ST0103 Brukerkurs i statistikk
Høst 2016

Forelesning 20, 28.10.2016

Hovedtyper av statistisk inferens

1) Estimering. *Hva er størrelsen på parameteren?*

— *Punktestimering:* Gitt ved ett enkelt tall

— *Intervallestimering:* Gitt ved et intervall der parameteren antas å ligge med høy sannsynlighet.

2) Hypotesetesting: *Velger mellom to konkurrerende påstander om størrelsen på parameteren*, for eksempel om den er større eller mindre enn en gitt verdi.

Punktestimering

Følgende egenskaper ønskes av et godt punktestimat:

— **Forventningsrett.** En observator kalles forventningsrett ('unbiased') hvis dens forventning er lik parameteren som skal estimeres. Hvis ikke, kalles den forventningsskjev ('biased'). Merk at $\bar{x}$ har forventning μ og er altså forventningsrett.

— **Liten standardfeil.** Merk at $\bar{x}$ har standardfeil $\sigma/\sqrt{n}$ som blir liten hvis n er stor (og σ ikke er for stor).

Intervallestimering

Intervallestimat Et intervall som med stor grad av konfidens (confidence) inneholder parameterverdien. Nedre og øvre grense i intervallet er observatorer beregnet fra utvalget (og er derfor tilfeldige variable).

Konfidensnivå Sannsynligheten for at intervallestimatet skal inneholde den ukjente parameteren. Skrives $1 - \alpha$ hvor α er et lite tall, f.eks. $\alpha = 0.05$ som gir $1 - \alpha = 0.95$.

Konfidensintervall Et intervallestimat med et spesifisert konfidensnivå $(1 - \alpha)$. Konfidensnivået oppgis ofte i prosent, dvs. f.eks. 95% istedenfor 0.95 og generelt $(1 - \alpha)100\%$.

6.4 Hypotesetesting

Sentrale termer:

- Hypotese: Påstand om at noe er sant.
- Hypotesetesting: Å velge mellom to konkurrerende hypoteser.
- Nullhypotese, H_0: Den hypotesen som er riktig inntil det motsatte er bevist (den konservative hypotesen).
- Alternativ hypotese, H_1: Den hypotesen vi prøver å bevise er riktig; årsaken til undersøkelsen.

Eksempel:
H_0: Klimaet har ikke endret seg
H_1: Klimaet har endret seg

H_0: Medisin A og B virker like bra
H_1: Medisin A virker bedre enn medisin B

Eksempel: vaskemiddel

Du ønsker å bestemme om du skal kjøpe et dyrt merkevare-vaskemiddel (f.eks. Omo), eller vaskemidlet som selges av lavpris-kjeden du handler hos (f.eks. First Price) som er mye billigere.

Ditt spørsmål er: *gir merkevare-vaskemidlet et bedre vaskeresultat enn lavpris-kjede-vaskemidlet?*

— Hva skal du nå sette som H_0 og H_1?

Eksempel: vaskemiddel

Vi fant
H_0: det er ingen forskjell i vaskekvalitet for merkevare- og lavpris-vaskemidlet.
H_1: merkevare-vaskemidlet gir bedre vaskeresultat.

Basert på en forkastningsregel (som vi snart skal lære mer om) kan vi enten forkaste eller beholde H_0.

— Hvis H_0 er sann, hva betyr det at vi forkaster H_0?

— Hvis H_0 er sann, hva betyr det at vi ikke forkaster H_0?

— Hvis H_1 er sann, hva betyr det at vi forkaster H_0?

— Hvis H_1 er sann, hva betyr det at vi ikke forkaster H_0?

Hvordan skal vi veie de to gale avgjørelsene? Er den ene viktigere å unngå enn den andre?

Hypotesetesting

La H_0 og H_1 være hypoteser om en gitt populasjon. Basert på et utvalg fra populasjonen skal vi lage en forkastningsregel – som går ut på om vi skal *forkaste* H_0 *eller ikke*.

To mulige avgjørelser:

1. Forkaste H_0 og påstå H_1 er sann.
2. Ikke forkaste H_0 (mangler bevis for å kunne påstå at H_0 er gal).

Dette gir fire situasjoner:

	H_0 sann	H_1 sann
Ikke forkast H_0	Korrekt avgjørelse	Type II-feil
Forkast H_0	Type I-feil	Korrekt avgjørelse

Eksempel: vaskemiddel

	Ingen forskjell mellom vaskemidlene H_0 er sann	Merkevare-midlet gir best resultat H_1 er sann
Vår beslutning:		
Behold H_0	Riktig konklusjon Vi kjøper det billige vaskemidlet, sparer penger, og får like godt resultat.	Gal konklusjon (Type II-feil) Vi kjøper det billige vaskemidlet, sparer penger, men får dårligere resultat.
Forkast H_0	Gal konklusjon (Type I-feil) Vi kjøper merkevare-midlet, bruker mer penger og får ikke noe bedre resultat.	Riktig konklusjon Vi kjøper merkevare-midlet, bruker mer penger, men får også et bedre resultat.

Hypotesetest og straffesak

I en straffesak er hypotesene:

— H_0: Tiltalte er uskyldig (riktig inntil det motsatte er bevist).

— H_1: Tiltalte er skyldig (prøver å bevise).

De typene feil vi kan gjøre er da

— Type I-feil: Justismord

— Type II-feil: Skyldig går fri.

I analogi med dette vil man i statistisk hypotesetesting betrakte **type I-feil som mest alvorlig.**

Type I og II feil og α og β

Mest alvorlig er type I-feil. Vi ønsker liten sannsynlighet for denne.
Vi krever

$$P(\text{type I-feil}) = \alpha$$

der α er et lite tall. α kalles *signifikansnivået* til testen og velges av brukeren. (Oppgis ofte i prosent, f.eks. 5%).

Vi definerer også

$$P(\text{type II-feil}) = \beta$$

$1 - \beta$ kalles *styrken* til testen og er sannsynligheten for korrekt forkastning av H_0.

Testobservator: En tilfeldig variabel (beregnet fra utvalget) som brukes til å treffe avgjørelsen.

6.1.3-4 Bør AluProd kjøpe det nye utstyret?

Problemstilling:

AluProd har et gammelt produksjonsutstyr som gir 11% defekte artikler.

Tester en ny maskin og vil kjøpe den dersom den gir en lavere prosent defekte.

La p være sannsynligheten for defekt med den *nye* maskinen. Da er p en ukjent parameter.

Hvorfor er følgende hypoteser rimelige å sette opp?

$$H_0 : p \geq 0.11, \quad H_1 : p < 0.11$$

AluProd tester maskinen ved å produsere $n = 160$ artikler. Dette ga $X = 13$ defekte.

Punktestimatet for p var dermed $\hat{p} = 13/160 = 0.081$.

Dette tyder kanskje på at H_0 bør forkastes, **men kan det lave resultatet skyldes tilfeldigheter?**

Vi beregnet den såkalte **signifikanssannsynligheten** (også kjent som **p-verdi**):

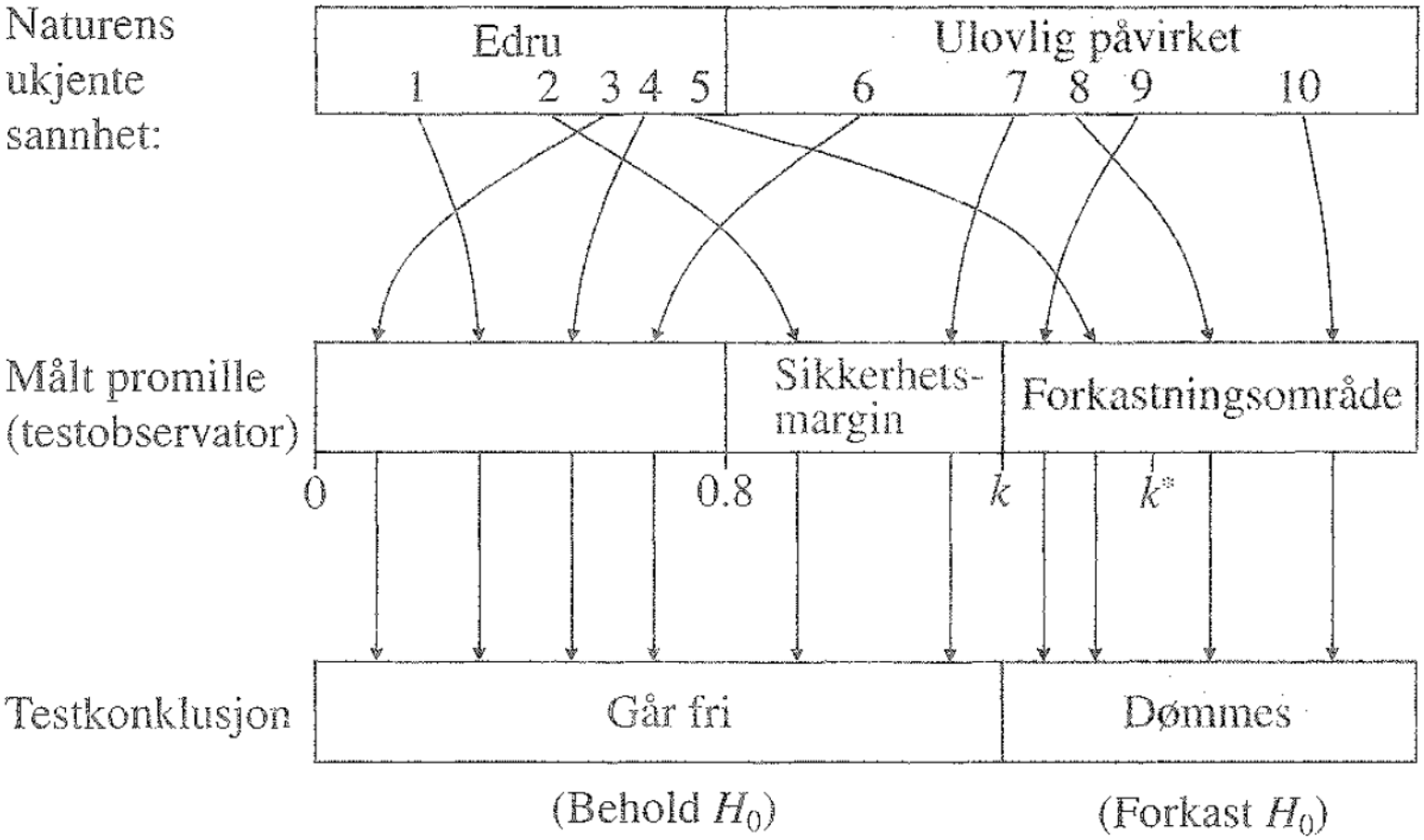
Den er sannsynligheten for å gjøre det så godt (eller bedre) enn vi gjorde, **dersom H_0 er sann**.

Vi bruker da at X (binomisk) er tilnærmet Normal(np, $\sqrt{np(1-p)}$), dvs. hvis $p = 0.11$ er X Normal(17.6,3.96), og beregner

$$P(X \leq 13) = \Phi\left(\frac{13.5 - 17.6}{3.96}\right) = \Phi(-1.04) = 0.15$$

Da denne sannynsligheten er forholdsvis stor (større enn et vanlig valgt **signifikansnivå** $\alpha = 0.05$), velger vi å beholde H_0.

Fra boka: Promillekontroll



Figur 6.12 Alkotest av ti båtførere. Noen av konklusjonene er feil