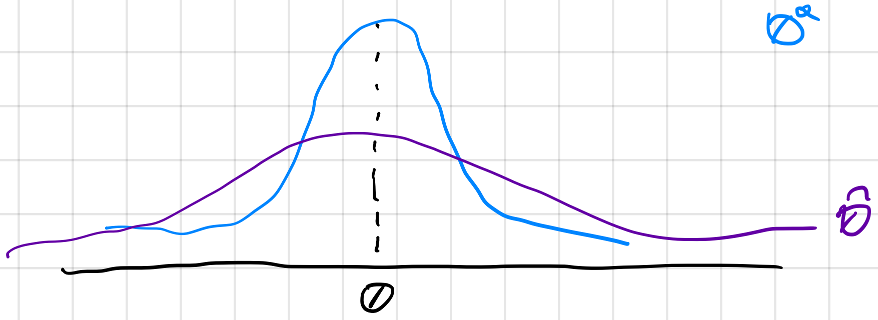
- här skiljer sig skattningens väntevärde $\theta' = E(\theta^*)$ kraftigt från θ.

Effektivitet

• Om två skattningar θ^*_{obs} och $\hat{\theta}_{obs}$ är väntevärdesriktiga och motsvarande stickprovsvariabler uppfyller

$$V(\theta^*) < V(\hat{\theta})$$ för alla $\theta \in \Omega_\theta$ så är θ^*_{obs} effektivare än $\hat{\theta}_{obs}$

- en väntevärdesriktig skattning med mindre varians är bättre



effektivitet är en jämförelse! kan säga att en skattning är mer effektiv än en annan, säger inte om skattningen är bra!

tenta 2019-12-19 # 7

Antag att X och Y är oberoende s.v. så att $E(X)=E(Y)=\theta$ och $V(X)=3\sigma^2$ och $V(Y)=2\sigma^2$. Låt x och y är observationer av X. resp. Y. Två skattningar av θ har föreslagits

$$\theta^*_{obs} = \frac{1}{2}(x+y) \quad \text{och} \quad \hat\theta_{obs} = \frac{1}{5}(2x+3y)$$

Väntevärdesriktighet? Effektivitet?

Lösning: Väntevärdesriktiga?

$$E(\theta^*) = \{=\theta ?\} = E\left[\frac{1}{2}X + \frac{1}{2}Y\right] = \{E(aX) = aE(X)\} =$$

$$= \frac{1}{2}E(X) + \frac{1}{2}E(Y) = \frac{1}{2}\theta + \frac{1}{2}\theta = \theta \quad \text{ok! vvr!}$$

$$E(\hat\theta) = E\left[\frac{2}{5}X + \frac{3}{5}Y\right] = \frac{2}{5}E(X) + \frac{3}{5}E(Y) = \frac{2}{5}\theta + \frac{3}{5}\theta = \theta \quad \text{ok!}$$

- Båda skattningar är väntevärdesriktiga. kollar variansen för att avgöra vilken skattning som är mer effektiv:

$$\underline{\underline{V(\theta^*)}} = V\left[\frac{1}{2}X + \frac{1}{2}Y\right] = \{V(aX) = a^2V(X)\} = \frac{1}{4}V(X) + \frac{1}{4}V(Y) =$$

$$= \frac{1}{4}3\sigma^2 + \frac{1}{4}2\sigma^2 = \frac{5}{4}\sigma^2 = \underline{\underline{1.25\sigma^2}}$$

$$V(\hat{\theta}) = V\left[\frac{2}{5}X + \frac{3}{5}Y\right] = \frac{4}{25}V(X) + \frac{9}{25}V(Y) =$$

$$= \frac{4}{25} \times 3\sigma^2 + \frac{9}{25} \times 2\sigma^2 = \frac{30}{25}\sigma^2 = 1.2\sigma^2$$

$V(\hat{\theta}) < V(\theta^*) \Rightarrow \hat{\theta}_{obs}$ är mer effektiv än θ^*_{obs}

Maximum likelihood-metoden (ML)

- bygger på att vi kan (eller antar att vi kan) sannolikhetsfunktionen eller täthetsfunktionen.

- eftersom vi har fått det mätdata man fick så bör sannolikheten vara stor att få just de värdena

$\Rightarrow$ maximerar sannolikheten.

. Det värde $\theta_{obs\,ML}$ för vilket $L(\theta)$ antar sitt största värde inom Ω_θ kallas för ML-skattningen av θ.

om oberoende: diskret: $L(\theta) = p_{X_1}(x_1;\theta) \cdot p_{X_2}(x_2;\theta) \cdots p_{X_n}(x_n;\theta)$

kontinuerliga fallet: $L(\theta) = f_{X_1}(x_1;\theta)\, f_{X_2}(x_2;\theta) \cdots f_{X_n}(x_n;\theta)$

går ut på att hitta maximum till funktionen:

$$\frac{dL(\theta)}{d\theta} = 0$$

- om en jobbig funktion $\Rightarrow$ logaritmera först

Minsta-kvadrat metoden (MK)

- Går ut på att minimera skillnaden mellan praktiska och teoretiska värdena.

$$Q = \sum_{i=1}^{n} (x_i - \mu_i(\theta_1 \theta_2 \dots \theta_n))^n$$

$x_1 \dots x_n$ är oberoende och $\theta_1 \dots \theta_n$ är okända parametrar.

går ut på att hitta minimum till funktionen

$$\frac{dQ}{d\theta} = 0$$

- om Q beror på flera okända parametrar så kollar man på partiella derivator dvs tar gradienten till funktionen.

tenta 2020-03-10 #7

Låt X_1, X_2 och X_3 vara tre oberoende lika fördelade s.v. sådana att $X_i \in Po(\mu)$. Beräkna ML-skattningen av μ då $x_1 = 4$, $x_2 = 10$ och $x_3 = 1$

Lösning: $X_i \in Po(\mu)$ $\begin{cases} x_1 = 4 = k_1 \\ x_2 = 10 = k_2 \\ x_3 = 1 = k_3 \end{cases}$

ML: $L(\mu) = \{\text{oberoende}\} = p_{X_1}(x_1; \mu)\, p_{X_2}(x_2; \mu)\, p_{X_3}(x_3; \mu)$

$$p_X(k) = \frac{\mu^k}{k!} e^{-\mu}$$

$$L(\mu) = \frac{\mu^4}{4!} e^{-\mu} \frac{\mu^{10}}{10!} e^{-\mu} \frac{\mu^1}{1!} e^{-\mu} = \frac{e^{-3\mu} \cdot \mu^{15}}{4! \cdot 10! \cdot 1!}$$

-hitter max genom att kolla på kritiska punkter till $L(\mu)$.

-derivatan kräver kedje och produkt regeln, har stora konstanter, stora potenser, enklare om vi logaritmerar först då max av $L(\mu)$ är samma som max av $\ln[L(\mu)]$.

$$\ln(ab) = \ln(a) + \ln(b)$$

$$\ln[L(\mu)] = \ln\left[\frac{\mu^{15} e^{-3\mu}}{10!4!}\right] = \ln(\mu^{15}) + \ln(e^{-3\mu}) - \ln(10!4!) =$$

$$= 15\ln(\mu) - 3\mu - \ln(10!4!)$$

max $\frac{d\ln[L(\mu)]}{d\mu} = 0$

$$\ln[L(\mu)]' = \frac{15}{\mu} - 3 = 0 \iff \frac{15}{\mu} = 3 \Rightarrow \mu_{obs,ML} = 5$$

tenta 2019-06-05

(9) Antag att vi har två oberoende observationer av x_1 och x_2 av $X_i \in Bin(n_i, p)$, $i = 1,2$. Ange ML skattningen av p.

$$L(p) = \binom{n_1}{x_1} p^{x_1} (1-p)^{n_1-x_1} \binom{n_2}{x_2} p^{x_2} (1-p)^{n_2-x_2}$$

eftersom vi har oberoende observationer och $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$

$$L(p) = \frac{n_1!\; n_2!}{(n_1-x_1)!\, x_1!\, (n_2-x_2)!\, x_2!}\; p^{x_1+x_2}\; (1-p)^{n_1+n_2-x_1-x_2}$$

n_1, n_2, x_1, x_2 är konstanter

-jobbig funktion, ta ln

$$L(p) = c\, p^{x_1+x_2} (1-p)^{n_1+n_2-x_1-x_2} \quad \text{där } c = \frac{n_1!\; n_2!}{(n_1-x_1)!\, x_1!\, (n_2-x_2)!\, x_2!}$$

$$\ln[L(p)] = \ln\left[c\, p^{x_1+x_2} (1-p)^{n_1+n_2-x_1-x_2}\right] =$$

$$= \ln(c) + \ln\left(p^{x_1+x_2}\right) + \ln\left[(1-p)^{n_1+n_2-x_1-x_2}\right] =$$

$$= \ln(c) + (x_1+x_2)\ln p + (n_1+n_2-x_1-x_2)\ln(1-p)$$

$$\ln[L(p)]' = \frac{x_1+x_2}{p} + \frac{n_1+n_2-x_1-x_2}{1-p}(-1) = 0$$

$$\frac{x_1+x_2}{p} = \frac{n_1+n_2-x_1-x_2}{1-p} \Rightarrow (x_1+x_2)(1-p) = p(n_1+n_2-x_1-x_2)$$

$$x_1+x_2-x_1p-x_2p = pn_1+pn_2-px_1-x_2p$$

$$p^{*}_{\text{obs m.}} = \frac{x_1+x_2}{n_1+n_2}$$

(10) Antag att vi har N st s.v. X_i som är oberoende och tillhör Γ-fördelningen $\Gamma(k, \theta)$

$$E(X_i) = \frac{k}{\theta} \quad \text{och} \quad V(X_i) = \frac{k}{\theta^2}$$

ML skattningen av θ är $\theta^*_{obs\ ML} = \frac{\sum^N X_i}{kN}$. Vad är variansen av denna skattning?

Lösning: $V(\theta^*) = V\left[\frac{\sum X_i}{kN}\right] = \frac{1}{k^2N^2} V(\sum X_i) = \frac{1}{k^2N^2} N V(X_i) =$

$$= \frac{1}{k^2 N} \cdot \frac{k}{\theta^2} = \frac{1}{kN\theta^2}$$

tenta 2020-01-08 #7

Låt x_i, $i = 1,2,3$ vara observationer av de oberoende s.v. X_i, $i=1,2,3$. De är gamma fördelade $\Gamma(2, \lambda)$ dvs har täthets funktionen:

$$f_{X_i}(x) \begin{cases} f_{X_i}(x) = \frac{\lambda^2}{1!} x e^{-\lambda x} & \text{om } x > 0,\ i = 1,2,3 \\ 0 & \text{om } x \leq 0 \end{cases}$$

för $\lambda > 0$. Detta medför att $E(X_i) = \frac{2}{\lambda}$ och $V(X_i) = \frac{2}{\lambda^2}$ $i = 1,2,3$

Antag att följande värden har erhållits: $x_1 = 1.4075$
$x_2 = 0.7142$
$x_3 = 0.4004$

Bestäm MK-skattningen av λ

Lösning: $Q = \sum_{i=1}^{n} (x_i - \mu_i(\lambda))^2$

$$\mu_i(\lambda) = \frac{2}{\lambda}$$

$$Q = \left(1.4075 - \frac{2}{\lambda}\right)^2 + \left(0.7142 - \frac{2}{\lambda}\right)^2 + \left(0.4004 - \frac{2}{\lambda}\right)^2$$

$\frac{dQ}{d\lambda} = 0$ ger min:

$$Q' = 2\left(1.4075 - \frac{2}{\lambda}\right)\frac{2}{\lambda^2} + 2\left(0.7142 - \frac{2}{\lambda}\right)\left(\frac{2}{\lambda^2}\right) +$$

$$+ 2\left(0.4004 - \frac{2}{\lambda}\right)\left(\frac{2}{\lambda^2}\right) =$$

$$= \frac{4}{\lambda^2}\left[1.4075 - \frac{2}{\lambda} + 0.7142 - \frac{2}{\lambda} + 0.4004 - \frac{2}{\lambda}\right] = 0$$

$\lambda > 0$, kan förenkla:

$$1.4075 + 0.7142 + 0.4004 = \frac{2}{\lambda} + \frac{2}{\lambda} + \frac{2}{\lambda}$$

$$2.37897 = \frac{6}{\lambda} \Rightarrow \lambda^*_{\text{obs mk}} = \frac{6}{2.37897} = \underline{\underline{2.379}}$$

tenta 2020-05-26 # 13 (10p)

För att undersöka en bilmodellens bromseffekt mättes bromssträckan efter inbromsning vid hastigheterna 20, 40, 60, 80 och 100 km/h.

Resultat

hastighet (v)	20	40	60	80	100
bromssträcka (x) (m)	12	25	52	78	119

Rent teoretiskt skulle bromssträckan vara en kvadratisk funktion av hastigheten plus en slumpmässig normalfördelad avvikelse och därför antas bromssträckan vid inbromsning vid hastigheten v vara $N(\gamma (\frac{v}{20})^2, \sigma^2)$ där γ är okänd.

Skatta γ med minsta-kvadrat metoden.

Lösning: MK: $Q = \sum_{i=1}^{n} (x_i - \mu_i(\theta))^2$

vårt fall: $\mu(\gamma) = E(X) = \gamma (\frac{v}{20})^2$

X_i är olika bromssträckor:

$X_1 = 12 \qquad \mu_1 = \gamma (\frac{20}{20})^2 = \gamma$

$X_2 = 25 \qquad \mu_2 = \gamma (\frac{40}{20})^2 = 4\gamma$

$X_3 = 52 \qquad \mu_3 = \gamma (\frac{60}{20})^2 = 9\gamma$

$X_4 = 78 \qquad \mu_4 = \gamma (\frac{80}{20})^2 = 16\gamma$

$X_5 = 119 \qquad \mu_5 = \gamma (\frac{100}{20})^2 = 25\gamma$

$$Q=(12-\gamma)^2+(25-4\gamma)^2+(52-9\gamma)^2+(78-16\gamma)^2+$$
$$+(119-25\gamma)^2$$

min av Q: $\frac{dQ}{d\gamma}=0$

$$Q'=2(12-\gamma)(-1)+2(25-4\gamma)(-4)+2(52-9\gamma)(-9)+$$
$$+2(78-16\gamma)(-16)+2(119-25\gamma)(-25)=$$
$$=-2(12-\gamma)-8(25-4\gamma)-18(52-9\gamma)-32(78-16\gamma)$$
$$-50(119-25\gamma)\overset{!}{=}0$$

$$-24+2\gamma-200+32\gamma-936+162\gamma-2496+512\gamma$$
$$-5950+1250\gamma=0$$

$$-9606+1958\gamma=0 \quad \Rightarrow \gamma=4.91$$

$$\gamma^*_{obs,\,mk}=4.91$$

Medelfel

- En skattning av $D(\theta^*)$ kallas medelfelet för θ^* och betecknas $d(\theta^*)$ (en siffra)
- Medelfelet för skattningen θ är alltså $D^*(\theta^*)_{obs}$

observerade skattningen av standard avvikelsen för skattningen av θ.

exempel

Antag att $X \in N(\mu, \sigma)$ och vi har $\mu^*_{obs} = \bar{x}$. Vad är medelfelet för denna skattning?

- vi söker är $D^*(\mu^*)_{obs}$

$$\mu^* = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$D(\mu^*) = D(\bar{X}) = \sqrt{V(\bar{X})}$$

$$V(\bar{X}) = V\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2}\, nV(X_i) = \frac{V(X_i)}{n} = \frac{\sigma^2}{n}$$

$$D(\bar{X}) = \sqrt{V(\bar{X})} = \frac{\sigma}{\sqrt{n}}$$

medelfel

$$D^*(\mu^*)_{obs} = \left(\frac{\sigma}{\sqrt{n}}\right)^*_{obs} = \frac{\sigma^*_{obs}}{\sqrt{n}} = \frac{s}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

exempel

Antag att $X \in Bin(n, p)$ $\quad p^*_{obs} = \frac{x}{n}$. Vad är $d(p^*)$?

$$d(p^*) = D^*(p^*)_{obs}$$

$$D(p^*) = D\left(\frac{X}{n}\right) = \sqrt{V\left(\frac{X}{n}\right)}$$

$$V\left(\frac{X}{n}\right) = \frac{1}{n^2}V(X) = \{Bin\} = \frac{1}{n^2}\, np(1-p) = \frac{p(1-p)}{n}$$

$$D(p^*) = \sqrt{\frac{p(1-p)}{n}}$$

medelfelet: $D^*(p^*)_{obs} = \left(\sqrt{\frac{p(1-p)}{n}}\right)^*_{obs} = \sqrt{\frac{p^*_{obs}(1-p^*_{obs})}{n}} =$

$= \left\{ p^*_{obs} = \frac{x}{n} \right\} = \sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}$

exempel

$X \in Po(\mu)$. Vad är medelfelet av μ^*?

vi vet att $\mu^*_{obs} = \bar{x}$

$D(\mu^*) = \sqrt{V(\mu^*)}$

$V(\mu^*) = V(\bar{X}) = V\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2} nV(X_i) = \frac{\mu}{n}$

$\Rightarrow D(\mu^*) = \sqrt{\frac{\mu}{n}}$

$D^*(\mu^*)_{obs} = \left(\sqrt{\frac{\mu}{n}}\right)^*_{obs} = \sqrt{\frac{\mu^*_{obs}}{n}} = \sqrt{\frac{\bar{x}}{n}}$

tenta 2018-10-24 #8

Låt $x = 38$ vara ett utfall av en s.v. $X \in Bin(600, p)$ där p är okänd. Bestäm medelfelet för skattningen $p^* = \frac{x}{n}$

$p^*_{obs} = \frac{x}{n} = \frac{38}{600}$

$V(p^*) = V\left(\frac{X}{n}\right) = \frac{1}{n^2} V(X) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$

$D(p^*) = \sqrt{\frac{p(1-p)}{n}} \Rightarrow D^*(p^*)_{obs} = \sqrt{\frac{p^*_{obs}(1-p^*_{obs})}{n}}$

$$\frac{X}{n} = p_{obs} = \frac{38}{600} \Rightarrow D^*_{obs} = \sqrt{\frac{\frac{38}{600}\left(1-\frac{38}{600}\right)}{600}} = 0.00994$$

tenta 2013-12-20 # 11

Låt X_i, $i=1,2,3$ vara oberoende s.v. med väntevärde μ och standardavvikelse σ. Anta att μ skattas med

$$\mu^*_{obs} = \frac{(x_1+x_2+2x_3)}{4}$$

och σ skattas med s. Låt utfallen på X_i vara $\begin{cases} x_1 = 5 \\ x_2 = 9 \\ x_3 = 7 \end{cases}$

Bestäm medelfelet för skattningen av μ.

Lösning: $V(\mu^*) = V\left(\frac{X_1+X_2+2X_3}{4}\right) = \frac{1}{16}\left[V(X_1) + V(X_2) + 4V(X_3)\right] =$

$$= \frac{6}{16} V(X_i) = \frac{6}{16}\sigma^2 = \frac{3}{8}\sigma^2$$

$$D(\mu^*) = \sqrt{\frac{3}{8}\sigma^2} = \sqrt{\frac{3}{8}}\sigma \Rightarrow D^*_{obs} = \sqrt{\frac{3}{2}}\, s.$$

måste hitta s:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

måste hitta $\bar{x}$ först. $\bar{x} = \frac{x_1+x_2+x_3}{3} = \frac{5+9+7}{3} = 7$

$$s^2 = \frac{1}{2}\left[(5-7)^2 + (9-7)^2 + (7-7)^2\right] = \frac{1}{2}(4+4+0) = 4$$

$$s = \sqrt{4} = 2. \rightarrow D^*_{obs} = \sqrt{\frac{3}{2}}\,2 = 1.22.$$

tenta 2019-03-12 #7

Antag att $X \in N(\mu, \sigma)$ och $Y \in N(2\mu, 2\sigma)$ och de s.v. X och Y är oberoende. Vi skattar μ med

$$\mu^*_{obs} = \frac{x}{2} + \frac{y}{4} \quad \text{och} \quad \sigma \text{ med}$$

$$\sigma^*_{obs} = \sqrt{(x - \mu^*_{obs})^2 + \left(\frac{y}{2} - \mu^*_{obs}\right)^2}$$

Vi har vidare fått de observerade värdena $x = 7$ och $y = 15$. Beräkna medelfelet av μ^* utgående från denna information.

$$\mu^*_{obs} = \frac{x}{2} + \frac{y}{4} = \frac{7}{2} + \frac{15}{4} = \frac{29}{4}$$

$$V(\mu^*) = V\left[\frac{X}{2} + \frac{Y}{4}\right] = \frac{1}{4}V(X) + \frac{1}{16}V(Y) = \left\{\begin{matrix} V(X) = \sigma^2 \\ V(Y) = (2\sigma)^2 \end{matrix}\right\} =$$

$$= \frac{1}{4}\sigma^2 + \frac{1}{16}(2\sigma)^2 = \frac{\sigma^2}{4} + \frac{4\sigma^2}{16} = \frac{\sigma^2}{2} \Rightarrow D(\mu^*) = \frac{\sigma}{\sqrt{2}}$$

medelfel: $D^*(\mu^*)_{obs} = \dfrac{\sigma^*_{obs}}{\sqrt{2}}$

$$\sigma^*_{obs} = \sqrt{(x - \mu^*_{obs})^2 + \left(\frac{y}{2} - \mu^*_{obs}\right)^2} = \sqrt{\left(7 - \frac{29}{4}\right)^2 + \left(\frac{15}{2} - \frac{29}{4}\right)^2} =$$

$$= 0.3536 \quad \Rightarrow D^*_{obs} = \frac{0.3536}{\sqrt{2}} = \frac{1}{4} = 0.25$$

tenta 2019-01-08 #11

Två undersökningar gjordes på två grupper av patienter, som hade blivit vaccinerade respektive inte vaccinerade. 500 vaccinerade patienter undersöktes där 49 blev sjuka. I den icke-vaccinerade gruppen var 600 undersökta där 58 blev sjuka. Låt p_1 vara andelen sjuka och vaccinerade och p_2 andelen sjuka och inte vaccinerade. Vad är medelfelet för skattningen av $p_1 - p_2$?

X_1 - vaccinerade och sjuka: $X_1 \sim Bin(500, p_1)$

$X_2 \sim Bin(600, p_2)$

$p^*_{1\,obs} = \frac{X_1}{n_1} = \frac{49}{500}$, $p^*_{2\,obs} = \frac{X_2}{n_2} = \frac{58}{600}$

$$V(p_1^* - p_2^*) = V\left(\frac{X_1}{n_1} - \frac{X_2}{n_2}\right) = \frac{1}{n_1^2} V(X_1) + \frac{1}{n_2^2} V(X_2) = \{Bin\text{-fördelning}\} =$$

$$= \frac{1}{n_1^2} n_1 p_1 (1-p_1) + \frac{1}{n_2^2} n_2 p_2 (1-p_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

$$D(p_1^* - p_2^*) = \sqrt{V(p_1^* - p_2^*)} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

$$D^*_{obs} = \sqrt{\frac{p^*_{1\,obs}(1-p^*_{1\,obs})}{n_1} + \frac{p^*_{2\,obs}(1-p^*_{2\,obs})}{n_2}} =$$

$$= \sqrt{\frac{\frac{49}{500}\left(1-\frac{49}{500}\right)}{500} + \frac{\frac{58}{600}\left(1-\frac{58}{600}\right)}{600}} = 0.018$$

Sammanfattning för medelfel

1. Hitta $V(\theta^*)$

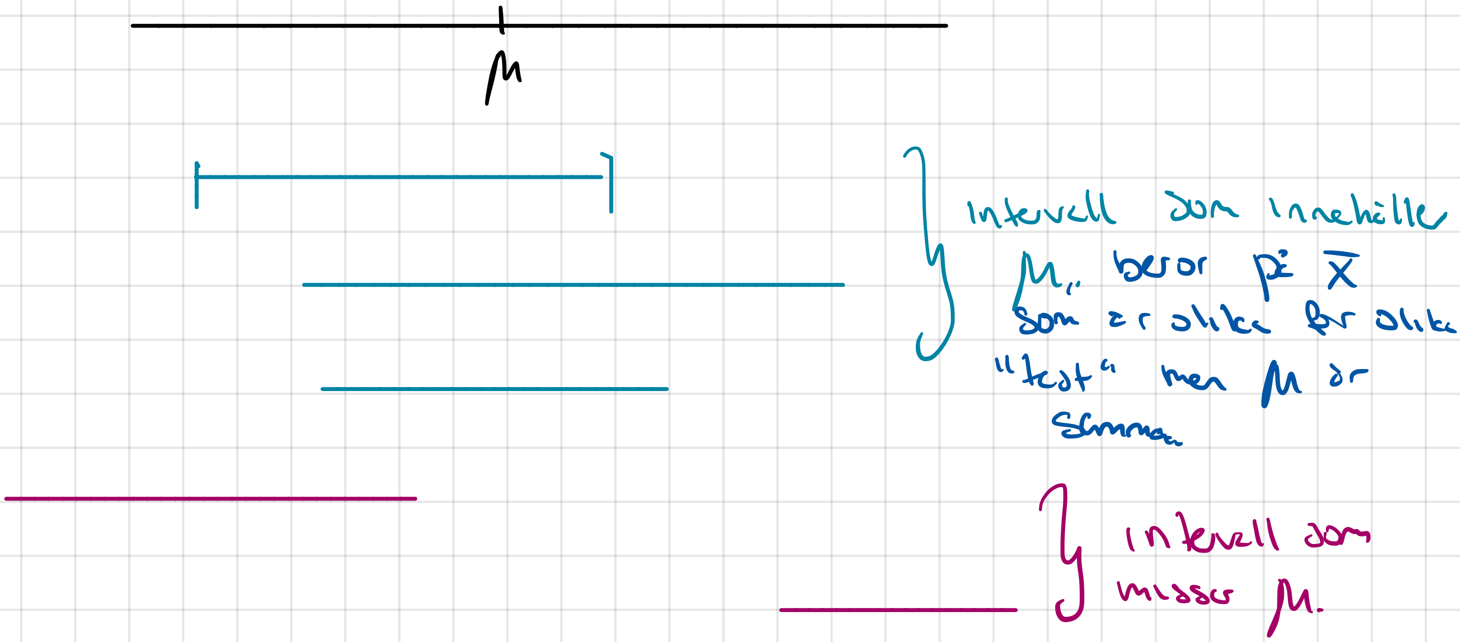
2. Hitta $D(\theta^*) = \sqrt{V(\theta^*)}$

3. Skatta $D(\theta^*) \Rightarrow D^*(\theta^*)_{obs}$

Konfidensintervall (konf-int)

- Definition: Ett intervall I_θ som med sannolikheten $1-\alpha$ täcker över θ kallas för ett konfidensintervall för θ med konfidensgraden $1-\alpha$

$$P(\theta \in I_\theta) = 1-\alpha$$



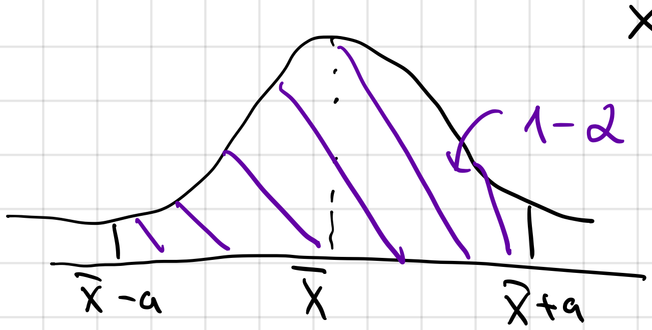
Förklaring till formulan i formelbladet §12.1

· Anta att X_i:na är oberoende och $X_i \in N(\mu, \sigma)$ med σ känd. Gör n st mätningar och vill bilda ett konf-int av graden $1-\alpha$ för μ.

- Skatta μ med $\bar{x}$ $\rightarrow \hat{\mu}_{obs} = \bar{x}$

$X_i \in N(\mu, \sigma) \Rightarrow \bar{X}$ också normalfördelad:

$$\bar{X} \in N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$



Vi vill hitta a.

$$P(\mu \text{ finns i intervallet}) = 1-\alpha \Leftrightarrow P(|\bar{X}-\mu| < a) = 1-\alpha$$

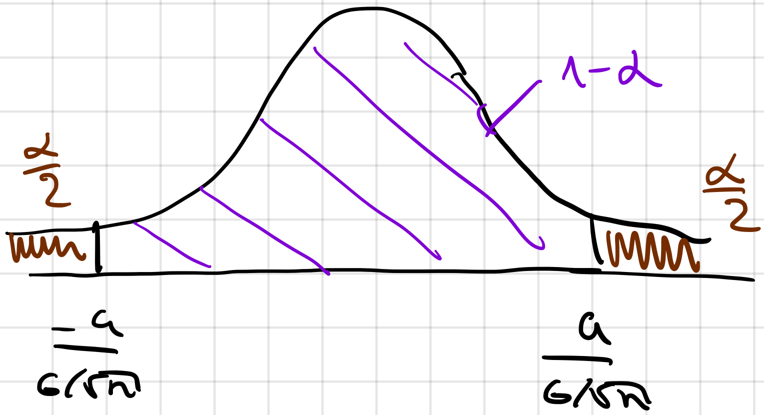
$$P(-a < \bar{X}-\mu < a) = 1-\alpha$$

$$P(\mu - a < \bar{X} < \mu + a) = 1-\alpha$$

gör om till $N(0,1)$ för att kunna använda tabell 2:

$$P\left[\frac{\mu-a-\mu}{\sigma/\sqrt{n}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{\mu+a-\mu}{\sigma/\sqrt{n}}\right] = 1-\alpha$$

$$P\left(\frac{-a}{\sigma/\sqrt{n}} < Y < \frac{a}{\sigma/\sqrt{n}}\right) = 1-\alpha \quad , Y \in N(0,1)$$



Vill använda tabell 2: $P\left(Y > \frac{a}{\sigma/\sqrt{n}}\right) = \frac{\alpha}{2}$

$$\frac{a}{\sigma/\sqrt{n}} = \lambda_{\frac{\alpha}{2}} \Rightarrow a = \frac{\sigma}{\sqrt{n}} \lambda_{\frac{\alpha}{2}}$$

$$I_\mu = \left(\bar{x} - \lambda_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + \lambda_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

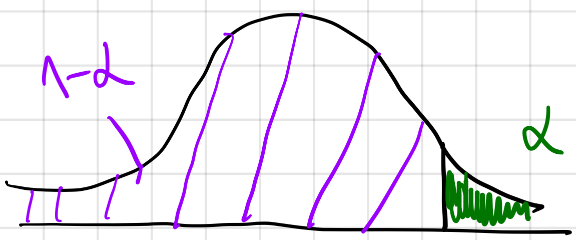
$$\Rightarrow \boxed{I_\mu = \bar{x} \pm \frac{\sigma}{\sqrt{n}} \lambda_{\frac{\alpha}{2}}}$$

ett stickprov, 2sidigt konf-int för väntevärdet

FS: §12.1 används när man söker konf-int för väntevärde när σ är känd.

Om vi har ensidigt intervall då?

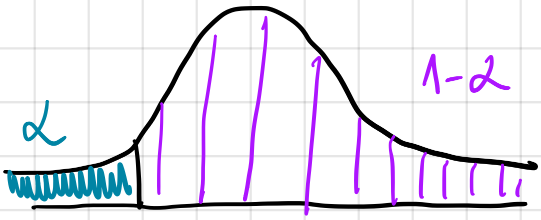
(1) Ensidigt uppåt begränsad konfidensintervall



$$I_\mu = \left(-\infty, \bar{x} + \frac{\sigma}{\sqrt{n}} \lambda_{\alpha}\right)$$

ensidigt !!!!!

② Ensidigt nedåt begränsad konfidensintervall:



$$I_\mu = \left(\bar{x} - \frac{\sigma}{\sqrt{n}}\lambda_\alpha\,,\ +\infty\right)$$

notera: $\boxed{\begin{array}{ll} \lambda_\alpha & \text{i ensidiga intervall} \\ \lambda_{\frac{\alpha}{2}} & \text{i tvåsidiga intervall} \end{array}}$

Konfidensintervallets bredd beror på:

1. En stor spridning hos mätdata (stor σ) ger bredare intervall
2. Ju fler observationer (stort n), ju smalare intervall
3. Ju större säkerhet vi vill ha, ju bredare intervall ($\lambda_{\frac{\alpha}{2}}$)

Konfidensintervall med σ okänd

- Måste skatta σ med s, $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sigma^*_{obs}$

har då: $P\left(\frac{-a}{S/\sqrt{n}} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < \frac{a}{S/\sqrt{n}}\right) = 1-\alpha$

- S är inte konstant, utan beror på flera s.v.

⇒ använder t-fördelningen

precis som $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \in N(0,1)$ så gäller att $\frac{\bar{X}-\mu}{S/\sqrt{n}} \in t(n-1)$

§ 11.1 d

t-fördelningen finns i tabell 3. tar n-1 istället för n, fördelningen är annars ganska lik normalfördelningen:

$$\boxed{I_\mu = \bar{x} \pm \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1)}$$

i §12.2 - används när σ okänd och vill ha konfidensintervall för μ.

$f = n-1$, frihetsgrader.

tenta 2020-04-16 #8

Antag att vi vill skatta väntevärdet μ för en normalfördelning med okänd standardavvikelse och tar därför ett slumpmässigt stickprov $x_1\ x_2 \cdots x_n$ av storlek $n=15$ av denna fördelning. Har $\bar{x} = 20$ och $s=5$. Ange ett 99% konfidensintervall för μ.

- både μ och σ är okända $\Rightarrow$ §12.2 i FS

$$I_\mu = \bar{x} \pm D^*_{obs}\, t_{\frac{\alpha}{2}}(n-1)$$

$$f = n-1 = 15-1 = 14$$

$$D^*_{obs} = \frac{s}{\sqrt{n}} = \frac{5}{\sqrt{15}} = 1.291$$

tvåsidigt intervall: $\alpha = 0.01 \Rightarrow \frac{\alpha}{2} = 0.005$

$$\Rightarrow t_{0.005}(14) = 2.98$$

$$I_\mu = 20 \pm 1.291 \times 2.98 = (16.15,\ 23.85)$$

tenta 2019-03-12 # 9

Vi mäter syrekoncentrationen i vatten och resultatet anses vara ett utfall av oberoende s.v. som följer $N(\mu, \sigma)$ där σ betecknar standardavvikelsen. Summan av de 31 mätvärden är 69.7 och stickprovsvariansen är 1.2772. Ange den övre gränsen till det ensidigt uppåt begränsade 95% -konfidensintervallet för μ.

μ och σ är okända $\Rightarrow$ §12.2

ensidigt (uppåt begränsad) $\Rightarrow$ $I_\mu = (-\infty, \bar{x} + D^{\alpha}_{obs}\, t_{\alpha}(n-1))$

ensidigt. Inte $\frac{\alpha}{2}$!

$$\bar{x} = \frac{69.7}{31} = 2.248$$

$$D^{\alpha}_{obs} = \frac{s}{\sqrt{n}} = \sqrt{\frac{1.2772}{31}} = 0.203$$

$$1-\alpha = 0.95 \Rightarrow \alpha = 0.05$$

$$t_{0.05}(31-1) = t_{0.05}(30) = 1.7$$

$$I_\mu = (-\infty, \overbrace{2.248 + 0.203 \times 1.7}^{\text{övre gräns}})$$

övre gränsen: 2.59

Summanfattning (I_μ)

- Om σ känd $\Rightarrow$ §12.1

  ensidigt: α

- Om σ okänd $\Rightarrow$ §12.2

  om tvåsidigt: $\frac{\alpha}{2}$

Skillnad mellan 2 stickprov

- Anta att vi har X_i:na som är oberoende och $N(\mu_x, \sigma_x)$ fördelade, samt vi har Y_i:na som är oberoende och $N(\mu_y, \sigma_y)$ fördelade

- Vi vill ta fram $I_{\mu_x - \mu_y}$, har vi då tre fall:

(1) σ_x och σ_y är kända

- använder §12.1 och §11.3

$$D(\bar{X} - \bar{Y}) = \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$$

$$\Rightarrow I_{\mu_x - \mu_y} = (\bar{x} - \bar{y}) \pm \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \lambda_{\frac{\alpha}{2}}$$

dvs enligt §12.1: $\theta^*_{obs} = \bar{x} - \bar{y}$

$\theta = \mu_x - \mu_y$

$D = D(\bar{x} - \bar{y})$ vilket finns i §11.3

(2) σ_x och σ_y är okända och olika

- väldigt stor osäkerhet, använder §12.3 approximativa metoden,

$\sigma^*_{x\,obs} = s_x$ och $\sigma^*_{y\,obs} = s_y$. har en approximativ konfidensgrad $1-\alpha$

$$I_{\mu_x - \mu_y} = (\bar{x} - \bar{y}) \pm \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}} \lambda_{\frac{\alpha}{2}}$$

③ σ_X och σ_Y är okända men lika

- här $\sigma_X = \sigma_Y = \sigma$ och är okända

§ 11.2 b): vikter ihop s_X och s_Y och får

$$s^2 = \frac{(n_x-1)s_x^2 + (n_y-1)s_y^2}{n_x + n_y - 2}$$

- Eftersom σ är okänd ⇒ § 12.2 för att ta fram konf-int:

$$I_{\mu_X - \mu_Y} = (\bar{x} - \bar{y}) \pm s\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}\, t_{\frac{\alpha}{2}}(n_x + n_y - 2)$$

$$s = \sqrt{\frac{(n_x-1)s_x^2 + (n_y-1)s_y^2}{n_x + n_y - 2}}$$

Stickprov i par

- Om X och Y är beroende av varandra och vi vill bilda ett konfidens intervall för $\mu_Y - \mu_X$ så måste vi först skapa en ny

s.v. Z som för varje i: $Z_i = Y_i - X_i$

$$Z_i \in N(\mu_z, \sigma_z)$$

tenta 2019-01-08 # 10

Två stickprov från två populationer och varje stickprov uppfattas som observationer på $N(\mu_i, \sigma_i)$ där vi antar att $\sigma_1 = \sigma_2$. Har fått följande sammanfattande mått:

Stickprov 1: $n_1 = 4$ $\bar{x}_1 = 1007.25$ $s_1 = 143.66$

Stickprov 2: $n_2 = 4$ $\bar{x}_2 = 817.75$ $s_2 = 73.627$

Beräkna den undre gränsen i ett 95%-igt tvåsidigt konfidens intervall för $\mu_1 - \mu_2$

Lösning:

- Har 2 populationer som inte beror på varandra ⇒ skillnad mellan 2 stickprov med samma varians som är okänd. (fall 3)

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{3(143.66)^2 + 3(73.627)^2}{4+4-2} =$$

$$= 13029.56 \implies s = \sqrt{13029.56} = 114.147$$

$$D^*_{ons} = s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 114.147\sqrt{\frac{1}{4} + \frac{1}{4}} = 80.714$$

$$\S\ 12.2: \quad I_{\mu_X - \mu_Y} = (\bar{x}_1 - \bar{x}_2) \pm D^*_{ons}\, t_{0.025}(n_1+n_2-2)$$

$$\frac{\alpha}{2}$$

undre gränsen: $\bar{x}_1 - \bar{x}_2 - D^*_{ons}\, t_{0.025}(6) = -8.25$

tenta 2020-01-08 #8

Kolesterolhalten i blodet mäts hos 5 personer före och efter en period med förändrad kost och motion. Antag att mätvärdena är normal fördelade. Ange med 95% säkerhet ett konfidens intervall för vilken genomsnittlig förändring av kolesterolhalten som kan påvisas om följande data finns

	1	2	3	4	5
före	262	272	284	298	294
efter	252	262	278	282	278

Lösning: mäter före och efter $\Rightarrow$ beroende av varandra $\Rightarrow$
stickprov i par!

Skapa: $z_i = x_i - y_i$ där x_i är före och y_i är efter

$z_1 = x_1 - y_1 = 262 - 252 = 10$

$z_2 = x_2 - y_2 = 272 - 262 = 10$

$z_3 = x_3 - y_3 = 284 - 278 = 6$

$z_4 = 298 - 282 = 16$

$z_5 = x_5 - y_5 = 294 - 278 = 16$

Skapar en lista: stat-edit - L_1 och sätter in värdena för z_i

räkna ut: list-math - mean(L1) $\Rightarrow \bar{z} = 11.6$

-stdDev(L1) = $s_z = 4.3359$

$Z \sim N(\mu, \frac{\sigma}{\sqrt{n}}) \sim \mu = \bar{x} = 11.6$

$\frac{s}{\sqrt{n}} = 1.9391$

både μ och σ okända $\Rightarrow$ §12.2

har 5 st data: $n-1=4$

$t_{0.025}(4) = 2.78$

$I_z = \bar{z} \pm \frac{s_z}{\sqrt{n}} t_{0.025}(4) = 11.6 \pm 1.9391 \times 2.78 = [6.21, 16.99]$

Approximativa konfidens-intervall

- Baseras på användning av normalapproximation
- I FS: §12.3 och den används när mätdata inte är normalfördelad men vårt θ^* är approximativt normalfördelad
   - får då ett konfidens intervall med approximativa konfidensgraden $1-\alpha$
- Vi kan använda §5 och §6 för att veta när vi kan approximera

$$I_\theta = \theta^*_{obs} \pm d(\theta^*)\lambda_{\frac{\alpha}{2}}$$

$d(\theta^*)$ är medelfelet, samma uträkning som innan!

Binomial-fördelning

- används tex vid kvalitetskontroll

- Vi har $X \sim Bin(n,p)$ och vi skattar p med $p^*_{obs} = \frac{x}{n}$.
- p^*_{obs} är en skattning $\Rightarrow$ vi vill bilda ett konfidens intervall för p.
- Vi har inte en Normal fördelning $\Rightarrow$ använder §12.3

θ^* i vårat fall är $\theta^*_{obs} = p^*_{obs} = \frac{x}{n} = \frac{\text{antal lyckade försök}}{\text{totalt antal försök}}$

$\theta^* = \frac{X}{n}$. om X är approximativt N-fördelad så är även θ^* N-fördelad.

$X \sim N$-fördelad om $np^*_{obs}(1-p^*_{obs}) \geq 10$ då kan §12.3 användas.

$$I_\theta = \theta^*_{obs} \pm D^*_{obs}\,\frac{\lambda}{2}$$

$\theta = p$, $\theta^*_{obs} = p^*_{obs} = \frac{x}{n}$

medelfel: $V\left(\frac{X}{n}\right) = \frac{1}{n^2} V(X) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$

$D\left(\frac{X}{n}\right) = \sqrt{\frac{p(1-p)}{n}} \Rightarrow D^*_{obs} = \left(\sqrt{\frac{p(1-p)}{n}}\right)^*_{obs} =$

$= \sqrt{\frac{p^*_{obs}(1-p^*_{obs})}{n}}$

$$\boxed{I_p = p^*_{obs} \pm \sqrt{\frac{p^*_{obs}(1-p^*_{obs})}{n}} \lambda_{\frac{\alpha}{2}}}$$

Konfidens intervall för skillnad mellan andelar

- Anta att vi har $X \in Bin(n_x, p_x)$ och $Y \in Bin(n_y, p_y)$ och vi vill bilda ett konfidensintervall för $p_x - p_y$

- Om $n_x p^*_x(1-p^*_x) \geq 10$ och $n_y p^*_y(1-p^*_y) \geq 10$ så kan vi använda §12.3

$I_\theta = \theta^*_{obs} \pm D^*_{obs} \lambda_{\frac{\alpha}{2}}$: $\theta = p_x - p_y$ och $\theta^*_{obs} = \frac{x}{n_x} - \frac{y}{n_y}$

$V\left(\frac{X}{n_x} - \frac{Y}{n_y}\right) = V\left(\frac{X}{n_x}\right) + V\left(\frac{Y}{n_y}\right) = \frac{1}{n_x^2} V(X) + \frac{1}{n_y^2} V(Y) =$

$= \frac{1}{n_x^2} n_x p_x(1-p_x) + \frac{1}{n_y^2} n_y p_y(1-p_y) = \frac{p_x(1-p_x)}{n_x} + \frac{p_y(1-p_y)}{n_y}$

$D\left(\frac{X}{n_x} - \frac{Y}{n_y}\right) = \sqrt{\frac{p_x(1-p_x)}{n_x} + \frac{p_y(1-p_y)}{n_y}} \Rightarrow D^*_{obs} = \sqrt{\frac{p^*_{x\,obs}(1-p^*_{x\,obs})}{n_x} + \frac{p^*_{y\,obs}(1-p^*_{y\,obs})}{n_y}}$

$I_{p_x - p_y} = p^*_{x\,obs} - p^*_{y\,obs} \pm \sqrt{\frac{p^*_{x\,obs}(1-p^*_{x\,obs})}{n_x} + \frac{p^*_{y\,obs}(1-p^*_{y\,obs})}{n_y}} \lambda_{\frac{\alpha}{2}}$

tenta 201 - 10-24 #10

Låt 20 personer utföra mätningar på en $N(\mu, \sigma)$-fördelad s.v. Därefter gör de var och en ett 95% konfidens-intervall för μ. Vilket är det troligaste värdet på det antal intervall som kommer att missa μ?

$$P(\text{missa}) = 1 - 0.95 = 0.05 \Rightarrow X \in Bin(20, 0.05)$$

$$E(X) = np = 20 \times 0.05 = 1$$

testtenta 2019 #10

Antag att $X \in Bin(200, p)$. Vi gör ett försök och får $x = 23$. Ange den övre gränsen för det ensidigt uppåt begränsade konf-int för p med approximativ konfidensgrad 95%

$$p^*_{obs} = \frac{x}{n} = \frac{23}{200}$$

$$V(p^*) = V\left(\frac{X}{n}\right) = \frac{1}{n^2} V(X) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$$

$$D^*_{obs} = \sqrt{\frac{p^*_{obs}(1-p^*_{obs})}{n}} = \sqrt{\frac{\frac{23}{200}\left(1 - \frac{23}{200}\right)}{200}} = 0.02256$$

ensidigt uppåt: $\lambda_\alpha = \lambda_{0.05} = 1.6449$

$$I_p^{(\text{övre})} = p^*_{obs} + D^*_{obs}\lambda_\alpha = \frac{23}{200} + 0.02256 \times 1.6449 = 0.152$$

tenta 2019-03-11 #9

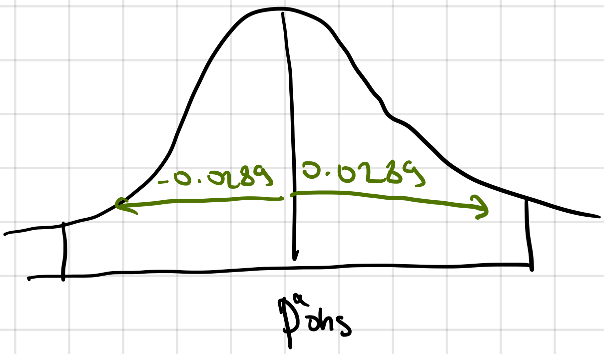
Antag att $X \in Bin(n, p)$ och vi vill bilda ett approximativt konf-int I_p för p. Vi gör $n = 500$ försök och får $x = 62$. Ange den tvåsidiga intervallets bredd på den approximativa graden 95%.

$$p^*_{obs} = \frac{x}{n} = \frac{62}{500}$$

$$V\left(\frac{X}{n}\right) = \frac{1}{n^2} V(X) = \frac{p(1-p)}{n} \Rightarrow D^*_{obs} = \sqrt{\frac{p^*_{obs}(1-p^*_{obs})}{n}}$$

$$\lambda_{0.025} = 1.96$$

$$I_p = \frac{62}{500} \pm \sqrt{\frac{\frac{62}{500}\left(1 - \frac{62}{500}\right)}{500}} \times 1.96 = \frac{62}{500} \pm 0.0289$$



bredd: $2 \times 0.0289 = 0.0578$

Poisson-fördelning

- Vi antar att $X_i \in Po(\mu)$ och skattar μ med $\bar{X}$
- Vi vill bilda ett konf-int för μ och vill använda §12.3

- måste ha θ^* är approximativt N-fördelad:

$$\theta = \mu \Rightarrow \theta_{obs} = \mu^*_{obs} = \bar{x} \Rightarrow \theta^* = \frac{1}{n} \sum_{i=1}^{n} x_i$$

§6 i FS: $Po(\mu) \sim N(\mu, \sqrt{\mu})$ om $\mu \geq 15$

enligt satsen: om $X_1 \in Po(\mu_1)$ och $X_2 \in Po(\mu_2)$ och X_1 och X_2 är oberoende så är

$$X_1 + X_2 \in Po(\mu_1 + \mu_2)$$

$$\sum X_i \in Po(\mu, \sqrt{\mu}) \Rightarrow \sum X_i \sim N(n\mu, \sqrt{n\mu})$$

om $n\mu \geq 15$

$\theta^*_{obs} = \bar{x}$, vad är $D^*_{obs} = ?$

$$V(\bar{X}) = V\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2} V\left(\sum X_i\right) = \frac{1}{n^2} n\mu = \frac{\mu}{n}$$

$$D^*_{obs} = \left(\sqrt{\frac{\mu}{n}}\right)^*_{obs} = \sqrt{\frac{\bar{x}}{n}}$$

$$\Rightarrow \boxed{I_\mu = \bar{x} \pm \sqrt{\frac{\bar{x}}{n}}\ \lambda_{\frac{\alpha}{2}}}$$

tenta 2019-12-19 #9

Antag att $X \in Po(\mu)$ där μ är okänt. Ange den övre gränsen för det ensidigt uppåt begränsade konfidens-intervallet för μ som har approximativa konfidensgraden 90%. Utgå från följande värden som är framtagna från fyra oberoende försök med $x_1 = 12$, $x_2 = 16$, $x_3 = 19$, $x_4 = 15$

$$\mu^*_{obs} = \bar{x} = \frac{12 + 16 + 19 + 15}{4} = 15.5 \qquad \mu \geq 15 \Rightarrow \sim N(\mu, \sqrt{\mu})$$

kan använda §12.3: $X \sim N(15.5,\ 3.937)$

$$I_\mu = (-\infty, \bar{x} + D^*_{obs}\ \lambda_{0.1})$$

för 0.1 då vi har ensidigt Intervall!

$\lambda_{0.1} = 1.2816$

$$D^*_{obs} = \sqrt{\frac{\bar{X}}{n}} = \frac{3.937}{\sqrt{4}} = \frac{3.937}{2} = 1.9685$$

övre gränsen: $\bar{x} + D^*_{obs} \lambda_{0.1} = 15.5 + 1.9685 \times 1.2816 =$

$$= \underline{\underline{18.02}}$$

tenta 2018-12-18 #12

Låt X_i, $i = 1, 2 \dots 5$ vara oberoende Po-fördelade s.v. med väntevärde μ. Man får $\bar{x} = 16$. Ange den nedre gränsen för det ensidigt nedåt begränsade konfidensintervallet för μ med den approximativa konfidensgraden 95%.

$$I_\mu = (\bar{x} - D^*_{obs} \lambda_{0.05}, +\infty)$$

undre

$$V(\bar{X}) = V\left(\frac{1}{5} \sum X_i\right) = \frac{1}{25} 5V(x) = \frac{1}{5} \mu$$

$$D^*_{obs} = \sqrt{\frac{\bar{x}}{5}} = \sqrt{\frac{16}{5}} = 1.789$$

$$\lambda_{0.05} = 1.6449$$

$$I_\mu = (\underbrace{16 - 1.789 \times 1.6449}, +\infty)$$

undre gräns: 13.06

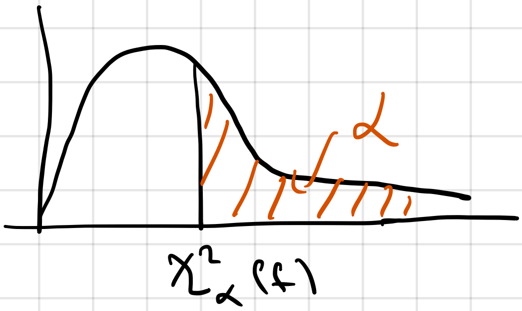
Konfidens intervall för σ och σ^2

- Använder oss av §12.4 och χ^2 - fördelning.

  - Om $X_1, X_2 \dots X_f$ är $\in N(0,1)$ -fördelade och oberoende så gäller det att

$$\sum_{i=1}^{f} X_i^2 \in \chi^2(f) \qquad f = n-1$$

  - tabell 4:



$$P(X > \chi^2_\alpha(f)) = \alpha$$

- Anta att X_i:na är oberoende och $N(\mu, \sigma)$ -fördelade

$$\sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma}\right)^2 \in \chi^2(n-1)$$

§11.2 b: $\dfrac{1}{\sigma^2} \displaystyle\sum_{i=1}^{n} (x_i - \bar{x})^2 = \dfrac{n-1}{\sigma^2} \underbrace{\dfrac{1}{n-1} \sum (x_i - \bar{x})^2}_{s^2} =$

$$= \frac{n-1}{\sigma^2} s^2 \in \chi^2(n-1)$$

tabell 4: $P\left(\dfrac{(n-1)s^2}{\sigma^2} > \chi^2_\alpha(n-1)\right) = \alpha$

$$P\left(\sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_\alpha(n-1)}}\right) = \alpha$$

vänstra gränsen: $P\left(\sigma < \sqrt{\dfrac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}}\right) = \dfrac{\alpha}{2}$

högra gränsen: $P\left(\sigma < \sqrt{\dfrac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}}\right) = 1-\dfrac{\alpha}{2}$

$$\Rightarrow \boxed{I_\sigma = \left(\sqrt{\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}}\right)}$$ §12.4

$$\boxed{I_{\sigma^2} = \left(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}, \frac{s^2(n-1)}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right)}$$

tenta 2019-01-03 #9

Antag att $X_1 \dots X_n$ utgör ett stickprov på $N(\mu, \sigma)$. Från 20 observationer erhålls följande värden: $\bar{x} = 0.46$ och $s = 0.43$. Ange den nedre gränsen för den tvåsidiga konfidens intervallet för σ med konfidensgraden 99%.

$\sigma^*_{obs} = s \Rightarrow$ använder §12.4

nedre gränsen: $\theta^*_{obs}\sqrt{\dfrac{f}{\chi^2_{\frac{\alpha}{2}}(f)}}$

$\sigma = \theta$

$\theta^*_{obs} = s$

$\dfrac{\alpha}{2} = \dfrac{1-0.99}{2} = 0.005$

$f = n-1 = 20-1 = 19$

$\chi^2_{0.005}(19) = 38.6$

$$\Rightarrow \quad \sigma^*_{obs}\sqrt{\frac{f}{\chi^2_{\frac{\alpha}{2}}(f)}} = s\sqrt{\frac{19}{\chi^2_{0.005}(19)}} = 0.43\sqrt{\frac{19}{38.6}} = 0.302$$

tenta 2019-05-28 #7

Antag att $X_1 \dots X_n$ utgör ett stickprov på $N(\mu, \sigma)$. Från 10 observationer erhålls följande värden: $\bar{x} = 2.13$ och $s = 1.03$. Ange den övre gränsen för den tvåsidiga konfidensintervallet för σ^2 med konfidensgraden 95%.

Skattar σ med $\sigma^*_{obs} = s$, använder §12.4

övre gränsen ges av $s\sqrt{\frac{f}{\chi^2_{1-\frac{\alpha}{2}}(f)}}$ för σ då vi skattar sigma med s.

$s = 1.03$, $1-\frac{\alpha}{2} = 0.975$, $f = n-1 = 10-1 = 9$

$$\sigma: \quad 1.03\sqrt{\frac{9}{\chi^2_{0.975}(9)}} = 1.88$$

för σ^2: $1.88^2 = 3.54$

Hypotesprövning

- Den hypotes som man vill testa om man ska förkasta eller inte kallas för nollhypotesen, H_0, ska alltid kunna förkastas.
- Mothypotesen till H_0 kallas för H_1
- Risknivå = signifikansnivå = α, högsta risken vi kan ta att förkasta H_0 om H_0 är sann.

$$P(\text{förkasta } H_0) \text{ om } H_0 \text{ sann} < \alpha$$

· p-värde: Sannolikheten att förkasta H_0 under förutsättningen att H_0 är sann givet att vi har fått ett visst värde på test-variabeln = observationen

$$p\text{-värdet} < \alpha \Rightarrow \text{förkasta } H_0 \text{, annars inte.}$$

· Styrkefunktionen: $h(\theta) = P(\text{förkasta } H_0)$ om θ är rätt parametervärde

· Styrkan hos testet för $\theta = \theta_1$: $P(\text{förkasta } H_0: \theta = \theta_0)$ om $H_1: \theta = \theta_1$ är sann.

Konfidensintervall metoden

· Om $H_0: \mu = a$, $H_1: \mu \neq a$ } tvåsidigt konf-int

– bilda ett konfidens intervall för μ och om a ligger i konf-int så kan vi inte förkasta H_0 på den risknivån

vid ensidig test: $H_0: \mu = a$

$H_1: \mu < a$ begränsas uppåt.

tenta 2019-10-23 # 9

Följande konfidens intervall med konfidensgraden 99%. för skillnaden mellan 2 stickprovs väntevärden är given:

$$I_{\mu_X - \mu_Y} = (0.12, 0.32)$$

Antag att stickproven kommer från $X_i \in N(\mu_X, \sigma_X)$ och $Y_i \in N(\mu_Y, \sigma_Y)$ samt alla s.v. är oberoende. Antag vidare att σ_X och σ_Y är kända. Man önskar testa $H_0: \mu_X = \mu_Y$ mot $H_1: \mu_X \neq \mu_Y$. Vilken slutsats kan vi dra om att förkasta H_0?

$H_0: \mu_X = \mu_Y$

$\mu_X - \mu_Y = 0$. Måste kolla om 0 finns i intervallet:

$I_{\mu_X - \mu_Y} = (0.12, 0.32)$ 99%

$0 \notin I_{\mu_X - \mu_Y} \Rightarrow$ kan förkasta H_0 på risknivån 1% och även 5%.

tenta 2019-08-12 #11

Vi antar att $X \in Exp(\lambda)$. Nollhypotesen är $H_0: \mu = 4$ där $\mu = E(X)$ och mothypotesen är $H_1: \mu > 4$. Har en observation $x = 8$. Bestäm testets p-värde.

p-värde = p (förkasta H_0) om H_0 sann. $= P(X \geq 8)$ om $\mu = 4$

$$P = \int_8^\infty \lambda e^{-\lambda x} dx$$

$$\lambda = \frac{1}{E(X)} = \frac{1}{4}$$

$$\text{p-värde} = \int_8^\infty \frac{1}{4} e^{-\frac{1}{4}x} dx = \left[-e^{-\frac{1}{4}x} \right]_8^\infty = e^{-\frac{1}{4} \cdot 8} = 0.135$$

tenta 2019-12-19 #1

Antag att $X \in ffg(p)$ och låt $H_0: p = 0.25$. Vill testa H_0 mot $H_1: p < 0.25$ och vill förkasta H_0 om vi observerar ett stort värde x på X. Antag att vi har observerat $x = 5$. Bestäm testets p-värde.

Stort X : $P(X \geq 5) = 1 - P(X \leq 4)$

$p = 1 - [p + p(1-p) + p(1-p)^2 + p(1-p)^3] = \{p = 0.25\} =$

$= 0.3164$

tenta 2019-01-08 #12

En forskare har gjort 10 försök som anses vara oberoende av varandra där sannolikheten för lyckat försök är p. Låt X stå för antalet lyckade försök. Forskaren önskar pröva $H_0: p = \frac{1}{2}$ mot $H_1: p > \frac{1}{2}$. Resultatet av de 10 försök var 8 lyckade. Beräkna p-värdet.

$n = 10, \quad p = \frac{1}{2} \Rightarrow X \sim Bin(10, 0.5)$ enligt H_0

$P(X \geq 8) = 1 - P(X < 8) = 1 - P(X \leq 7) = 0.0547$

tenta 2018-10-24 #12

Givet är 7 oberoende observationer av en s.v. X som är $N(\mu, 0.4)$. Medelvärdet beräknat på 7 observationer är 0.719. Vill testa $H_0: \mu = 1$ mot $H_1: \mu < 1$. Vad är testets p-värde?

p-värde $= P(X > \mu) = \{$gör om till $N(0,1)\} = P\left(\frac{X - \bar{x}}{\sigma/\sqrt{n}}, \frac{\mu - \bar{x}}{\sigma/\sqrt{n}}\right) =$

$= P\left(Y > \frac{1 - 0.719}{0.4/\sqrt{7}}\right) = P(Y > 1.86) = 1 - P(Y \leq 1.86) = 0.0314$

χ^2 - test: test av en given fördelning

- Används när nollhypotesen är att vi har en viss sannolikhetsfunktion

$$H_0: p(A_1) = p_1 \quad p(A_2) = p_2 \ldots$$ summan av alla är 1.

§14.3: $$Q = \sum_{i=1}^{r} \frac{(x_i - np_i)^2}{np_i}$$ villkor för test: alla $np_i > 5$

r - antal resulteler.

om $Q > \chi^2_\alpha(f)$, $f = r-1$ förkasta H_0 på risknivån α

om $Q \leq \chi^2_\alpha(f)$, förkasta ej H_0 på risknivån α

- tabell 4 för χ^2- fördelningsvärdena.

tenta 2019-01-08 #15 (10p)

Man misstänkte att ett roulettbord på ett kasino var manipulerat och genomförde ett test med 8000 försök. Om rouletten är korrekt ska röd, svart och grön komma i proportionerna 18 : 18 : 1. Testresultatet gav röd: 3751, svart: 4018, grön: 231. Avgör med felrisken 1% om rouletten är korrekt.

- H_0 nollhypotesen att vi har en viss sannolikhetsfunktion.

⇒ test av en given fördelning.

$n = 8000$

A_1 = röd $\quad X_1 = 3751$

A_2 = svart $\quad X_2 = 4018$

A_3 = grön $\quad X_3 = 231$

$r = 3$

$18+18+1=37$

H_0: $p(A_1) = \frac{18}{37}$ $\quad p(A_2) = p_2 = \frac{18}{37}$ $\quad p(A_3) = p_3 = \frac{1}{37}$

Villkor för χ^2-test: alla $np_i > 5$

np_3 (grön) har lägst sannolikhetsfunktion:

$8000 \times \frac{1}{37} > 5 \Rightarrow$ χ^2-test ok att användas.

$$Q = \sum_{i=1}^{r} \frac{(x_i - np_i)^2}{np_i} = \frac{\left(3751 - \frac{18}{37}\times 8000\right)^2}{\frac{18}{37}\times 8000} +$$

$$+ \frac{\left(4018 - \frac{18}{37}\times 8000\right)^2}{\frac{18}{37}\times 8000} + \frac{\left(231 - \frac{1}{37}\times 8000\right)^2}{\frac{1}{37}\times 8000} = 10.20$$

· Ska vi förkasta H_0?

$$\chi^2_\alpha(r-1) = \chi^2_{0.01}(2) = 9.21$$

$Q > \chi^2_\alpha(r-1) \Rightarrow$ Q = stort, förkasta H_0 på risknivå 1%.

χ^2-test: test av en given fördelning med skattade parametrar

- Vill testa $H_0: p(A_1) = p_1(\theta) \ldots$ för en okänd parameter θ.

- Skatta θ först, helst använda ML

$$Q' = \sum_{i=1}^{r} \frac{(x_i - np_i^*)^2}{np_i^*}$$

approximativt $\chi^2(r-k-1)$ fördelad där k är antalet skattade parametrar.

villkor $np_i^* \geq 5$

$Q > \chi^2_\alpha(r-k-1) \rightarrow$ förkasta H_0 på risknivå α

$Q \leq \chi^2_\alpha(r-k-1) \Rightarrow$ förkasta inte H_0 på risknivå α.

tenta 2020-01-08 #11

Man vill undersöka om följande data kommer från en Bin-fördelning med parametern $n=3$ för något värde på p.

	0	1	2	3	totalt
antal observationer	15	54	28	3	100

Om $p^*_{obs} = 0.3967$ används så får man $Q = 6.5329$.
Vad kan man säga om att förkasta H_0?

1% och 5%

Lösning: χ^2-test med en skattad variabel

$r = 4$, $k = 1$ (endast p^*_{obs} är skattad)

$\chi^2_{0.01}(r-k-1) = \chi^2_{0.01}(4-1-1) = \chi^2_{0.01}(2) = 9.21$

$\chi^2_{0.05}(2) = 5.99$

$Q > \chi^2_{0.05}(2) \Rightarrow$ kan förkasta H_0 på 5% risknivå

$Q < \chi^2_{0.01}(2) \Rightarrow$ kan inte förkasta på 1% risknivå.

Tenta 2019-06-05 #12

Man misstänker att ett roulettbord på ett kasino var manipulerat och genom förde ett test med 8000 försök. Om korrekt: 18:18:1. Testet gav: röd = 3771, svart = 4002, grön = 227. H_0 att vi teorer sannolikheter, $Q = 7.4$. P-värdet?

test av en given fördelning: $\chi^2(r-1) = \chi^2(2)$

för p-värde kolla tabell 4

$\chi^2_{0.025}(2) = 7.38$

$\chi^2_{0.01}(2) = 9.21$

$\chi^2_{0.025} < Q < \chi^2_{0.01}$ $\quad$ Q närmare $\chi^2_{0.025}(2)$

p-värde mellan 1% och 2.5%, lite närmare till 2.5%.

χ^2-test: homogenitetstest:

- Används när nollhypotesen är att vi har samma sannolikhetsfunktion i alla grupper

$$Q = \sum_{i=1}^{s}\sum_{j=1}^{r} \frac{\left(x_{ij} - \frac{n_i m_j}{N}\right)^2}{\frac{n_i m_j}{N}}$$

villkor: alla $\frac{n_i m_j}{N} \geq 5$

om $Q > \chi^2_\alpha[(r-1)(s-1)]$ förkasta H_0

tenta 2020-03-10 #11

studerar 570 personer

	typ 1	typ 2	placebo
~~biverkning~~	27	17	10
~~ingen biverkning~~	373	73	70

För att testa hypotesen H_0: Det finns ingen skillnad mellan grupperna, fick $Q = 13.62$. Slutsats?

- har homogenitetstest: $s = 2$, $r = 3$ $= \chi^2[(3-1)(2-1)] = \chi^2(2)$

$\chi^2_{0.01}(2) = 9.21$

$\chi^2_{0.001}(2) = 13.8$

$Q = 13.62 > \chi^2_{0.01}$ $\Rightarrow$ förkasta H_0 på risknivå 1%.

$Q < \chi^2_{0.001}$ $\Rightarrow$ kan inte förkasta på risknivå 0.1%.

χ^2-test: oberoendetest

· Om vi vill undersöka om två egenskaper A och B är oberoende

- gör till som homogenitetstest

H_0: antar att oberoende.