Matematisk Statistik och Diskret Matematik, MVE051/MSG810, VT22

Timo Vilkas

Chalmers - Göteborgs Universitet

Föreläsning 9
28 april 2022

Definitioner

- Andra typen av inferentiell statistik är **hypotesprövning**.
- En **hypotes** är ett enkelt påstående/antagande om populationen. (t.ex. fördelningens form, parametrar)
- Ofta gäller en hypotes parametrar av populationens fördelning (medelvärde, varians etc.)
- Hypotesprövning innebär att man försöker avgöra om påståendet passar in på datan eller ej.
- Det finns alltid två hypoteser:
  - **Alternativet** eller **mothypotesen** *(eng. alternative/research hypothesis)* som föreslås av forskaren/experimentatorn. Denna betecknas med H_1 eller H_A.
  - Och **nollhypotesen** *(eng. null hypothesis)* som är i någon mån den förutfattade meningen och är negationen av den alternativa hypotesen. Denna betecknas med H_0.

- Om hypotesen gäller parametern θ kan nollhypotesen t.ex. vara $\theta = \theta_0$, $\theta \geq \theta_0$ eller $\theta \leq \theta_0$ för ett givet värde $\theta_0 \in \mathbb{R}$. Den motsvarande alternativa hypotesen är då $\theta \neq \theta_0$, $\theta < \theta_0$ resp. $\theta > \theta_0$.
- Det man gör är att försöka avgöra hur rimlig nollhypotesen är för den givna datamängden.
- Om den bedöms orimlig, förkastas H_0 och accepteras H_1 (alternativet) istället.
- Likhet med det så kallade noll värdet θ_0 ingår alltid i H_0.
- Hypotesen kan också syfta på fördelningens form, t.ex. H_0: Datan är ett stickprov ur en normalfördelning.

Exempel

1. Allmän uppfattning: "Majoriteten av studenterna på Chalmers kommer från Göteborg." Låt q vara andelen av Chalmers studenter som kommer från Göteborg och omnejd. Då kan vi skriva
$$H_0: \quad q \geq 0.5 \quad \text{och} \quad H_1: \quad q < 0.5$$
2. Ett företag som håller på med opinionsundersökning påstår att färre än 40% av alla röstberättigade skulle gå och rösta om valet var nu på söndag.
Låt r vara andelen av röstberättigade som skulle rösta om valet var nästa helg. Då är det
$$H_0: \; r = 0.4 \quad \text{och} \quad H_1: \; r < 0.4$$

Example

3. Vi kastar ett mynt och undrar om någon har tummat på det, eller om det är ett symmetriskt mynt, som landar på båda sidor med samma sannolikhet precis som den ska. Låt oss skriva $p = \mathbb{P}(\text{klave})$.

$$H_0: \quad p = 0.5 \quad \text{and} \quad H_1: \quad p \neq 0.5$$

Möjliga resultat av en hypotesprövning

- Att vi antingen behåller nollhypotesen eller förkastar den och antar alternativet istället leder till fyra olika fall:
  1. H_0 förkastas trots att den stämmer: **typ I-fel**.
  2. H_0 förkastas och den stämde ej: korrekt.
  3. H_0 förkastas ej trots att H_1 stämmer: **typ II-fel**.
  4. H_0 förkastas ej och H_1 stämmer ej: korrekt.
- Sannolikheten för typ I-fel kallas **signifikans** och betecknas med α. Sannolikheten för typ II-fel betecknas med β.

		Condition of Null Hypothesis	
		True	False
Possible Action	Fail to reject H_0	Correct action	Type II error
	Reject H_0	Type I error	Correct action

Anmärkning

- *I vanliga fall kan man välja/kontrollera sannolikheten för typ I-fel och välja α liten. Däremot är β vanligtvis okänt och svårt att ha kontroll över.*
- *Ett mått på testets effektivitet (eng. power of the test) är $1-\beta$ (som är sannolikheten att H_0 förkastas givet att H_1 stämmer).*

Exempel 8.3.3, 8.3.4, 8.3.5 i boken ger en bättre förståelse av typ I-/typ-II fel och testets effektivitet, men det kommer vi inte fokussera på.

Steg i hypotesprövning

1. **Data:** Vi behöver data (ett stickprov) för att kunna utföra en hypotesprövning.
2. **Antaganden:** Ofta har man några antaganden med avseende på populationen, t.ex. fördelningens typ, motsvarande parametrar (så som μ or σ, etc.), som inte ifrågasätts.
3. **Hypoteser:** Bestäm vilket påstånde/antagande du vill pröva (H_0, nollhypotes) mot ett alternativ (H_1).
4. **Teststatistik:** Välj en statistik som kan beräknas med hjälp av ett stickprov. Utfallet på denna används för att antingen förkasta eller behålla nollhypotesen. Ett exempel på en sådan teststatistik är

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}.$$

Steg i hypotesprövning

5. **Teststatistikens fördelning:** Fördelningen av teststatistiken bestäms *under antagandet att θ_0 är rätt värde för parametern* θ.
6. **Beslutsregel:** Med hjälp av teststatistikens fördelning bestämmer man två regioner: en mängd där nollhypotesen känns som ett orimligt antagande *(eng. rejection/critical region)* och en mängd där nollhypotesen känns rimlig *(eng. non-rejection region)*.
   Beroende på i vilken mängd utfallet på vår teststatistik landar väljer vi att förkasta nollhypotesen (om utfallet ligger i den kritiska regionen) eller inte göra det (om utfallet ligger i "non-rejection" regionen). Hur mängderna väljs beror på både typen av hypotes och **signifikansnivån** (se längre ner).

i de allra flesta fall:
- standard normal
- χ^2 med $(n-1)$ frihetsgr.
- t-fördelning —"—

Hypotesprövning

Steg i hypotesprövning

7. **Beräkna teststatistiken:** Vi använder stickprovet för att beräkna teststatistiken.
8. **Statistiskt beslut:** Vi förkastar eller behåller nollhypotesen beroende på i vilken region utfallet på vår teststatistik landar.
9. **Slutsats:** Om H_0 blir förkastad, drar vi slutsatsen att alternativet H_1 stämmer. Om H_0 inte förkastas, blir slutsatsen att H_0 **kan** stämma.

T. Vilkas Matematisk statistik & diskret matematik, MVE051/MSG810

Hypotesprövning för medelvärdet

Stickprov ur en normalfördelad population med känd varians

Antag att populationen är normalfördelad med okänt medelvärde μ och känd varians σ^2. Vi vill pröva en hypotes om medelvärdet μ.

- Som bekant är stickprovsmedelvärdet i denna situation också normalfördelat (med väntevärde μ och varians σ^2/n).
- Som teststatistik väljer vi

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}.$$

- Z är då standard normalfördelad.

Exempel

Låt oss kika på de olika stegen av hypotesprövning (och val av motsvarande regioner) när det gäller populationens medelvärde genom att gå genom ett enkelt exempel:

Hypotesen att medelvärdet av åldern i en viss population är 30 år ska prövas. Antaganden är att åldern är approximativt normalfördelad med varians $\sigma^2 = 20$. Man tar fram ett stickprov på $n = 10$ individer och räknar ut stickprovsmedelvärdet $\overline{x} = 27$.

Hypoteser:

$$H_0 : \quad \mu = 30 \quad \text{och} \quad H_1 : \quad \mu \neq 30$$ → tvåsidigt test

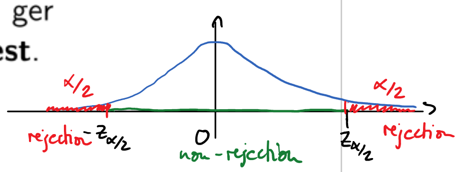
Teststatistik:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

Exempel

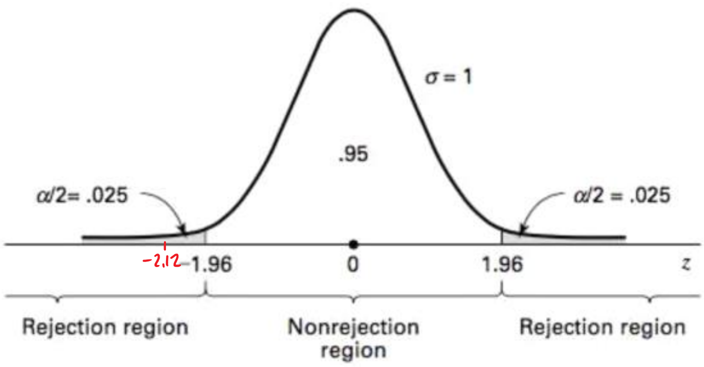
Teststatistikens fördelning: I situationen ovan är Z standard normalfördelad.

Beslutsregel: Antag att testets signifikansnivå väljs som $\alpha = 0.05$. Nollhypotesen känns orimligt om teststatistiken är antingen tillräckligt liten eller tillräckligt stor. Man väljer därför en kritisk region med sannolikhet $\alpha/2$ både för stora och små värden. Sedan gäller det att beräkna motsvarande **kritiska värden**, dvs. trösklar z_1 och z_2 så att $\mathbb{P}(Z < z_1) = 0.025$ och $\mathbb{P}(Z > z_2) = 0.025$. Symmetrin av normalfördelningen ger $z_2 = z_{\alpha/2}$ och $z_1 = -z_2$. Detta kallas ett **två-sidigt test**.



Hypotesprövning

Exempel



Beräkning av teststatistiken: $z = \frac{27-30}{\sqrt{20/10}} = -2.12.$

Statistiskt beslut: Då $-2.12 \leq 1.96$ (z ligger i den kritiska regionen), förkastar vi H_0 med signifikansnivån $\alpha = 0.05$.

Slutsats: Att μ är lika med 30 är ett orimligt antagande.

Hypotesprövning vs. Konfidensintervall

Om man istället beräknar ett 95% konfidensintervall för μ i exemplet ovan, fås

$$\left(27 - 1.96\sqrt{20/10}, 27 + 1.96\sqrt{20/10}\right),$$

alltså

$$\left(24.2286, 29.7714\right)$$

Man ser att medelvärdet $\mu = 30$ inte ligger i intervallet, vilket betyder att vi kan förkasta denna hypotes.

Ensidigt test

Vänstra svansen

Om nollhypotesen i exemplet ovan går ut på att medelvärdet av åldern är större eller lika än 30 istället, använder vi ett **ensidigt** test:

Hypoteser: $H_0: \quad \mu \geq 30 \quad$ och $\quad H_1: \quad \mu < 30$

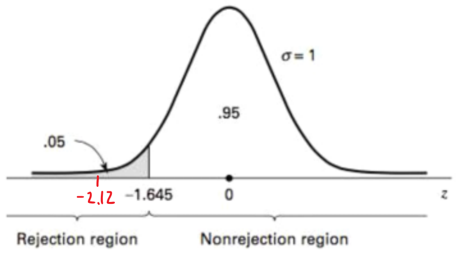
Teststatistiken, dess fördelning och värde, som räknas fram med hjälp av stickprovet, är samma som ovan. Däremot ligger den kritiska regionen nu bara på (för) små värden och ändrar därmed beslutsregeln: Nu kommer bara små värden på teststatistiken leda till att nollhypotesen förkastas.

Låt oss välja samma signifikansnivå $\alpha = 0.05$. Vi bestämmer nu alltså bara en tröskel z_1 så att alla värden $z \leq z_1$ leder till att nollhypotesen förkastas.

Ensidigt test

Vänstra svansen

Värdet z_1 måste då väljas så att $\mathbb{P}(Z \leq z_1) = 0.05$ för $Z \sim \mathcal{N}(0,1)$, alltså $z_1 = z_{1-\alpha} = z_{0.95} = -1.645$.



Detta kallas ett **ensidigt test** med avseende på **vänstra svansen** *(eng. left-tailed test)*.
Då utfallet på teststatistiken $z = -2.12$ ligger i den kritiska regionen, förkastar vi H_0 och kan påstå att $\mu < 30$ med en signifikansnivå på 0.05.

Hypotesprövning

Ensidigt test

Högra svansen

På samma sätt kan vi definera ett **ensidigt test** med avseende på **högra svansen** *(eng. right-tailed test)*, som är en hypotesprövning med avseende på $H_1 : \ \theta > \theta_0$. Om vi väljer i exemplet ovan

$$H_0 : \quad \mu \leq 30 \quad \text{och} \quad H_1 : \quad \mu > 30$$

ligger den kritiska regionen till höger om tröskeln $z_\alpha = 1.645$. Dvs. att vi förkastar H_0 om utfallet på teststatistiken är större än 1.645. I exemplet ovan, där $z = -2.12$, skulle vi alltså behålla H_0, och påstå att medelvärdet rimligtvis är mindre eller lika än 30.

T. Vilkas Matematisk statistik & diskret matematik, MVE051/MSG810

En- eller tvåsidigt test?

$H_0 : \theta = \theta_0$ $H_1 : \theta \neq \theta_0$	$H_0 : \theta \geq \theta_0$ $H_1 : \theta < \theta_0$	$H_0 : \theta \leq \theta_0$ $H_1 : \theta > \theta_0$
tvåsidigt	ensidigt (vänster)	ensidigt (höger)

Som sagt är det en konvention att likhet med nollvärdet ingår i nollhypotesen H_0 (då vi egentligen utgår ifrån att parametern θ är lika med θ_0). Ibland skrivs därför hypoteserna istället som:

$H_0 : \theta = \theta_0$ $H_1 : \theta \neq \theta_0$	$H_0 : \theta = \theta_0$ $H_1 : \theta < \theta_0$	$H_0 : \theta = \theta_0$ $H_1 : \theta > \theta_0$
tvåsidigt	ensidigt (vänster)	ensidigt (höger)

P-värde

Istället för att välja en signifikansnivå och sedan bestämma om nollhypotesen bör förkastas eller ej (med hänsyn till den givna signifikansnivån) kan man bestämma ett så kallad **P-värde** istället. Denna strategi kallas **signifikanstest** *(eng. significance testing)*.

- P-värdet för en test är det minsta värdet för α så att nollhypotesen skulle förkastas.
- Om $\alpha \geq P$ förkastar vi nollhypotesen.
- Om $\alpha < P$ förkastar vi nollhypotesen inte.
- Observerar att P-värdet beror på utfallet av teststatistiken, och ändras därför när man använder ett annat stickprov.
- Hur beräknar man ett P-värde?

Beräkning av P-värdet

Låt z_0 vara utfallet av en teststatistik Z vars fördelning är symmetrisk kring $z = 0$.

- **Tvåsidigt test:** För ett tvåsidigt test ges P-värdet som

$$P = \begin{cases} 2\,\mathbb{P}(Z \leq z_0) & \text{om } z_0 < 0, \\ 2\,\mathbb{P}(Z \geq z_0) & \text{om } z_0 > 0. \end{cases}$$

OBS: Bara för symmetrisk fördelning av teststatistiken, som $Z \sim N(0,1)$ här (men också t.ex. t-fördelning)

- **Ensidigt test:** Om hypotesen avser vänstra svansen ges motsvarande P-värde som

$$P = \mathbb{P}(Z \leq z_0),$$

om det gäller högra svansen blir det istället

$$P = \mathbb{P}(Z \geq z_0).$$

Exempel

Låt oss betrakta exemplet ovan ännu en gång.

- Då teststatistiken Z har en symmetrisk fördelning kring 0 och $z_0 = -2.12$ ges P-värdet för den tvåsidiga hypotesen ($H_0: \ \mu = 30$) som $P = 2\,\mathbb{P}(Z \leq -2.12) = 2 \cdot 0.017 = 0.034$. Då $\alpha = 0.05 > P$, förkastade vi nollhypotesen till denna signifikansnivå.
- I det ensidiga testet med avseende på vänstra svansen är $P = \mathbb{P}(Z \leq -2.12) = 0.017$ som återigen är mindre än α, vilket är anledningen varför $H_0: \ \mu \geq 30$ förkastades.
- I det ensidiga testet med avseende på högra svansen är $P = \mathbb{P}(Z \geq -2.12) = 1 - \mathbb{P}(Z < -2.12) = 1 - 0.017 = 0.983 > 0.05 = \alpha$. Därför valde vi att inte förkasta nollhypotesen $H_0: \ \mu \leq 30$.

Hypotesprövning för medelvärdet

Stickprov ur en normalfördelad population med okänd varians

- Om populationen är normalfördelad med okänd varians, lämpar sig teststatistiken
$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$
för hypotesprövning med avseende på medelvärdet.
- Som bekant har denna då en t-fördelning (med $n-1$ frihetsgrader, där n är stickprovets storlek).
- För given signifikansnivå α är tröskelvärden för ett tvåsidigt test $t_{\alpha/2}$ och $t_{(1-\alpha/2)}$
- För ett ensidigt test med avseende på vänstra eller högra svansen blir de $t_{1-\alpha}$ resp. t_α.
- Alla andra steg av hypotesprövningen är samma som för känd varians.
- Också P-värdet beräknas på samma sätt: Bara Z ersätts av T (och dess utfall är då t_0 istället för z_0).

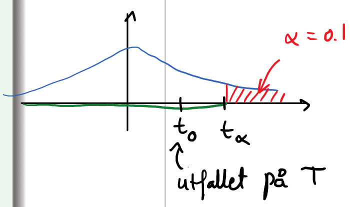
Exempel

När det gäller nivån på mikrovågstrålning man utsätts för anses i staterna 10 mikrowatt per kvadratcentimeter som tröskeln till oacceptabelt. I en by nära en stor tv programm sändare misstänks strålningen vara starkare än det som anses ofarligt. Ett stickprov på 25 mätningar tas fram med stickprovsmedelvärde $\overline{x} = 10.3$ och stickprovsvarians $s = 2$. Man antar att strålningsnivån är normalfördelad, dock både μ och σ^2 okända.
Genomför en hypotesprövning med signifikansnivå 10% för att avgöra om sändaren ger upphov till skadliga strålningsnivåer eller inte, alltså $H_0 : \mu \leq 10$ och $H_1 : \mu > 10$.

- Uppenbarligen handlar det sig om ett ensidigt test (med avseende på högra svansen).
- Då variansen är okänd använder vi $T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$ som teststatistik.

Example

- Utfallet av teststatistiken är $t_0 = \frac{\overline{x}-\mu}{s/\sqrt{n}} = \frac{10.3-10}{2/\sqrt{25}} = 0.75$.
- Tröskelvärdet ges av $t_\alpha = t_{0.1} = 1.318$.
- Då $t_0 < t_{0.1}$ ligger utfallet inte i den kritiska regionen, vilket leder till att vi inte kan förkasta H_0. Dvs. att medelvärdet mycket väl kan vara mindre eller lika 10 och att stickprovet inte stödjer alternativet.
- Motsvarande P-värde ges av $P = \mathbb{P}(T > 0.75)$. Tabell VI (s.699) är inte detaljerad nog för att ge ett exakt t-värde, men i raden med $\gamma = 24$ (frihetsgrader), kan man avläsa att 0.75 ligger mellan $t = 0.685$ och $t = 1.318$ som motsvarar $\mathbb{P}(T \leq t) = 0.75$ resp. 0.9. P-värdet ligger alltså mellan $0.1 = 1 - \mathbb{P}(T \leq 1.318)$ och $0.25 = 1 - \mathbb{P}(T \leq 0.685)$. Då $P > \alpha = 0.1$ kan vi inte förkasta H_0.



Hypotesprövning för medelvärdet

Stickprov ur en icke-normalfördelad population

- Om fördelningen av populationen inte är en normalfördelning eller okänd, dock n (stickprovets storlek) tillräckligt stort, kan vi använda centrala gränsvärdessatsen för att dra slutsatsen att $\overline{X}$ är approximativt normalfördelad med väntevärde μ och varians σ^2/n.
- För tillräckligt stora n kan vi alltså genomföra en hypotesprövning på samma sätt som ovan (där population var normalfördelad).