

Seminar 2

Eirik Heen

23-09-2002

Seminar 2

I dette seminaret skal vi fokusere på hypotesetesting.

Oppgave 1

For datasettet ‘late_shipments’ er det to variabler som forteller oss om forsendelsen har vært forsinket eller ikke, dette er ‘late’ og ‘late_delivery’.

Oppgave 1.1

Med til bakelegg repliker dataene for forsentkomninger. For hver replikasjon bruk prop=1, altså at du trekker like mange ganger som observasjoner i datasettet. Repliker 1’000 ganger, og plott resultatet i et histogram.

Oppgave 1.2

Fra replikasjonen i oppg1.1 regn ut gjennomsnittet og standardavviket til gjennomsnittene.

Oppgave 1.3

Fra utvalget (altså observasjonene i det originale datasettet) regn ut, gjennomsnitt, standardavvik og standard error. Husk standard error skrives $se = \frac{\sigma}{\sqrt{n}}$.

Oppgave 1.4

Sammenlikn standard error oppg1.3 med standardavviket til gjennomsnittet oppg1.2

Oppgave 1.5

Ut fra dataen ønsker vi å vi forsikre “kundene” som bruker denne shipping metoden at ikke mer enn 8% av forsendelsene er forsene. Anta et signifikans nivå på 1%, altså $\alpha = 0.01$.

Oppgave 1.5.1 Start med å sett opp nullhypotesen og alternativs hypotesen. For hvilken p – verdi sammenliknet med signifikans nivået skal vi beholde nullhypotesen eller forkaste nullhypotesen?

Oppgave 1.5.2 Gjennomfør en t-test og konkluder ut fra dette.

Oppgave 1.5.3 Siden forsentkomming er en dikotom variable (altså kun 0 eller 1), kan det være bedre å bruke en proposjons test. Gjennomfør en proposjons test og konkluder. Hvilken av testene burde brukes (samenklin med oppg1.5.2)?

Oppgave 1.5.4 Når du skal gjennomføre en hypotese testing. Hvilken feil kan du gjøre?

Oppgave 2

I denne oppgaven skal vi se litt nærmere på demokratense president valg i USA.

Oppgave 2.1

Få litt oversikt over datasettet og hva det inneholder.

Oppgave 2.2

Lag et boxplot av prosentvis stemmer de demoktratiske presidentene har fått i 2008, 2012 og 2016. Er det noe forskjell?

Oppgave 2.3

Test om det er noe forskjell mellom stemmer motatt mellom 2008 og 2012, 2008 og 2016, 2012 og 2016. Gjennomfør en uparret t-test?

Oppgave 2.4

Gjennomfør en parret t-test? Hvorfor er det så stor forskjell? Og hvordan burde vi bruke?

Oppgave 3