HD Dataanalyse

Note 6: Analyse af én gruppe (hypotesetest)

Copenhagen Business School

Emne i dette notesæt

Når vi analyserer en variabel i et datamateriale, er formålet at nå frem til klare og letforståelige konklusioner om variablen.

Note 6 ser på ***hvilke konklusioner vi kan drage om en variabel*** ved at opstille en ***metode bestående af 5 trin***, der giver mulighed for at formulere og efterprøve forskellige påstande om variablen.

Idéen er først at formulere en teoretisk model, der giver en generel beskrivelse af variablen (Trin 1: *"Antagelser"*), og dernæst på baggrund af modellen formulere en specifik påstand om variablen (Trin 2: *"Hypoteser"*).

Herefter kan vi efterprøve påstanden ved at sammenholde vores datamateriale med, hvordan variablen ifølge den teoretiske model burde opføre sig, hvis ellers den formulerede påstand er korrekt (Trin 3: *"Teststørrelse"*).

Ved at se på om der er stor eller lille forskel på datamateriale og model (Trin 4: *"P-værdi"*), kan vi slutteligt konkludere, om der er belæg for den formulerede påstand eller ej (Trin 5: *"Konklusion"*).

Emne i dette notesæt

Eksempel: Ølsalg

Ser vi på prisen for 1 stk. Grøn Tuborg (33 cl glasflaske) i supermarkedskæden Føtex, så kunne et eksempel på en konklusion om denne variabel være:

"Den forventede pris på 1 stk. Grøn Tuborg i Føtex er 3,50 kr."

Spørgsmålet er, om vores datamateriale bestående af 157 ugentlige observationer af prisen på Grøn Tuborg i Føtex understøtter denne påstand? Eller om datamaterialet måske snarere indikerer, at påstanden er forkert?

Vi har tidligere set, at vores bedste gæt på den forventede pris på Grøn Tuborg i Føtex er $\hat{\mu} = 3,44$ kr. Det gæt er imidlertid behæftet med usikkerhed, og afhængig af størrelsen af denne usikkerhed er spørgsmålet, hvad vi reelt kan sige om vores forventning til prisen på Grøn Tuborg?

Hvilke påstande om forventningen til prisen på Grøn Tuborg er i overensstemmelse med vores datamateriale? Er der eksempelvis på baggrund af datamaterialets 157 observationer belæg for at hævde, at den forventede pris på Grøn Tuborg...

- er 3,40 kr.?
- højest er 3,60 kr.?
- mindst er 3,40 kr.?

Det vi skal se på i note 6 er, hvordan vi kan afgøre, om der er belæg for en given påstand – i dette eksempel om forventningen til prisen på Grøn Tuborg – på baggrund af observationerne i vores datamateriale.

Emne i dette notesæt

Eksempel: Skat

I spørgeskemaundersøgelsen om velfærd og skat har vi tidligere set på svarene på spørgsmålet *"Er topskatten for høj?"*. Et eksempel på en konklusion om denne variabel kunne være:

"Andelen af danskere, der mener at topskatten er for høj, er 35%"

Spørgsmålet er, om vores datamateriale bestående af svar fra 975 personer understøtter denne påstand? Eller om datamaterialet måske snarere indikerer, at påstanden er forkert?

Vi har tidligere set, at vores bedste gæt på andelen af befolkningen, der mener topskatten er for høj, er $\hat{p} = 35{,}5\%$. Det gæt er imidlertid behæftet med usikkerhed, og afhængig af størrelsen af denne usikkerhed er spørgsmålet, hvad vi reelt kan sige om andelen af befolkningen, der mener at topskatten er for høj?

Hvilke påstande om denne andel er i overensstemmelse med vores datamateriale? Er der eksempelvis på baggrund af datamaterialets 975 observationer belæg for at hævde, at andelen af befolkningen, der mener topskatten er for høj...

- er 30%?
- højest er 40%?
- mindst er 30%?

Det vi skal se på i note 6 er, hvordan vi kan afgøre, om der er belæg for en given påstand – i dette eksempel om andelen af befolkningen, der mener topskatten er for høj – på baggrund af observationerne i vores datamateriale.

Indholdsfortegnelse

METODEN BAG HYPOTESETEST

Hypotesetest om en middelværdi μ i en normalfordeling og hypotesetest om en andel p i en binomialfordeling foregår efter nøjagtig ***samme metode***. Vi beskriver derfor i dette afsnit for overskueligheden skyld blot metoden med udgangspunkt i test af hypoteser om middelværdien μ.

Når vi skal undersøge en påstand om den forventede værdi μ af en normalfordelt variabel $N(\mu, \sigma)$, tager vi udgangspunkt i et datamateriale bestående af n observationer. Formålet er – på baggrund af datamaterialet – at undersøge en specifik påstand om den ukendte middelværdi μ.

Metoden kaldes **hypotesetest om μ** og består af følgende 5 trin:

- TRIN 1: *Antagelser* *("assumptions")*
- TRIN 2: *Hypoteser* *("hypotheses")*
- TRIN 3: *Teststørrelse* *("test statistic")*
- TRIN 4: *P-værdi* *("P-value")*
- TRIN 5: *Konklusion* *("conclusion")*

Eksempel: Ølsalg

Lad os her, som illustration af metoden, undersøge påstanden om, at den forventede pris μ på Grøn Tuborg i Føtex er 3,50 kr.

Trin 1: Antagelser

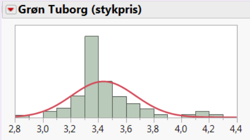
Vi antager, at vores datamateriale består af n indbyrdes uafhængige observationer, der alle stammer fra den samme normalfordeling $N(\mu, \sigma)$.

I praksis kan vi undersøge om denne antagelse er rimelig ved at tegne et histogram af datamaterialet og se på, om histogrammet har en "klokkeform".

Eksempel: Ølsalg

Vi antager her, at de 157 ugentlige observationer af prisen på Grøn Tuborg i Føtex er indbyrdes uafhængige og alle kan beskrives af den samme normalfordeling $N(\mu, \sigma)$.

Som vi tidligere har set, viser et histogram af de 157 prisobservationer, at normalfordelingen med en vis rimelighed kan bruges til beskrivelse af prisen på Grøn Tuborg i Føtex.



TRIN 2: HYPOTESER

Den påstand vi ønsker at undersøge om den ukendte middelværdi μ udtrykker vi ved hjælp af den såkaldte **nulhypotese** H_0 *("null hypothesis")*.

Hypoteser formuleres altid i par, dvs. at nulhypotesen altid suppleres af en såkaldt **alternativhypotese** H_a *("alternative hypothesis")*, som udtrykker det modsatte af nulhypotesen. Så snart man har formuleret sin nulhypotese, så er alternativhypotesen automatisk givet som "det modsatte af nulhypotesen".

Idéen er, at enten er datamaterialet i overensstemmelse med nulhypotesen, eller også er det i overensstemmelse med alternativhypotesen. Hvorvidt det ene eller det andet er tilfældet afgør vi ved at se på en såkaldt teststørrelse.

Eksempel: Ølsalg

Påstanden om, at den forventede pris μ på Grøn Tuborg i Føtex er 3,50 kr., kan vi udtrykke ved at angive henholdsvis nul- og alternativhypotese som

$$H_0: \ \mu = 3{,}50 \qquad\qquad H_a: \ \mu \neq 3{,}50$$

Nulhypotesen (H_0) udtrykker vores påstand om den forventede pris på Grøn Tuborg, mens alternativhypotesen (H_a) udtrykker det modsatte af vores påstand.

Trin 3: Teststørrelse

Teststørrelsen bruger vi til at sammenligne vores datamateriale med den formulerede påstand H_0. Vi ønsker at vurdere, om datamaterialet ser ud til at være i overensstemmelse med påstanden eller ej.

Som teststørrelse bruger vi udtrykket

$$Z = \frac{\hat{\mu} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}}$$

Her repræsenterer $\hat{\mu}$ det bedste gæt på værdien af den ukendte middelværdi ***baseret på datamaterialet***, mens μ repræsenterer den sande middelværdi ***baseret på den formulerede påstand***.

Vi tænker dermed på tælleren $\hat{\mu} - \mu$ i teststørrelsen Z som et mål for om ***datamateriale*** og ***påstand*** er i overensstemmelse med hinanden eller ej:

$$\hat{\mu} - \mu = \text{“datamaterialets gæt på } \mu\text{”} \; - \; \text{“den påståede værdi af } \mu\text{”}$$

Trin 3: Teststørrelse

Hvis påstanden H_0 er sand (falsk), vil vi forvente, at $\hat{\mu}$ og μ ligger forholdsvis tæt på (langt fra) hinanden.

En lille (stor) værdi af Z er derfor udtryk for, at der er lille (stor) forskel på vores datamateriale og den formulerede påstand.

Normeringen $\frac{\hat{\sigma}}{\sqrt{n}}$ i vores teststørrelse

$$Z = \frac{\hat{\mu} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}}$$

kan man tænke på som en teknisk bekvemmelighed, der gør, at vi ved hvilken fordeling, der beskriver teststørrelsen Z, og at vi dermed kan regne sandsynligheder ud om Z.

For at vurdere om den beregnede værdi af teststørrelsen Z er lille eller stor, dvs. om datamaterialet er i god eller dårlig overenstemmelse med påstanden H_0, har vi behov for at kunne afgøre, hvornår en værdi af Z skal opfattes som "lille" henholdsvis "stor". Det gør vi ved hjælp af den såkaldte P-værdi.

Trin 3: Teststørrelse

Eksempel: Ølsalg

I tilfældet med prisen på Grøn Tuborg i Føtex estimeres normalfordelingens parametre til

$$\hat{\mu} = 3,438 \qquad \hat{\sigma} = 0,230$$

på baggrund af et datamateriale bestående af $n = 157$ observationer.

Summary Statistics	
Mean	3,4381017
Std Dev	0,2299489
Std Err Mean	0,0183519
Upper 95% Mean	3,474352
Lower 95% Mean	3,4018514
N	157

Under forudsætning af at nulhypotesen

$$H_0 : \mu = 3,50$$

er sand, kan værdien af teststørrelsen beregnes til

$$Z = \frac{\hat{\mu} - \mu_0}{\frac{\hat{\sigma}}{\sqrt{n}}} = \frac{3,438 - 3,50}{\frac{0,230}{\sqrt{157}}} = -3,37$$

Test Mean	
Hypothesized Value	3,5
Actual Estimate	3,4381
DF	156
Std Dev	0,22995

	t Test
Test Statistic	-3,3729

Trin 4: P-værdi

Hvis vi antager, at påstanden H_0 er korrekt, så har vi tidligere set (note 5), at teststørrelsen Z er beskrevet ved en t-fordeling med $n-1$ frihedsgrader.

Dermed kan vi ***under antagelse af at påstanden H_0 er sand*** beregne sandsynligheden for, at et datamateriale vil være i dårligere overensstemmelse med påstanden H_0 (som målt ved teststørrelsen Z) end tilfældet er med vores observerede datamateriale.

Denne sandsynlighed – kaldet **P-værdien** – vil vi herefter bruge som mål for, om datamaterialet ser ud til at være i overensstemmelse med påstanden H_0 eller ej. P-værdien kaldes også for **signifikanssandsynligheden**.

Hvis P-værdien er stor (lille), betyder det, at der er stor (lille) sandsynlighed for at observere et datamateriale, som er i dårligere overensstemmelse med påstanden H_0 end tilfældet er med vores datamateriale.

Trin 4: P-værdi

Med andre ord:

En stor (lille) P-værdi betyder, at det er meget sandsynligt (usandsynligt) at observere et datamateriale, der passer dårligere med påstanden H_0 end vores givne datamateriale ***under forudsætning af at påstanden*** H_0 ***rent faktisk er korrekt***. Meget tyder derfor på, at påstanden H_0 nok er (ikke er) korrekt.

Derfor vil vi opfatte en stor (lille) P-værdi som evidens for (imod) påstanden H_0.

Eksempel: Ølsalg

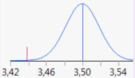
I tilfældet med prisen på Grøn Tuborg i Føtex er teststørrelsen Z beskrevet ved en t-fordeling med $n-1 = 156$ frihedsgrader.

Teststørrelsen blev beregnet til -3,37 og på baggrund af det kan P-værdien beregnes til 0,09% (vi skal senere i denne note se, præcis hvordan P-værdien beregnes).

Test Mean

Hypothesized Value	3,5
Actual Estimate	3,4381
DF	156
Std Dev	0,22995

	t Test
Test Statistic	-3,3729
Prob > |t|	0,0009*
Prob > t	0,9995
Prob < t	0,0005*



Trin 5: Konklusion

På baggrund af indholdet af trin 1 – trin 4 kan vi afgøre, hvorvidt vores datamateriale underbygger påstanden formuleret i form af nulhypotesen H_0 eller ej.

Som praktisk beslutningsregel vil vi sammenholde den beregnede P-værdi med et på forhånd valgt **signifikansniveau** α *("significance level")* (typisk bruger vi $\alpha = 5\%$).

Hvis P-værdien er...

- ***mindre end*** α vil vi ikke tro på, at påstanden H_0 er sand. Vi vil derfor forkaste påstanden H_0 og i stedet tro på, at alternativhypotesen H_a er sand
- ***større end*** α vil vi ikke afvise, at påstanden H_0 kan være sand. Vi vil derfor ikke forkaste påstanden H_0

Trin 5: Konklusion

Eksempel: Ølsalg

Hvis vi anvender et signifikansniveau på $\alpha = 5\%$, vil vi forkaste nulhypotesen H_0 (fordi P-værdi $= 0{,}09\% < 5\% = \alpha$)

$$H_0: \ \mu = 3{,}50$$

dvs. forkaste at den forventede pris på Grøn Tuborg i Føtex er 3,50 kr. Vi vil i stedet tro på indholdet af alternativhypotesen

$$H_a: \ \mu \neq 3{,}50$$

dvs. tro på at den forventede pris på Grøn Tuborg i Føtex ikke er 3,50 kr.

På baggrund af det udførte hypotesetest er konklusionen, at der ikke på baggrund af vores datamateriale bestående af 157 ugentlige observationer af prisen på Grøn Tuborg i Føtex er belæg for en påstand om, at den forventede pris på Grøn Tuborg i Føtex er 3,50 kr.

Et naturligt opfølgende spørgsmål at stille sig selv er, hvilke påstande der så *er* belæg for på baggrund af datamaterialet?!

Udover at prøve at lave hypotesetest med andre værdier end lige 3,50 i nulhypotesen H_0, så findes der også andre måder at formulere nulhypotesen på end den, vi indtil videre har set på, og andre hypoteser kan muligvis give os andre indsigter i, hvad der er belæg for at hævde på baggrund af vores datamateriale.

Formulering af hypoteser

Hver gang man laver et hypotesetest om μ, skal man igennem alle fem trin: Trin 1 – Trin 5. Indholdet af Trin 1 (Antagelser), Trin 3 (Teststørrelse) og Trin 5 (Konklusion) er dog altid det samme.

Der findes imidlertid forskellige måder at formulere sine hypoteser på, og valget af hypotese påvirker også beregningen af P-værdien. Derfor kan der være forskel på indholdet af Trin 2 (Hypoteser) og Trin 4 (P-værdi) fra test til test.

Inden vi ser nærmere på de forskellige mulige hypoteser og de tilhørende beregninger af P-værdier, så lad os først lige opsummere de tre trin, der altid er de samme.

Formulering af hypoteser

Resultat [Hypotesetest om μ – Trin 1, 3, 5]

Trin 1: *Antagelser*

$X_1,\ldots,X_n$ er indbyrdes uafhængige observationer af en variabel, der er approksimativt normalfordelt $N(\mu,\sigma)$.

Trin 3: *Teststørrelse*

Test af nulhypotesen H_0 udføres ved hjælp af teststørrelsen $Z = \frac{\hat{\mu}-\mu_0}{\frac{\hat{\sigma}}{\sqrt{n}}}$, hvor μ_0 er talværdien specificeret i nulhypotesen H_0. Under forudsætning af at nulhypotesen H_0 er sand, er teststørrelsen Z approksimativt beskrevet ved en t-fordeling med $n-1$ frihedsgrader.

Trin 5: *Konklusion*

Hvis P-værdien er...

- *mindre end signifikansniveauet α forkaster vi nulhypotesen H_0*
- *større end signifikansniveauet α forkaster vi ikke nulhypotesen H_0*

Bemærk: Testmetoden viser sig at give pålidelige resultater, også selv om datamaterialet afviger en smule fra normalfordelingsantagelsen. Derfor kræver vi i Trin 1 kun, at datamaterialets observationer er approksimativt (dvs. sådan cirka) normalfordelte.

Formulering af hypoteser

Som nævnt kan hypoteser (nul- og alternativ) formuleres på flere forskellige måder. Vi skal i dette kursus se på tre forskellige muligheder.

Resultat [Hypotesetest om μ – Trin 2]

Når vi laver hypotesetest om μ, vil vi altid anvende én ud af tre nedenstående formuleringer af nulhypotesen H_0 og alternativhypotesen H_a:

- $H_0 : \mu = \mu_0 \quad \text{og} \quad H_a : \mu \neq \mu_0$
  (Nulhypotese: "Middelværdien er lig μ_0")
- $H_0 : \mu \leq \mu_0 \quad \text{og} \quad H_a : \mu > \mu_0$
  (Nulhypotese: "Middelværdien er mindre end μ_0")
- $H_0 : \mu \geq \mu_0 \quad \text{og} \quad H_a : \mu < \mu_0$
  (Nulhypotese: "Middelværdien er større end μ_0")

Formulering af hypoteser

Metode
Test om μ (praksis)
Formulering
P-værdi
Opsummering
Test om μ (intuition)
Test om p (praksis)
Test om p (intuition)
H_0 og α
OPSUMMERING

Bemærk:

- μ_0 er blot notation for den talværdi, man ønsker at undersøge i forbindelse med sit hypotesetest
- $\neq$ betyder "forskellig fra". Udtrykket $\mu \neq \mu_0$ betyder således, at enten er $\mu > \mu_0$ eller også er $\mu < \mu_0$
- Test af nulhypotesen $H_0 : \mu = \mu_0$ (dvs. hvor $H_a : \mu \neq \mu_0$) kaldes for et **2-sidet test** *("two-tailed")*, mens de to øvrige kaldes for **1-sidet test** *("one-tailed")*
- Ved et 2-sidet test ser man på afvigelser fra nulhypotesen til begge (dvs. 2) sider i alternativhypotesen ($H_a : \mu \neq \mu_0$), mens man ved et 1-sidet test kun ser på afvigelser fra nulhypotesen til den ene side (dvs. enten $H_a : \mu > \mu_0$ eller $H_a : \mu < \mu_0$)
- I formuleringen af en nulhypotese om μ der involverer ulighedstegn, er det ligegyldigt, om man erstatter $\leq$ med $<$ og tilsvarende erstatter $\geq$ med $>$. Det vil sige at det er ligegyldigt, om man skriver

  $H_0 : \mu \leq \mu_0$ og $H_a : \mu > \mu_0$ eller $H_0 : \mu < \mu_0$ og $H_a : \mu \geq \mu_0$

  og tilsvarende om man skriver

  $H_0 : \mu \geq \mu_0$ og $H_a : \mu < \mu_0$ eller $H_0 : \mu > \mu_0$ og $H_a : \mu \leq \mu_0$

BEREGNING AF P-VÆRDI

For at kunne drage en konklusion på et hypotesetest (i TRIN 5) kræver det, at vi først har beregnet en P-værdi (i TRIN 4).

Beregning af P-værdien bygger – uanset valget af hypotese – på den værdi af teststørrelsen $Z = (\hat{\mu} - \mu_0)/(\hat{\sigma}/\sqrt{n})$, vi har beregnet i TRIN 3.

P-værdien bruges til at vurdere, om datamaterialet er foreneligt med nulhypotesen H_0. Det gøres ved at se på, hvor stor sandsynligheden er for at observere noget, der er ***mere i modstrid med nulhypotesen*** H_0 ***end vores datamateriale er***.

P-værdien beregnes derfor som sandsynligheden for at få en værdi af teststørrelsen Z, der er ***mere i modstrid med nulhypotesen*** H_0 end værdien af teststørrelsen beregnet på baggrund af datamaterialet.

Hvis P-værdien er lille (stor), er det udtryk for, at datamateriale og nulhypotese stemmer dårligt (fint) overens.

Hvordan P-værdien helt specifikt beregnes afhænger af det konkrete valg af nul- og alternativhypotese.

Beregning af P-værdi

Når P-værdien er beregnet, kan man drage konklusionen på hypotesetestet.

Hvis den beregnede P-værdi er ***lille***...

- er det tegn på, at det er en meget usædvanlig værdi af vores teststørrelse, vi har beregnet på baggrund af datamaterialet (under forudsætning af at nulhypotesen H_0 er sand)
- virker det rimeligt at tro, at nulhypotesen nok ikke er sand
- vil vi i trin Trin 5 vælge at ***forkaste*** nulhypotesen H_0, såfremt P-værdien er lille nok (= mindre end det valgte signifikansniveau α)

Hvis den beregnede P-værdi er ***stor***...

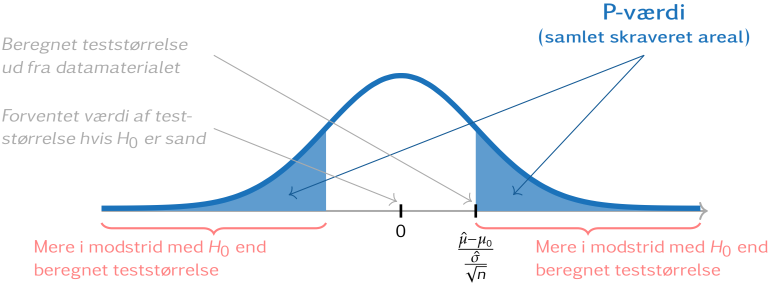
- er det tegn på, at det ikke er en usædvanlig værdi af vores teststørrelse, vi har beregnet på baggrund af datamaterialet (under forudsætning af at nulhypotesen H_0 er sand)
- virker det rimeligt at tro, at nulhypotesen nok er sand
- vil vi i trin Trin 5 vælge ***ikke at forkaste*** nulhypotesen H_0, såfremt P-værdien er stor nok (= større end det valgte signifikansniveau α).

BEREGNING AF P-VÆRDI

Test af $H_0 : \mu = \mu_0$

HVIS nulhypotesen H_0 er sand...

- vil vi regne med, at vores estimat $\hat{\mu}$ ligger tæt på talværdien μ_0, og at teststørrelsen $Z = (\hat{\mu} - \mu_0)/(\hat{\sigma}/\sqrt{n})$ derfor ligger tæt på 0
- vil store positive eller negative værdier af teststørrelsen Z være i modstrid med nulhypotesen H_0
- beregner vi P-værdien som sandsynligheden for, at teststørrelsen Z ligger længere væk fra 0 (i enten positiv eller negativ retning) end værdien beregnet ud fra datamaterialet



BEREGNING AF P-VÆRDI

Eksempel: Ølsalg

Vi ser fortsat på prisen for 1 stk. Grøn Tuborg (33 cl glasflaske) i supermarkedskæden Føtex, der kan beskrives ved en normalfordeling med estimerede parametre $\hat{\mu} = 3,438$ og $\hat{\sigma} = 0,230$ baseret på $n = 157$ observationer.

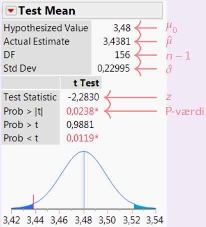
Hvis vi betragter hypoteserne

$$H_0 : \mu = 3{,}48 \qquad H_a : \mu \neq 3{,}48$$

(dvs. $\mu_0 = 3{,}48$) så svarer det til at undersøge, om den forventede pris kan antages at være 3,48 kr. eller om den ikke kan.

Under forudsætning af at nulhypotesen H_0 er sand kan teststørrelsen beregnes til

$$z = \frac{\hat{\mu} - \mu_0}{\frac{\hat{\sigma}}{\sqrt{n}}} = \frac{3,438 - 3,48}{\frac{0,230}{\sqrt{157}}} = -2{,}28$$



Herefter kan P-værdien beregnes som sandsynligheden for at observere noget, der er i dårligere overensstemmelse med H_0 end værdien $-2{,}28$, dvs. som

$$P(Z > 2{,}28) + P(Z < -2{,}28) = 2{,}38\%$$

hvor Z er beskrevet ved en t-fordeling med $n - 1 = 156$ frihedsgrader.

Ved et signifikansniveau på $\alpha = 5\%$ ***forkaster*** vi således nulhypotesen H_0 (fordi P-værdi $= 2{,}38\% < 5\% = \alpha$). Der er således ***ikke*** på baggrund af datamaterialet belæg for at ***hævde***, at den forventede pris på Grøn Tuborg er 3,48 kr.

JMP-video [Analyze -> Distribution]

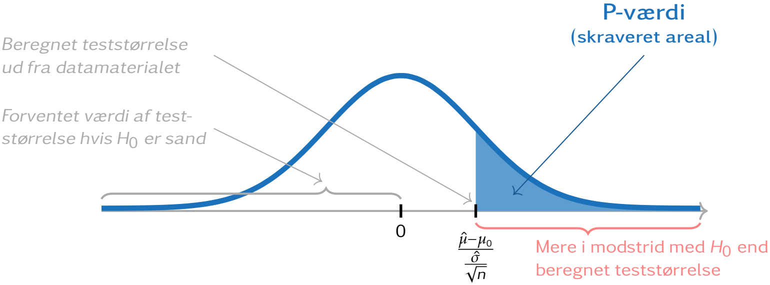
Beregning af P-værdi

Metode
Test om μ (praksis)
Formulering
P-værdi
Opsummering
Test om μ (intuition)
Test om p (praksis)
Test om p (intuition)
H_0 og α
OPSUMMERING

Test af $H_0 : \mu \leq \mu_0$

HVIS nulhypotesen H_0 er sand...

- vil vi regne med, at vores estimat $\hat{\mu}$ er mindre end talværdien μ_0, og at teststørrelsen $Z = (\hat{\mu} - \mu_0)/(\hat{\sigma}/\sqrt{n})$ derfor er mindre end 0
- vil store positive værdier af teststørrelsen Z være i modstrid med nulhypotesen H_0
- beregner vi P-værdien som sandsynligheden for, at teststørrelsen Z ligger længere væk fra 0 (i positiv retning) end værdien beregnet ud fra datamaterialet



Beregning af P-værdi

Eksempel: Ølsalg

Vi ser fortsat på prisen for 1 stk. Grøn Tuborg (33 cl glasflaske) i supermarkedskæden Føtex, der kan beskrives ved en normalfordeling med estimerede parametre $\hat{\mu} = 3,438$ og $\hat{\sigma} = 0,230$ baseret på $n = 157$ observationer.

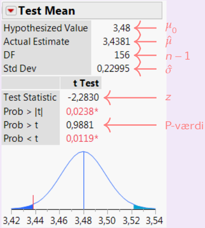
Hvis vi betragter hypoteserne

$$H_0 : \mu \leq 3{,}48 \qquad H_a : \mu > 3{,}48$$

(dvs. $\mu_0 = 3{,}48$) så svarer det til at undersøge, om den forventede pris kan antages at være lavere end 3,48 kr. eller om den ikke kan.

Under forudsætning af at nulhypotesen H_0 er sand kan teststørrelsen beregnes til

$$z = \frac{\hat{\mu} - \mu_0}{\frac{\hat{\sigma}}{\sqrt{n}}} = \frac{3,438 - 3,48}{\frac{0{,}230}{\sqrt{157}}} = -2{,}28$$



Herefter kan P-værdien beregnes som sandsynligheden for at observere noget, der er i dårligere overensstemmelse med H_0 end værdien $-2{,}28$, dvs. som

$$P(Z > -2{,}28) = 98{,}81\%$$

hvor Z er beskrevet ved en t-fordeling med $n - 1 = 156$ frihedsgrader.

Ved et signifikansniveau på $\alpha = 5\%$ ***forkaster*** vi således ***ikke*** nulhypotesen H_0 (fordi P-værdi $= 98{,}81\% > 5\% = \alpha$). Der er således ***ikke*** på baggrund af datamaterialet belæg for at ***afvise***, at den forventede pris på Grøn Tuborg er lavere end 3,48 kr.

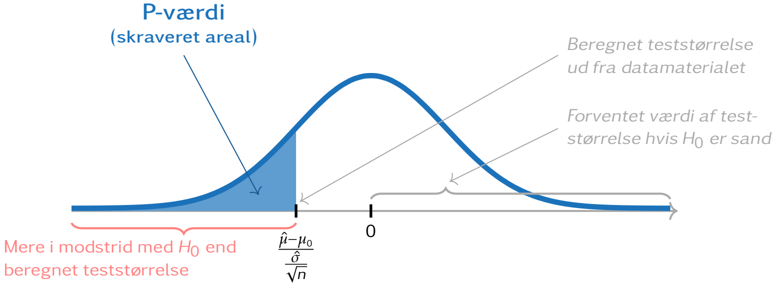
JMP-video [Analyze -> Distribution]

Beregning af P-værdi

Test af $H_0 : \mu \geq \mu_0$

HVIS nulhypotesen H_0 er sand...

- vil vi regne med, at vores estimat $\hat{\mu}$ er større end talværdien μ_0, og at teststørrelsen $Z = (\hat{\mu} - \mu_0)/(\hat{\sigma}/\sqrt{n})$ derfor er større end 0
- vil store negative værdier af teststørrelsen Z være i modstrid med nulhypotesen H_0
- beregner vi P-værdien som sandsynligheden for, at teststørrelsen Z ligger længere væk fra 0 (i negativ retning) end værdien beregnet ud fra datamaterialet



Beregning af P-værdi

Eksempel: Ølsalg

Vi ser fortsat på prisen for 1 stk. Grøn Tuborg (33 cl glasflaske) i supermarkedskæden Føtex, der kan beskrives ved en normalfordeling med estimerede parametre $\hat{\mu} = 3{,}438$ og $\hat{\sigma} = 0{,}230$ baseret på $n = 157$ observationer.

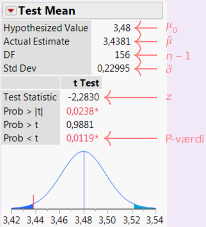
Hvis vi betragter hypoteserne

$$H_0 : \mu \geq 3{,}48 \qquad H_a : \mu < 3{,}48$$

(dvs. $\mu_0 = 3{,}48$) så svarer det til at undersøge, om den forventede pris kan antages at være højere end 3,48 kr. eller om den ikke kan.

Under forudsætning af at nulhypotesen H_0 er sand kan teststørrelsen beregnes til

$$z = \frac{\hat{\mu} - \mu_0}{\frac{\hat{\sigma}}{\sqrt{n}}} = \frac{3{,}438 - 3{,}48}{\frac{0{,}230}{\sqrt{157}}} = -2{,}28$$



Herefter kan P-værdien beregnes som sandsynligheden for at observere noget, der er i dårligere overensstemmelse med H_0 end værdien $-2{,}28$, dvs. som

$$P(Z < -2{,}28) = 1{,}19\%$$

hvor Z er beskrevet ved en t-fordeling med $n - 1 = 156$ frihedsgrader.

Ved et signifikansniveau på $\alpha = 5\%$ ***forkaster*** vi således nulhypotesen H_0 (fordi P-værdi $= 1{,}19\% < 5\% = \alpha$). Der er således ***ikke*** på baggrund af datamaterialet belæg for at ***hævde***, at den forventede pris på Grøn Tuborg er højere end 3,48 kr.

JMP-video [Analyze -> Distribution]

Opsummering

Beregningen af P-værdi ved hypotesetest om μ kan vi opsummere i nedenstående oversigt.

Resultat [Hypotesetest om μ – Trin 4]

Når vi laver hypotesetest om μ, er teststørrelsen

$$Z = \frac{\hat{\mu} - \mu_0}{\frac{\hat{\sigma}}{\sqrt{n}}}$$

beskrevet ved en t-fordeling med $n-1$ frihedsgrader under forudsætning af, at nulhypotesen H_0 er sand.

Ved analyse af...

- $H_0 : \mu = \mu_0$ *og* $H_a : \mu \neq \mu_0$ *beregnes P-værdien som* $P(|Z| > |z|)$
- $H_0 : \mu \leq \mu_0$ *og* $H_a : \mu > \mu_0$ *beregnes P-værdien som* $P(Z > z)$
- $H_0 : \mu \geq \mu_0$ *og* $H_a : \mu < \mu_0$ *beregnes P-værdien som* $P(Z < z)$

hvor z er værdien af teststørrelsen beregnet på baggrund af datamaterialet.

Opsummering

Bemærk:

- Værdien af teststørrelsen z beregnet på baggrund af datamaterialet er ***den samme*** uanset hvilken nulhypotese, der analyseres
- Fordelingen af teststørrelsen Z er den samme uanset hvilken nulhypotese, der analyseres
- Beregningen af P-værdien hørende til nulhypotesen $H_0 : \mu = \mu_0$ skrives på lidt forskellige måder (der alle er identiske):

$$P(Z < |z|) + P(Z < -|z|) = P(|Z| > |z|) = 2 \cdot P(Z > |z|)$$

hvor $|z|$ betegner den absolutte værdi af talværdien z

Hypotesetest om μ (intuition)

Ved hypotesetest om μ beregnes P-værdien ved hjælp af fordelingen af teststørrelsen $Z = \frac{\hat{\mu}-\mu_0}{\frac{\hat{\sigma}}{\sqrt{n}}}$ under forudsætning af, at den valgte nulhypotese H_0 er sand.

For at forstå mere omkring intuitionen bag hypotesetest om μ, kan det være nyttigt i stedet at se på fordelingen af middelværdiestimatet $\hat{\mu}$ under forudsætning af at den valgte nulhypotese H_0 er sand.

Teststørrelsen Z og middelværdiestimatet $\hat{\mu}$ er blot omskrivninger af hinanden, så det ændrer principielt ikke på noget, om vi ser på fordelingen af Z eller fordelingen af $\hat{\mu}$.

Det er imidlertid lettere at forstå konsekvenserne for, hvornår vi forkaster henholdsvis ikke forkaster en nulhypotese, hvis vi ser på fordelingen af $\hat{\mu}$.

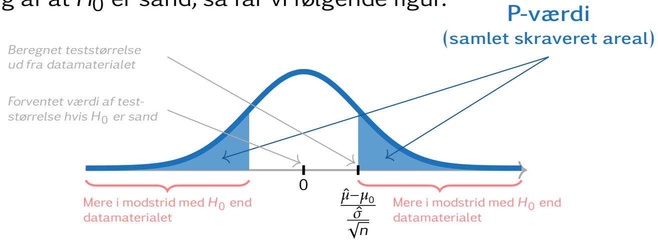
Vi tager i det følgende udgangspunkt i hypoteserne

$$H_0 : \mu = \mu_0 \qquad\qquad H_a : \mu \neq \mu_0$$

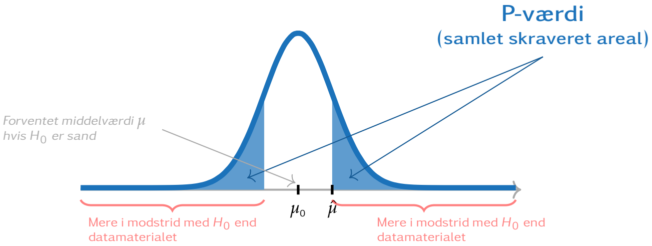
men man kan lave lignende overvejelser for de øvrige to typer af nul- og alternativhypotese.

HYPOTESETEST OM μ (INTUITION)

Ser vi først, som hidtil, på fordelingen af teststørrelsen $Z = (\hat{\mu} - \mu_0)/(\hat{\sigma}/\sqrt{n})$ under forudsætning af at H_0 er sand, så får vi følgende figur.



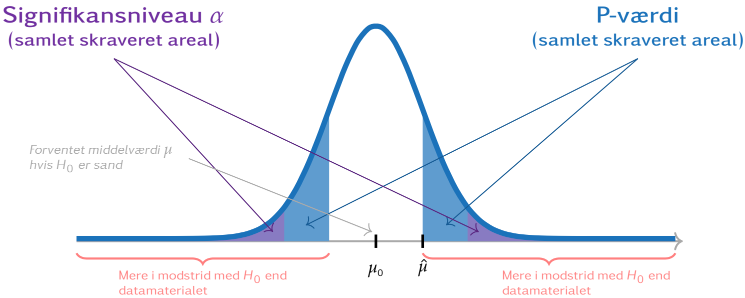
Ser vi nu i stedet på fordelingen af middelværdiestimatet $\hat{\mu}$, så får vi i stedet nedenstående figur. Reelt er det kun værdierne på 1. aksen, der har ændret sig.



Konklusionen på et hypotesetest afgøres af, om den beregnede P-værdi er større eller mindre end det valgte signifikansniveau α. Det kan vi illustrere ved også at indtegne signifikansniveauet i figuren.

Vi forkaster nulhypotesen H_0, når P-værdien er mindre end α, dvs. netop når det blå område er mindre end det lilla område.

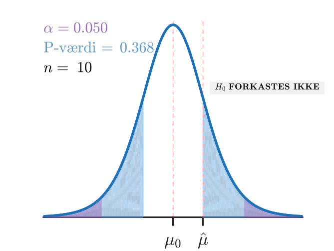
Vi forkaster ikke nulhypotesen H_0, når P-værdien er større end α, dvs. netop når det blå område er større end det lilla område.



HYPOTESETEST OM μ (INTUITION)

Nedenstående animation viser, hvordan forholdet mellem P-værdi og signifikansniveau α ændrer sig, når ***antallet af observationer*** n i datamaterialet øges.

ANIMATION: Fordeling af $\hat{\mu}$ under H_0 - ændring af antal observationer n



Start animation

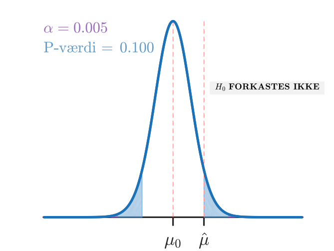
Jo ***flere*** observationer n, desto smallere bliver fordelingen af $\hat{\mu}$, og desto ***mindre*** bliver P-værdien dermed (bemærk: her holdes $\hat{\mu}$ uændret og α holdes fast på 5%).

Intuitionen er, at jo flere observationer (dvs. jo mere information) vi har til rådighed, desto mere præcist er vi i stand til at gætte på værdien af μ, og desto mindre afvigelse fra μ_0 skal der til for at forkaste nulhypotesen H_0.

Metode
Test om μ (praksis)
Test om μ (intuition)
Test om p (praksis)
Test om p (intuition)
H_0 og α
OPSUMMERING

Nedenstående animation viser, hvordan forholdet mellem P-værdi og signifikansniveau α ændrer sig, i takt med at ***signifikansniveauet*** øges.

ANIMATION: Fordeling af $\hat{\mu}$ under H_0 - ændring af signifikansniveau α



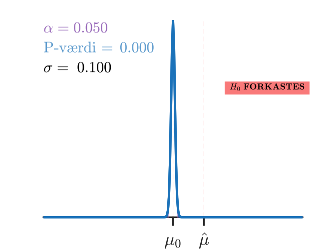
Start animation

Jo ***højere*** signifikansniveau α, desto ***mindre*** usandsynlig skal nulhypotesen H_0 være, for at vi forkaster den.

Intuitionen er, at jo lempeligere kravet er til hvor usandsynlig H_0 skal være, for at vi forkaster H_0 (dvs. jo større α), desto mindre afvigelse fra μ_0 skal der til for at forkaste nulhypotesen H_0 (fordi mindre afvigelser er mere sandsynlige).

Nedenstående animation viser, hvordan forholdet mellem P-værdi og signifikansniveau α ændrer sig, i takt med at ***standardafvigelsen*** σ øges.

ANIMATION: Fordeling af $\hat{\mu}$ under H_0 - ændring af standardafvigelse σ



Start animation

Jo ***højere*** standardafvigelse σ, desto bredere bliver fordelingen af $\hat{\mu}$, og desto ***større*** bliver P-værdien dermed (bemærk: her holdes $\hat{\mu}$ uændret og α holdes fast på 5%).

Intuitionen er, at jo mere usikkerhed der er om vores observationer, desto mindre præcist er vi i stand til at gætte på værdien af μ, og desto større afvigelse fra μ_0 skal der til for at forkaste nulhypotesen H_0.

Hypotesetest om μ (intuition)

Eksempel: Ølsalg

Vi ser igen på prisen på Grøn Tuborg med estimerede parametre $\hat{\mu} = 3{,}438$ og $\hat{\sigma} = 0{,}230$ baseret på et datamateriale med $n = 157$ observationer. Vi har set, at nulhypotesen $H_0 : \mu = 3,48$ giver en P-værdi på 2,38% og dermed forkastes ved et signifikansniveau på $\alpha = 5\%$.

Hvis datamaterialet indeholdt 300 observationer, ville P-værdien i stedet blive 0,17% og nulhypotesen $H_0 : \mu = 3{,}48$ ville ved et signifikansniveau på $\alpha = 5\%$ kunne forkastes med endnu større overbevisning (fordi P-værdien nu er lavere).

Test Inputs

Hypothesized Mean	3,48
Sample Average	3,438
Sample Standard Deviation	0,230
Sample Size	300
Significance Level (alpha)	0,05

Test Results

Result	Value
Standard Error of the Mean	0,0133
t-score	-3,1629
t Critical Values	+/- 1,9679
Observed Significance (p-value)	0,0017

Reject Null Hypothesis

Det øgede antal observationer betyder, at estimatet $\hat{\mu} = 3{,}438$ er mere præcist bestemt og derfor er i dårligere overensstemmelse med værdien 3,48 under nulhypotesen.

Hvis variationen fra uge til uge i prisen på Grøn Tuborg var dobbelt så stor ($\hat{\sigma} = 0,460$ i stedet for $\hat{\sigma} = 0,230$), så ville P-værdien i stedet blive 25,44% og nulhypotesen $H_0 : \mu = 3{,}48$ ville ved et signifikansniveau på $\alpha = 5\%$ ikke kunne forkastes.

Test Inputs

Hypothesized Mean	3,48
Sample Average	3,438
Sample Standard Deviation	0,460
Sample Size	157
Significance Level (alpha)	0,05

Test Results

Result	Value
Standard Error of the Mean	0,0367
t-score	-1,144
t Critical Values	+/- 1,9753
Observed Significance (p-value)	0,2544

Fail to Reject Null Hypothesis

Den større variation i priserne medfører større usikkerhed omkring præcisionen af vores estimat $\hat{\mu} = 3{,}438$, som derfor er i bedre overensstemmelse med værdien 3,48 under nulhypotesen.

JMP-video [Help -> Sample Data -> Calculators -> Hypothesis Test for One Mean]

Formulering af hypoteser

Hver gang man laver et hypotesetest om p, skal man igennem alle fem trin: Trin 1 – Trin 5. Indholdet af Trin 1 (Antagelser), Trin 3 (Teststørrelse) og Trin 5 (Konklusion) er dog altid det samme.

Der findes imidlertid forskellige måder at formulere sine hypoteser på, og valget af hypotese påvirker også beregningen af P-værdien. Derfor kan der være forskel på indholdet af Trin 2 (Hypoteser) og Trin 4 (P-værdi) fra test til test.

Inden vi ser nærmere på de forskellige mulige hypoteser og de tilhørende beregninger af P-værdier, så lad os først lige opsummere de tre trin, der altid er de samme.

Formulering af hypoteser

Resultat [Hypotesetest om p – Trin 1, 3, 5]

Trin 1: *Antagelser*

$X_1, \ldots, X_n$ er indbyrdes uafhængige observationer af en variabel med to mulige udfald: 1 og 0, og lad p betegne sandsynligheden for at få udfaldet 1. Endvideres antages at $n\hat{p} > 15$ og $n(1-\hat{p}) > 15$.

Trin 3: *Teststørrelse*

Test af nulhypotesen H_0 udføres ved hjælp af teststørrelsen $Z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$, hvor p_0 er talværdien specificeret i nulhypotesen H_0. Under forudsætning af at nulhypotesen H_0 er sand, er teststørrelsen Z approksimativt beskrevet ved en standardnormalfordeling $N(0,1)$.

Trin 5: *Konklusion*

Hvis P-værdien er...

- *mindre end signifikansniveauet α forkaster vi nulhypotesen H_0*
- *større end signifikansniveauet α forkaster vi ikke nulhypotesen H_0*

Bemærk: Testmetoden viser sig at give pålidelige resultater, hvis blot datamaterialet har en vis mængde observationer af både 1'er og 0'er. Det er årsagen til, at vi i Trin 1 kræver at $n\hat{p} > 15$ og $n(1-\hat{p}) > 15$.

Formulering af hypoteser

Som nævnt kan hypoteser (nul- og alternativ) formuleres på flere forskellige måder. Vi skal i dette kursus se på tre forskellige muligheder.

Resultat [Hypotesetest om p – Trin 2]

Når vi laver hypotesetest om p, vil vi altid anvende én ud af tre nedenstående formuleringer af nulhypotesen H_0 og alternativhypotesen H_a:

- $H_0 : p = p_0$ og $H_a : p \neq p_0$
  (Nulhypotese: "Andelen er lig p_0")
- $H_0 : p \leq p_0$ og $H_a : p > p_0$
  (Nulhypotese: "Andelen er mindre end p_0")
- $H_0 : p \geq p_0$ og $H_a : p < p_0$
  (Nulhypotese: "Andelen er større end p_0")

FORMULERING AF HYPOTESER

Metode
Test om μ (praksis)
Test om μ (intuition)
Test om p (praksis)
Formulering
P-værdi
Opsummering
Test om p (intuition)
H_0 og α
OPSUMMERING

BEMÆRK:

- p_0 er blot notation for den talværdi, man ønsker at undersøge i forbindelse med sit hypotesetest
- $\neq$ betyder "forskellig fra". Udtrykket $p \neq p_0$ betyder således, at enten er $p > p_0$ eller også er $p < p_0$
- Test af nulhypotesen $H_0 : p = p_0$ (dvs. hvor $H_a : p \neq p_0$) kaldes for et **2-sidet test** *("two-tailed")*, mens de to øvrige kaldes for **1-sidet test** *("one-tailed")*
- Ved et 2-sidet test ser man på afvigelser fra nulhypotesen til begge (dvs. 2) sider i alternativhypotesen ($H_a : p \neq p_0$), mens man ved et 1-sidet test kun ser på afvigelser fra nulhypotesen til den ene side (dvs. enten $H_a : p > p_0$ eller $H_a : p < p_0$)
- I formuleringen af en nulhypotese om p der involverer ulighedstegn, er det ligegyldigt, om man erstatter $\leq$ med $<$ og tilsvarende erstatter $\geq$ med $>$. Det vil sige at det er ligegyldigt, om man skriver

  $$H_0 : p \leq p_0 \quad \text{og} \quad H_a : p > p_0 \qquad \text{eller} \qquad H_0 : p < p_0 \quad \text{og} \quad H_a : p \geq p_0$$

  og tilsvarende om man skriver

  $$H_0 : p \geq p_0 \quad \text{og} \quad H_a : p < p_0 \qquad \text{eller} \qquad H_0 : p > p_0 \quad \text{og} \quad H_a : p \leq p_0$$

Beregning af P-værdi

For at kunne drage en konklusion på et hypotesetest (i Trin 5) kræver det, at vi først har beregnet en P-værdi (i Trin 4).

Beregning af P-værdien bygger – uanset valget af hypotese – på den værdi af teststørrelsen $Z = (\hat{p} - p_0)/\sqrt{p_0(1-p_0)/n}$, vi har beregnet i Trin 3.

P-værdien bruges til at vurdere, om datamaterialet er foreneligt med nulhypotesen H_0. Det gøres ved at se på, hvor stor sandsynligheden er for at observere noget, der er ***mere i modstrid med nulhypotesen*** H_0 ***end vores datamateriale er***.

P-værdien beregnes derfor som sandsynligheden for at få en værdi af teststørrelsen Z, der er ***mere i modstrid med nulhypotesen*** H_0 end værdien af teststørrelsen beregnet på baggrund af datamaterialet.

Hvis P-værdien er lille (stor), er det udtryk for, at datamateriale og nulhypotese stemmer dårligt (fint) overens.

Hvordan P-værdien helt specifikt beregnes afhænger af det konkrete valg af nul- og alternativhypotese.

Beregning af P-værdi

Når P-værdien er beregnet, kan man drage konklusionen på hypotesetestet.

Hvis den beregnede P-værdi er ***lille***...

- er det tegn på, at det er en meget usædvanlig værdi af vores teststørrelse, vi har beregnet på baggrund af datamaterialet (under forudsætning af at nulhypotesen H_0 er sand)
- virker det rimeligt at tro, at nulhypotesen nok ikke er sand
- vil vi i trin Trin 5 vælge at ***forkaste*** nulhypotesen H_0, såfremt P-værdien er lille nok (= mindre end det valgte signifikansniveau α)

Hvis den beregnede P-værdi er ***stor***...

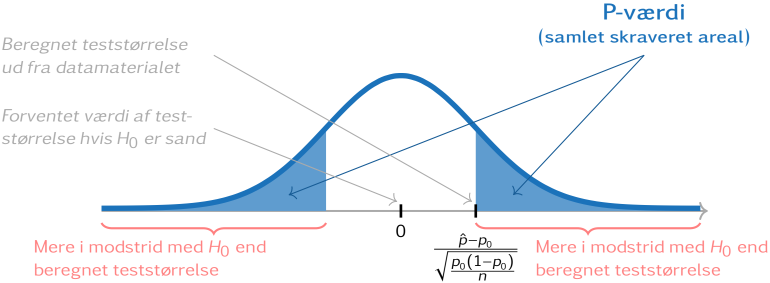
- er det tegn på, at det ikke er en usædvanlig værdi af vores teststørrelse, vi har beregnet på baggrund af datamaterialet (under forudsætning af at nulhypotesen H_0 er sand)
- virker det rimeligt at tro, at nulhypotesen nok er sand
- vil vi i trin Trin 5 vælge ***ikke at forkaste*** nulhypotesen H_0, såfremt P-værdien er stor nok (= større end det valgte signifikansniveau α).

Beregning af P-værdi

Test af $H_0 : p = p_0$

HVIS nulhypotesen H_0 er sand...

- vil vi regne med, at vores estimat $\hat{p}$ ligger tæt på talværdien p_0, og at teststørrelsen $Z = (\hat{p} - p_0)/\sqrt{p_0(1-p_0)/n}$ derfor ligger tæt på 0
- vil store positive eller negative værdier af teststørrelsen Z være i modstrid med nulhypotesen H_0
- beregner vi P-værdien som sandsynligheden for, at teststørrelsen Z ligger længere væk fra 0 (i enten positiv eller negativ retning) end værdien beregnet ud fra datamaterialet



Beregning af P-værdi

Eksempel: Skat

Vi ser fortsat på de $n = 975$ indkomne svar på spørgsmålet *"Er topskatten for høj?"*, hvoraf en andel på $\hat{p} = 0{,}355\%$ har svaret, at de mener topskatten er for høj.

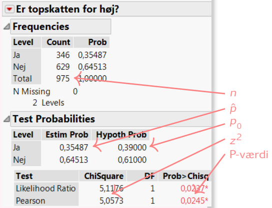
Hvis vi betragter hypoteserne

$$H_0: \; p = 0{,}39 \qquad\qquad H_a: \; p \neq 0{,}39$$

(dvs. $p_0 = 0{,}39$) så svarer det til at undersøge om andelen af befolkningen, der mener topskatten er for høj, kan antages at være 39% eller om den ikke kan.

Under forudsætning af at nulhypotesen H_0 er sand kan teststørrelsen beregnes til

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0{,}355 - 0{,}39}{\sqrt{\frac{0{,}39(1-0{,}39)}{975}}} = -2{,}25 \quad \left(= -\sqrt{5{,}0573}\right)$$



Er topskatten for høj?

Frequencies

Level	Count	Prob
Ja	346	0,35487
Nej	629	0,64513
Total	975	1,00000

N Missing 0
2 Levels

Test Probabilities

Level	Estim Prob	Hypoth Prob
Ja	0,35487	0,39000
Nej	0,64513	0,61000

Test	ChiSquare	DF	Prob>Chisq
Likelihood Ratio	5,1176	1	0,0237*
Pearson	5,0573	1	0,0245*

Herefter kan P-værdien beregnes som sandsynligheden for at observere noget, der er i dårligere overensstemmelse med H_0 end værdien $-2{,}25$, dvs. som

$$P(Z > 2{,}25) + P(Z < -2{,}25) = 2{,}45\%$$

hvor Z er approksimativt beskrevet ved en standardnormalfordeling $N(0,1)$.

Ved et signifikansniveau på $\alpha = 5\%$ ***forkaster*** vi således nulhypotesen H_0 (fordi P-værdi $= 2{,}45\% < 5\% = \alpha$). Der er således ***ikke*** på baggrund af datamaterialet belæg for at ***hævde***, at andelen af befolkningen, der mener topskatten er for høj, er 39%.

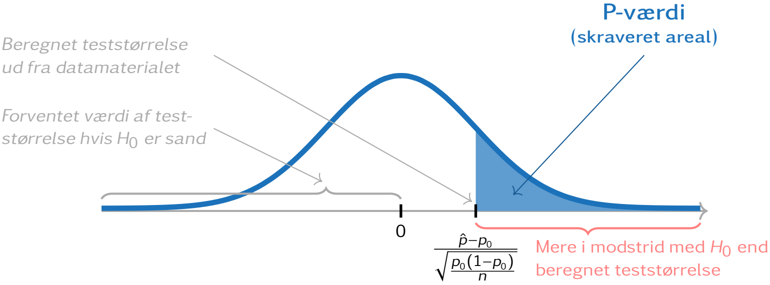
JMP-video [Analyze -> Distribution]

BEREGNING AF P-VÆRDI

Test af $H_0 : p \leq p_0$

HVIS nulhypotesen H_0 er sand...

- vil vi regne med, at vores estimat $\hat{p}$ er mindre end talværdien p_0, og at teststørrelsen $Z = (\hat{p} - p_0)/\sqrt{p_0(1-p_0)/n}$ derfor er mindre end 0
- vil store positive værdier af teststørrelsen Z være i modstrid med nulhypotesen H_0
- beregner vi P-værdien som sandsynligheden for, at teststørrelsen Z ligger længere væk fra 0 (i positiv retning) end værdien beregnet ud fra datamaterialet



Beregning af P-værdi

Eksempel: Skat

Vi ser fortsat på de $n = 975$ indkomne svar på spørgsmålet *"Er topskatten for høj?"*, hvoraf en andel på $\hat{p} = 0{,}355\%$ har svaret, at de mener topskatten er for høj.

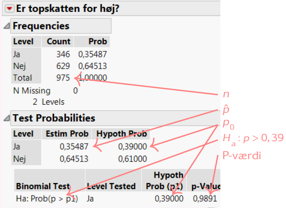
Hvis vi betragter hypoteserne

$$H_0 : p \leq 0{,}39 \qquad H_a : p > 0{,}39$$

(dvs. $p_0 = 0{,}39$) så svarer det til at undersøge om andelen af befolkningen, der mener topskatten er for høj, kan antages at være mindre end 39% eller om den ikke kan.

Under forudsætning af at nulhypotesen H_0 er sand kan teststørrelsen beregnes til

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0{,}355 - 0{,}39}{\sqrt{\frac{0{,}39(1-0{,}39)}{975}}} = -2{,}25 \quad (= -\sqrt{5{,}0573})$$



Herefter kan P-værdien beregnes som sandsynligheden for at observere noget, der er i dårligere overensstemmelse med H_0 end værdien $-2{,}25$, dvs. som

$$P(Z > -2{,}25) = 98{,}91\%$$

hvor Z er approksimativt beskrevet ved en standardnormalfordeling $N(0,1)$.

Ved et signifikansniveau på $\alpha = 5\%$ ***forkaster*** vi således ***ikke*** nulhypotesen H_0 (fordi P-værdi $= 98{,}91\% > 5\% = \alpha$). Der er således ***ikke*** på baggrund af datamaterialet belæg for at ***afvise***, at andelen af befolkningen, der mener topskatten er for høj, er mindre end 39%.

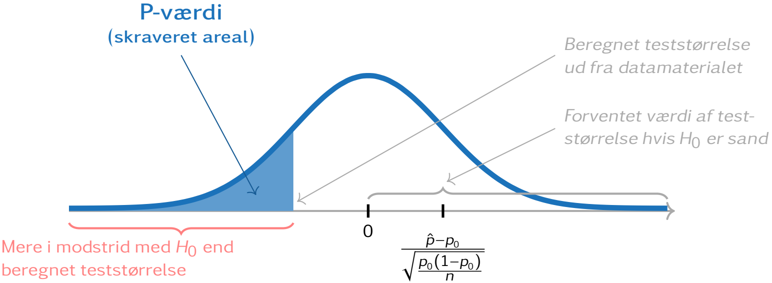
JMP-video [Analyze -> Distribution]

Beregning af P-værdi

Test af $H_0 : p \geq p_0$

HVIS nulhypotesen H_0 er sand...

- vil vi regne med, at vores estimat $\hat{p}$ er større end talværdien p_0, og at teststørrelsen $Z = (\hat{p} - p_0)/\sqrt{p_0(1-p_0)/n}$ derfor er større end 0
- vil store negative værdier af teststørrelsen Z være i modstrid med nulhypotesen H_0
- beregner vi P-værdien som sandsynligheden for, at teststørrelsen Z ligger længere væk fra 0 (i negativ retning) end værdien beregnet ud fra datamaterialet



BEREGNING AF P-VÆRDI

Eksempel: Skat

Vi ser fortsat på de $n = 975$ indkomne svar på spørgsmålet *"Er topskatten for høj?"*, hvoraf en andel på $\hat{p} = 0{,}355\%$ har svaret, at de mener topskatten er for høj.

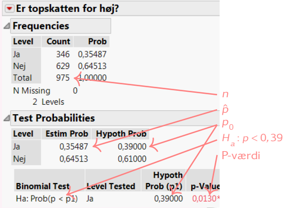
Hvis vi betragter hypoteserne

$$H_0 : p \geq 0{,}39 \qquad H_a : p < 0{,}39$$

(dvs. $p_0 = 0{,}39$) så svarer det til at undersøge om andelen af befolkningen, der mener topskatten er for høj, kan antages at være større end 39% eller om den ikke kan.

Under forudsætning af at nulhypotesen H_0 er sand kan teststørrelsen beregnes til

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0{,}355 - 0{,}39}{\sqrt{\frac{0{,}39(1-0{,}39)}{975}}} = -2{,}25 \quad (= -\sqrt{5{,}0573})$$



Herefter kan P-værdien beregnes som sandsynligheden for at observere noget, der er i dårligere overensstemmelse med H_0 end værdien $-2{,}25$, dvs. som

$$P(Z < -2{,}25) = 1{,}30\%$$

hvor Z er approksimativt beskrevet ved en standardnormalfordeling $N(0,1)$.

Ved et signifikansniveau på $\alpha = 5\%$ ***forkaster*** vi således nulhypotesen H_0 (fordi P-værdi $= 1{,}30\% < 5\% = \alpha$). Der er således ***ikke*** på baggrund af datamaterialet belæg for at ***hævde***, at andelen af befolkningen, der mener topskatten er for høj, er større end 39%.

JMP-video [Analyze -> Distribution]

Opsummering

Beregningen af P-værdi ved hypotesetest om p kan vi opsummere i nedenstående oversigt.

Resultat [Hypotesetest om p – Trin 4]

Når vi laver hypotesetest om p, er teststørrelsen

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

approksimativt beskrevet ved en standardnormalfordeling $N(0,1)$ under forudsætning af, at nulhypotesen H_0 er sand.

Ved analyse af...

- $H_0 : p = p_0$ og $H_a : p \neq p_0$ *beregnes P-værdien som* $P(|Z| > |z|)$
- $H_0 : p \leq p_0$ og $H_a : p > p_0$ *beregnes P-værdien som* $P(Z > z)$
- $H_0 : p \geq p_0$ og $H_a : p < p_0$ *beregnes P-værdien som* $P(Z < z)$

hvor z er værdien af teststørrelsen beregnet på baggrund af datamaterialet.

OPSUMMERING

BEMÆRK:

- Værdien af teststørrelsen z beregnet på baggrund af datamaterialet er ***den samme*** uanset hvilken nulhypotese, der analyseres
- Fordelingen af teststørrelsen Z er den samme uanset hvilken nulhypotese, der analyseres
- Beregningen af P-værdien hørende til nulhypotesen $H_0 : p = p_0$ skrives på lidt forskellige måder (der alle er identiske):

$$P(Z < |z|) + P(Z < -|z|) = P(|Z| > |z|) = 2 \cdot P(Z > |z|)$$

hvor $|z|$ betegner den absolutte værdi af talværdien z

Hypotesetest om p (intuition)

Ved hypotesetest om p beregnes P-værdien ved hjælp af fordelingen af teststørrelsen $Z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ under forudsætning af, at den valgte nulhypotese H_0 er sand.

For at forstå mere omkring intuitionen bag hypotesetest om p, kan det være nyttigt i stedet at se på fordelingen af estimatet $\hat{p}$ under forudsætning af at den valgte nulhypotese H_0 er sand.

Teststørrelsen Z og estimatet $\hat{p}$ er blot omskrivninger af hinanden, så det ændrer principielt ikke på noget, om vi ser på fordelingen af Z eller fordelingen af $\hat{p}$.

Det er imidlertid lettere at forstå konsekvenserne for, hvornår vi forkaster henholdsvis ikke forkaster en nulhypotese, hvis vi ser på fordelingen af $\hat{p}$.

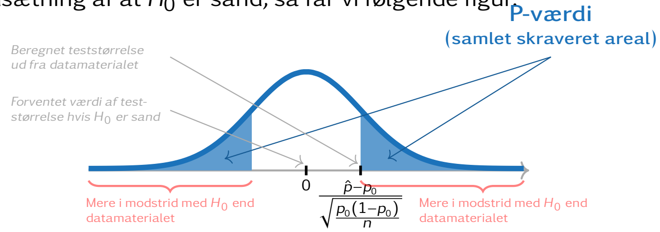
Vi tager i det følgende udgangspunkt i hypoteserne

$$H_0: \; p = p_0 \qquad\qquad H_a: \; p \neq p_0$$

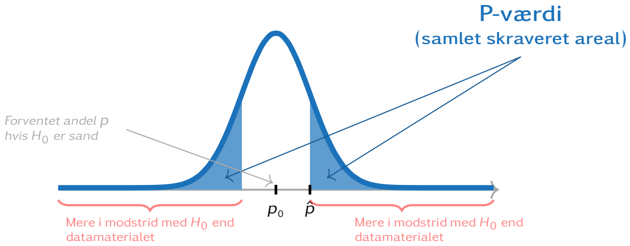
men man kan lave lignende overvejelser for de øvrige to typer af nul- og alternativhypotese.

Hypotesetest om p (intuition)

Ser vi først, som hidtil, på fordelingen af teststørrelsen $Z = (\hat{p} - p_0)/\sqrt{p_0(1-p_0)/n}$ under forudsætning af at H_0 er sand, så får vi følgende figur.



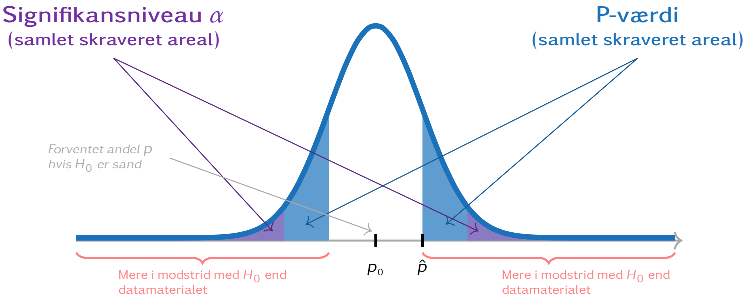
Ser vi nu i stedet på fordelingen af estimatet $\hat{p}$, så får vi i stedet nedenstående figur. Reelt er det kun værdierne på 1. aksen, der har ændret sig.



Konklusionen på et hypotesetest afgøres af, om den beregnede P-værdi er større eller mindre end det valgte signifikansniveau α. Det kan vi illustrere ved også at indtegne signifikansniveauet i figuren.

Vi forkaster nulhypotesen H_0, når **P-værdien** er mindre end α, dvs. netop når det **blå område** er mindre end det **lilla område**.

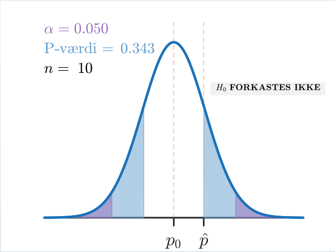
Vi forkaster ikke nulhypotesen H_0, når **P-værdien** er større end α, dvs. netop når det **blå område** er større end det **lilla område**.



HYPOTESETEST OM p (INTUITION)

Nedenstående animation viser, hvordan forholdet mellem P-værdi og signifikansniveau α ændrer sig, når ***antallet af observationer*** n i datamaterialet øges.

ANIMATION: Fordeling af $\hat{p}$ under H_0 - ændring af antal observationer n



Start animation

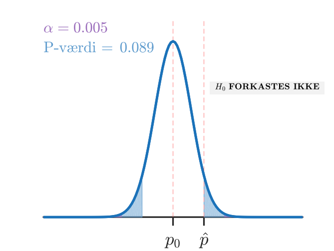
Jo ***flere*** observationer n, desto smallere bliver fordelingen af $\hat{p}$, og desto ***mindre*** bliver P-værdien dermed (bemærk: her holdes $\hat{p}$ uændret og α holdes fast på 5%).

Intuitionen er, at jo flere observationer (dvs. jo mere information) vi har til rådighed, desto mere præcist er vi i stand til at gætte på værdien af p, og desto mindre afvigelse fra p_0 skal der til for at forkaste nulhypotesen H_0.

Hypotesetest om p (intuition)

Nedenstående animation viser, hvordan forholdet mellem P-værdi og signifikansniveau α ændrer sig, i takt med at ***signifikansniveauet*** øges.

ANIMATION: Fordeling af $\hat{p}$ under H_0 - ændring af signifikansniveau α



Start animation

Jo ***højere*** signifikansniveau α, desto ***mindre*** usandsynlig skal nulhypotesen H_0 være, for at vi forkaster den.

Intuitionen er, at jo lempeligere kravet er til hvor usandsynlig H_0 skal være, for at vi forkaster H_0 (dvs. jo større α), desto mindre afvigelse fra p_0 skal der til for at forkaste nulhypotesen H_0 (fordi mindre afvigelser er mere sandsynlige).

Hypotesetest om p (intuition)

Eksempel: Skat

Vi ser fortsat på de $n = 975$ indkomne svar på spørgsmålet *"Er topskatten for høj?"*, hvoraf en andel på $\hat{p} = 346/975 = 0{,}355$ har svaret, at de mener topskatten er for høj.

Vi har set, at nulhypotesen $H_0 : p = 0{,}39$ giver en P-værdi på 2,45% og dermed forkastes ved et signifikansniveau på $\alpha = 5\%$.

Hvis datamaterialet i stedet var dobbelt så stort ($= 2 \cdot 975 = 1.950$ observationer), og hvor ligeså mange som hidtil mener, at topskatten er for høj ($= 2 \cdot 346 = 692$), så ville teststørrelsen i stedet blive $z = -3{,}18$ og den tilhørende P-værdi blive 0,15%. Nulhypotesen $H_0 : p = 0{,}39$ ville dermed ved et signifikansniveau på $\alpha = 5\%$ kunne forkastes med endnu større overbevisning (fordi P-værdien nu er lavere).

Det øgede antal observationer betyder, at estimatet $\hat{p} = 0{,}355$ er mere præcist bestemt og derfor er i dårligere overensstemmelse med værdien 0,39 under nulhypotesen.

Hypothesized Proportion 0,39
Number of Successes 692
Sample Size 1950
Significance Level (alpha) 0,05

Test Results

Result	Value
Estimated Proportion	0,3549
Standard Error of Proportion	0,011
z Critical Values	+/- 1,96
Test Statistic	-3,1804
Observed Significance (p-Value)	0,0015

Reject Null Hypothesis

JMP-video [Help -> Sample Data -> Calculators -> Hypothesis Test for One Proportion]

Valg af nulhypotese H_0

Den grundlæggende tanke bag hypotesetest er at forsøge at drage en specifik konklusion på baggrund af observationerne i et datamateriale.

Rent metodisk fungerer hypotesetest ved, at man formulerer en nulhypotese H_0, som man herefter enten forkaster eller ikke forkaster:

- Hvis man ***forkaster*** en nulhypotese H_0 betyder det, at man ***ikke tror på*** udsagnet i H_0.
- Hvis man ***ikke forkaster*** en nulhypotese H_0 betyder det, at man ***ikke kan afvise***, at udsagnet i H_0 kan være korrekt. Men det er ***IKKE*** det samme som at sige, at indholdet af H_0 så nødvendigvis også *er* korrekt. Mønstrene i datamaterialet er blot ikke tydelige nok til, at man kan afvise udsagnet i H_0.

Det er et stærkere udsagn at kunne forkaste en nulhypotese end ikke at kunne forkaste en nulhypotese:

- Når man forkaster en nulhypotese, siger man, at nulhypotesen ser ud til at være forkert
- Når man ikke forkaster en nulhypotese, så siger man kun, at man ikke kan afvise, at nulhypotesen kan være korrekt (beviserne i datamaterialet imod nulhypotesen er ikke stærke nok til, at nulhypotesen kan afvises)

VALG AF NULHYPOTESE H_0

Tankegangen omkring at acceptere/ikke acceptere en nulhypotese har en klar analogi til retssystemet:

- Udgangspunktet er en nulhypotese om, at den anklagede er uskyldig
- At forkaste nulhypotesen svarer til, at man forkaster påstanden om, at den anklagede er uskyldig. Dvs. det svarer til, at den anklagede erklæres skyldig.
- Ikke at forkaste nulhypotesen svarer til, at man ikke forkaster påstanden om, at den anklagede er uskyldig. Dvs. det svarer til, at den anklagede erklæres uskyldig. Det er ikke det samme som, at den anklagede rent faktisk *er* uskyldig. Det er blot udtryk for, at beviserne imod den anklagede (nulhypotesen) ikke er stærke nok
- Det er således mere alvorligt, at den anklagede erklæres skyldig end at vedkommende erklæres uskyldig (f.eks. pga. manglende beviser)
- Tankegangen følger princippet om, at *"man er uskyldig indtil det modsatte er bevist"*

VALG AF NULHYPOTESE H_0

Når vi laver hypotesetest vil vi allerhelst forkaste en nulhypotese (fordi det giver os det stærkeste udsagn).

Vi vil derfor typisk vælge vores nul- og alternativhypoteser sådan, at ***nulhypotesen*** H_0 ***udtrykker den påstand, vi ønsker at afvise***.

Eksempel: Ølsalg

Hvis vi eksempelvis ønsker at argumentere for, at forventningen til prisen på Grøn Tuborg i Føtex er lavere end 3,48 kr., så skal vi formulere vores nul- og alternativhypotese som

$$H_0: \mu \geq 3,48 \qquad H_a: \mu < 3,48$$

Så kan vi nemlig, hvis ellers datamaterialet tillader det, afvise nulhypotesen H_0 om, at den forventede pris er højere end 3,48 kr. og dermed konkludere, at prisen (i forventning/gennemsnit) er lavere end 3,48 kr.

Hvis vi i stedet formulerer hypoteserne som

$$H_0: \mu \leq 3,48 \qquad H_a: \mu > 3,48$$

så kan vi i bedste fald håbe på, ikke at kunne afvise nulhypotesen H_0, dvs. ikke kunne afvise at prisen kan være lavere end 3,48 kr. Men det er et svagere udsagn end ovenfor, givet at vi ønsker at argumentere for, at prisen (i forventning) faktisk er lavere end 3,48 kr.

VALG AF SIGNIFIKANSNIVEAU α

Vi kan ikke bruge et hypotesetest til at "bevise noget som helst med". Derimod kan vi bruge hypotesetest til at afvise påstande (nulhypoteser) som "højst usandsynlige".

Det betyder, at de konklusioner vi når frem til gennem hypotesetest ikke med 100% sikkerhed kan siges at være ubestrideligt korrekte. De kan derimod siges med stor sandsynlighed at være korrekte.

Når vi laver hypotesetest, er der to fejl, vi risikerer at begå:

- **Type I-fejl**:
  Vi forkaster en nulhypotese H_0, selv om den faktisk er sand
- **Type II-fejl**:
  Vi undlader at forkaste en nulhypotese H_0, selv om den faktisk er forkert

Som mål for om en nulhypotese H_0 er så usandsynlig, at vi ikke tror på den, bruger vi signifikansniveauet α, der typisk sættes til 5%.

Ved at beregne og sammenligne P-værdien med det valgte signifikansniveau afgør vi, om vi forkaster eller ikke forkaster nulhypotesen H_0.

Valg af signifikansniveau α

P-værdien beregnes ***under forudsætning af at nulhypotesen H_0 er sand***. Hvis vi bruger $\alpha = 5\%$ som signifikansniveau i vores hypotesetest, så betyder det, at i 5% af alle de hypotesetest, vi laver, vil vi fejlagtigt forkaste en nulhypotese H_0, selv om den faktisk er sand.

P-værdien er IKKE sandsynligheden for, at nulhypotesen H_0 er sand. P-værdien er derimod sandsynligheden for at observere noget, der er mere i modstrid med nulhypotesen H_0 end tilfældet er, givet at nulhypotesen er sand.

Jo lavere P-værdien er, desto større evidens er der imod nulhypotesen H_0. Det taler alt andet lige for at vælge en lille værdi som signifikansniveau α. En lav værdi af α betyder imidlertid også, at der er mange nulhypoteser vi ikke vil kunne forkaste, selv om der faktisk er klar evidens imod dem. Det typiske kompromis er at sætte $\alpha = 5\%$.

Kort opsummering af dette notesæt:

Hypotesetest om middelværdi μ i normalfordeling

TRIN 1: *Antagelser*

$X_1, \ldots, X_n$ er indbyrdes uafhængige observationer, af en variabel der er approksimativt normalfordelt $N(\mu, \sigma)$.

TRIN 2: *Hypoteser*

Der findes tre mulige valg af nul- og alternativhypotese:

- $H_0 : \mu = \mu_0$ og $H_a : \mu \neq \mu_0$
- $H_0 : \mu \leq \mu_0$ og $H_a : \mu > \mu_0$
- $H_0 : \mu \geq \mu_0$ og $H_a : \mu < \mu_0$

TRIN 3: *Teststørrelse*

Teststørrelsen $Z = (\hat{\mu} - \mu_0)/(\hat{\sigma}/\sqrt{n})$ er approksimativt beskrevet ved en t-fordeling med $n-1$ frihedsgrader, under forudsætning af at nulhypotesen H_0 er sand.

TRIN 4: *P-værdi*

Ved analyse af...

- $H_0 : \mu = \mu_0$ og $H_a : \mu \neq \mu_0$ beregnes P-værdien som $P(|Z| > |z|)$
- $H_0 : \mu \leq \mu_0$ og $H_a : \mu > \mu_0$ beregnes P-værdien som $P(Z > z)$
- $H_0 : \mu \geq \mu_0$ og $H_a : \mu < \mu_0$ beregnes P-værdien som $P(Z < z)$

hvor z er værdien af teststørrelsen beregnet på baggrund af datamaterialet.

TRIN 5: *Konklusion*

Hvis P-værdien er mindre (større) end signifikansniveauet α forkaster vi (forkaster vi ikke) nulhypotesen H_0.

OPSUMMERING

Hypotesetest om andel p i binomialfordeling

Trin 1: *Antagelser*

$X_1, \ldots, X_n$ er indbyrdes uafhængige observationer af en variabel med to mulige udfald: 1 og 0, og p betegner sandsynligheden for at få udfaldet 1. Endvidere antages at $n\hat{p} > 15$ og $n(1-\hat{p}) > 15$.

Trin 2: *Hypoteser*

Der findes tre mulige valg af nul- og alternativhypotese:

- $H_0 : p = p_0$ og $H_a : p \neq p_0$
- $H_0 : p \leq p_0$ og $H_a : p > p_0$
- $H_0 : p \geq p_0$ og $H_a : p < p_0$

Trin 3: *Teststørrelse*

Teststørrelsen $Z = (\hat{p} - p_0)/\sqrt{p_0(1-p_0)/n}$ er approksimativt beskrevet ved en standardnormalfordeling $N(0,1)$, under forudsætning af at nulhypotesen H_0 er sand.

Trin 4: *P-værdi*

Ved analyse af...

- $H_0 : p = p_0$ og $H_a : p \neq p_0$ beregnes P-værdien som $P(|Z| > |z|)$
- $H_0 : p \leq p_0$ og $H_a : p > p_0$ beregnes P-værdien som $P(Z > z)$
- $H_0 : p \geq p_0$ og $H_a : p < p_0$ beregnes P-værdien som $P(Z < z)$

hvor z er værdien af teststørrelsen beregnet på baggrund af datamaterialet.

Trin 5: *Konklusion*

Hvis P-værdien er mindre (større) end signifikansniveauet α forkaster vi (forkaster vi ikke) nulhypotesen H_0.

INDEKS

Nye funktionaliteter i dette notesæt:

- *Analyze -> Distribution:*
  - Test af hypotese om μ eller p (baseret på data)
- *Help -> Sample Data -> Calculators -> Hypothesis Test for One Mean:*
  - Test af hypotese om μ (baseret på input $\mu_0, \hat{\mu}, \hat{\sigma}, n, \alpha$)
- *Help -> Sample Data -> Calculators -> Hypothesis Test for One Proportion:*
  - Test af hypotese om p (baseret på input $p_0, \hat{p}, n, \alpha$)

JMP-videoer:

s. 23: ▶ [Analyze -> Distribution] (test af $H_0 : \mu = \mu_0$)

s. 25: ▶ [Analyze -> Distribution] (test af $H_0 : \mu \leq \mu_0$)

s. 27: ▶ [Analyze -> Distribution] (test af $H_0 : \mu \geq \mu_0$)

s. 37: ▶ [Help -> Sample Data -> Calculators -> Hypothesis Test for One Mean] (test af H_0 om μ)

s. 46: ▶ [Analyze -> Distribution] ($H_0 : p = p_0$)

s. 48: ▶ [Analyze -> Distribution] ($H_0 : p \leq p_0$)

s. 50: ▶ [Analyze -> Distribution] ($H_0 : p \geq p_0$)

s. 59: ▶ [Help -> Sample Data -> Calculators -> Hypothesis Test for One Proportion] (test af H_0 om p)