

Plan for dagens undervising

- ★ Informere om avslutningsspørreundersøkelsen i TMA4240
- ★ Repetisjon/oppsummering
- ★ Noen råd før og på eksamen
- ★ Statistikk er mer enn TMA4240

Avslutningsspørreundersøkelse

- ★ Husk å svare på avslutningsspørreundersøkelsen
 - dere er nå eksperter på å være student i TMA4240
 - vi kan hjelpe med å forbedre undervisningen neste semester
- ★ URL til spørreundersøkelsen: <https://nettskjema.no/a/380337>
 - eventuelt: se melding på hjemmesiden til TMA4240

Repetisjon TMA4240

November 2023

Begreper vi har diskutert

- ★ Stokastisk forsøk
- ★ Utfallsrom
- ★ Hendelse
- ★ Venndiagram
- ★ Sannsynlighet
- ★ Betinget sannsynlighet
- ★ Uavhengige hendelser
- ★ Stokastisk variabel
- ★ Punktsannsynlighet
- ★ Sannsynlighetstetthet
- ★ Kumulativ fordeling
- ★ Simultanfordeling
- ★ Betinget fordeling
- ★ Uavhengige stokastiske variabler
- ★ Stokastisk (Monte Carlo) simulering
- ★ Forventingsverdi
- ★ Varians
- ★ Standardavvik
- ★ Kovarians
- ★ Korrelasjon
- ★ Bernoulliprosess
- ★ Poissonprosess
- ★ Transformasjon
- ★ Feilforplantningsformelen
- ★ Momentgenererende funksjoner
- ★ Populasjon

Begreper vi har diskutert

- ★ Utvalg
- ★ Tilfeldig utvalg
- ★ Observator
- ★ Estimator
- ★ Forventningsrett estimator
- ★ Forventningsskjev estimator
- ★ Mest effisient estimator
- ★ Rimelighetsfunksjon
- ★ Sannsynlighetsmaksimeringsprinsippet
- ★ Sannsynlighetsmaksimerings-estimator
- ★ Konfidensintervall
- ★ Ett-utvalg
- ★ To-utvalg
- ★ Prediksjonsintervall
- ★ Hypotesetesting
- ★ Nullhypotese
- ★ Alternativ hypotese
- ★ Testobservator
- ★ Signifikansnivå
- ★ Forkastningsområde
- ★ p -verdi
- ★ Ensidig test
- ★ Tosidig test
- ★ Enkel lineær regresjon
- ★ Minste kvadraters metode

Hva har vi gjort med disse begrepene?

- ★ Definert (og ofte motivert definisjonen)
- ★ Diskutert tolkning
- ★ Introdusert regneregler
- ★ Regnet på eksempler
- ★ Tolket svaret i regneeksempler

Hva har vi gjort med disse begrepene?

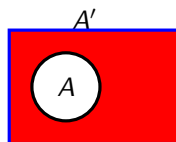
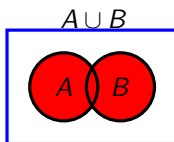
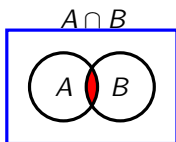
- ★ Definert (og ofte motivert definisjonen)
- ★ Diskutert tolkning
- ★ Introdusert regneregler
- ★ Regnet på eksempler
- ★ Tolket svaret i regneeksempler

Hva forventer vi av dere?

- ★ Skal kunne anvende begrepene til å svare på spørsmål
- ★ Skal kunne regne
- ★ Skal kunne tolke svarene
- ★ Skal kunne forklare begrepene

Stokastisk forsøk, utfallsrom og venndiagram

- ★ Stokastisk forsøk: Et forsøk hvor resultatet er underlagt tilfeldigheter
- ★ Utfallsrom: Mengden av alle mulige resultater, S
- ★ Enkeltutfall: Et mulig utfall, $e \in S$
- ★ Hendelse: En delmengde av utfallsrommet, $A, B, C, \dots$
- ★ Venndiagram: Illustrerer hendelser og operasjoner på hendelser



- ★ Operasjoner på hendelser
 - Snitt
 - Union
 - Komplement
- ★ Disjunkte hendelser

Sannsynlighet

★ Sannsynlighet: Reell funksjon definert på hendelser, P

1. $0 \leq P(A) \leq 1$
2. $P(S) = 1$
3. Hvis $A_1, A_2, \dots$ parvis disjunkte:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

– tenk på sannsynlighet som areal i venndiagram

★ Tolkning: må tenke oss at vi gjentar forsøket uendelig mange ganger

Sannsynlighet

★ Sannsynlighet: Reell funksjon definert på hendelser, P

1. $0 \leq P(A) \leq 1$
2. $P(S) = 1$
3. Hvis $A_1, A_2, \dots$ parvis disjunkte:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

– tenk på sannsynlighet som areal i venndiagram

★ Tolkning: må tenke oss at vi gjentar forsøket uendelig mange ganger

★ Uniform sannsynlighetsmodell:

- $S = \{e_1, e_2, \dots, e_m\}$
- $P(\{e_1\}) = P(\{e_2\}) = \dots = P(\{e_m\})$

$$P(A) = \frac{g}{m} = \frac{\text{antall gunstige}}{\text{antall mulige}}$$

Kombinatorikk/Telleregler

- ★ Multiplikasjonssetningen: $n = n_1 \cdot n_2 \cdot \dots \cdot n_p$
- ★ Urnemodell: Trekker kuler fra en urne
 - ordnet utvalg eller ikke-ordnet utvalg
 - trekker med tilbakelegging eller trekker uten tilbakelegging
- ★ Inndeling i grupper

Kombinatorikk/Telleregler

- ★ Multiplikasjonssetningen: $n = n_1 \cdot n_2 \cdot \dots \cdot n_p$
- ★ Urnemodell: Trekker kuler fra en urne
 - ordnet utvalg eller ikke-ordnet utvalg
 - trekker med tilbakelegging eller trekker uten tilbakelegging
- ★ Inndeling i grupper
- ★ I formelsamling, nederst på side 32:
 - $n! = 1 \cdot 2 \cdot \dots \cdot n$ er antall måter å permutere (ordne) n elementer
 - ${}_nP_r = \frac{n!}{(n-r)!} = n \cdot (n-1) \cdot \dots \cdot (n-r+1)$ er antall ordnede utvalg når r elementer velges blant n uten tilbakelegging
 - $\binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{{}_nP_r}{r!} = {}_nC_r$ er antall ikke-ordnede utvalg når r elementer velges blant n uten tilbakelegging

Regneregler for sannsynlighet

- ★ Additiv regel: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
 - A, B disjunkte: $P(A \cup B) = P(A) + P(B)$
- ★ Komplementærsetningen $P(A) = 1 - P(A')$
- ★ Betinget sannsynlighet: $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- ★ Multiplikasjonssetningen: $P(A \cap B) = P(B)P(A|B)$
- ★ Uavhengighet: A, B uavhengige $\Leftrightarrow P(A|B) = P(A)$
 - $P(A \cap B) = P(A) \cdot P(B)$

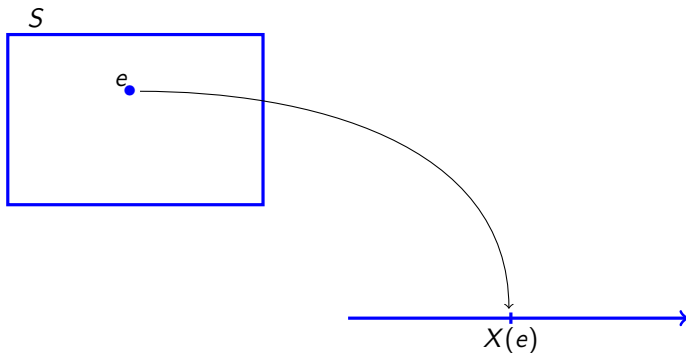
Regneregler for sannsynlighet

- ★ Additiv regel: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
 - A, B disjunkte: $P(A \cup B) = P(A) + P(B)$
- ★ Komplementærsetningen $P(A) = 1 - P(A')$
- ★ Betinget sannsynlighet: $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- ★ Multiplikasjonssetningen: $P(A \cap B) = P(B)P(A|B)$
- ★ Uavhengighet: A, B uavhengige $\Leftrightarrow P(A|B) = P(A)$
 - $P(A \cap B) = P(A) \cdot P(B)$
- ★ Bayes regel: $P(A|B) = \frac{P(A)P(B|A)}{P(B)}$
- ★ Setning om total sannsynlighet: $B_1, B_2, \dots, B_n$ partisjon av S

$$P(A) = \sum_{i=1}^n P(B_i \cap A) = \sum_{i=1}^n P(A|B_i)P(B_i)$$

Stokastiske variabler og sannsynlighetsfordelinger

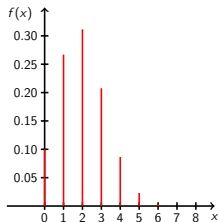
- ★ Stokastisk variabel: Funksjon fra utfallsrom til den reelle tallinja



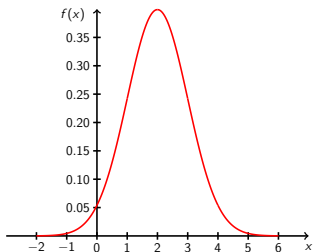
- ★ Diskret og kontinuerlig stokastisk variabel
 - diskret: Endelig eller tellbart uendelig antall mulige verdier
 - kontinuerlig: Ikke-tellbart uendelig mange mulige verdier

$f(x)$: Punktsannsynlighet og sannsynlighetstetthet

- ★ Punktsannsynlighet, $f(x) = P(X = x)$

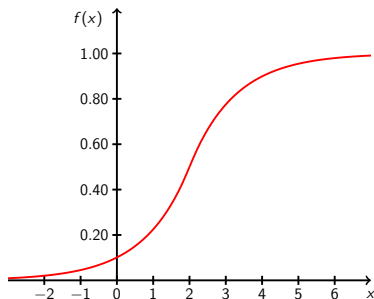
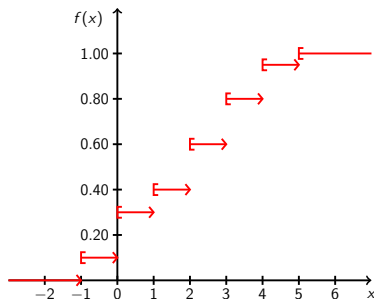


- ★ Sannsynlighetstetthet, $f(x)$: $P(a < X \leq b) = \int_a^b f(x)dx$



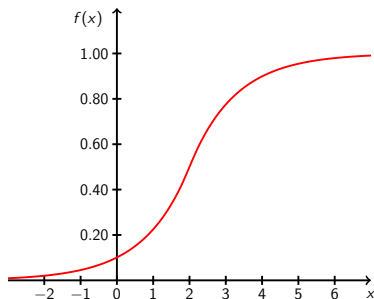
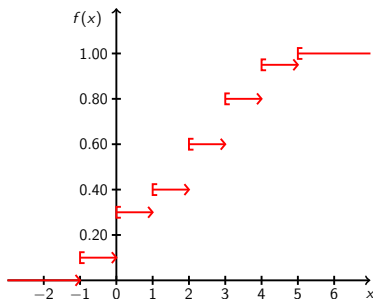
$F(x)$: Kumulativ sannsynlighetsfordeling

- ★ Kumulativ fordeling, $F(x) = P(X \leq x)$



$F(x)$: Kumulativ sannsynlighetsfordeling

- ★ Kumulativ fordeling, $F(x) = P(X \leq x)$



- ★ Merk: Det er en en-til-en sammenheng mellom $f(x)$ og $F(x)$

Simultanfordeling

- ★ Simultan punktsannsynlighet, $f(x, y)$

$$f(x, y) = P(X = x, Y = y)$$

- ★ Simultan sannsynlighetstetthet, $f(x, y)$

$$P((X, Y) \in A) = \iint_A f(x, y) dx dy$$

- ★ Marginalfordeling

- summerer/integrerer bort variabelen vi ikke er interessert i

- ★ Betinget fordeling

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

- ★ X og Y er uavhengige stokastiske variabler hvis

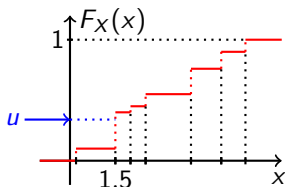
$$f_{XY}(x, y) = f_X(x) \cdot f_Y(y)$$

Stokastisk (Monte Carlo) simulering

- ★ Datamaskin kan generere realisasjoner fra en angitt fordeling $f(x)$
 - «lar datamaskinen kaste terninger»
- ★ Hvis vi har $F(x) = P(X \leq x)$ kan vi generere realisasjon ved

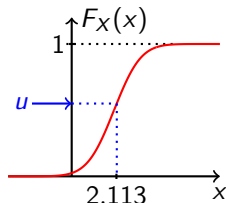
$$U \sim \text{Unif}[0, 1]$$

$$X = \min\{x : F_X(x) \geq U\}$$



$$U \sim \text{Unif}[0, 1]$$

$$X = F_X^{-1}(U)$$

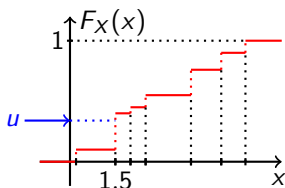


Stokastisk (Monte Carlo) simulering

- ★ Datamaskin kan generere realisasjoner fra en angitt fordeling $f(x)$
 - «lar datamaskinen kaste terninger»
- ★ Hvis vi har $F(x) = P(X \leq x)$ kan vi generere realisasjon ved

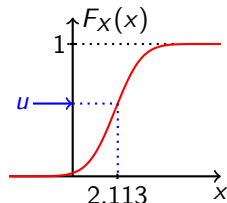
$$U \sim \text{Unif}[0, 1]$$

$$X = \min\{x : F_x(x) \geq U\}$$



$$U \sim \text{Unif}[0, 1]$$

$$X = F_X^{-1}(U)$$



- ★ For noen fordelinger: Kan bruke hvordan fordelingen er definert
 - binomisk
 - t-fordeling

Hva kan vi bruke stokastisk simulering til?

- ★ Estimere sannsynligheter (som vi ikke kan regne ut analytisk)

$$p = P(X \geq 2.4), \quad \hat{p} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(x_i \geq 2.4)$$

Hva kan vi bruke stokastisk simulering til?

- ★ Estimere sannsynligheter (som vi ikke kan regne ut analytisk)

$$p = P(X \geq 2.4), \quad \hat{p} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(x_i \geq 2.4)$$

- ★ Estimere forventningsverdier og varianser

$$\hat{E}[X] = \frac{1}{n} \sum_{i=1}^n x_i \quad \widehat{\text{Var}}[X] = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Hva kan vi bruke stokastisk simulering til?

- ★ Estimere sannsynligheter (som vi ikke kan regne ut analytisk)

$$p = P(X \geq 2.4), \quad \hat{p} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(x_i \geq 2.4)$$

- ★ Estimere forventningsverdier og varianser

$$\hat{E}[X] = \frac{1}{n} \sum_{i=1}^n x_i \quad \widehat{\text{Var}}[X] = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

- ★ Estimere sannsynlighetsfordelinger (ved histogrammer)

Hva kan vi bruke stokastisk simulering til?

- ★ Estimere sannsynligheter (som vi ikke kan regne ut analytisk)

$$p = P(X \geq 2.4), \quad \hat{p} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(x_i \geq 2.4)$$

- ★ Estimere forventningsverdier og varianser

$$\hat{E}[X] = \frac{1}{n} \sum_{i=1}^n x_i \quad \widehat{\text{Var}}[X] = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

- ★ Estimere sannsynlighetsfordelinger (ved histogrammer)
- ★ Estimere kvantiler/kritiske verdier

Hva kan vi bruke stokastisk simulering til?

- ★ Estimere sannsynligheter (som vi ikke kan regne ut analytisk)

$$p = P(X \geq 2.4), \quad \hat{p} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(x_i \geq 2.4)$$

- ★ Estimere forventningsverdier og varianser

$$\hat{E}[X] = \frac{1}{n} \sum_{i=1}^n x_i \quad \widehat{\text{Var}}[X] = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

- ★ Estimere sannsynlighetsfordelinger (ved histogrammer)
- ★ Estimere kvantiler/kritiske verdier
- ★ Estimere p-verdier
 - husk: en p-verdi er en sannsynlighet
 - vi brukte dette for tosidig permutasjonstest

Forventingsverdi og varians

- ★ Forventingsverdi, $\mu = E[X]$

$$E[X] = \begin{cases} \sum_x xf(x) & \text{hvis } X \text{ er diskret SV} \\ \int_{-\infty}^{\infty} xf(x)dx & \text{hvis } X \text{ er kontinuertlig SV} \end{cases}$$

- Tolkning: gjennomsnitt av uendelig mange observasjoner
- regneregler: $E[\cdot]$ er en lineær operator

- ★ Varians, $\sigma^2 = \text{Var}[X] = E[(X - \mu)^2]$

- Tolkning: Mål på spredning/variasjon
- Regneregler:

$$\text{Var}[X] = E[X^2] - E[X]^2$$

$$\text{Var}[aX + b] = a^2\text{Var}[X]$$

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y]$$

- ★ Kovarians og korrelasjon

Viktige diskrete sannsynlighetsfordelinger

★ Binomisk fordeling

- n uavhengige forsøk
- hvert forsøk gir suksess eller fiasko
- sannsynlighet p for suksess i hvert forsøk
- X : antall suksess

$$X \sim b(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}; x = 0, 1, \dots, n$$

- multinomisk fordeling: k mulige resultat i hvert forsøk

★ Negativt binomisk fordeling

- som ved binomisk fordeling, men gjør forsøk inntil k suksesser
- X : antall trekninger

$$X \sim b^*(x; k, p) = \binom{x-1}{k-1} p^k (1-p)^{x-k}; x = k, k+1, \dots$$

- Spesielt: $k = 1$ kalles geometrisk fordeling

$$X \sim g(x; p) = (1-p)^{x-1} p; x = 1, 2, \dots$$

Viktige diskrete sannsynlighetsfordelinger

- ★ Poissonfordeling

- X : antall hendelser i $[0, t]$ i en poissonprosess

$$X \sim p(x; \lambda t) = \frac{e^{-\lambda t} (\lambda t)^x}{x!}; x = 0, 1, \dots$$

Viktige diskrete sannsynlighetsfordelinger

- ★ Mye finnes i formelsamlinga
 - formel for $f(x)$
 - formel for $E[X]$
 - formel for $\text{Var}[X]$
 - formel for $M_X(t)$
 - tabell over $F(x) = P(X \leq x)$ (ikke for geometrisk fordeling)

Viktige diskrete sannsynlighetsfordelinger

★ Mye finnes i formelsamlinga

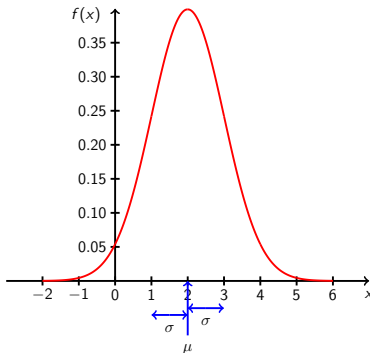
- formel for $f(x)$
- formel for $E[X]$
- formel for $\text{Var}[X]$
- formel for $M_X(t)$
- tabell over $F(x) = P(X \leq x)$ (ikke for geometrisk fordeling)

★ Approksimasjoner

- binomisk $\approx$ poisson hvis n stor og p liten
 - intuitivt rimelig, kan vises matematisk
- binomisk $\approx$ normal hvis n stor
 - spesialtilfelle av sentralgranseteoremet

Viktige kontinuerlige sannsynlighetsfordelinger

★ Normalfordeling: $X \sim n(x; \mu, \sigma)$



- en lineær funksjon (lineærkombinasjon) av uavhengige normalfordelte variabler er normalfordelt
- standardisering:

$$X \sim n(x; \mu, \sigma) \Leftrightarrow Z = \frac{X - \mu}{\sigma} \sim n(z; 0, 1)$$

Viktige kontinuerlige sannsynlighetsfordelinger

- ★ Eksponensialfordeling

- tid til første hendelse i en poissonfordeling/tid mellom etterfølgende hendelser i en poissonfordeling

$$X \sim f(x; \lambda) = \lambda e^{-\lambda x}; x \geq 0$$

- ★ Gammafordeling: $X \sim f(x; \alpha, \beta) = \dots$

- $\alpha = 1, \beta = \frac{1}{\lambda}$: eksponensialfordeling
- $\alpha = \frac{\nu}{2}, \beta = 2$: χ^2_{ν} -fordeling
- tid til k -te hendelse i en poissonprosess: $\alpha = k, \beta = \frac{1}{\lambda}$

- ★ χ^2 -fordeling

- ★ Student t -fordeling

Viktige kontinuerlige sannsynlighetsfordelinger

★ Mye finnes i formelsamlinga

- formel for $f(x)$
- formel for $E[X]$
- formel for $\text{Var}[X]$
- formel for $M_X(t)$
- tabell over $F(x) = P(X \leq x)$ (ikke for eksponensialfordeling)

Sentralgrenseteoremet

- ★ Anta $X_1, X_2, \dots, X_n$ uavhengige og identisk fordelt
- ★ Anta $\mu = E[X_i]$ og $\sigma^2 = \text{Var}[X_i]$
- ★ La $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ og

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

- ★ Resultat: Når $n \rightarrow \infty$ vil fordelingen til Z konvergere mot $n(z; 0, 1)$
- ★ Dette betyr at når n er stor har vi

$$\bar{X} \approx n \left(\bar{x}; \mu, \sqrt{\frac{\sigma^2}{n}} \right)$$

Transformasjoner av stokastiske variabler

- ★ Anta $X \sim f(x)$ og definer Y ved

$$Y = u(X) \Leftrightarrow X = w(Y)$$

- ★ Hvis X (og Y) er diskrete stokastiske variabler:

$$g(y) = f(w(y))$$

- ★ Hvis X (og Y) er kontinuerlige stokastiske variabler:

$$g(y) = f(w(y)) \cdot |w'(y)|$$

Transformasjoner av stokastiske variabler

- ★ Anta $X \sim f(x)$ og definer Y ved

$$Y = u(X) \Leftrightarrow X = w(Y)$$

- ★ Hvis X (og Y) er diskrete stokastiske variabler:

$$g(y) = f(w(y))$$

- ★ Hvis X (og Y) er kontinuerlige stokastiske variabler:

$$g(y) = f(w(y)) \cdot |w'(y)|$$

- ★ Merk:

- formelen for kontinuerlig SV står i formelsamlinga
- alternativt kan man regne ut $G(y) = P(Y \leq y)$ i stedet, og så bruke den til å finne $g(y)$

Feilforplantningsformelen

- ★ Situasjon: Er interessert i

$$V = g(X, Y, Z)$$

- ønsker å approksimere $E[V]$ og $\text{Var}[V]$
- ★ Bruker Taylors formel for rekkeutvikling av $g(x, y, z)$
 - rekkeutvikler x om $\mu_X = E[X]$, y om $\mu_Y = E[Y]$, $\dots$
 - approksimerer ved kun å ta med opp til første ordens ledd
- ★ Får at

$$\begin{aligned} E[V] &\approx g(\mu_X, \mu_Y, \mu_Z) \\ \text{Var}[V] &\approx \left(\frac{\partial g}{\partial x}(\mu_X, \mu_Y, \mu_Z) \right)^2 \text{Var}[X] + \left(\frac{\partial g}{\partial y}(\mu_X, \mu_Y, \mu_Z) \right)^2 \text{Var}[Y] \\ &\quad + \left(\frac{\partial g}{\partial z}(\mu_X, \mu_Y, \mu_Z) \right)^2 \text{Var}[Z] \end{aligned}$$

- ★ Den siste av disse formlene kalles (Gauss) feilforplantningsformel

Momentgenererende funksjoner

★ Definisjon: $M_X(t) = E[e^{tX}]$

★ Regneregler:

- $M_{X+a}(t) = e^{at} M_X(t)$
- $M_{aX}(t) = M_X(at)$
- Hvis $X_1, X_2, \dots, X_n$ er uavhengige er

$$M_{X_1+X_2+\dots+X_n}(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdot \dots \cdot M_{X_n}(t)$$

★ Egenskaper:

- $M_X^{(r)}(0) = E[X^r]$
- X og Y har samme fordeling dersom $M_X(t) \equiv M_Y(t)$

Momentgenererende funksjoner

★ Definisjon: $M_X(t) = E[e^{tX}]$

★ Regneregler:

- $M_{X+a}(t) = e^{at} M_X(t)$
- $M_{aX}(t) = M_X(at)$
- Hvis $X_1, X_2, \dots, X_n$ er uavhengige er

$$M_{X_1+X_2+\dots+X_n}(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdot \dots \cdot M_{X_n}(t)$$

★ Egenskaper:

- $M_X^{(r)}(0) = E[X^r]$
- X og Y har samme fordeling dersom $M_X(t) \equiv M_Y(t)$

★ Vi har brukt momentgenererende funksjoner til å vise en del sammenhenger mellom fordelinger

- lineær funksjon av uavh. normalfordelte variabler er normalfordelt
- sum av uavhengige poissonfordelte variabler er poissonfordelt

Estimator

- ★ $X_1, X_2, \dots, X_n$ er et tilfeldig utvalg fra $f(x)$ -populasjonen
 - $X_1, X_2, \dots, X_n$ er uavhengige og $X_i \sim f(x)$ for $i = 1, 2, \dots, n$
- ★ Observator: En observerbar funksjon av SV er en observator
 - ikke funksjon av ukjente parametre
- ★ Estimator: Observator som brukes til å anslå verdien til parameter
 - parameter: θ, μ, λ, p
 - estimator: $\hat{\theta}, \hat{\mu}, \hat{\lambda}, \hat{p}$

Estimator

- ★ $X_1, X_2, \dots, X_n$ er et tilfeldig utvalg fra $f(x)$ -populasjonen
 - $X_1, X_2, \dots, X_n$ er uavhengige og $X_i \sim f(x)$ for $i = 1, 2, \dots, n$
- ★ Observator: En observerbar funksjon av SV er en observator
 - ikke funksjon av ukjente parametre
- ★ Estimator: Observator som brukes til å anslå verdien til parameter
 - parameter: θ, μ, λ, p
 - estimator: $\hat{\theta}, \hat{\mu}, \hat{\lambda}, \hat{p}$
- ★ Merk: En estimator er en stokastisk variabel
 - har sannsynlighetsfordeling
 - har forventningsverdi
 - har varians

Viktige egenskaper for estimatorer

- ★ $\hat{\theta}$ er en forventningsrett estimator for θ dersom

$$E[\hat{\theta}] = \theta$$

- hvis ikke er den forventningsskjev
- ★ Foretrekker en estimator som er forventningsrett
 - tenker oss at vi gjentar forsøket uendelig mange ganger
- ★ Av flere forv.rette estimatore foretrekker vi den med minst varians
 - tenker oss at vi gjentar forsøket uendelig mange ganger

Viktige egenskaper for estimatorer

- ★ $\hat{\theta}$ er en forventningsrett estimator for θ dersom

$$E[\hat{\theta}] = \theta$$

- hvis ikke er den forventningsskjev
- ★ Foretrekker en estimator som er forventningsrett
 - tenker oss at vi gjentar forsøket uendelig mange ganger
- ★ Av flere forv.rette estimatore foretrekker vi den med minst varians
 - tenker oss at vi gjentar forsøket uendelig mange ganger
- ★ Hvordan bestemme om $\text{Var}[\hat{\theta}]$ eller $\text{Var}[\tilde{\theta}]$ er størst?
 - se på $\text{Var}[\hat{\theta}] - \text{Var}[\tilde{\theta}]$
 - se på $\text{Var}[\hat{\theta}]/\text{Var}[\tilde{\theta}]$
 - merk: kan avhenge av verdien til θ

Sannsynlighetmaksimeringsestimator (SME)

- ★ Idé: Maksimer sannsynligheten for å observere det vi har observert
- ★ Rimelighetsfunksjon: Sannsynligheten for å observere det vi har observert, som funksjon av ukjent parameter θ

- typisk:
$$L(\theta) = f(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

- ★ For å finne SME må vi maksimere $L(\theta)$ m.h.p. θ

- typisk:
$$l(\theta) = \ln L(\theta) = \sum_{i=1}^n \ln f(x_i; \theta)$$

$$l'(\theta) = 0 \Rightarrow \theta = u(x_1, x_2, \dots, x_n)$$

$$\hat{\theta} = u(X_1, X_2, \dots, X_n)$$

Sannsynlighetsmaksimeringestimator (SME)

- ★ Idé: Maksimer sannsynligheten for å observere det vi har observert
- ★ Rimelighetsfunksjon: Sannsynligheten for å observere det vi har observert, som funksjon av ukjent parameter θ

- typisk:
$$L(\theta) = f(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

- ★ For å finne SME må vi maksimere $L(\theta)$ m.h.p. θ

- typisk:
$$l(\theta) = \ln L(\theta) = \sum_{i=1}^n \ln f(x_i; \theta)$$

$$l'(\theta) = 0 \Rightarrow \theta = u(x_1, x_2, \dots, x_n)$$

$$\hat{\theta} = u(X_1, X_2, \dots, X_n)$$

- ★ Husk: Det "typiske" er ikke alltid korrekt, så man må også tenke!

Konfidensintervall

- ★ Anta $X_1, X_2, \dots, X_n$ tilfeldig utvalg fra $f(x; \theta)$ -populasjonen
- ★ Intervallet $[\hat{\theta}_L(x_1, x_2, \dots, x_n), \hat{\theta}_U(x_1, x_2, \dots, x_n)]$ er da et $(1 - \alpha) \cdot 100\%$ konfidensintervall for θ dersom

$$P(\hat{\theta}_L(X_1, X_2, \dots, X_n) \leq \theta \leq \hat{\theta}_U(X_1, X_2, \dots, X_n)) = 1 - \alpha$$

Konfidensintervall

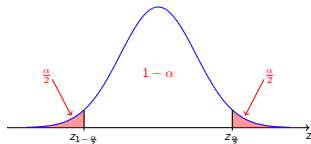
- ★ Anta $X_1, X_2, \dots, X_n$ tilfeldig utvalg fra $f(x; \theta)$ -populasjonen
- ★ Intervallet $[\hat{\theta}_L(x_1, x_2, \dots, x_n), \hat{\theta}_U(x_1, x_2, \dots, x_n)]$ er da et $(1 - \alpha) \cdot 100\%$ konfidensintervall for θ dersom

$$P(\hat{\theta}_L(X_1, X_2, \dots, X_n) \leq \theta \leq \hat{\theta}_U(X_1, X_2, \dots, X_n)) = 1 - \alpha$$

- ★ Hva betyr dette?
 - husk tolkningen av sannsynlighet
 - det er $X_1, X_2, \dots, X_n$ som er stokastiske, θ er (ukjent) tall
 - må tenke oss at vi gjentar forsøket uendelig mange ganger

Hvordan utlede et konfidensintervall

- ★ La $X_1, X_2, \dots, X_n$ være tilfeldig utvalg fra $f(x; \theta)$ -populasjonen
- ★ Finn pivotal $Z = g(X_1, X_2, \dots, X_n, \theta)$
 - kun θ er ukjent
 - har kjent fordeling
- ★ Bestemmer $\frac{\alpha}{2}$ og $1 - \frac{\alpha}{2}$ -kvantilene



$$P(z_{1-\frac{\alpha}{2}} \leq g(X_1, X_2, \dots, X_n, \theta) \leq z_{\frac{\alpha}{2}}) = 1 - \alpha$$

- ★ Løser hver ulikhet m.h.p. θ , setter sammen igjen med θ i midten

$$P(\hat{\theta}_L(X_1, X_2, \dots, X_n) \leq \theta \leq \hat{\theta}_U(X_1, X_2, \dots, X_n)) = 1 - \alpha$$

- ★ Konkluder: $(1 - \alpha) \cdot 100\%$ konfidensintervall for θ er

$$[\hat{\theta}_L(x_1, x_2, \dots, x_n), \hat{\theta}_U(x_1, x_2, \dots, x_n)]$$

Hvilke konfidensintervall har vi diskutert?

★ Ett-utvalg normalfordeling

- Konf.int. for μ , σ^2 kjent: $Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim n(z; 0, 1)$
- Konf.int. for μ , σ^2 ukjent: $T = \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} \sim t_{n-1}$
- Konf.int. for σ^2 , μ kjent: $V = \frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi_n^2$
- Konf.int. for σ^2 , μ ukjent: $V = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$

★ To-utvalg normalfordeling

- Konf.int. for $\mu_1 - \mu_2$, σ_1^2 og σ_2^2 kjente

★ Ett-utvalg, binomisk fordeling

- Konf.int. for p

★ Regresjonsmodell: α , β

Hvilke konfidensintervall har vi diskutert?

- ★ Ett-utvalg normalfordeling

- Konf.int. for μ , σ^2 kjent: $Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim n(z; 0, 1)$
- Konf.int. for μ , σ^2 ukjent: $T = \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} \sim t_{n-1}$
- Konf.int. for σ^2 , μ kjent: $V = \frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi_n^2$
- Konf.int. for σ^2 , μ ukjent: $V = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$

- ★ To-utvalg normalfordeling

- Konf.int. for $\mu_1 - \mu_2$, σ_1^2 og σ_2^2 kjente

- ★ Ett-utvalg, binomisk fordeling

- Konf.int. for p

- ★ Regresjonsmodell: α , β

- ★ Merk: Vi har lært en fremgangsmåte

- ikke bare disse spesialtilfellene

Prediksjonsintervall

- ★ Anta $X_1, X_2, \dots, X_n$ tilfeldig utvalg fra $f(x; \theta)$ -populasjonen. X^* er en ny observasjon fra samme fordeling uavhengig av de gamle
- ★ Intervallet $[X_L^*(x_1, x_2, \dots, x_n), X_U^*(x_1, x_2, \dots, x_n)]$ er da et $(1 - \alpha) \cdot 100\%$ prediksjonsintervall for X^* dersom

$$P(X_L^*(X_1, X_2, \dots, X_n) \leq X^* \leq X_U^*(X_1, X_2, \dots, X_n)) = 1 - \alpha$$

Prediksjonsintervall

- ★ Anta $X_1, X_2, \dots, X_n$ tilfeldig utvalg fra $f(x; \theta)$ -populasjonen. X^* er en ny observasjon fra samme fordeling uavhengig av de gamle
- ★ Intervallet $[X_L^*(x_1, x_2, \dots, x_n), X_U^*(x_1, x_2, \dots, x_n)]$ er da et $(1 - \alpha) \cdot 100\%$ prediksjonsintervall for X^* dersom

$$P(X_L^*(X_1, X_2, \dots, X_n) \leq X^* \leq X_U^*(X_1, X_2, \dots, X_n)) = 1 - \alpha$$

- ★ Hva betyr dette?
 - husk tolkningen av sannsynlighet
 - både $X_1, X_2, \dots, X_n$ og X^* er stokastiske
 - må tenke oss at vi gjentar forsøket uendelig mange ganger

Prediksjonsintervall

- ★ Anta $X_1, X_2, \dots, X_n$ tilfeldig utvalg fra $f(x; \theta)$ -populasjonen. X^* er en ny observasjon fra samme fordeling uavhengig av de gamle
- ★ Intervallet $[X_L^*(x_1, x_2, \dots, x_n), X_U^*(x_1, x_2, \dots, x_n)]$ er da et $(1 - \alpha) \cdot 100\%$ prediksjonsintervall for X^* dersom

$$P(X_L^*(X_1, X_2, \dots, X_n) \leq X^* \leq X_U^*(X_1, X_2, \dots, X_n)) = 1 - \alpha$$

- ★ Hva betyr dette?
 - husk tolkningen av sannsynlighet
 - både $X_1, X_2, \dots, X_n$ og X^* er stokastiske
 - må tenke oss at vi gjentar forsøket uendelig mange ganger
- ★ Hvordan utlede et prediksjonsintervall
 - Starte med $X^* - \bar{X}$ (eller noe lignende) og regne tilsvarende som for konfidensintervall

Hypotesetesting

- ★ To hypoteser
 - nullhypotese, H_0 : Hypotesen vi skal undersøke om det er grunnlag for å forkaste
 - alternativ hypotese, H_1 : Det vi ønsker å undersøke om det er grunnlag for å påstå
- ★ To typer feil

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

- ★ Testobservator, Z : En observator som kan brukes til å formulere en beslutningsregel, med en kjent fordeling når H_0 er riktig
- ★ Beslutningsregel: Forkaster H_0 dersom $Z > k$ (for eksempel)
- ★ Bestemmer kritisk verdi k fra kravet

$$P(\text{Forkast } H_0 | H_0 \text{ er riktig}) = \alpha$$

Hypotesetesting

- ★ To hypoteser
 - nullhypotese, H_0 : Hypotesen vi skal undersøke om det er grunnlag for å forkaste
 - alternativ hypotese, H_1 : Det vi ønsker å undersøke om det er grunnlag for å påstå
- ★ To typer feil

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

- ★ Testobservator, Z : En observator som kan brukes til å formulere en beslutningsregel, med en kjent fordeling når H_0 er riktig
- ★ Beslutningsregel: Forkaster H_0 dersom $Z > k$ (for eksempel)
- ★ Bestemmer kritisk verdi k fra kravet

$$P(\text{Forkast } H_0 | H_0 \text{ er riktig}) = \alpha$$

- ★ p -verdi: Sannsynligheten for å observere det vi har observert eller noe mer ekstremt, når H_0 er sann

Hypotesetesting ved forkastningsområdemetoden

- ★ Anta $X_1, X_2, \dots, X_n$ tilfeldig utvalg fra $n(x; \mu, \sigma)$ -populasjonen
- ★ Anta $n = 30$ og at observerte verdier gir $\bar{x} = 121.6$ og $s^2 = 8.83^2$
- ★ Ønsker å teste

$$H_0 : \mu = 120 \quad \text{mot} \quad H_1 : \mu > 120$$

- ★ Estimator for μ : $\hat{\mu} = \bar{X}$
- ★ Vi vet at

$$\frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} \sim t_{n-1} \quad (\text{både under } H_0 \text{ og } H_1)$$

- ★ Testobservator:

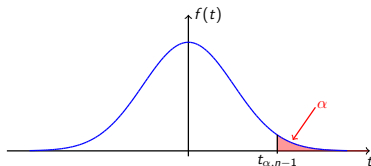
$$T = \frac{\bar{X} - 120}{\sqrt{\frac{S^2}{n}}}$$

- ★ Når H_0 er sann er $T \sim t_{n-1}$

Hypotesetesting ved forkastningsområdemetoden (forts.)

- ★ Testobservator: T
- ★ Vet at $T \sim t_{n-1}$ når H_0 er sann
- ★ Forkastningsregel: Forkaster H_0 dersom $T > k$, der k bestemmes fra kravet

$$P(\text{Forkast } H_0 | H_0 \text{ er sann}) = P(T > k | H_0 \text{ er sann}) = \alpha$$



- ★ Fra figuren ser vi at $k = t_{\alpha, n-1}$
- ★ Forkastningsregelen blir dermed: Forkast H_0 dersom $T > t_{\alpha, n-1}$
- ★ Setter inn tall, og konkluderer

Hypotesetesting ved å regne ut p -verdi

- ★ Anta $X_1, X_2, \dots, X_n$ tilfeldig utvalg fra $n(x; \mu, \sigma)$ -populasjonen
- ★ Anta $n = 30$ og at observerte verdier gir $\bar{x} = 121.6$ og $s^2 = 8.83^2$
- ★ Ønsker å teste

$$H_0 : \mu = 120 \quad \text{mot} \quad H_1 : \mu > 120$$

- ★ Estimator for μ : $\hat{\mu} = \bar{X}$
- ★ Vi vet at

$$\frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} \sim t_{n-1} \quad (\text{både under } H_0 \text{ og } H_1)$$

- ★ Testobservator:

$$T = \frac{\bar{X} - 120}{\sqrt{\frac{S^2}{n}}}$$

- ★ Når H_0 er sann er $T \sim t_{n-1}$

Hypotesetesting ved å regne ut p -verdi (forts.)

- ★ Testobservator: T
- ★ Vet at $T \sim t_{n-1}$ når H_0 er sann
- ★ Forkastningsregel: Forkaster H_0 dersom $T > k$
- ★ Observert verdi for testobservatoren:

$$t = \frac{\bar{x} - 120}{\sqrt{\frac{s^2}{n}}} = \frac{121.6 - 120}{\sqrt{\frac{8.83^2}{30}}} = 0.992$$

- ★ p -verdien er da: $p = P(T > 0.992 | \text{når } H_0 \text{ er sann}) = 0.1647$

Hypotesetesting — hvilke situasjoner har vi sett på?

- ★ Ett-utvalg normalfordeling

- Hypotesetest for μ , σ^2 kjent: $Z = \frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \sim n(z; 0, 1)$
- Hypotesetest for μ , σ^2 ukjent: $T = \frac{\bar{X} - \mu_0}{\sqrt{\frac{S^2}{n}}} \sim t_{n-1}$
- Hypotesetest for σ^2 , μ kjent: $V = \frac{n\hat{\sigma}^2}{\sigma_0^2} \sim \chi_n^2$
- Hypotesetest for σ^2 , μ ukjent: $V = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi_{n-1}^2$

- ★ To-utvalg normalfordeling, $H_0 : \mu_1 = \mu_2$

- Hypotesetest når σ_1^2 og σ_2^2 kjente

- ★ Ett-utvalg, binomisk fordeling, $H_0 : p = p_0$

- Konf.int. for p

- ★ Fortegnstesten

- ★ Tosidig permutasjonstest

- ★ Regresjonsmodell: α , β

Hypotesetesting — hvilke situasjoner har vi sett på?

- ★ Ett-utvalg normalfordeling
 - Hypotesetest for μ , σ^2 kjent: $Z = \frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \sim n(z; 0, 1)$
 - Hypotesetest for μ , σ^2 ukjent: $T = \frac{\bar{X} - \mu_0}{\sqrt{\frac{S^2}{n}}} \sim t_{n-1}$
 - Hypotesetest for σ^2 , μ kjent: $V = \frac{n\hat{\sigma}^2}{\sigma_0^2} \sim \chi_n^2$
 - Hypotesetest for σ^2 , μ ukjent: $V = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi_{n-1}^2$
- ★ To-utvalg normalfordeling, $H_0 : \mu_1 = \mu_2$
 - Hypotesetest når σ_1^2 og σ_2^2 kjente
- ★ Ett-utvalg, binomisk fordeling, $H_0 : p = p_0$
 - Konf.int. for p
- ★ Fortegnstesten
- ★ Tosidig permutasjonstest
- ★ Regresjonsmodell: α , β
- ★ Merk: Vi har lært en fremgangsmåte
 - ikke bare disse spesialtilfellene

Hypotesetesting — teststyrke

- ★ Anta $X_1, X_2, \dots, X_n$ tilfeldig utvalg fra $n(x; \mu, \sigma_0)$ -populasjonen
- ★ Ønsker å teste: $H_0 : \mu = 3.5$ mot $H_1 : \mu < 3.5$
- ★ Estimator for μ : $\hat{\mu} = \bar{X}$
- ★ Vi vet at

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma_0^2}{n}}} \sim n(z; 0, 1) \quad (\text{både under } H_0 \text{ og } H_1)$$

- ★ Testobservator:

$$Z = \frac{\bar{X} - 3.5}{\sqrt{\frac{\sigma_0^2}{n}}}$$

- ★ Når H_0 er sann er $Z \sim n(z; 0, 1)$
- ★ Forkastningskriterium: Forkaster H_0 når $Z < -z_\alpha$
- ★ Spørsmål:
 - Hva blir $P(\text{Forkast } H_0 | \mu = 3.3)$?

Hypotesetesting — teststyrke

- ★ Anta $X_1, X_2, \dots, X_n$ tilfeldig utvalg fra $n(x; \mu, \sigma_0)$ -populasjonen
- ★ Ønsker å teste: $H_0 : \mu = 3.5$ mot $H_1 : \mu < 3.5$
- ★ Estimator for μ : $\hat{\mu} = \bar{X}$
- ★ Vi vet at

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma_0^2}{n}}} \sim n(z; 0, 1) \quad (\text{både under } H_0 \text{ og } H_1)$$

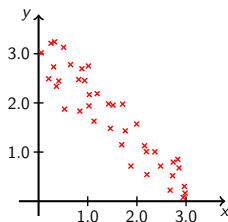
- ★ Testobservator:

$$Z = \frac{\bar{X} - 3.5}{\sqrt{\frac{\sigma_0^2}{n}}}$$

- ★ Når H_0 er sann er $Z \sim n(z; 0, 1)$
- ★ Forkastningskriterium: Forkaster H_0 når $Z < -z_\alpha$
- ★ Spørsmål:
 - Hva blir $P(\text{Forkast } H_0 | \mu = 3.3)$?
 - Hvor stor må n være for at $P(\text{Forkast } H_0 | \mu = 3.3) \geq 0.2$?

Enkel lineær regresjon

- ★ Har observasjonspaar $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$

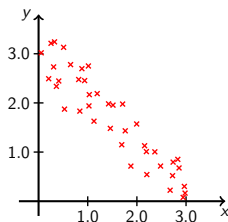


- ★ Modell: $Y_1, Y_2, \dots, Y_n$ er uavhengige og

$$Y_i \sim n(y_i; \alpha + \beta x_i, \sigma)$$

Enkel lineær regresjon

- ★ Har observasjonspaar $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$



- ★ Modell: $Y_1, Y_2, \dots, Y_n$ er uavhengige og

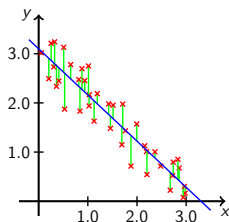
$$Y_i \sim n(y_i; \alpha + \beta x_i, \sigma)$$

- ★ Minste kvadraters metode: Finner estimatorer for α og β ved å minimere

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2$$

Enkel lineær regresjon

- ★ Har observasjonspaar $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$



- ★ Modell: $Y_1, Y_2, \dots, Y_n$ er uavhengige og

$$Y_i \sim n(y_i; \alpha + \beta x_i, \sigma)$$

- ★ Minste kvadraters metode: Finner estimatorer for α og β ved å minimere

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2$$

Hva mangler i denne repetisjonen

- ★ Mange detaljer — utledninger
- ★ Regning på konkrete situasjoner
 - mange eksamensoppgaver har fokus her

Noen råd før og på eksamen

Noen råd FØR eksamen

- ★ Få oversikt over emnet
 - mer enn en lang rekke av spesialtilfeller
- ★ Regn tidligere eksamensoppgaver
 - bruk tillatte hjelpemidler
 - mai og august 2020 var bestått/ikke-bestått: annerledes!
 - november 2020, mai og august 2021 var hjemmeeksamen: varianter av oppgaver
 - bruk løsningsforslagene fornuftig!!!
 - husk: pensum er litt endret i år!
- ★ Spør hverandre, kom på fysisk statistikklab (gjerne i grupper)
 - denne uka er siste uke med vanlig fysisk statistikklab!
 - tid og sted for fysisk statistikk før eksamen kommer senere
 - still spørsmål på digital statistikklab

Noen råd PÅ eksamensdagen

- ★ Husk å ta med tillatte hjelpemidler
- ★ Vær ute i (veldig) god tid
- ★ Alle deloppgaver (1a,1b,2a,2b,2c,3a,3b) teller like mye
 - alle er ikke like vanskelige
 - alle innebærer ikke like mye arbeid
- ★ Faglærere går rundt på eksamen
 - det er lov å spørre om hva du vil
 - vi kan ikke svare på alle spørsmål
- ★ Les oppgaveteksten nøye:
 - det er forskjell på “skriv opp” og “utled”
 - det er forskjell på “forklar” og “bevis”
- ★ **Skriv mellomregning** slik at det er mulig å forstå hva du har gjort/hvordan du har tenkt — før besvarelsen fornuftig
- ★ Skriv slik at det er mulig **for andre** å lese håndskriften din!
- ★ Pass på at du besvarer (alle) spørsmålene!

Tillatte hjelpemidler på eksamen

- ★ Tabeller og formler i statistikk
- ★ Bestemt, enkel kalkulator
- ★ Ett stemplet gult A5-ark med egne håndskrevne (skrevet rett på arket) formler og notater
 - du kan skrive på alle sider og med alle farger

Statistikk er mer enn TMA4240

Hva har vi gjort i TMA4240?

- ★ Sannsynlighetsregning:
 - definert sannsynlighet og stokastiske variabler
 - innført regneregler for sannsynligheter og sannsynlighetsfordelinger
 - definert forventning og varians, og sett på regneregler for disse
- ★ Statistisk inferens:
 - stort sett: $X_1, \dots, X_n$ tilfeldig utvalg
 - punktestimator, konfidensintervall, hypotesetesting
- ★ Hva mer finnes av statistikk?

Hva har vi gjort i TMA4240?

- ★ Sannsynlighetsregning:
 - definert sannsynlighet og stokastiske variabler
 - innført regneregler for sannsynligheter og sannsynlighetsfordelinger
 - definert forventning og varians, og sett på regneregler for disse
- ★ Statistisk inferens:
 - stort sett: $X_1, \dots, X_n$ tilfeldig utvalg
 - punktestimator, konfidensintervall, hypotesetesting
- ★ Hva mer finnes av statistikk?
 - andre forsøksopplegg, ikke identisk fordelt og/eller avhengige

Hva har vi gjort i TMA4240?

- ★ Sannsynlighetsregning:
 - definert sannsynlighet og stokastiske variabler
 - innført regneregler for sannsynligheter og sannsynlighetsfordelinger
 - definert forventning og varians, og sett på regneregler for disse
- ★ Statistisk inferens:
 - stort sett: $X_1, \dots, X_n$ tilfeldig utvalg
 - punktestimator, konfidensintervall, hypotesetesting
- ★ Hva mer finnes av statistikk?
 - andre forsøksopplegg, ikke identisk fordelt og/eller avhengige
 - ikke-parametrisk statistikk

Hva har vi gjort i TMA4240?

- ★ Sannsynlighetsregning:
 - definert sannsynlighet og stokastiske variabler
 - innført regneregler for sannsynligheter og sannsynlighetsfordelinger
 - definert forventning og varians, og sett på regneregler for disse
- ★ Statistisk inferens:
 - stort sett: $X_1, \dots, X_n$ tilfeldig utvalg
 - punktestimator, konfidensintervall, hypotesetesting
- ★ Hva mer finnes av statistikk?
 - andre forsøksopplegg, ikke identisk fordelt og/eller avhengige
 - ikke-parametrisk statistikk
 - modeller/problemer hvor man ikke har analytiske svar

Hva har vi gjort i TMA4240?

- ★ Sannsynlighetsregning:
 - definert sannsynlighet og stokastiske variabler
 - innført regneregler for sannsynligheter og sannsynlighetsfordelinger
 - definert forventning og varians, og sett på regneregler for disse
- ★ Statistisk inferens:
 - stort sett: $X_1, \dots, X_n$ tilfeldig utvalg
 - punktestimator, konfidensintervall, hypotesetesting
- ★ Hva mer finnes av statistikk?
 - andre forsøksopplegg, ikke identisk fordelt og/eller avhengige
 - ikke-parametrisk statistikk
 - modeller/problemer hvor man ikke har analytiske svar
 - bayesiansk statistikk (dere har lært frekventistisk statistikk)

Naturlige neste statistikk-kurs

- ★ TMA4265 Stokastisk modellering
 - sannsynlighetsregning
 - regning på prosesser som utvikler seg i tid, markovegenskap
 - mye kan løses analytisk
- ★ TMA4255 Anvendt statistikk
 - statistisk inferens
 - multippel lineær regresjon, modellvalg, forsøksplanlegging, ikke-parametriske metoder
- ★ TMA4267 Lineære statistiske modeller
 - statistisk inferens
 - mye av det samme som TMA4255, men undervises mer matematisk
- ★ TMA4268 Statistisk læring
 - regresjon og klassifikasjon, modellvalg/regularisering, ikke-linearitet
 - bruker funksjoner i R