Plan framover

- Konfidensintervall (KI), samanlikning to andelar (pengespelet)
- Hypotesetesting (i dag og **onsdag 16.mars 12-14 i R8**)
- KI og hypotesetesting (**onsdag 30.mars 12-14 i R8**)
  - Hypotesetesting samanlikning av to utval
  - Samanheng KI og hypotesetesting
  - KI og hypotesetest for varians.

Modell:

- Spm.lagar A: Tilfeldig deltakar klarar < 5 med sanns. q_1.
- Spm.lagar B: Tilfeldig deltakar klarar < 5 med sanns. q_2.

Data:

- Spm.lagar A: n_1 deltakarar, Z_1 klarte < 5
- Spm.lagar B: n_2 deltakarar, Z_2 klarte < 5

Estimatorar:

- $\hat{q}_1 = \frac{Z_1}{n_1}$
- $\hat{q}_2 = \frac{Z_2}{n_2}$

Oppgåve 1c):

- Utlei eit tilnærma 95% konfidensintervall for $q_1 - q_2$ basert på normaltilnærming til binomisk fordeling.
- Beregn intervallet når $n_1 = n_2 = 64$, og $z_1 = 34$ og $z_2 = 18$.
- Gjer KI grunnlag for å seie at oppgåvene frå A og B har ulik vanskelighetsgrad? Begrunn svaret.

REP Sentralgrenseteoremet(SGT)

Sentralgrenseteoremet, teorem 8.2

La X_i, $i = 1, 2, \ldots, n$ vere uavhengige identisk fordelte (u.i.f.) stokastiske vaiable med $E(X_i) = \mu$ og $Var(X_i) = \sigma^2 < \infty$ (endeleg varians). La $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ og $Z = \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}}$. Då

$$Z \rightarrow N(0,1)$$

når $n \rightarrow \infty$.
Dette tilsvarer

$$\bar{X} \rightarrow N(\mu, \frac{\sigma^2}{n})$$

når $n \rightarrow \infty$.

PS: Gjeld uansett fordeling for X_i

Typiske spørsmål

- Kva er gjennomsnittshøgda for NTNU-studiner. Estimering
- Er eg høgare enn gjennomsnittleg kvinneleg NTNU-student. Hypotesetest
- Kva er forskjellen i kvalitet på laks for lagringsmetode 1 og lagringsmetode 2. Estimering
- Er kvaliteten på laks ved lagringsmetode 2 betre enn ved lagringsmetode 1. Hypotesetest

Rettsvesen hypotese

Tiltalte er uskyldig inntil det motsatte er bevist.

Hypoteser

- H_0: Tiltalte er uskyldig
- H_1: Tiltalte er skyldig

Moglege beslutningar

- Forkastar H_0, og aksepterer H_1.
  Tiltalte er skyldig
- Forkaster ikkje H_0.
  Kan ikkje motbevise at tiltalte er uskyldig, ergo er han uskyldig.
  Er ikkje tilstrekkeleg usannsynleg at tiltalte er uskyldig.

Pengespelet, eksamen juni 07

- Spm.lagar A: Tilfeldig deltakar klarar < 5 med sanns. q_1.
- Spm.lagar B: Tilfeldig deltakar klarar < 5 med sanns. q_2.

Ulik vanskelighetsgrad?

Hypoteser

- H_0: $q_1 = q_2$
- H_1: $q_1 \neq q_2$

Moglege beslutningar

- Forkastar H_0, og aksepterer H_1.
  Det er ulik vanskelighetsgrad. Set i gong tiltak.
- Forkaster ikkje H_0.
  Kan ikkje bevise at $q_1 \neq q_2$. Går ut frå at $q_1 = q_2$

Fiskekvalitet

- μ_1: Gj.snitt kvalitet på fisk, tradisjonell lagring.
- μ_2: Gj.snitt kvalitet på fisk, ny lagring.

Ny metode betre kvalitet?

Hypoteser

- H_0: $\mu_1 = \mu_2$
- H_1: $\mu_2 > \mu_1$

Moglege beslutningar

- Forkastar H_0, og aksepterer H_1.
  Det er ulik kvalitet. Innfører ny lagringsmetode.
- Forkaster ikkje H_0.
  Kan ikkje bevise at $\mu_2 > \mu_1$. Fortsett med eksisterande lagring.

Høgdehypotese

Hypotese

Mannlege NTNU-studentar er høgare enn landsgjennomsnittet.

H_0 **og** H_1**?**

Høgdehypotese forts.

Hypoteser

- H_0: $\mu = \mu_0 = 179.8$.
- H_1: $\mu > \mu_0$

Hypotesetest

Velger $\alpha = 0.05$ Antar normalfordelte data, med kjent $\sigma^2 = 6.5^2$ og $n = 87$ data. Vårt estimat $\mu^* = \bar{x} = 182.8$.

- Dersom H_0 er sann: $\bar{X} \sim N(\mu_0, \sigma^2/n) = N(183.1, 0.70^2)$.
- $p - verdi = Pr(\text{Vårt estimat eller noko meir ekstremt}|H_0)$
- $p - verdi = Pr(\bar{X} > 183.1) = Pr(Z > 4.75) = 1 - P(Z < 4.92) = 9.8 \cdot 10^{-7}$
- $p - verdi < \alpha \Rightarrow$ forkastrar H_0, konkluderer med at H_1 er sann.

Eksempel NTNU kvinner, case 1 og 2

Høgde kvinnelege NTNU-studentar

DATA

- CASE 1
  - $n = 68$
  - $\bar{x} = 169.1$, kjent varians $\sigma^2 = 6.0^2$
- CASE 2
  - n=5
  - $\bar{x} = 169.1$, kjent varians $\sigma^2 = 6.0^2$

ANTAR

- $X_i \overset{u.i.f.}{\sim} N(\mu, \sigma^2)$ for $i = 1, 2, \ldots, n$

ESTIMATOR

- $\hat{\mu} = \bar{X} \sim N(\mu, \sigma^2/n)$
- CASE 1: $Var(\hat{\mu}) = 0.7^2$
- CASE 2: $Var(\hat{\mu}) = 2.7^2$

Viktige observatorar

Dersom X_i, $i = 1, 2, \ldots, n$ er uavhengig identisk fordelt med $E(X_i) = \mu$ og $Var(X_i) = \sigma^2$, eller dersom $X_i \overset{u.i.f.}{\sim} N(\mu, \sigma^2)$ og kjent varians, så

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{n\to\infty}{\to} N(0, 1)$$

Dersom $X_i \overset{u.i.f.}{\sim} N(\mu, \sigma^2)$ og $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

Student-t fordelt med $\nu = n - 1$ fridomsgrader.

Dersom $X_i \overset{u.i.f.}{\sim} N(\mu, \sigma^2)$ og $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

$$X^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Kji-kvadrat fordelt med $\nu = n - 1$ fridomsgrader.

Styrkefunksjon høgdehypotese kvinner