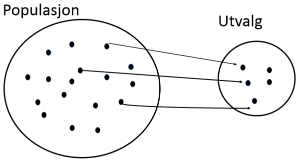
Populasjon, utvalg og observator I

- Populasjon: består av alle enheter man er interessert i (vi ofte antar at vi vet fordeling)
  - normalpopulasjon
  - poissonpopulasjon
  - $f(x)$-populasjon
- Utvalg: delmengde av en populasjon



Populasjon, utvalg og observator II

- **Tilfeldig utvalg**: $X_1, X_2, \ldots, X_n$ er uavhengige og identisk fordelte
  - $\mathrm{E}(X_i) = \mu$ og $\mathrm{Var}(X_i) = \sigma^2$ for $i = 1, \ldots, n$
- **Observator**: observerbar funksjon av SV (som er et tilfeldig utvalg)
  - Utvalgs Gjennomsnitt, $\bar{X} = \frac{1}{n} \sum_i^n X_i$
  - Utvalgs Varians, $S^2 = \frac{1}{n-1} \sum_i^n (X_i - \bar{X})^2$
- **Utvalgsfordeling**: sannsynlighetsfordelingen til en observator
  - fordeling for $\bar{X}$:
    - $\mathrm{E}(\bar{X}) = \mu$ og $\mathrm{Var}(\bar{X}) = \sigma^2/n$ alltid
    - Hvis $X_i \sim N(\mu, \sigma^2) \Rightarrow \bar{X} \sim N(\mu, \sigma^2/n)$

Theorem

La $X_1, X_2, \ldots, X_n$ vaere uavhengege og normal fordelt SV. Anta at

$$\mathrm{E}(X_1) = \mu_1, \ldots, \mathrm{E}(X_n) = \mu_n$$

$$\mathrm{Var}(X_1) = \sigma_1^2, \ldots, \mathrm{Var}(X_n) = \sigma_n^2$$

La $Y = a_1 X_1 + a_2 X_2 + \cdots + a_n + b X_n$ da er $Y \sim N(\mu_Y, \sigma_Y^2)$ med

$$\mu_Y = a_1 \mu_1 + a_2 \mu_2 + \cdots + a_n \mu_n + b$$

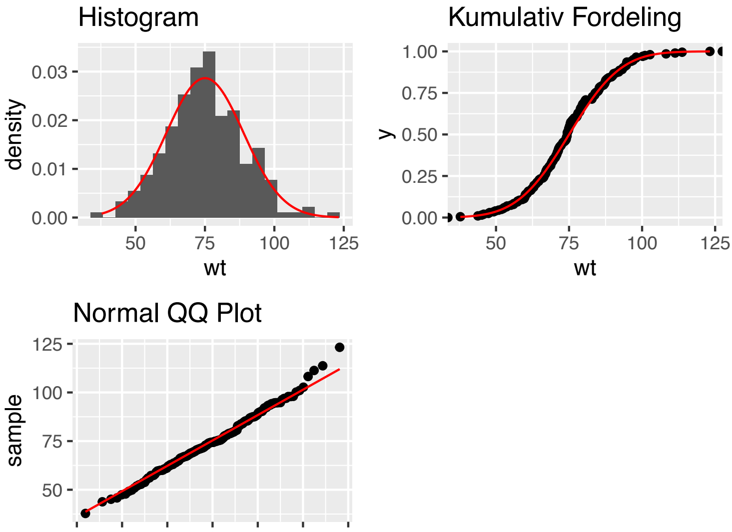
$$\sigma_Y^2 = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \cdots + a_n^2 \sigma_n^2$$

Populasjon, utvalg og observator II

- **Tilfeldig utvalg**: $X_1, X_2, \ldots, X_n$ er uavhengige og identisk fordelte
  - $\mathrm{E}(X_i) = \mu$ og $\mathrm{Var}(X_i) = \sigma^2$ for $i = 1, \ldots, n$
- **Observator**: observerbar funksjon av SV (som er et tilfeldig utvalg)
  - Utvalgs Gjennomsnitt, $\bar{X} = \frac{1}{n} \sum_i^n X_i$
  - Utvalgs Varians, $S^2 = \frac{1}{n-1} \sum_i^n (X_i - \bar{X})^2$
- **Utvalgsfordeling**: sannsynlighetsfordelingen til en observator
  - fordeling for $\bar{X}$:
    - $\mathrm{E}(\bar{X}) = \mu$ og $\mathrm{Var}(\bar{X}) = \sigma^2/n$ alltid
    - Hvis $X_i \sim N(\mu, \sigma^2) \Rightarrow \bar{X} \sim N(\mu, \sigma^2/n)$
    - Er datane normalfordelt?

Er datane normal fordelt?

Plott for å vurdere om observerte verdier $x_1, x_2, \ldots, x_n$ synes å komme fra en normalpopulasjon



Populasjon, utvalg og observator II

- **Tilfeldig utvalg**: $X_1, X_2, \ldots, X_n$ er uavhengige og identisk fordelte
  - $\mathrm{E}(X_i) = \mu$ og $\mathrm{Var}(X_i) = \sigma^2$ for $i = 1, \ldots, n$
- **Observator**: observerbar funksjon av SV (som er et tilfeldig utvalg)
  - Utvalgs Gjennomsnitt, $\bar{X} = \frac{1}{n}\sum_i^n X_i$
  - Utvalgs Varians, $S^2 = \frac{1}{n-1}\sum_i^n (X_i - \bar{X})^2$
- **Utvalgsfordeling**: sannsynlighetsfordelingen til en observator
  - fordeling for $\bar{X}$:
    - $\mathrm{E}(\bar{X}) = \mu$ og $\mathrm{Var}(\bar{X}) = \sigma^2/n$ alltid
    - Hvis $X_i \sim N(\mu, \sigma^2) \Rightarrow \bar{X} \sim N(\mu, \sigma^2/n)$
    - Er datane normalfordelt?
    - Hva hvis de IKKE er normalfordelt?

Sentralgrense teorem

La $X_1, X_2, \ldots, X_n$ være tilfeldig utvalg fra fordeling med $E(X_i) = \mu$ og $Var[X_i] = \sigma^2$. Da vil fordelingen til

$$Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

konvergere mot en standard normal fordeling når $n \to \infty$.

- Når n er stor har vi dermed at

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

Sentralgrense teorem

La $X_1, X_2, \ldots, X_n$ være tilfeldig utvalg fra fordeling med $E(X_i) = \mu$ og $Var[X_i] = \sigma^2$. Da vil fordelingen til

$$Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

konvergere mot en standard normal fordeling når $n \to \infty$.

- Når n er stor har vi dermed at

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

- Det betyr at (når n er stor)

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

- unntatt for svært skjeve fordelinger har man en god approksimasjon når $n \geq 30$

Fordeling til utvalgsgjennomsnitt

$X_1, X_2, \dots, X_n$ u.i.f. fra $f(x)$ med $\mathrm{E}(X_i) = \mu$ og $\mathrm{Var}(X_i) = \sigma^2$.

La $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ da

- Hvis $f(x) = N(\mu, \sigma^2) \Rightarrow \bar{X} \sim N(\mu, \sigma^2/n)$
- Hvis $f(x)$ ikke er $N(\mu, \sigma^2) \Rightarrow \bar{X} \xrightarrow[n\to\infty]{} N(\mu, \sigma^2/n)$

Illustrasjon fra simulering

- **Situasjon**
  - $X_1, X_2, \ldots, X_9$ er et tilfeldig utvalg fra $N(\mu, \sigma^2)$-populasjon -verdien til μ er ukjent -skal estimere (anslå) verdien til μ -vil sammenligne
    - Gjennomsnitt $\bar{x} = \frac{1}{n} \sum X_i$
    - Median $\tilde{X}$
- Generere 9 observasjoner $X_i \sim N(2.4, 0.1)$
- Regne $\bar{x}$ og $\tilde{x}$

Oppsummering

- Hva er en estimator?
- Hva mener vi med “god” estimator?
- Hvordan finner vi en estimator?

Oppsummering

- Hva er en estimator?
  - Anta at vi har et tilfeldig utvalg $X_1, X_2, \ldots, X_n$ fra $f(x;\theta)$-populasjonen, der verdien til parameteren θ er ukjent. En estimator for θ er da en observator som benyttes til å anslå verdien til θ.
  - En estimator er en stokastic variabel med sin egen fordeling
- Hva mener vi med "god" estimator?
- Hvordan finner vi en estimator?

Oppsummering

- Hva er en estimator?
- Vi har to estimatorer $\hat{\theta}_1$ og $\hat{\theta}_2$ hvilken er best?
    - Ide: tenker oss at vi kan gjenta forsøket (uendelig) mange ganger. Vi foretrekker den estimator son vanligvis treffer naermest den sanne verdi θ
        - Vi vet ikke verdi til θ
        - Vi vanligvis gjør forsøket bare en gang!
- Hvordan finner vi en estimator?

Oppsummering

- Hva er en estimator?
- Vi har to estimatorer $\hat{\theta}_1$ og $\hat{\theta}_2$ hvilken er best?
    - Ide: tenker oss at vi kan gjenta forsøket (uendelig) mange ganger. Vi foretrekker den estimator son vanligvis treffer naermest den sanne verdi θ
        - Vi vet ikke verdi til θ
        - Vi vanligvis gjør forsøket bare en gang!
    - Def: En observator $\hat{\theta}$ er en forventingsrett (unbiased) estimator for θ hvis $E[\hat{\theta}] = \theta$. Hvis ikke sier vi at $\hat{\theta}$ er forventingsskjev (biased)
    - Def:Av flere forventingsrette estimatorer sier vi at den med minst varians er mest effisient
- Hvordan finner vi en estimator?