10 Hypotesetesting

OPPGAVEHEFTE · STATISTIKK

Overskriftene viser til delkapitlene i boka. Først kommer en oppsummering av hele kapittelet. Under hver overskrift står gjennomregnede eksempler med løsning, og deretter oppgaver du skal løse uten hjelpemidler. Fasiten står bakerst. Bruk tabellene i bilag A og $\alpha = 0{,}05$ der ikke annet står.

Formelark

HYPOTESETESTER

Fem steg Hypoteser, signifikansnivå, testobservator, p-verdi, konklusjon i konteksten.

Binomisk test $X \sim B(n, p_0)$, p-verdi $P(X \geqslant x)$ eller $P(X \leqslant x)$

Test for andel

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

Test for gjennomsnitt

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Tosidig p-verdi

$$2 \cdot P(Z \geqslant |z|)$$

Kritiske verdier ($\alpha = 0{,}05$) Ensidig: 1,645. Tosidig: 1,96.

> - Forkast H_0 når p-verdien $\leqslant \alpha$.
> - Velg H_1 og α før du ser på dataene.

TOLKNING

p-verdi Sannsynligheten for et minst like ekstremt resultat når H_0 er sann. Ikke sannsynligheten for at H_0 er sann.

Type I og type II Forkaste sann H_0 (sannsynlighet α), beholde usann H_0.

> - Å beholde H_0 beviser den ikke.
> - Statistisk signifikant betyr ikke stort eller viktig.
> - Mange tester gir signifikante resultater ved en tilfeldighet.

10.1.1 Kan det være tilfeldig?

10.1 Hva er en nullhypotese?

10.2 Skriv H_0 og H_1.

a) Sara påstår at hun kan gjette fargen på et kort (rødt eller svart) bedre enn tilfeldig.

b) En lege tror at færre enn 30 % av pasientene får bivirkninger av et nytt middel.

10.3 Hva er en p-verdi?

10.4 I 10 000 simuleringer under H_0 ga 230 et resultat minst like ekstremt som det observerte. Hva er p-verdien omtrent?

10.5 p-verdien er 0,31. Hva konkluderer du?

10.6 Jonas sier: «p-verdien er 0,03, så det er 3 % sannsynlighet for at H_0 er sann.» Hva er feil?

10.7 Hva blir utskriften?

Python

```
sim = [3, 6, 5, 7, 2, 5, 4, 6, 8, 5]
antall = 0
for s in sim:
    if s >= 7:
        antall = antall + 1
print(antall / len(sim))
```

10.8 Hvorfor betyr ikke en stor p-verdi at H_0 er bevist?

10.1.2 Hypotesetest med binomisk fordeling

EKSEMPEL *Colatesten*

Gjennomfør testen av Ida med $\alpha = 0{,}05$. Hun fikk 8 riktige av 10.

Løsning

1. H_0: $p = 0{,}5$ (Ida gjetter). H_1: $p > 0{,}5$ (hun kan smake forskjell).
2. $\alpha = 0{,}05$.
3. $X \sim B(10; 0{,}5)$ under H_0, og $x = 8$.
4. $P(X \geqslant 8) = 1 - P(X \leqslant 7)$, som er $1 - 0{,}9453 = 0{,}0547$.
5. $0{,}0547 > 0{,}05$, så vi beholder H_0. Testen gir ikke nok bevis for at Ida kan smake forskjell, men resultatet er nær grensen. En ny test med flere glass vil gi et klarere svar.

■

EKSEMPEL *Defekte varer*

En produsent sier at høyst 10 % av varene er defekte. En kunde kontrollerer 20 tilfeldige varer og finner 5 defekte. Test påstanden med $\alpha = 0{,}05$.

Løsning

1. H_0: $p = 0{,}1$. H_1: $p > 0{,}1$.
2. $\alpha = 0{,}05$.
3. $X \sim B(20; 0{,}1)$ under H_0, og $x = 5$.
4. $P(X \geqslant 5) = 1 - P(X \leqslant 4)$, som er $1 - 0{,}9568 = 0{,}0432$.
5. $0{,}0432 \leqslant 0{,}05$, så vi forkaster H_0. Kontrollen tyder på at mer enn 10 % av varene er defekte.

■

EKSEMPEL *Færre feil*

Med den gamle metoden feiler halvparten av forsøkene i et laboratorium. En ny metode blir prøvd 15 ganger og feiler 3 ganger. Er den nye metoden bedre? Bruk $\alpha = 0{,}05$.

Løsning

1. H_0: $p = 0{,}5$. H_1: $p < 0{,}5$.
2. $\alpha = 0{,}05$.
3. $X \sim B(15; 0{,}5)$ under H_0, og $x = 3$.
4. $P(X \leqslant 3) = 0{,}0176$ (tabellen).
5. $0{,}0176 \leqslant 0{,}05$. Vi forkaster H_0: Den nye metoden ser ut til å gi færre feil.

■

10.9 Sara gjetter fargen på 12 kort og får 10 riktige. Test H_0: $p = 0{,}5$ mot H_1: $p > 0{,}5$.

10.10 I et forsøk med 8 myntkast fikk Ola 7 kron. Test om mynten gir for mye kron.

10.11 Normalt gir 30 % av pasientene en bivirkning. Med et nytt middel får 8 av 15 bivirkningen. Test om andelen er høyere.

10.12 Tidligere strøk 20 % på en prøve. Etter et nytt opplegg strøk 1 av 20. Test om andelen som stryker, har gått ned.

10.13 Høyst 5 % av varene skal være defekte. I en kontroll av 20 varer er 3 defekte. Test påstanden.

10.14 Skriv de fem stegene i en hypotesetest.

10.15 $X \sim B(20; 0{,}5)$ under H_0, og H_1 er $p > 0{,}5$. Hvor mange suksesser må vi minst få for å forkaste H_0?

10.16 Hva blir utskriften?

Python

```
from scipy.stats import binom
print(round(1 - binom.cdf(9, 12, 0.5), 4))
```

10.17 Hvorfor skal H_1 og α velges før man ser på dataene?

10.1.3 Hypotesetest med normalfordeling

EKSEMPEL *Flere gutter enn jenter*

I 2025 ble det født 28 388 gutter av 55 401 barn. Test H_0: $p = 0{,}5$ mot H_1: $p > 0{,}5$ med $\alpha = 0{,}05$.

Løsning

$\hat{p} = \frac{28\,388}{55\,401} \approx 0{,}5124$, og $\sqrt{\frac{0{,}5 \cdot 0{,}5}{55\,401}} \approx 0{,}00212$.

$$z = \frac{0{,}5124 - 0{,}5}{0{,}00212} \approx 5{,}84$$

Tabellen slutter ved 3,49, så p-verdien er under 0,0003 (Python gir $2{,}6 \cdot 10^{-9}$). Vi forkaster H_0: Det blir født flere gutter enn jenter, og det skyldes ikke tilfeldigheter. ∎

EKSEMPEL *Tromsø*

I Tromsø ble det født 431 gutter av 854 barn. Test H_0: $p = 0{,}5$ mot H_1: $p > 0{,}5$ med $\alpha = 0{,}05$.

Løsning

$z = \frac{0{,}5047 - 0{,}5}{\sqrt{0{,}25/854}} = \frac{0{,}0047}{0{,}0171} \approx 0{,}27$. p-verdien er $1 - 0{,}6064 = 0{,}3936$.

Vi beholder H_0. Dataene fra Tromsø alene gir ikke grunnlag for å si at det blir født flere gutter enn jenter. Det betyr ikke at $p = 0{,}5$ i Tromsø: Utvalget er for lite til å oppdage en så liten forskjell (10.2.2). ∎

EKSEMPEL *Melkekartonger*

Et meieri sier at kartongene inneholder 1000 ml i gjennomsnitt. Forbrukerrådet måler 50 tilfeldige kartonger og finner $\bar{x} = 998{,}6$ ml og $s = 4{,}2$ ml. Test om kartongene inneholder for lite, med $\alpha = 0{,}05$.

Løsning

1. H_0: $\mu = 1000$. H_1: $\mu < 1000$.
2. $\alpha = 0{,}05$.
3. $\text{SE} = \frac{4{,}2}{\sqrt{50}} \approx 0{,}594$ og $z = \frac{998{,}6 - 1000}{0{,}594} \approx -2{,}36$.
4. p-verdi: $P(Z \leqslant -2{,}36) = 1 - 0{,}9909 = 0{,}0091$.
5. $0{,}0091 \leqslant 0{,}05$. Vi forkaster H_0: Kartongene inneholder i gjennomsnitt mindre enn 1000 ml.

∎

10.18 220 av 400 spurte foretrekker merke A. Test om flere enn halvparten foretrekker A.

10.19 I en by med mange velgere sier 282 av 600 spurte ja til et forslag. Test om andelen er mindre enn 50 %.

10.20 Tidligere brukte 25 % av elevene sykkel til skolen. I en undersøkelse blant 300 elever bruker 60 sykkel. Test om andelen har gått ned.

10.21 En maskin skal fylle 100 g. 64 pakker gir $\bar{x} = 103$ g og $s = 12$ g. Test om maskinen fyller for mye.

10.22 36 batterier har $\bar{x} = 48{,}5$ timer og $s = 6$. Produsenten sier $\mu = 50$. Test om batteriene varer kortere.

10.23 Hvorfor bruker vi p_0 og ikke $\hat{p}$ i standardfeilen i testen for andel?

10.24 Hvilke forutsetninger må være oppfylt for z-testen for andel?

10.25 Hva blir utskriften?

Python

```
xbar = 103
mu0 = 100
s = 12
n = 64
print((xbar - mu0) / (s / n**0.5))
```

10.26 Kan du bruke z-testen for andel når $n = 30$ og $p_0 = 0{,}1$? Hva gjør du i så fall?

10.2.1 Signifikansnivå og ensidige og tosidige tester

EKSEMPEL *Er mynten skjev?*

100 kast ga 59 kron. Test H_0: $p = 0{,}5$ mot H_1: $p \neq 0{,}5$ med $\alpha = 0{,}05$. Hva ville konklusjonen vært med en ensidig test?

Løsning

$z = \frac{0{,}09}{\sqrt{0{,}25/100}} = \frac{0{,}09}{0{,}05} = 1{,}8$.

Tosidig: p-verdien er $2 \cdot (1 - 0{,}9641) = 0{,}0718 > 0{,}05$. Vi beholder H_0.

Ensidig ($p > 0{,}5$): p-verdien er $0{,}0359 < 0{,}05$, og vi ville forkastet H_0. Det viser hvorfor retningen må velges på forhånd. Uten en grunn til å vente for mye kron er det den tosidige testen som er riktig. ∎

EKSEMPEL *Valg av signifikansnivå*

Et legemiddelfirma tester om et nytt middel virker bedre enn det gamle. Hvorfor kan det være lurt å bruke $\alpha = 0{,}01$ i stedet for 0,05?

Løsning

Med $\alpha = 0{,}05$ forkaster vi en sann H_0 i 5 % av testene. Da kan et middel som ikke er bedre, bli tatt i bruk, med kostnader og mulige bivirkninger. Når en slik feil er alvorlig, krever vi sterkere bevis, altså et lavere signifikansnivå. Prisen er at det blir

vanskeligere å oppdage et middel som faktisk er bedre (10.2.2).

■

10.27 Når bruker vi en tosidig test?

10.28 $z = 2{,}5$. Finn p-verdien for en ensidig og en tosidig test.

10.29 $z = -1{,}75$ i en tosidig test. Forkaster du H_0?

10.30 Hva er de kritiske verdiene for $\alpha = 0{,}05$, ensidig og tosidig?

10.31 Et 95 % konfidensintervall for en andel er $[0{,}52; 0{,}60]$. Hva blir konklusjonen av en tosidig test av H_0: $p = 0{,}5$?

10.32 Finn den kritiske verdien for en tosidig test med $\alpha = 0{,}10$.

10.33 Kim velger ensidig test etter å ha sett at dataene gikk i én retning. Hva er problemet?

10.34 Hva blir utskriften?

Python

```
from scipy.stats import norm
z = 2.5
print(round(2 * (1 - norm.cdf(z)), 4))
```

10.2.2 Feil og tolkning

EKSEMPEL *Kan Ida smake forskjell?*

Anta at Ida faktisk kan smake forskjell og svarer riktig med sannsynlighet 0,8. Med 10 glass og $\alpha = 0{,}05$ forkaster vi H_0: $p = 0{,}5$ bare hvis hun får minst 9 riktige ($P(X \geqslant 9) = 0{,}0107$ under H_0, mens $P(X \geqslant 8) = 0{,}0547$ er for stor). Hva er sannsynligheten for type II-feil?

Løsning

Med $p = 0{,}8$ er $P(X \geqslant 9) = 0{,}376$ (regnet med Python). Testen oppdager altså evnen hennes bare i 38 % av tilfellene. Sannsynligheten for type II-feil er $1 - 0{,}376 = 0{,}624$.

Med 20 glass (forkast ved minst 15) blir sannsynligheten for å oppdage evnen 0,80, og med 30 glass (forkast ved minst 20) 0,97. Sannsynligheten for å oppdage en virkelig effekt kalles testens **styrke**. Små tester har liten styrke.

■

EKSEMPEL *Signifikant, men lite*

Testen i 10.1.3 viste at det blir født flere gutter enn jenter ($z = 5{,}84$). Melkekartongtesten viste at kartongene inneholder mindre enn 1000 ml ($p = 0{,}009$). Er resultatene viktige?

Løsning

Begge er statistisk signifikante, men avvikene er små. Andelen gutter er 51,2 % i stedet for 50 %. Det er en kjent biologisk effekt, men den betyr lite i praksis. Kartongene inneholder i gjennomsnitt 1,4 ml for lite, under en teskje. Det er neppe viktig for kundene, selv om meieriet bør justere maskinen.

Med svært store utvalg blir selv bittesmå avvik signifikante. Vurder derfor alltid størrelsen på avviket, gjerne med et konfidensintervall.

■

EKSEMPEL *Gutter i kommunene*

En journalist tester for hver av de 356 kommunene med fødsler i 2025 om andelen gutter er forskjellig fra landsgjennomsnittet 0,514, med $\alpha = 0{,}05$. Hun finner 15 kommuner med signifikant avvik og skriver: «I disse 15 kommunene blir det født uvanlig mange eller få gutter.» Vurder.

Løsning

Hvis sannsynligheten for gutt er 0,514 i alle kommunene, venter vi at omtrent 5 % av 356, altså rundt 18 kommuner, får signifikant avvik av rene tilfeldigheter. 15 er akkurat det vi venter. Artikkelen viser bare tilfeldig variasjon (som i 6.2.3). Med mange tester er det nesten sikkert at noen blir signifikante: Med 20 uavhengige tester av sanne nullhypoteser er sannsynligheten for minst én signifikant $1 - 0{,}95^{20} \approx 0{,}64$.

■

10.35 Hva er en type I-feil og en type II-feil?

10.36 Hva er sannsynligheten for type I-feil når $\alpha = 0{,}05$ og H_0 er sann?

10.37 Hvordan kan vi gjøre både type I- og type II-feil mindre sannsynlige?

10.38 En brannalarm skal teste H_0: «Det brenner ikke.» Hva er type I- og type II-feil her? Hvilken er verst?

10.39 Et resultat er statistisk signifikant. Betyr det at effekten er stor?

10.40 En forsker tester 20 uavhengige hypoteser der alle nullhypotesene er sanne, med $\alpha = 0{,}05$. Hva er sannsynligheten for minst ett signifikant resultat?

10.41 Hva er p-hacking, og hvorfor er det et problem?

10.42 En liten studie med 12 deltakere finner ingen signifikant effekt av et treningsprogram. Kan vi konkludere med at programmet ikke virker?

10.2.3 Undersøkelse: Hypotesetest på reelle data

10.43 102 av 357 kommuner har et folketall som begynner på 1. Test H_0: $p = \frac{1}{9}$ mot H_1: $p > \frac{1}{9}$.

10.44 Test om andelen i oppgaven over er forskjellig fra Benfords 0,301.

10.45 I 2025 ble det født 28 388 gutter av 55 401 barn. Test om andelen gutter er forskjellig fra 0,5, og vurder om resultatet er praktisk viktig.

10.46 Hvorfor er det problematisk å teste alle 356 kommunene hver for seg og rapportere de som har signifikant avvik i andelen gutter?

Fasit

10.1.1 Kan det være tilfeldig?

10.1
Påstanden om at det ikke er noen effekt, at resultatet bare skyldes tilfeldigheter. Den skrives med likhetstegn, for eksempel H_0: $p = 0{,}5$.

10.2
a) H_0: $p = 0{,}5$. H_1: $p > 0{,}5$.
b) H_0: $p = 0{,}3$. H_1: $p < 0{,}3$.

10.3
Sannsynligheten for å få et resultat som er minst like ekstremt som det observerte, hvis H_0 er sann.

10.4
$\frac{230}{10\,000} = 0{,}023$

10.5
Resultatet kan godt skyldes tilfeldigheter. Vi beholder H_0.

10.6
p-verdien er sannsynligheten for dataene (eller noe mer ekstremt) gitt at H_0 er sann, ikke sannsynligheten for at H_0 er sann.

10.7
0.2

10.8
En stor p-verdi betyr bare at dataene passer med H_0. Det kan være en virkelig effekt som utvalget er for lite til å oppdage.

10.1.2 Hypotesetest med binomisk fordeling

10.9
$X \sim B(12; 0{,}5)$. p-verdi: $P(X \geqslant 10) = 1 - 0{,}9807 = 0{,}0193$. Forkast H_0: Sara ser ut til å gjette bedre enn tilfeldig.

10.10
H_0: $p = 0{,}5$. H_1: $p > 0{,}5$. p-verdi: $P(X \geqslant 7) = 1 - 0{,}9648 = 0{,}0352$. Forkast H_0.

10.11
H_0: $p = 0{,}3$. H_1: $p > 0{,}3$. p-verdi: $P(X \geqslant 8) = 1 - 0{,}9500 = 0{,}0500$. $0{,}05 \leqslant 0{,}05$, så vi forkaster H_0, men det er helt på grensen.

10.12
H_0: $p = 0{,}2$. H_1: $p < 0{,}2$. p-verdi: $P(X \leqslant 1) = 0{,}0692$. Behold H_0. Det er ikke nok bevis for en nedgang.

10.13
H_0: $p = 0{,}05$. H_1: $p > 0{,}05$. p-verdi: $P(X \geqslant 3) = 1 - 0{,}9245 = 0{,}0755$. Behold H_0.

10.14
Hypoteser, signifikansnivå, testobservator og fordelingen under H_0, p-verdi, konklusjon i konteksten.

10.15
$P(X \geqslant 14) = 0{,}0577$ og $P(X \geqslant 15) = 0{,}0207$. Minst 15.

10.16
0.0193, altså $P(X \geqslant 10)$ for $X \sim B(12; 0{,}5)$.

10.17
Ellers kan man velge retning og grense som gir det resultatet man ønsker, og sannsynligheten for å forkaste en sann H_0 blir større enn α.

10.1.3 Hypotesetest med normalfordeling

10.18
$\hat{p} = 0{,}55$. $z = \frac{0{,}05}{\sqrt{0{,}25/400}} = 2$. p-verdi $1 - 0{,}9772 = 0{,}0228$. Forkast H_0.

10.19
$\hat{p} = 0{,}47$. $z = \frac{-0{,}03}{\sqrt{0{,}25/600}} \approx -1{,}47$. p-verdi $1 - 0{,}9292 = 0{,}0708$. Behold H_0.

10.20
$\hat{p} = 0{,}2$. $z = \frac{-0{,}05}{\sqrt{0{,}1875/300}} = -2$. p-verdi 0,0228. Forkast H_0.

10.21
$z = \frac{3}{12/8} = 2$. p-verdi 0,0228. Forkast H_0.

10.22
$z = \frac{-1{,}5}{1} = -1{,}5$. p-verdi $1 - 0{,}9332 = 0{,}0668$. Behold H_0.

10.23
Testen regner ut hvor sannsynlig dataene er hvis H_0 er sann, og da er $p = p_0$.

10.24
Tilfeldig utvalg, $np_0 \geqslant 10$ og $n(1 - p_0) \geqslant 10$.

10.25
2.0

10.26
Nei, $np_0 = 3$ er for lite. Bruk binomisk test.

10.2.1 Signifikansnivå og ensidige og tosidige tester

10.27
Når avvik i begge retninger er interessante, og vi ikke har grunn på forhånd til å vente et avvik i én bestemt retning.

10.28
Ensidig: $1 - 0{,}9938 = 0{,}0062$. Tosidig: 0,0124.

10.29
p-verdi $2 \cdot (1 - 0{,}9599) = 0{,}0802$. Nei, behold H_0.

10.30
1,645 ensidig og 1,96 tosidig.

10.31
0,5 ligger utenfor intervallet. Forkast H_0.

10.32
0,05 i hver hale, så $z = 1{,}645$.

10.33
p-verdien blir halvparten så stor som den burde, og sannsynligheten for å forkaste en sann H_0 blir dobbelt så stor som α.

10.34
0.0124

10.2.2 Feil og tolkning

10.35
Type I: forkaste en sann H_0. Type II: beholde en usann H_0.

10.36
0,05

10.37
Ved å bruke et større utvalg.

10.38
Type I: Alarmen går uten at det brenner (falsk alarm). Type II: Alarmen går ikke når det brenner. Type II er verst, så alarmer lages følsomme, med mange falske alarmer.

10.39
Nei. Med store utvalg blir også små effekter signifikante. Størrelsen må vurderes for seg.

10.40
$1 - 0{,}95^{20} \approx 0{,}64$

10.41
Å teste mange hypoteser (eller analysere data på mange måter) og bare rapportere de signifikante resultatene. Da blir mange av funnene type I-feil, og leseren tror resultatene er sterkere enn de er.

10.42
Nei. Med så få deltakere er styrken liten, og en virkelig effekt kan lett bli oversett (type II-feil).

10.2.3 Undersøkelse: Hypotesetest på reelle data

10.43
$\hat{p} \approx 0{,}286$. $z = \dfrac{0{,}286 - 0{,}111}{\sqrt{0{,}111 \cdot 0{,}889/357}} \approx$ 10,5. Forkast H_0.

10.44
$$z = \frac{0{,}286 - 0{,}301}{\sqrt{0{,}301 \cdot 0{,}699/357}} \approx -0{,}63.$$
Tosidig p-verdi $2 \cdot (1 - 0{,}7357) \approx 0{,}53$. Behold H_0: Dataene passer med Benfords lov for sifferet 1.

10.45
$\hat{p} \approx 0{,}5124$. $z \approx 5{,}84$. p-verdien er svært liten, så vi forkaster H_0. Avviket er 1,2 prosentpoeng, som er lite i praksis, men en kjent og stabil effekt.

10.46
Med mange tester blir omtrent 5 % signifikante av rene tilfeldigheter, rundt 18 kommuner. De signifikante kommunene kan være type I-feil.