6.2 Signifikanstester

- Konfidensintervaller er nyttige når vi ønsker å estimere en populasjonsparameter
- Signifikanstester er nyttige dersom vi ønsker å teste en hypotese om en parameter i en populasjon
- Bruker observerte data til å teste hypotesen om populasjonen
- Typisk prosedyre
  - Beregn sannsynlighet for observert utfall av observator (eller noe mer ekstremt) gitt antatt hypotese
  - Hvis sannsynlighet liten, forkast hypotese

Eksempel studielån

- Gjennomsnittlig studielån:

$\bar{x}_1$=$21200 ved private college

$\bar{x}_2$=$17100 ved offentlige college

- Forskjell $4100 mellom private og offentlige college, reell eller tilfeldig?
- P(Forskjell mellom $\bar{x}_1$ og $\bar{x}_2$ ≥ 4100, gitt at det ikke er noen forskjell mellom forventningene i populasjonen) =0.17
- Ganske høy sannsynlighet, altså ikke så veldig overraskende resultat, data gir ikke grunnlag for å påstå at det er forskjeller i lånenivå mellom private og offentlige college

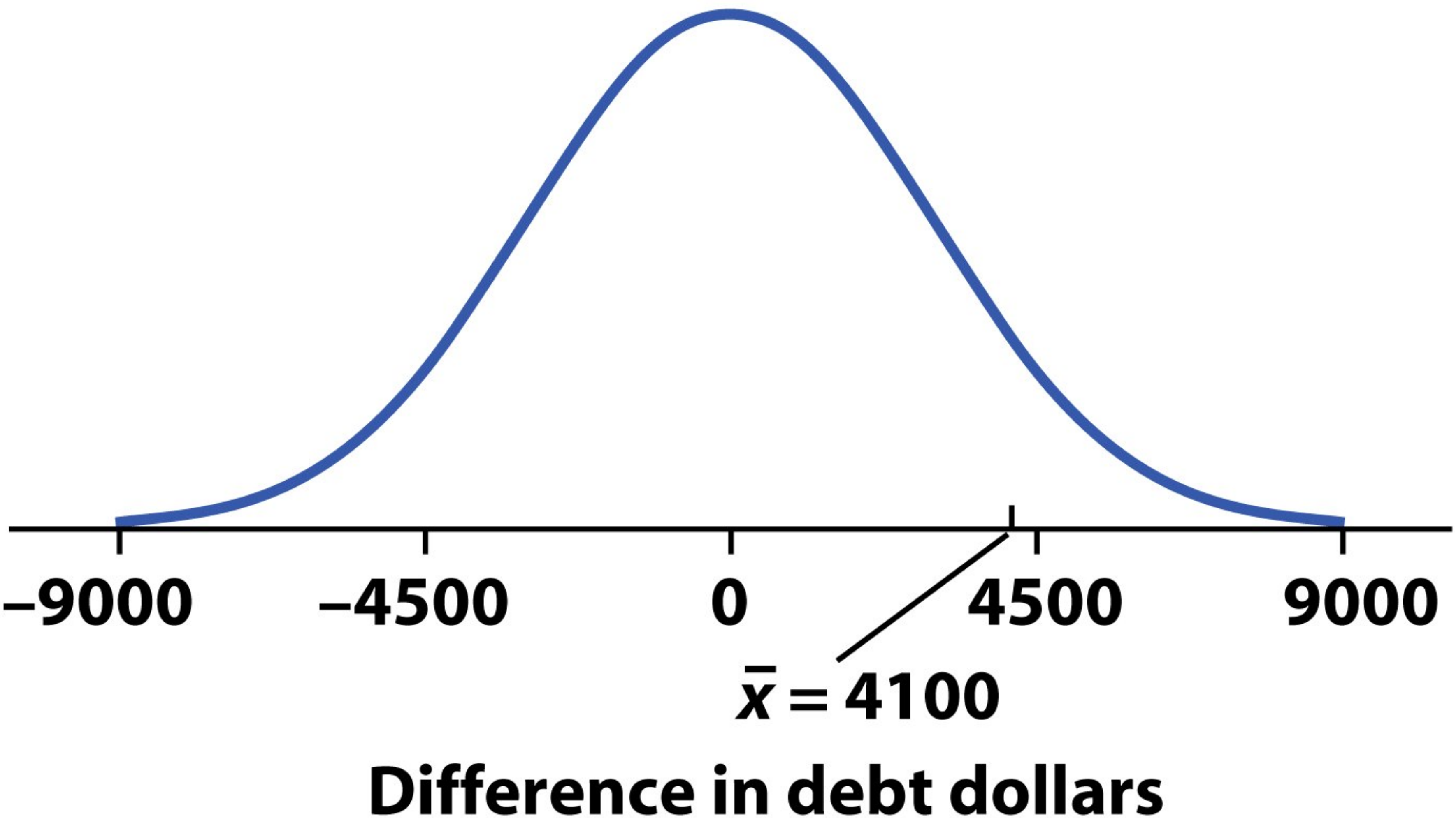


Figure 6-7
Introduction to the Practice of Statistics, Fifth Edition

Eksempel studielån

- Gjennomsnittlig studielån

  i 1997 er $\bar{x}_1$ =\$11400

  i 2002 er $\bar{x}_2$ =\$18900

  - Forskjell \$7500 mellom 2002 og 1997
  - P(Forskjell mellom $\bar{x}_1$ og $\bar{x}_2$ ≥ 7500, gitt at det ikke er noen forskjell mellom forventningene i populasjonen) =0.00004
  - Data gir grunnlag for å forkaste hypotese om at det ikke er forskjeller i lånenivå mellom 2002 og 1997

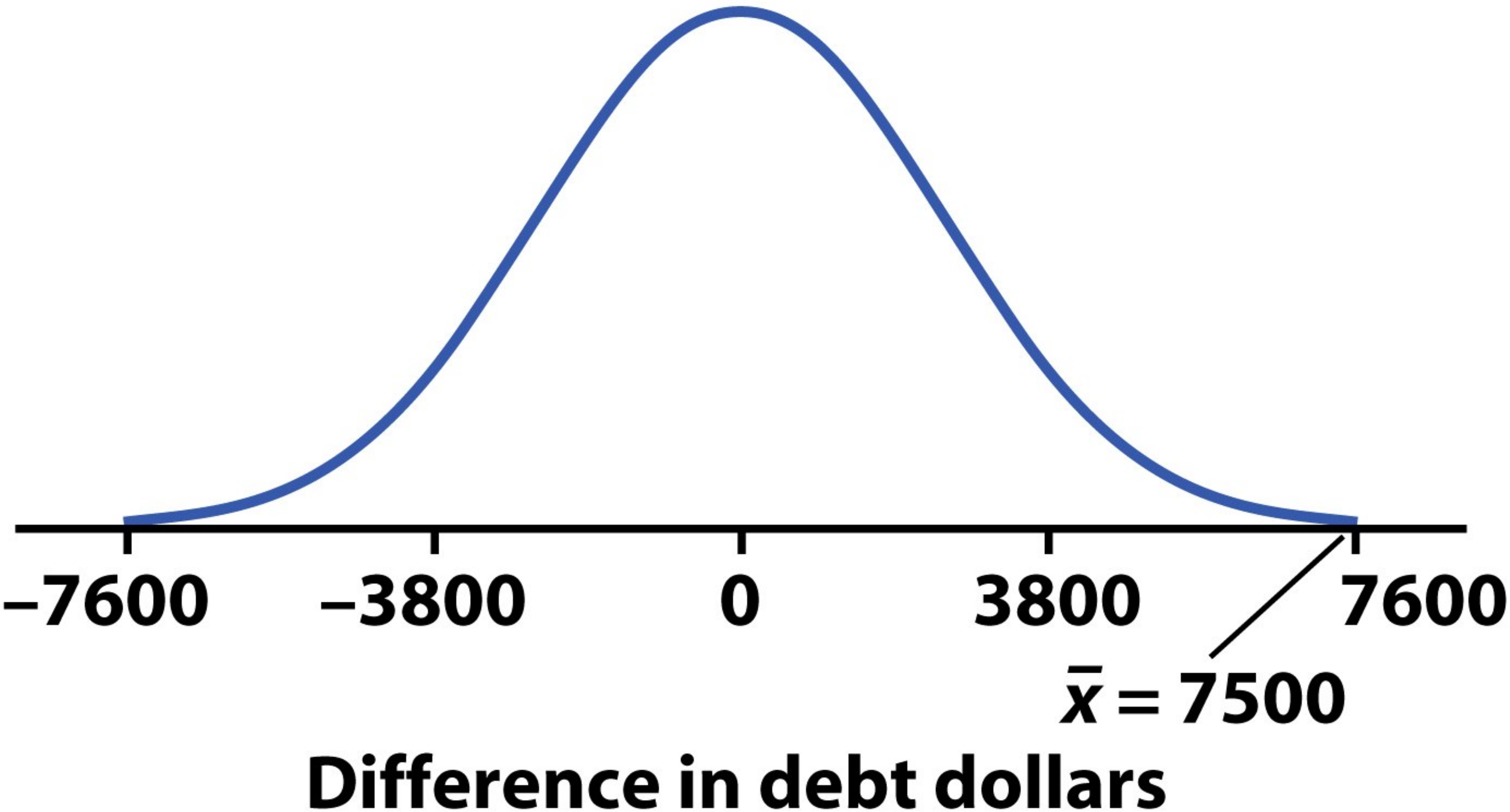


Figure 6-8
Introduction to the Practice of Statistics, Fifth Edition

Hovedtrinn

Spørsmål: Forskjell mellom nivåer (forventninger)

Ser om data er *kompatibel med hypotesen om at det ikke er noen forskjell*

Hvis ikke overraskende forskjell (eks. sanns. 0.17)

- Ikke grunnlag i data for å påstå at det er en forskjell

Hvis overraskende stor forskjell (eks. sanns. 0.00004)

- Forkast antagelse om ingen forskjell

Hypoteser

Null-hypotese H_0

- Påstand som ønskes testet
- Typisk: Ingen effekt, Ingen forskjell mellom de ukjente forventningene
- Eks. H_0: $\mu = \mu_1 - \mu_2 = 0$, eventuelt $\mu_1 = \mu_2$

Signifikanstest: Designet for å angi bevisstyrke *mot* H_0

Alternativ hypotese H_a

- Det er en effekt / forskjell
- Eks. H_a: $\mu \neq 0$ eventuelt $\mu_1 \neq \mu_2$

NB: Hypoteser er alltid utsagn/antagelser om populasjoner (eller modeller), ikke et spesielt utfall. Derfor må H_0 og H_a alltid formuleres utifra de ukjente populasjonsparametrene

Hypoteser

- Fordi H_a utrykker effekten vi ønsker å finne ut om er tilstede, er det lurt å *starte med å formulere* H_a og deretter sette opp H_0 som utsagnet om at den ønskede effekten ikke er tilstede
- Å formulere H_a er ofte den vanskeligste oppgaven. Spesielt er det ofte ikke opplagt om denne skal være

  ensidig (f.eks. H_a: $\mu>0$) eller *tosidig* (f.eks. H_a: $\mu\neq0$)

  - uttrykker om parameteren er forskjellig fra nullhypoteseverdien i en bestemt retning eller ikke
  - Man får ikke lov til å se på data for å formulere H_a (juks!)
  - Dersom man ikke har god begrunnelse på forhånd om retning på effekten, skal man velge tosidig H_a

Testobservator

Baserer test på en observator som estimerer parameteren vi er interessert i (ofte den samme som vi ville brukt til et konfidensintervall for parameteren)

- Eks.: $\bar{x}_1 - \bar{x}_2$ estimerer $\mu = \mu_1 - \mu_2$

- Verdier langt fra parameterverdi spesifisert av H_0 gir bevis mot H_0
- H_a angir hvilken retning som teller:
  - Ensidig H_a: $\mu_1 > \mu_2$ angir at vi må ha stor $\bar{x}_1 - \bar{x}_2$ som bevis mot H_0
  - Ensidig H_a: $\mu_1 < \mu_2$ angir at vi må ha liten $\bar{x}_1 - \bar{x}_2$ som bevis mot H_0
  - Tosidig H_a: $\mu_1 \neq \mu_2$ angir at vi må ha stor $|\bar{x}_1 - \bar{x}_2|$ som bevis mot H_0

Standardisert test-observator

For å undersøke hvor langt estimatet er fra parameterverdien spesifisert av H_0 , standardiserer vi estimatet

$$z = \frac{estimat - parameterverdi\ under\ H_0}{standard\ avvik\ \ for\ estimat}$$

Eksempel studielån

Forskjell på de ukjente, sanne forventningene til størrelse på lån ved private og offentlige college?

H_0 : Det er ingen forskjell i de sanne forventningene
H_a: Det er en forskjell i de sanne forventningene

Dvs: $H_0 : \mu_1 = \mu_2$ mot $H_a : \mu_1 \neq \mu_2$ (tosidig alternativ)

Vi får oppgitt: $\sigma_{\bar{x}1-\bar{x}2} = 3000$

Signifikanstest: P-verdi

- *P-verdi*: Sannsynligheten for at et utfall er like ekstremt eller mer ekstremt enn faktisk utfall (beregnet ved å anta at parameterverdien gitt av H_0 er sann)
  - Ekstremt: Lang fra hva vi ville forvente hvis H_0 var sann. Retning på hva som regnes som ekstremt: Bestemmes av H_a og H_0
  - *Jo mindre P-verdien er, jo sterkere bevis har vi mot H_0*

Eksempel studielån

Forskjell mellom private og offentlige college?
H0: $\mu1-\mu2=0$ mot Ha: $\mu1-\mu2\neq 0$
Estimat for $\mu1-\mu2$: $\bar{x}1-\bar{x}2$ (= 4100)

Ekstremt: Langt fra hva vi ville forvente hvis H_0 var sann, dvs store verdier av både positive og negative forskjeller, dvs store verdier av absoluttverdien $|\bar{x}_1-\bar{x}_2|$ teller som bevis mot H_0

Utfall er like ekstremt eller mer ekstremt enn faktisk utfall: Like store eller større enn observert verdi av absoluttverdien $|\bar{x}1-\bar{x}2|$, som betyr like store eller større enn observert verdi av $|z|$, der z er standardiserert test-observator (gitt $H_0 : \mu_1 - \mu_2 = 0$): $z=(4100-0)/3000 = 1.37$

Eksempel studielån

P-verdi: Sannsynligheten for at et utfall er like ekstremt eller mer ekstremt enn faktisk utfall (beregnet ved å anta at parameterverdien gitt av H_0 er sann)

Standardiserert,observert test-observator (gitt $H_0 : \mu_1 - \mu_2 = 0$):

$z = (4100-0)/3000 = 1.37$

Utfall er like ekstremt eller mer ekstremt enn faktisk utfall: Like store eller større enn observert verdi av absoluttverdien $|\bar{x}_1 - \bar{x}_2|$, som betyr like store eller større enn observert verdi av $|z|$

Z tilnærmet N(0,1): *P-verdi* =

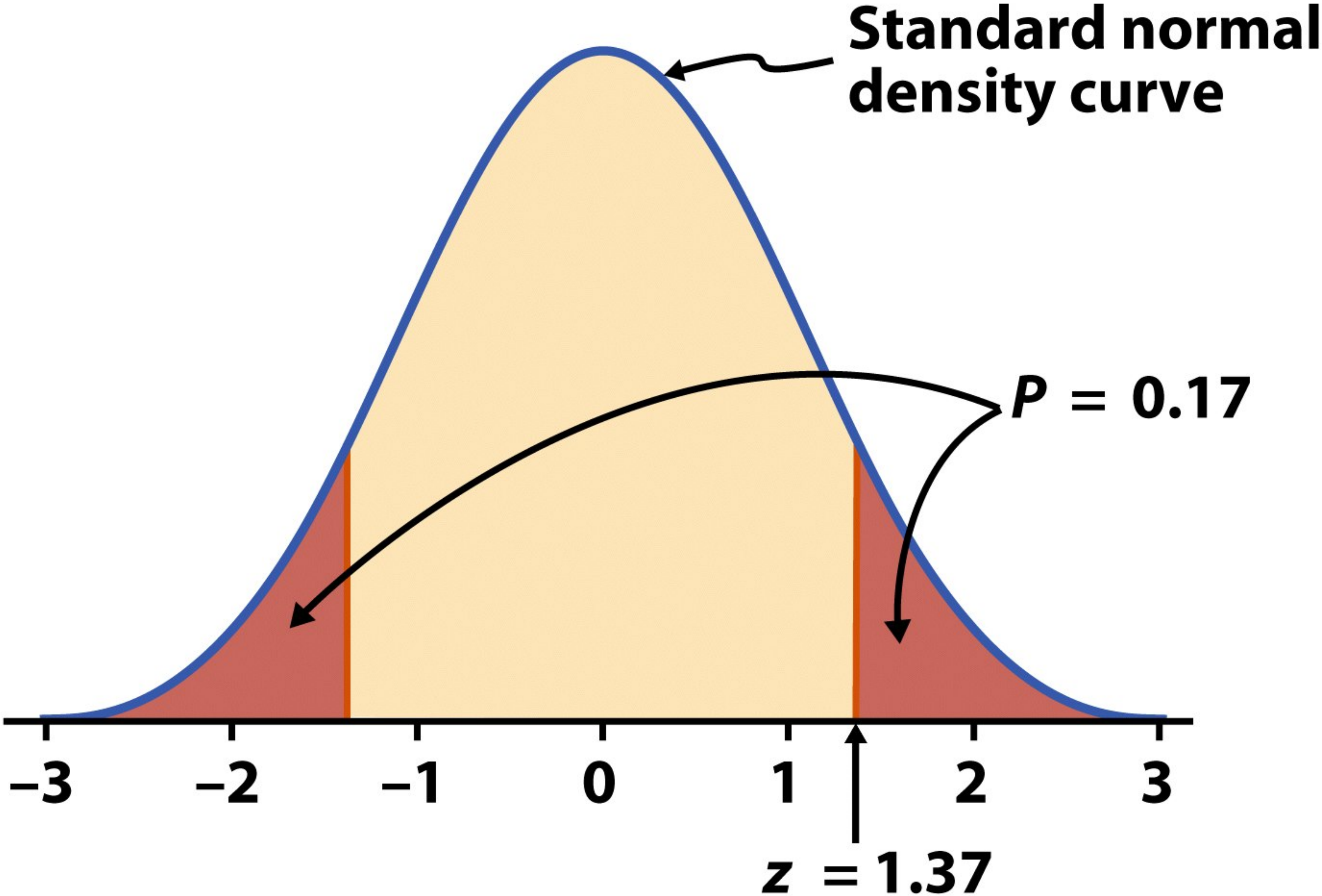


Figure 6-9
Introduction to the Practice of Statistics, Fifth Edition

Statistisk signifikans

Hvordan konkludere?

- Forkaster H_0 når *P-verdi* er liten nok

Signifikansnivå α: Grenseverdi for når vi forkaster

- Forkaster H_0 når *P-verdi* ≤ α
- Ikke grunnlag for å forkaste H_0 når P-verdi>α
- Typisk: α=0.05 (eller 0.01)

Signifikant betyr at bevisene mot nullhypotesen oppfyller standarden satt av α. Typisk utsagn er «Resultatene er signifikante (P < 0.05)»

- Hvis vi velger signifikansnivå α=0.05 krever vi at dataene gir bevis mot H_0 som bare vil skje i 5% av tilfellene hvis H_0 er sann
- Hvis vi velger signifikansnivå α=0.01 krever vi at dataene gir bevis mot H_0 som bare vil skje i 1% av tilfellene hvis H_0 er sann
  - *Vi krever altså sterkere bevis mot H_0 hvis vi velger α=0.01 enn hvis vi velger α=0.05*

Signifikanstest

1. Formuler H_0 og H_a
2. Beregn test-observator
3. Finn P-verdi
4. Formuler en konklusjon

NB: Bruker ofte datamaskin til å finne P-verdi, men en datamaskin

- Kan ikke formulere H_0 og H_a
- Kan ikke tolke P-verdien for deg
- Kan ikke bedømme om forutsetningene for å bruke testen er oppfylt

Eksempel studielån

Forskjell mellom private og offentlige college?

H0: $\mu 1- \mu 2=0$ mot Ha: $\mu 1-\mu 2\neq 0$

Estimat for $\mu 1- \mu 2$: $\bar{x}1-\bar{x}2$ (= 4100)

Standardiserer: z=(4100-0)/3000 = 1.37

Z tilnærmet N(0,1):

- *P-verdi* = P(|Z|>1.37) = P(Z<-1.37) + P(Z>1.37)

= 2* P(Z<-1.37) = 2*0.0853 = 0.1706

Eksempel på tolkning og konklusjon:

Det er 17% sjanse for å observere en forskjell like ekstrem eller mer ekstrem som den observerte forskjellen på $4100 hvis den sanne populasjons-forskjellen mellom forventningene er 0. P-verdien sier oss altså at det observerte utfallet ikke er spesielt ekstremt. Vi kan si at de observerte dataene ikke gir grunnlag for å konkludere at det er en signifkant forskjell i forventet studielån mellom private og offentlige skoler

Eksempel studielån

Økning i størrelsen på studielån fra 1997 til 2002?
H_0: $\mu_2 - \mu_1 = 0$ mot ensidig H_a: $\mu_2 - \mu_1 > 0$,

Estimat for $\mu_2 - \mu_1$: $\bar{x}_2 - \bar{x}_1$ (= 7500)

Vi får oppgitt: $\sigma_{\bar{x}1-\bar{x}2} = 1900$

Standardiserer (parameterverdi under H0 er $\mu_2 - \mu_1 = 0$):
$z = (7500-0)/3000 = 3.95$
Z tilnærmet N(0,1):

- *P-verdi* = P(Z> 3.95) = 0.00004

Eksempel på tolkning og konklusjon: Det er en 4 til 100.000 sjanse for å observere en forskjell like stor eller større som den observerte forskjellen på $7500 hvis den sanne populasjons-forskjellen mellom forventningene er 0. *P-verdien sier oss altså at det observerte utfallet er ekstremt sjeldent. Vi konkluderer at de observerte dataene gir grunnlag for å konkludere at nullhypotesen er feil*. Dataene viser tydelig at forventet studielån har økt mellom 1997 og 2002 (P<0.001)

Tester for populasjonsforventning

H_0:$\mu=\mu_0$

Data: $x_1,\ldots,x_n$

Estimator for μ: $\bar{x}$

Testobservator: $z=(\bar{x}-\mu_0)/\sigma_{\bar{x}}=(\bar{x}-\mu_0)/(\sigma\sqrt{n})$

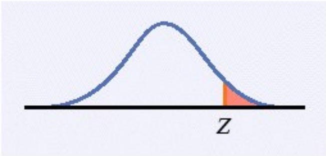
z TEST FOR A POPULATION MEAN

To test the hypothesis H_0: $\mu = \mu_0$ based on an SRS of size n from a population with unknown mean μ and known standard deviation σ, compute the test statistic

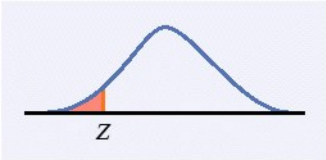
$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

In terms of a standard normal random variable Z, the P-value for a test of H_0 against

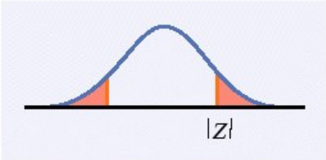
H_a: $\mu > \mu_0$ is $P(Z \geq z)$



H_a: $\mu < \mu_0$ is $P(Z \leq z)$



H_a: $\mu \neq \mu_0$ is $2P(Z \geq |z|)$



These P-values are exact if the population distribution is normal and are approximately correct for large n in other cases.

Definition, pg 410

Introduction to the Practice of Statistics, Fifth Edition

Eksempel: Blodtrykk

Blodtrykk menn alder 35-44 har forventning 128 og standardavvik σ=15

En bedriftslege ønsker å undersøke: Har mannlige toppledere (alder 35-44) i bedriften annerledes forventet blodtrykk enn den generelle populasjonen av menn på samme alder (som er 128)?

Målinger av blodtrykket til 72 toppledere: $\bar{x}$ =126.07
Bedriftslegen har ingen grunn til å anta noen retning på forskjellen. Vi kan anta at standardavviket for mannlige toppledere også er σ=15.

- $H_0:\mu=\mu_0=128$, $H_a:\mu\neq\mu_0$ der μ_0 =128
- $z=(126.07-128)/(15/\sqrt{72})=-1.09$
- P-verdi =

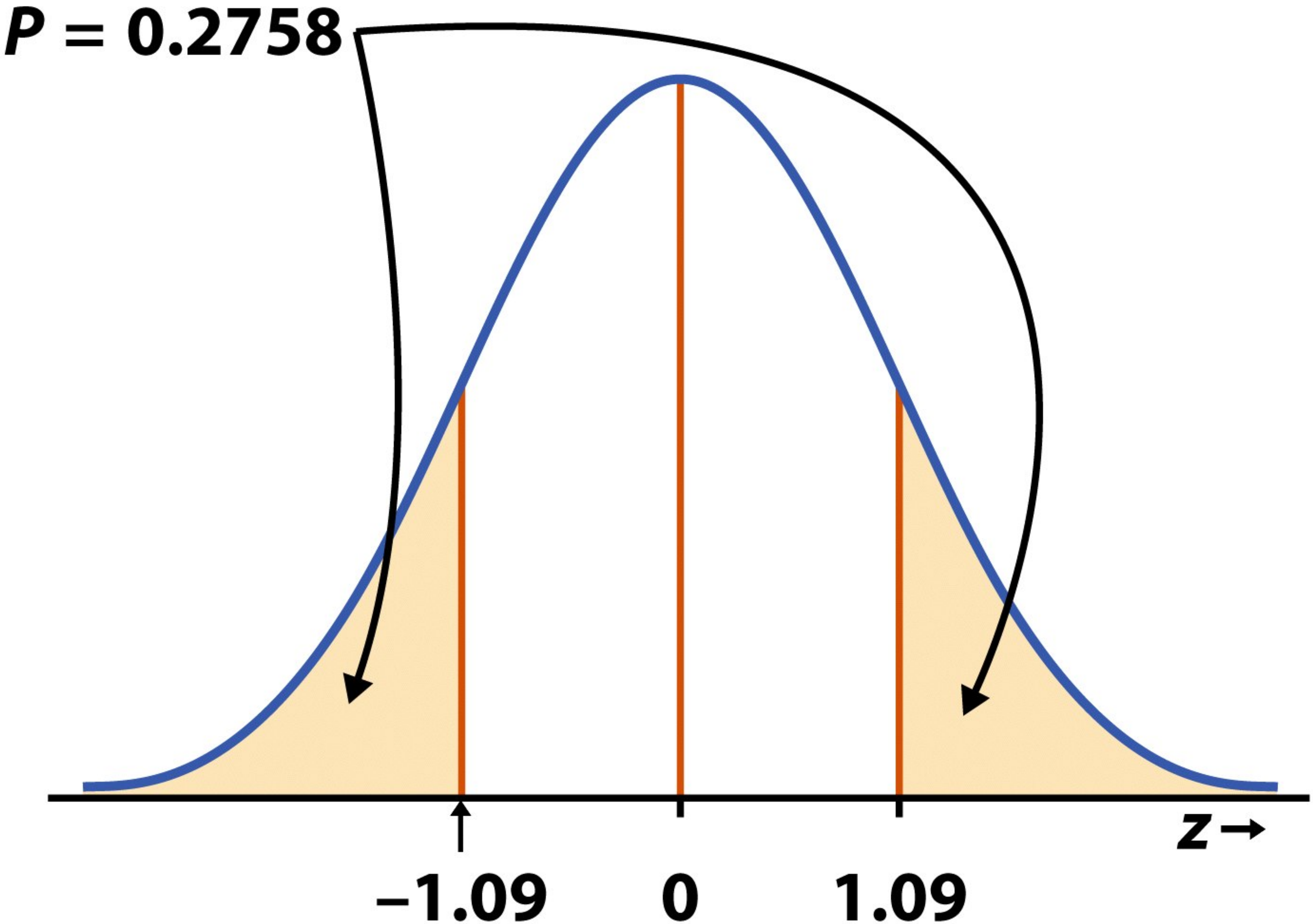


Figure 6-11
Introduction to the Practice of Statistics, Fifth Edition

Eksempel: Blodtrykk

- P-verdi = 0.2758
- Tolkning og konklusjon: Et SRS med n=27 fra den den generelle populasjonen av menn på samme alder vil 27% av gangene gi et gjennomsnittlig blodtrykk like langt eller lenger fra 128 som det observerte $\bar{x}$=126.07. Den observerte $\bar{x}$ gir altså ikke bevis for at topplederene er annerledes enn andre menn på samme alder. Ingen grunn til å forkaste H_0

To-sidige tester og konfidensintervall

Konfidensintervall med konfidens C:

$[\bar{x}-z^*\sigma/\sqrt{n}, \bar{x}+z^*\sigma/\sqrt{n}]$

Dette er laget fra de observerte dataene for ett utvalg. Verdier av μ utenfor intervall er ikke kompatible med de observerte data.

Mulig test-prosedyre:

– Forkaste H_0: $\mu=\mu_0$ hvis μ_0 ikke er i konfidensintervallet

Kan vises: Ekvivalent med signifikanstest

En tosidig signifikanstest med nivå α *som forkaster H0:*$\mu = \mu_0$ *er ekvivalent med at* μ_0 *faller utenfor konfidensintervallet for* μ *med nivå* $C=1-\alpha$

Noen ganger enklere å konstruere konfidensintervaller

Eksempel: To-sidige tester og konfidensintervall

- Analyse av konsentrasjonen av farmasøytisk produkt. Ikke presis målemetode, repeterte analyser av samme prøve vil gi litt forskjellig svar. Vi vet at konsentrasjon $N(\mu,\sigma=0.0068)$-fordelt
- Laboratoriet har blitt spurt om å evaluere en påstand om at konsentrasjonen i en prøve er 0.86%
  - $H_0: \mu=0.86$ $H_a: \mu \neq 0.86$
  - Tre målinger av prøven ga resultatene 0.8403, 0.8363 og 0.8447
  - $\bar{x}=0.8404$
  - Standardiserer: $z=(0.8404-0.86)/(0.0068/\sqrt{3})=-4.99$
  - $P=2P(Z>|-4.99|) < 0.0004$
  - Forkaster på H_0 på signifikansnivå 0.01

Eksempel: To-sidige tester og konfidensintervall

- I stedet for å gjøre signifikanstesten, kan vi alternativt lage et konfidensintervall og se om μ_0=0.86 er innenfor eller utenfor dette intervallet. For signifkansnivå 0.01 må vi lage et 99% konfidensintervall (fordi 1-0.01=0.99)

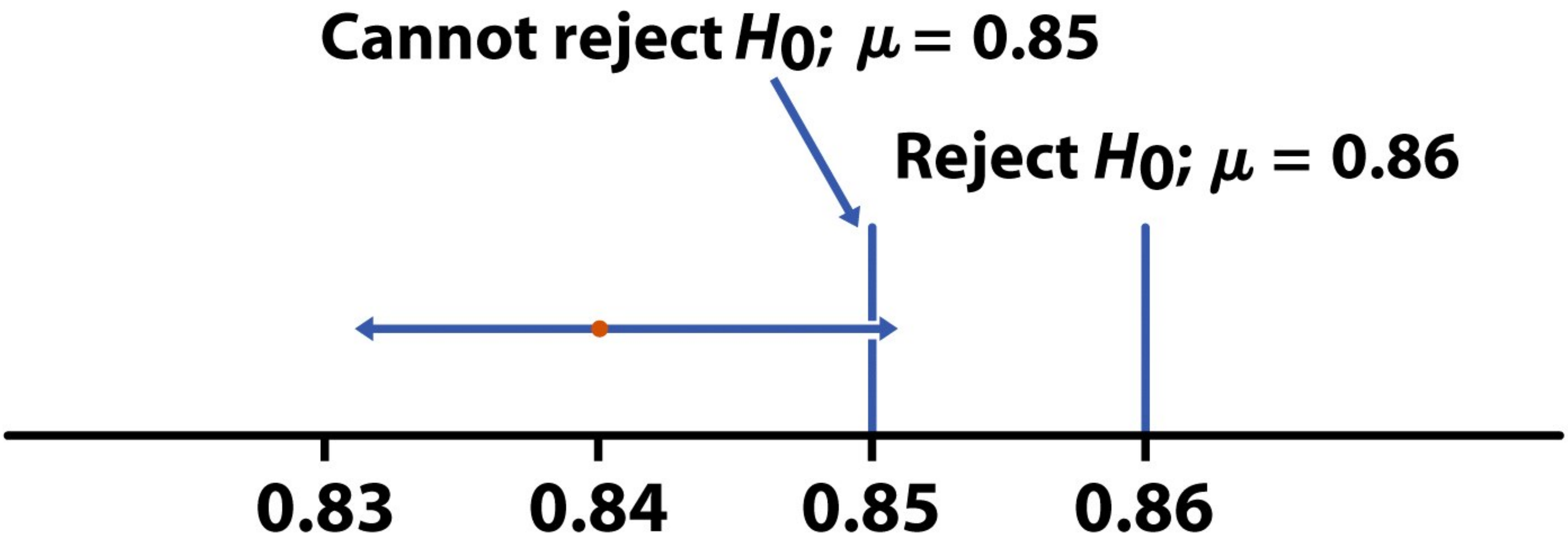


Figure 6-13
Introduction to the Practice of Statistics, Fifth Edition

P-verdier mot fast nivå α

To mulige måter å rapportere:

- Hvis P-verdi < α, si at H_0 er forkastet på nivå α
- Angi P-verdi direkte

Testing farmasøytisk produkt: Konsentrasjon $N(\mu,\sigma=0.0068)$

- $H0: \mu=0.86$ Ha: $\mu \neq 0.86$
  - Observasjoner 0.8403, 0.8363, 0.8447
  - $\bar{x}=0.8404$
  - $z=(0.8404-0.86)/(0.0068/\sqrt{3})=-4.99$
  - $P=2P(Z>|-4.99|)<0.0004$

- Signifikant på 0.05 nivå, 0.01 nivå, og 0.001 nivå. Faktisk signifikant helt ned til 0.0004-nivå

Generelt: P-verdi er det laveste signifikansnivå som gir forkastning.

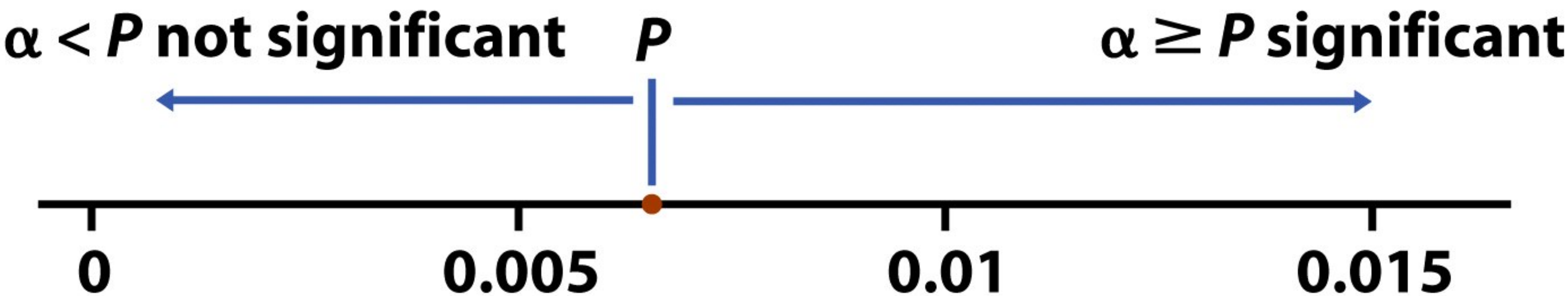


Figure 6-14
Introduction to the Practice of Statistics, Fifth Edition

6.3 Bruk og misbruk av tester

Utføre test er enkelt (software), å bruke test vanskeligere

- Kun gyldig under visse forutsetninger
- Konfidensnivå: Ingen klar grense, vanlig å rapportere P-verdi
- Forkastning H_0: Effekt statistisk signifikant
  - Men effekten kan være liten! (hvis *n* er stor)
- Ingen forkastning behøver ikke bety H_0 er sann
- Ofte mulig å gjøre mange mulige tester
  - P-verdi relatert til å gjøre *en* test
  - Hvis man gjør mange tester må justeringer gjøres

Eksempel liten effekt som er signifikant: Testing av en korrelasjon

- To variable
  - H_0: Ingen korrelasjon mellom de to variablene ($\rho = 0$)
  - H_a: Det er korrelasjon mellom de to variablene ($\rho \neq 0$)
- 400 observasjoner, observerer r=0.1
- Signifikanstest viser at det er en statistisk signifikant korrelasjon mellom de to variablene på signifikansnivå $\alpha=0.05$
- Observert variasjon i den ene variabelen forklart av den andre variabelen: $r^2=0.1^2=0.01$

Hiv-behandling

- Behandlingsgruppe og kontrollgruppe
- Insidensrateforhold I:
  - Forhold mellom smitterate i behandlingsgruppe i forhold til kontrollgruppe
- H0: I=1 (lik smitterate smitterate i behandlingsgruppe og kontrollgruppe)
- 95% konfidensintervall: [0.63,1.58]
- Kan ikke forkaste nullhypotesen om at smitterate er lik i behandlingsgruppe og kontrollgruppe, men det betyr ikke at nullhypotesen er sann!
  - Konfidensintervallet indikerer at behandling både kan gi forbedringer og forverringer!
  - Mer data er nødvendig for å finne ut hva som er riktig

Genomiske eksperimenter

- Ønsker å finne gen som forklarer sykdom
- Mange titusener mulige gener
- Kan utføre en test på hvert gen
  - 10 000 tester, α=0.05, dvs ca 5% av 10 000 tester vil være signifikante bare pga tilfeldigheter
  - Dvs forventer at 500 tester vil være signifikante kun pga. tilfeldigheter!

- Forskning:
  - Hvordan behandle mange tester simultant