



MVE051/MSG810 - Statistik och diskret matematik (LP4 2022-2023)

Chalmers och Göteborgs universitet

Föreläsning 9 – Hypotestest

May 2, 2023

Dagens föreläsning

- Hypotestest
 - Nollhypotes
 - Alternativhypotes
 - Typ-II-fel och typ-II-fel
 - Signifikans
 - Styrka
 - Hypotestest vs. konfidensintervall
 - p -värde

Hypotestest

Nollhypotes

Någonting som vi vill visa inte stämmer, ex.

$$H_0: \theta = 0.$$

Typiskt ett antagande om populationens fördelning, ex. dess parametrar.

Alternativhypotes

Ett alternativ till nollhypotesen, ex.

- $H_1: \theta \neq 0$ (tvåsidigt test)
- $H_1: \theta > 0$ (ensidigt test)
- $H_1: \theta < 0$ (ensidigt test)

Hypotesprövning

- Välj nollhypotes och alternativhypotes
- Bestäm en regel för när H_0 ska förkastas. Ofta är regeln att ett tal som går att beräkna från datan har ett ovanligt/extremt värde. Talet kallas för en teststatistika.
- Använd data och din regel för att bestämma om du ska förkasta H_0 eller inte.

Hypotestest

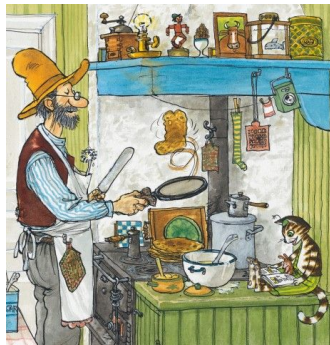
Exempel

I en tidning står att hönsäggs vikt kan antas vara normalfördelad med medelvärde $\mu = 54$ och standardavvikelse $\sigma = 1$.

Pettson tycker att det låter mycket, och väger därför äggen från hans höns under en dag för att kunna testa hypotesen

$$H_0: \mu = 54 \quad \text{mot} \quad H_1: \mu < 54.$$

Totalt väger han 4 ägg. Pettson bestämmer sig för att förkasta H_0 om alla ägg väger mindre än 54 gram.



Hypotestest (fel av typ I)

Typ-I-fel

Antag att teststatistikan får ett värde som skiljer sig (mycket) från 0 och vi väljer vi att förkasta H_0 . Då är antingen

1. H_0 sann och vi har fått ett ovanligt stickprov, eller
2. H_1 sann.

Om H_0 är sann har man alltså tagit ett felaktigt beslut.

Exempel (Pettson)

Pettson gör ett typ-I-fel om

- tidningen har rätt, och
- han drar slutsatsen att $\mu < 54$ efter att ha gjort sitt test.

Hypotestest (fel av typ II)

Typ-II-fel

Om värdet av teststatistikan inte var tillräckligt stort eller litet för att förkasta H_0 så skulle man inte förkasta H_0 . Då är antingen

1. H_0 sann, eller
2. H_1 sann, men vi har fått ett ovanligt stickprov.

Om H_1 är sann har man alltså tagit ett felaktigt beslut.

Typ-I och Typ-II fel

	Acceptera H_0	Förkasta H_0
H_0 korrekt		typ-I-fel
H_1 korrekt	typ-II-fel	

Signifikans och styrka

Signifikans

Ett tests signifikans är sannolikheten att man gör ett typ-I-fel om man använder det, dvs.

$$\alpha = P(\text{typ-I-fel}) = P(\text{förkasta } H_1 \mid H_0 \text{ sann})$$

Teststyrka

Sannolikheten för typ-II-fel är

$$\beta = P(\text{typ-II-fel}) = P(\text{förkasta } H_1 \mid H_1 \text{ sann})$$

Sannolikheten β används för att definiera teststyrkan

$$P(\text{förkasta } H_0 \mid H_1 \text{ sann}) = 1 - P(\text{förkasta } H_1 \mid H_1 \text{ sann}) = 1 - \beta.$$

Signifikans och styrka

	Acceptera H_0	Förkasta H_0
H_0 korrekt		typ-I-fel
H_1 korrekt	typ-II-fel	

	Acceptera H_0	Förkasta H_0
H_0 korrekt	$1 - \alpha$	α
H_1 korrekt	β	$1 - \beta$

Signifikans och styrka

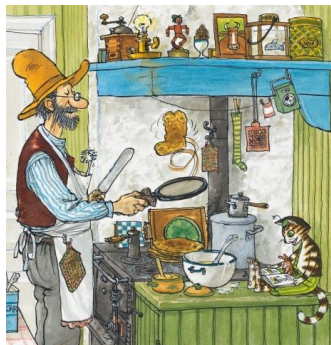
Exempel

I en tidning står att hönsäggs vikt kan antas vara normalfördelad med medelvärde $\mu = 54$ och standardavvikelse $\sigma = 1$.

Pettson tycker att det låter mycket, och väger därför äggen från hans höns under en dag för att kunna testa hypotesen

$$H_0: \mu = 54 \quad \text{mot} \quad H_1: \mu < 54.$$

Totalt väger han 4 ägg. Pettson bestämmer sig för att förkasta H_0 om alla ägg väger mindre än 54 gram.



1. Vad är signifikansnivån på Pettsons test?
2. Anta att Pettsons höns verkligen lägger ovanligt små ägg med viktfordelning $N(52, 1)$. Vad är styrkan av Pettsons test?

Signifikans och styrka

Signifikans

Låt $X \sim N(54, 1)$ vara vikten av ett slumpvist valt ägg. Då är

$$\begin{aligned}\text{signifikans} &= \alpha = P(\text{typ-I-fel}) = P(\text{förkasta } H_0 \mid H_0 \text{ sann}) \\ &= P(X < 54)^4 = (1/2)^4 = 1/16 \approx 0.0625\end{aligned}$$

Styrka

Låt $Y \sim N(52, 1)$ vara vikten av ett slumpvist valt ägg. Då är

$$\begin{aligned}\text{styrka} &= 1 - \beta = 1 - P(\text{typ-II-fel}) = P(\text{förkasta } H_0 \mid H_1 \text{ sann}) \\ &= P(Y < 54)^4 = \Phi\left(\frac{54 - 52}{1}\right)^4 = \Phi(2)^4 \approx 0.912058\end{aligned}$$

Var är ett bra test?

- För ett bra test är signifikansnivån α liten och styrkan $1 - \beta$ stor, dvs. sannolikheten för båda sorternas fel är liten.
- Ofta konstrueras test så att de har hög signifikansnivå.
- För att få hög styrka behöver man oftast ha ett väldigt stort stickprov.
- Det är oftast svårt/omöjligt att beräkna styrkan eftersom att det kräver att man vet vad ex. μ verkligen är. Ofta beräknar man i så fall styrkan som en funktion av det okända parametern μ istället.

Hypotestest

Data

Säg att vi vill undersöka om kvinnors medellängd i Sverige är 166cm. Man antar att längden är normalfördelad. Vi har ett stickprov med n längder.

Hypoteser

$$H_0: \mu = 166 \quad \text{mot} \quad H_1: \mu \neq 166$$

1. Om H_0 stämmer har ingen förändring har skett, dvs. medellängden av kvinnor har inte ändrats.
2. Om H_1 stämmer har medellängden ändrats. Det är den här hypotesen som forskaren tror på.

Vi antar att H_0 är sann tills vi hittar tillräckligt mycket bevis för att förkasta H_0 .

Hypotestest

Teststatistika

Som teststatistika vill vi ha estimator av någonting som vi vet fördelningen för om H_0 är sann.

Om H_0 är sann så är $\mu = 166$ och

$$\bar{X} \sim N(166, \sigma/\sqrt{n}) \Rightarrow T = \frac{\bar{X} - 166}{s/\sqrt{n}} \sim t_{n-1}$$

För att testa om H_0 är sann är använder vi \bar{x} som en skattning för det sanna populationsmedelvärdet, och beräknar teststatistikan

$$t = \frac{\bar{x} - 166}{s/\sqrt{n}}.$$

Om vi har ett stickprov med $n = 177$, $\bar{x} = 165.0$, $s = 6.05$, får vi

$$t = \frac{\bar{x} - 166}{s/\sqrt{n}} = \frac{165.0 - 166}{6.05/\sqrt{177}} = -2.20.$$

Hypotestest

Vi fick $z = -2.20$. Är detta ett tillräckligt litet värde på z för att förkasta H_0 ?

Om nollhypotesen är sann borde $T \sim t_{n-1}$, så t borde vara någorlunda nära 0. För att kunna förkasta nollhypotesen bör t antingen vara betydligt mindre eller betydligt större än 0. Hur mycket mindre/större?

Signifikans

Om $\alpha \in (0, 1)$ så är

$$\begin{aligned} 1 - \alpha &= P(t_{n-1, (1-\alpha)/2} < T < t_{n-1, (1+\alpha)/2}) \\ &= P(-t_{n-1, (1+\alpha)/2} < T < t_{n-1, (1+\alpha)/2}) = P(|T| \leq t_{n-1, (1+\alpha)/2}) \end{aligned}$$

Om vi förkastar H_0 när

$$|t| > t_{n-1, 1-\alpha/2}$$

så är

$$P(\text{förkasta } H_0 \mid H_0 \text{ sann}) = P(|T| > t_{n-1, 1-\alpha/2}) = \alpha$$

dvs. α är testets signifikansnivå.

Hypotestest

Testprocedur

- Vi testar hypotesen

$$H_0: \mu = 166 \quad \text{mot} \quad H_1: \mu \neq 166.$$

- Om H_0 är sann är $T = (\bar{X} - 166)/(s/\sqrt{n}) \sim t_{n-1}$.
- Vi bestämmer oss att använda signifikansnivå $\alpha = 0.05$. Då förkastar vi H_0 om $|t| > t_{n-1,0.975} = 1.974$
- Vi räknar ut teststatistikan $t = (\bar{x} - 166)/(s/\sqrt{n}) = -2.2$.
- Vi har

$$|t| = 2.20 > 1.974 = z_{n-1,0.975}$$

och vi förkastar därför H_0 på signifikansnivå 0.05.

- Slutsats: Det verkar som medellängden av kvinnor i Sverige har förändrats.

Kritiskt område

Intervallet där H_0 förkastas kallas för det kritiska området. I vårt fall är det

$$(-\infty, -z_{n-1,0.975}) \cup (z_{n-1,0.975}, \infty) = (-\infty, -1.974) \cup (1.974, \infty).$$

Val av alternativhypotes

Om nollhypotesen är

$$H_0: \mu = \mu_0 \quad (\text{medellängden har inte förändrats})$$

kan alternativhypotesen vara en av följande:

- Om vi tror att medellängden av kvinnor har förändrats men vet inte åt vilken riktning, väljer vi

$$H_1: \mu \neq \mu_0. \quad (\text{medellängden har förändrats})$$

- Om vi tror att medellängden av kvinnor har minskat, väljer vi

$$H_1: \mu < \mu_0. \quad (\text{medellängden har minskat})$$

- Om vi tror att medellängden av kvinnor har ökat, väljer vi

$$H_1: \mu > \mu_0. \quad (\text{medellängden har ökat})$$

Val av alternativhypotes

Anta att vi tror att medellängden av kvinnor har ökat och testar

$$\begin{cases} H_0: \mu = 166 \\ H_1: \mu > 166. \end{cases}$$

- Teststatistikan och dess värde blir samma, dvs

$$t = (\bar{x} - 166)/(s/\sqrt{n}) = -2.2.$$

- Vi är nu bara intresserade av om värdet är signifikant större än nollhypotesvärdet 166. Vi förkastar alltså H_0 på signifikansnivå 0.05 om t är större än $t_{n-1,0.95} = 1.654$.
- Eftersom att

$$-2.2 \not> 1.654$$

kan vi inte förkasta H_0 .

Med andra ord verkar medellängden av kvinnor inte ha blivit högre, eller så har vi inte tagit ett tillräckligt stort stickprov för att kunna se den eventuella skillnad i medellängd som finns. Ju mindre skillnad vi vill kunna upptäcka, desto större stickprov behöver vi.

Val av alternativhypotes

Man bör följa följande riktlinjer:

- Hypotes och alternativhypotesen ska bestämmas innan man tittar på data.
- Använd bara en enkelsidig alternativhypotes $H_1: \mu < \mu_0$ eller $H_1: \mu > \mu_0$ om den andra riktningen är
 - omöjlig,
 - osannolik, eller
 - irrelevant.

Annars används ett tvåsidigt test med $H_1: \mu \neq \mu_0$.

Hypotestest vs konfidensintervall

När vi testade $H_0: \mu = \mu_0$ mot $H_1: \mu \neq \mu_0$ och vi antog att \bar{X} var normalfördelad så använde vi statistikan

$$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

och förkastade H_0 på signifikansnivå α om $|T| > t_{n-1, (1+\alpha)/2}$.

$$\begin{aligned} |T| < t_{n-1, (1+\alpha)/2} &\Leftrightarrow -t_{n-1, (1+\alpha)/2} < \frac{\bar{X} - \mu_0}{s/\sqrt{n}} < t_{n-1, (1+\alpha)/2} \\ &\Leftrightarrow \mu_0 \in \underbrace{\left[\bar{X} - t_{n-1, (1+\alpha)/2} \cdot s/\sqrt{n}, \bar{X} + t_{n-1, (1+\alpha)/2} \cdot s/\sqrt{n} \right]}_{CI_{1-\alpha}} \end{aligned}$$

Vi förkastar alltså H_0 på signifikansnivå α om $\mu_0 \notin CI_\alpha$.

Hypotestest vs konfidensintervall

Exempel (kvinnors medellängd)

Det 95%-iga konfidensintervallet för längden blir

$$\bar{x} \pm t_{n-1, 0.975} \cdot (s/\sqrt{n}) = 165 \pm 1.96 \cdot 6.05/177 = 165 \pm 0.89 = (164.11, 165.89)$$

Eftersom att 166 inte ligger i konfidensintervallet ser vi direkt se att $H_0: \mu = \mu_0$ kan förkastas på signifikansnivå 5%.

p -värde

I stället för att undersöka om värdet på en teststatistika är inom det kritiska området kan man beräkna ett så kallat p -värde. Man gör då följande:

1. Man väljer man en teststatistika och räknar dess värde.
2. Man beräknar sannolikheten att observera ett värde som är minst lika extremt som det som vår teststatistika har givet att H_0 är sann. Den här sannolikheten kallas p -värde.

p -värde

I vårt tidigare exempel (medellängden av kvinnor) var teststatistikan $t = -2.20$. Vi kan då beräkna sannolikheten att observera ett värde som är minst lika litet, dvs

$$p = P(T \leq -2.20 | H_0 \text{ sann}) = 0.0139.$$

Om vår alternativhypotes är tvåsidig är det istället naturligt att beräkna

$$p = P(|T| \leq -2.20 | H_0 \text{ sann}) = 2 \cdot P(T \leq -2.20 | H_0 \text{ sann}) = 0.0278.$$

Om p -värdet är mindre än 0.05 kan vi förkasta H_0 på signifikansnivå 0.05.

Jämförelse av två populationer

Antag att vi har två olika stickprov som vi vill jämföra:

- Data från kvinnor resp. män
- Data från en sjö, insamlad vid två olika tillfällen
- Vikt av ägg före och efter byte av foder

Jämförelse av två populationer

Oparad data

Det första stickprovet kommer från en population med fördelning $N(\mu_1, \sigma_1)$ och att det andra kommer från en population med fördelning $N(\mu_2, \sigma_2)$, och att stickproven är oberoende.

Exempel: Pettson delar in hönorna i två grupper och ger hälften ett annat foder. Pettson samlar in 10 ägg från varje grupp.

Parad data

Det första stickprovet kommer från en population med fördelning $N(\mu_1, \sigma_1)$ och att det andra kommer från en population med fördelning $N(\mu_2, \sigma_2)$, med samma individer/mätpunkter finns med i båda stickproven.

Exempel: Pettson väljer ut 10 hönor och samlar in ett ägg från varje höna. Han byter sedan foder och samlar en vecka senare in ägg från samma hönor. Han räknar ut hur mycket vikten har ökat för varje höna

Jämförelse av två populationer (oparad data, kända varianser)

Vi antar att

- det 1:a stickprovet kommer från en population med fördelning $N(\mu_1, \sigma_1)$,
- det 2:a stickprovet kommer från en population med fördelning $N(\mu_2, \sigma_2)$,
- stickproven är oberoende,
- vi vet vad σ_1 och σ_2 är.

Statistika

$$\bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2})$$

Konfidensintervall för $\mu_1 - \mu_2$

$$\bar{x}_1 - \bar{x}_2 \pm z_{(1+\alpha)/2} \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$$

Jämförelse av två populationer (oparad data, okänd men lika varians)

Vi antar att

- det 1:a stickprovet kommer från en population med fördelning $N(\mu_1, \sigma_1)$,
- det 2:a stickprovet kommer från en population med fördelning $N(\mu_2, \sigma_2)$,
- stickproven är oberoende,
- $\sigma_1 = \sigma_2$

Poolad stickprovsstandardavvikelse

Vi skattar $\sigma = \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2} = \sigma \sqrt{1/n_1 + 1/n_2}$ med

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Statistika

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{1/n_1 + 1/n_2}} \sim t_{n_1 + n_2 - 1}$$

Konfidensintervall för $\mu_1 - \mu_2$

$$\bar{x}_1 - \bar{x}_2 \pm t_{n_1 + n_2 - 2, (1+\alpha)/2} s_p \sqrt{1/n_1 + 1/n_2}$$

Jämförelse av två populationer (parad data, okänd men lika varians)

Vi antar att

- stickprovet består av n par av mätningar (X_i, X'_i) ,
- $D = X - X' \sim N(\mu, \sigma)$,
- stickproven är oberoende.

Statistika

Om skillnaderna är normalfördelade eller om stickprovet är tillräckligt stort och man vill testa $H_0: \mu = \mu_0$ kan man använda statistikan

$$\frac{\bar{d} - \mu_0}{s/\sqrt{n}},$$

där s är stickprovsstandardavvikelsen av differanserna. Den stokastiska motsvarigheten har en t_{n-1} -fördelning om H_0 är sann.

Konfidensintervall för μ

$$\bar{d} \pm t_{n-1, (1+\alpha)/2} s / \sqrt{n}$$

Jämförelse av två populationer

Exempel

I en tidning står att hönsäggs vikt kan antas vara normalfördelad med medelvärde $\mu = 54$ och standardavvikelse $\sigma = 1$. Av sitt första experiment har Pettson dragit slutsatsen att hönornas ägg är ovanligt små. Han bestämmer dig därför för att testa om de lägger mer ägg om han matar dem med köttbullar.



Pettson och Findus genomför två olika experiment:

1. Pettson delar in 20 hönor i två lika stora grupper och Findus matar grupp 2 med köttbullar varje dag i en vecka. Grupp 1 får inga köttbullar. Efter en vecka väger de ett ägg per höna.
2. Pettson och Findus väger ett ägg från varje höna i en grupp med 10 hönor och håller koll på vilket ägg som kom från vilken höna. Findus matar sedan alla hönor med köttbullar i en vecka. Efter en vecka vägs ett ägg per höna igen.

Jämförelse av två populationer

Kan de dra slutsatsen att hönor lägger större ägg om de får äta köttbullar? Vad behöver Pettson och Findus anta?

Experiment 1 (oparad data)

$$H_0: \mu_2 - \mu_1 = 0 \quad \text{mot} \quad H_1: \mu_2 - \mu_1 \neq 0 \quad (\text{eller } \mu_1 - \mu_2 > 0) \quad \alpha = 0.95$$

Stickprov 1: {52., 49.8, 51.6, 50.7, 50.2, 48.6, 54., 48.7, 53.1, 47.5}

Stickprov 2: {48.7, 52., 54.4, 51.5, 52., 47.9, 57., 51.8, 56.3, 48.1}

$$\bar{x}_1 = 50.6, \quad \bar{x}_2 = 52.0, \quad s_1 = 2.1, \quad s_2 = 3.2, \quad n_1 = n_2 = 10, \quad s_p = 2.7$$

$$t = \frac{(\bar{x}_2 - \bar{x}_1) - 0}{s_p \sqrt{1/n_1 + 1/n_2}} = \frac{1.4}{2.7 \sqrt{1/10 + 1/10}} = 1.16$$

$$\begin{cases} t_{n_1+n_2-1, 0.975} = t_{19, 0.975} = 2.093 \\ t_{n_1+n_2-1, 0.95} = t_{19, 0.95} = 1.729 \end{cases}$$

Slutsats: $|t| = 1.16 < 1.729$ och $t = 1.16 < 2.093$ så vi förkastar inte

$H_0: \mu_1 - \mu_2 = 0$ för varken $H_1: \mu_2 - \mu_1 = 0$ ($p = 0.26$) eller

$H_1: \mu_2 - \mu_1 > 0$ ($p = 0.13$).

Jämförelse av två populationer

Kan de dra slutsatsen att hönor lägger större ägg om de får äta köttbullar? Vad behöver Pettson och Findus anta?

Experiment 2 (parad data)

$$H_0: \mu = 0 \quad \text{mot} \quad H_1: \mu \neq 0 \text{ (eller } \mu > 0) \quad \alpha = 0.95$$

Höna	1	2	3	4	5	6	7	8	9	10
x_1	52.0	49.8	51.6	50.7	50.2	48.6	54.0	48.7	53.1	47.5
x_2	48.7	52.0	54.4	51.5	52.0	47.9	57.0	51.8	56.3	48.1
d_i	-3.3	2.2	2.8	0.8	1.8	-0.7	3.	3.1	3.2	0.6

$$\bar{d} = 1.35, s = 2.08, n = 10$$

$$t = \frac{\bar{d} - 0}{s/\sqrt{n}} = \frac{1.35}{2.08/\sqrt{10}} = 2.05$$

$$t_{n-1,0.975} = t_{9,0.975} = 2.262 \quad t_{n-1,0.95} = t_{9,0.95} = 1.833$$

Slutsats: $|t| = |2.05| = 2.05 < 2.262$ så vi förkastar inte $H_0: \mu = 0$ om $H_1: \mu \neq 0$ ($p = 0.07$) men eftersom att $t = 2.05 > 1.833$ hade vi förkastat $H_0: \mu = 0$ om $H_1: \mu > 0$ ($p = 0.05$).