Høgskolen i Østfold

EKSAMEN

Emnekode:	Emnenavn:
SFB10711	Metodekurs 1, skriftlig eksamen vår - statistikk
Dato:	**Eksamenstid:**
22.11.2019	4 timer
Hjelpemidler:	**Faglærere:**
Formelsamling (vedlagt) Godkjent kalkulator	Henrik Sætra

Om eksamensoppgaven og poengberegning:

Oppgavesettet består av 4 sider inklusiv denne forsiden. Kontroller at oppgavesettet er komplett før du begynner å besvare oppgaven. I tillegg kommer 16 sider med formelsamling.

Alle oppgavene skal besvares og teller som angitt ved sensurering.

Sensurfrist: 12.12.2019

Karakterene er tilgjengelige for studenter i Studentweb.

Eksamen

SFB10711 Metodekurs I – deleksamen 2

Høsten 2019 - ny/utsatt eksamen

1 15%

Du jobber i et selskap som produserer spesialdesignede datamaskiner.

Dere har nå fått et vareparti med 225 varer, og sannsynligheten for at en vare er defekt er 20%.

a) Hva er sannsynligheten for at høyst 50 varer er defekte?

b) Hva er sannsynligheten for at minst 35 varer er defekte?

c) Finn $E[X]$, $Var[X]$ og $\sigma[X]$.

En undersøkelse viser at gjennomsnittlig årlig inntekt for deres kunder (μ) er 750.000 kroner med et standardavvik (σ) på 100.000.

d) Hva er sannsynligheten for at en tilfeldig valgt kunde tjener mindre enn 650.000 kroner?

e) Hva er sannsynligheten for at en tilfeldig valgt kunde tjener mellom 700.000 og 800.000 kroner?

2 10%

Dere skal gjennomføre en spørreundersøkelse, og som en del av forberedelsene til denne har dere testet spørreskjemaet på 20 personer. Dere spurte bl.a. respondentene om hvor mange app'er til telefonene sine de hadde kjøpt i år. Gjennomsnittet $\bar{X}$ ble 4,7 apper, og det var 20 respondenter. S_X^2 var 21,8.

a) Lag et 95% konfidensintervall for gjennomsnittet $\bar{X}$.

Senere sendte dere ut spørreundersøkelsen til et større utvalg, og nå ble gjennomsnittet 4,5. Dere vet at standardavviket i populasjonen σ er 3. 150 personer ble nå undersøkt.

b) Lag et nytt 95% konfidensintervall for gjennomsnittet $\bar{X}$.

3 15%

I forbindelse med en datamesse dere skal arrangere, blir du bedt om å analysere salget av mat. I messens matutsalg selger dere to menyer med mat – en til 100 kroner og en til 150 kroner. 60% av kundene som kjøper en meny kjøper den til 100 kroner, og 40% kjøper den som koster 150 kroner.

a) Hva er forventet gjennomsnittlig pris kundene betaler for en meny ($E[X]$), og hva er variansen og standardavviket til X?

Når dere har samlet inn salgstall fra matbodene setter dere opp følgende tabell som viser simultansannsynlighetene for forskjellige priser (X) og omsetning (Y):

Menypris * Antall solgt Crosstabulation

% of Total

		Antall solgt			
		50.00	100.00	150.00	200.00
Menypris	100.00	6.7%	20.0%	13.3%	20.0%
	150.00	13.3%	20.0%	6.7%	

Figur 1: Simultan sannsynlighetsfordeling

Oppgitte verdier: $E[Y] = 120$ og $Var[Y] = 2600$.

b) Regn ut korrelasjonskoeffisienten $\rho[X, Y]$.

4 15%

Du skal nå besøke en festival. Av besøkende til festivalen, er 60% menn. Statistikken viser videre at 30% av mannlige besøkende kjøper teltplass, og 20% av kvinnene gjør det samme. For begge kjønn viser det seg at 70% av de som betaler for teltplass også kjøper et kort som heter *Teltmat* (som kun er tilgjengelig for besøkende med teltplass), som gir de teltende tilgang til frokost- og lunsjbuffet.

a) Hvor stor andel av de besøkende (totalt) kjøper kortet *Teltmat*?

Det ene året var *Kvelertak* hovedattraksjonen på festivalen. Dere gjennomførte en spørreundersøkelse dette året. 60% var menn, og 35% svarte at hovedgrunnen til at de kom til festivalen var for å se *Kvelertak*. 30% av de som svarte at hovedgrunnen til at de kom var *Kvelertak*, var kvinner.

b) Hva er sannsynligheten for at hovedgrunnen for at man kom var *Kvelertak*, dersom man er kvinne?

Det er to likeverdige hovedscener på festivalen, og dere har 7 band som er aktuelle for å headline hver sin hovedscene på festivalens avslutningskveld.

c) Hvor mange mulige kombinasjoner av headlinere på avslutningskvelden finnes?

9 band har vunnet en lokal konkurranse, som lar dem spille på en mindre scene i løpet av festivalen. 3 av disse skal spille på åpningsdagen, og dere har bestemt dere for å sette opp rekkefølgen av bandene tilfeldig.

d) Hvor mange mulige bandoppsett finnes for denne scenen på åpningsdagen?

5 10%

Du har nå blitt sendt på en datamesse, og du har fått ansvar for å sette opp standen deres. Dere lager 6 forskjellige datamaskiner, og du skal stille ut en av hver modell. Disse skal settes ved siden av hverandre på rekke.

a) Hvor mange måter kan sortere de seks typene på?

På datamaskinen du skal bruke på messen blir du bedt om å lage en pin-kode, den skal bestå av 4 siffer (0-9).

b) Hvor mange mulige pin-koder kan du lage?

c) Hvor mange pin-koder kunne du laget om koden var 6 siffer lang?

6 15%

Dere har tidligere antatt at minst 50% av befolkningen årlig leser en av deres databøker. Nå frykter ledelsen at denne andelen er lavere enn 50%, og du blir bedt om å undersøke dette. Grunnlaget er den siste spørreundersøkelsen, hvor 63 av 150 svarte at de hadde lest en av deres bøker de siste året.

a) Lag et 95% konfidensintervall for $\hat{p}$ som estimerer andelen i populasjonen som har lest minst en av deres bøker.

b) Foreta en hypotesetest for å avgjøre om andelen som leser minst en av deres bøker har minket. Sett opp H_0, H_A og bruk 95% sikkerhetsnivå.

c) Beregn p-verdien til hypotesetesten.

7 20%

Salget (X) av en datakomponent dere lager (X-drive) i årene 2014-2018 er oppgitt nedenfor:

2014	2015	2016	2017	2018
275	225	300	325	375

a) Hva er gjennomsnittet ($\bar{X}$), variansen (S_X^2) og standardavviket (S_X) for salget av denne harddisken?

Dere har så sett på salget til en annen type harddisk en konkurrent lager og selger, for samme perioder (Y). Tallet for denne (Y-Drive) tall er 450, 450, 400, 350 og 350. $\bar{Y} = 400$ og $S_Y^2 = 2500$.

b) Beregn korrelasjonskoeffisienten R_{XY} for boksalget til X-drive (X) og Y-drive (Y).

c) Hva kan du si om forholdet mellom salget til de to harddiskene, når ledelsen spør deg om hva disse tallene egentlig *betyr*?

Formelsamling

Henrik Skaug Sætra

22.4.2019

α	`\alpha`	θ	`\theta`	σ	`\sigma`	τ	`\tau`
β	`\beta`	ρ	`\rho`	λ	`\lambda`	χ	`\chi`
δ	`\delta`	μ	`\mu`	Ω	`\Omega`	$\sum$	`\sum`
$\pm$	`\pm`	$\in$	`\in`	$\ni$	`\ni`	∞	`\infty`
$\subset$	`\subset`	$\supset$	`\supset`	$\cap$	`\cap`	$\cup$	`\cup`
$\hat{p}$	`\hat{p}`	$\bar{X}$	`\bar{X}`	$\approx$	`\approx`	$\vert$	`\vert`
$\frac{x_1+x_2}{n-1}$	`\frac{x_1+x_2}{n-1}`	$\binom{N}{s}$	`\binom{N}{s}`				

Tabell 1: Koder for Inspera/Canvas

1 Deskriptiv statistikk

1.1 Kvartiler og median

Første kvartil: $\frac{n+1}{4}$

Andre kvartil (median): $\frac{n+1}{2}$

Tredje kvartil: $\frac{3(n+1)}{4}$

Vi finner kvartilene med formlene ovenfor, men vi kan få en rest som er enten 0, 0.25, 0.5 eller 0.75 - hva gjør vi da?

1. Hvis resten er null, bruker vi observasjon k
2. Hvis resten er 0.25, starter vi i observasjon nr, $(k - 0.25)$ og flytter oss 25% av avstanden til neste observasjon
3. Hvis resten er 0.5, starter vi i observasjon nr, $(k - 0.5)$ og flytter oss 50% av avstanden til neste observasjon
4. Hvis resten er 0.75, starter vi i observasjon nr, $(k - 0.75)$ og flytter oss 75% av avstanden til neste observasjon

1.2 Modalprosent

Modalprosent: $\frac{modus}{n}$

1.3 Gjennomsnitt

Gjennomsnitt:

$$\bar{X} = \frac{1}{n}(X_1 + \ldots + X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1}$$

1.4 Utvalgsvarians og utvalgsstandardavvik

Utvalgsvarians:

$$\begin{aligned} S_X^2 &= \frac{1}{n-1}((X_1 - \bar{X})^2 + \ldots + (X_n - \bar{X})^2) \\ &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \end{aligned} \tag{2}$$

Utvalgsstandardavvik:

$$S_X = \sqrt{S_X^2} \tag{3}$$

1.5 Utvalgskovarians

Utvalgskovarians:

$$\begin{aligned} S_{XY} &= \frac{1}{n-1}(X_1 - \bar{X})(Y_1 - \bar{Y}) + \ldots + (X_n - \bar{X})(Y_n - \bar{Y}) \\ &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) \end{aligned} \tag{4}$$

Korrelasjonskoeffisienten:

$$R_{XY} = \frac{S_{XY}}{S_X \cdot S_Y} \tag{5}$$

2 Sannsynlighetsregning

Utfallsrom: $\Omega =$ Mengden av alle mulige utfall

Begivenhet: En delmengde av Ω

Sannsynligheten for en begivenhet: $P(A) = \sum_{u \in A} p(u)$

Uniform sannsynlighetsmodell: $P(A) = \frac{a}{n} = \frac{\textbf{gunstige}}{\textbf{mulige}}$

Komplement: $A^C =$ A skjer ikke

Delmengde: $B \subset A =$ B er en delmengde av A

Ikke delmengde: $B \not\subset A =$ B medfører ikke alltid A

Tom delmengde: $A \cap B = \emptyset =$ A og B inntreffer aldri samtidig

Sannsynlighetsmodell:

1. $0 \leq p_i \leq 1, i = 1m \ldots n$

2. $p_1 + p_2 + \ldots + p_n = 1$

Addisjonsformler:

$$P(A) + P(A^C) = 1, \qquad P(A^C) = 1 - P(A) \tag{6}$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{7}$$
$$\begin{aligned} P(A \cup B \cup C) = &P(A) + P(B) + P(C) \\ &- P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &+ P(A \cap B \cap C) \end{aligned} \tag{8}$$

3 Kombinatorikk

Ordnet med tilbakelegging:

$$N^s = N \cdot N \cdot N \cdot \ldots \cdot N \tag{9}$$

Ordnet uten tilbakelegging:

$$\begin{aligned} (N)_s &= N \cdot (N-1) \cdot \ldots \cdot (N-s+1) \\ &= \frac{N!}{(N-s)!} \end{aligned} \tag{10}$$

Uordnet uten tilbakelegging:

$$\begin{aligned} \binom{N}{s} &= \frac{N \cdot (N-1) \cdot \ldots \cdot (N-s+1)}{s!} \\ &= \frac{N!}{s!(N-s)!} \end{aligned} \tag{11}$$

Uordnet med tilbakelegging:

$$\binom{N+s-1}{s} \tag{12}$$

4 Betinget sannsynlighet

Definisjon av betinget sannsynlighet:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{13}$$

Bayes' regel:

$$P(B|A) = P(A|B) \cdot \frac{P(B)}{P(A)} \tag{14}$$

Oppsplittingsprinsippet (loven om total sannsynlighet): (når $\Omega = \cup \ldots \cup B_n$ og $B_1, \ldots, B_n$ alle er disjunkte)

$$P(A) = P(A|B_1) \cdot P(B_1) + \ldots + P(A|B_n) \cdot P(B_n) \tag{15}$$
$$P(A) = P(A|B) \cdot P(B) + P(A|B^C) \cdot P(B^C) \tag{16}$$

To begivenheter er uavhengige når:

$$P(A \cap B) = P(A) \cdot P(B) \tag{17}$$

5 Tilfeldige variabler, forventning og varians

Tilfeldig variabel: Enhver funksjon som til ethvert mulig utfall definerer et bestemt reelt tall.

Fordelingsfunksjonen til X:

$$P(x) = P(X = x) \tag{18}$$

Den kumulative fordelingsfunksjonen til X:

$$F(x) = P(X \leq x) = \sum_{x_r \leq x} P(X \leq x_r) \tag{19}$$

Forventningen til X:

$$E[X] = \sum_{x} x \cdot P(X = x) \tag{20}$$

Forventningen til h(X):

$$E[h(X)] = \sum_{x} h(x) \cdot P(X = x) \tag{21}$$

Variansen til en diskret variabel X:

$$Var[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2 \tag{22}$$

Standardavviket til en tilfeldig variabel X:

$$\sigma[X] = \sqrt{Var[X]} \tag{23}$$

Regneregler for forventning og varians (a og b konstanter):

$$\begin{aligned} E[X+Y] = E[X] + E[Y], \qquad E[a] = a, \qquad E[b \cdot X] = b \cdot E[X] \\ Var[a + bX] = b^2 Var[X] \end{aligned} \tag{24}$$

6 Simultane sannsynlighetsfordelinger

Simultanfordelingen for to tilfeldige variabler X og Y:

$$p(x, y) = P(X = x, Y = y) \tag{25}$$

Marginalsannsynligheten til X:

$$P(X = x) = \sum_{y} p(x, y) \tag{26}$$

Marginalsannsynligheten til Y:

$$P(Y = y) = \sum_{x} p(x, y) \tag{27}$$

Forventningen til $h(X, Y)$):

$$E[h(X, Y)] = \sum_{x,y} h(x, y) p(x, y) \tag{28}$$

Kovariansen til X og Y:

$$\begin{aligned} Cov[X,Y] =& E[(X-E[X])(Y-E[Y])] \\ =& E[X\cdot Y]-E[X]\cdot E[Y] \end{aligned} \tag{29}$$

Hvis X og Y er uavhengige, da er:

$$\begin{aligned} &p(x,y) = P(X=x)\cdot P(Y=y) \\ &Cov[X,Y]=0 \\ &Var[X+Y] = Var[X]+Var[Y] \end{aligned} \tag{30}$$

Generelt gjelder:

$$Var[X+Y] = Var[X]+Var[Y]+2Cov[X,Y] \tag{31}$$

Korrelasjonskoeffisienten:

$$\rho[X,Y] = \frac{Cov[X,Y]}{\sqrt{Var[X]}\cdot\sqrt{Var[Y]}} \in [-1,1] \tag{32}$$

Hvis $X_1, X_2, \ldots X_n$ alle er uavhengige, da er:

$$Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var[X_i] \tag{33}$$

7 Sentrale sannsynlighetsfordelinger

X = Binomisk(n,p) :

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad E[X]=np, \qquad Var[X]=np(1-p) \tag{34}$$

X = Hypergeometrisk(N,M,n):

$$\begin{aligned} P(X=x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \qquad E[X] = n\frac{M}{N}, \\ Var[X] = \frac{N-n}{N-1}\cdot n\frac{M}{N}\cdot\left(1-\frac{M}{N}\right) \end{aligned} \tag{35}$$

X = Poisson λ :

$$P(X=x) = \frac{\lambda^x}{x!}e^{-\lambda}, \qquad E[X]=\lambda, \qquad Var[X]=\lambda \tag{36}$$

X = N[0,1] = Standard normalfordeling. Kontinuerlig fordeling:

$$P(X\in[a,b]) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}dx, \qquad E[X]=0, \qquad Var[X]=1 \tag{37}$$

Kumulativ fordelingsfunksjon for normalfordelingen (verdiene finnes i tabell):

$$G(z) = P(X\le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}dx \tag{38}$$

Her gjelder:

$$G(-z) = 1 - G(z), \qquad P(a \leq X \leq b) = G(b) - G(a) \tag{39}$$

Kumulativ fordelingsfunksjon generelt:

$$G(z) = P(X \leq z) = \int_{-\infty}^{z} h(x)dx \tag{40}$$

Sentralgrensesetningen:

$$\lim_{n\to\infty} P\left(\frac{S_n - E[S_n]}{\sigma[S_n]} \leq z\right) = G(z) \tag{41}$$

Kriterier for når vi kan bruke normaltilnærming:

Binomisk	$np(1-p) \geq 5$	OK
Binomisk	$np(1-p) \geq 10$	God
Hypergeometrisk	$N \geq 20n$ og $n\frac{M}{N}\left(1-\frac{M}{N}\right) \geq 5$	OK
Hypergeometrisk	$N \geq 20n$ og $n\frac{M}{N}\left(1-\frac{M}{N}\right) \geq 10$	God
Poisson	$\lambda \geq 5$	OK
Poisson	$\lambda \geq 10$	God

Hvis X er binomisk og vi bruker normalfordelingen:

$$P(X \leq x) \approx G\left(\frac{x - n \cdot p}{\sqrt{n \cdot p(1-p)}}\right) \tag{42}$$

Mange utregninger er basert på følgende prinsipp::

$$P(S \leq s) = P\left(\frac{S - E[S]}{\sigma[S]} \leq \frac{s - E[S]}{\sigma[S]}\right) \approx G\left(\frac{s - E[S]}{\sigma[S]}\right) \tag{43}$$

Sentralgrensesetningen for gjennomsnitt:

$$S \approx N[n\mu, n\sigma^2], \quad \bar{X} = N\left[\mu, \frac{\sigma^2}{n}\right] \tag{44}$$

$$E[S] = \mu n, \quad Var[S] = n\sigma^2, \quad E[\bar{X}], \quad Var[\bar{X}] = \frac{\sigma^2}{n} \tag{45}$$

Heltallskorreksjon: I situasjoner der S og s skal være hele tall, er:

$$P(S \leq s) = P(S \leq s + 0,5) \approx G\left(\frac{s + 0,5 - E[S]}{\sigma[S]}\right) \tag{46}$$

8 Estimering og estimatorer

Forventningsrett estimator for et tall θ: En tilfeldig variabel $\hat{\theta}$ med $E[\hat{\theta}] = \theta$

Forventningsskjev estimator for et tall θ: En tilfeldig variabel $\hat{\theta}$ med $E[\hat{\theta}] \neq \theta$

Rapportering av resultater: En rapporterer gjerne resultater på formen:

$$observasjon \pm standardfeil = \hat{\theta} \pm \sigma[\hat{\theta}] \tag{47}$$

Målemodell:

$$\bar{X} \pm \frac{S_X}{\sqrt{n}} \tag{48}$$

Standardfeil for $\bar{X}$:

$$\sigma_{\bar{X}} = \frac{S_X}{\sqrt{n}} \tag{49}$$

Binomisk modell med $\hat{p} = \frac{X}{n}$:

$$\hat{p} \pm \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{50}$$

$(1-\alpha)100\%$ konfidensintervall: Et intervall $[\theta_1, \theta_2]$ slik at det er $(1-\alpha)100\%$ sannsynlighet for at konstanten $\theta \in [\theta_1, \theta_2]$

Normalfordeling, forventningsrett estimator og kjent standardavvik:

$$\begin{aligned}
P(|\hat{\theta}-\theta| \leq d) = P\left(\left|\frac{\hat{\theta}-\theta}{\sigma(\hat{\theta})}\right|\right) &\approx G(z) - G(-z) && (51)\\
&= G(z) - (1 - G(z)) && (52)\\
&= 2G(z) - 1 && (53)
\end{aligned}$$

Tolkning av rapport:

- 68,3% sannsynlighet for at intervallet mellom $\hat{\theta} \pm \sigma[\hat{\theta}]$ inneholder θ
- 95,4% sannsynlighet for at intervallet mellom $\hat{\theta} \pm 2\sigma[\hat{\theta}]$ inneholder θ
- 99,7% sannsynlighet for at intervallet mellom $\hat{\theta} \pm 3\sigma[\hat{\theta}]$ inneholder θ
- $\hat{\theta} \pm 1,645\sigma[\hat{\theta}]$ er et 90% konfidensintervall
- $\hat{\theta} \pm 1,96\sigma[\hat{\theta}]$ er et 95% konfidensintervall
- $\hat{\theta} \pm 2,58\sigma[\hat{\theta}]$ er et 99% konfidensintervall
- Generelt: $\hat{\theta} \pm z\sigma[\hat{\theta}]$ er et $2G(z)-1$ konfidensintervall

I en binomisk modell der $np(1-p) \geq 10$, bruker vi tilnærming til normalfordelingen, og et 95 % konfidensintervall er gitt ved:

$$\hat{p} \pm 1,96 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{54}$$

I en målemodell der vi har mange observasjoner, bruker vi normalfordelingen, og et 95 % konfidensintervall er gitt ved:

$$\bar{X} \pm 1,96 \cdot \frac{S_X}{\sqrt{n}} \tag{55}$$

Ved få observasjoner, og tilfeldig normalfordelt variabel, kan vi bruke t-fordelingen for å bestemme konfidensintervaller. Finn først $t_{\alpha/2}^{n-1}$ slik at:

$$P\left(T_{n-1} \geq t_{\alpha/2}^{n-1}\right) = \frac{\alpha}{2} \tag{56}$$

Et $(1-\alpha)100\%$ konfidensintervall er da gitt ved grensene

$$\bar{X} \pm t_{\alpha/2}^{n-1} \frac{S_X}{\sqrt{n}} \tag{57}$$

1 I lotterimodellen er $\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y^i$ en forventningsrett estimator for gjennomsnittet av hele populasjonen $\bar{v}$, og

$$Var[\bar{Y}] = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n} \tag{58}$$

9 Hypotesetesting

En hypotesetest har følgende elementer:

- En nullhypotese H_0
- En alternativ hypotese H_A
- En testobservator $\hat{\theta}$
- Et forkastningsområde A

Testen gjennomføres ved at en observerer $\hat{\theta}$. En skal forkaste nullhypotesen H_0 til fordel for H_A dersom $\hat{\theta} \in A$.

Signifikansnivå: En hypotesetest har $\alpha\%$ signifikansnivå når det er høyst $\alpha\%$ sjanse for at en forkaster H_0 når H_0 er korrekt (forkastningsfeil).

Styrke/godtakingsfeil: En alternativ θ i H_A har styrke $\gamma(\theta) = 1-\beta\%$ dersom det er høyst $\beta\%$ sjanse for å godta H_0 når H_0 er feil mens alternativet er korrekt godtakingsfeil. Stor styrke gir altså lite sannsynlighet for godtakingsfeil.

P-verdi: P-verdien for en test måles fra den observerte verdien for $\hat{\theta}$. Med utgangspunkt i at nullhypotesen er korrekt, skal en regne ut hvor ofte en kan få observasjoner som avviker minst like mye fra H_0 som $\hat{\theta}$. Hvis signifikansnivået α er mindre enn P-verdien, skal vi beholde nullhypotesen. Omvendt vil nullhypotesen bli forkastet på ethvert signifikansnivå større eller like P-verdien.

9.1 Noen spesielle hypotesetester

Test for sannsynlighet i en binomisk forsøksrekke:

- Er sannsynligheten for suksess forskjellig fra før?
- Modellantagelser: $H_0 : p = p_0, \quad p \leq p_0, \quad p \geq p_0$
- Alternativer: $H_A : p \neq p_0, \quad p > p_0, \quad p < p_0$
- Testobservator: $Z = \frac{X - np_0}{\sqrt{np_0(1-p_0)}}$
- Grense: $P(Z \geq z_{\alpha/2}) = \alpha/2$ eller $P(Z \geq z_\alpha) = \alpha$

t-test for forventning

- Er forventningen i et utvalg forskjellig fra før?
- Modellantagelser: $H_0 : \mu = \mu_0, \quad \mu \leq \mu_0, \quad \mu \geq \mu_0$
- Alternativer: $H_A : \mu \neq \mu_0, \quad \mu > \mu_0, \quad \mu < \mu_0$
- Testobservator: $T = \frac{\bar{X} - \mu_0}{S[\bar{X}]}$

- Grense: $P(T_v \geq t_{\alpha/2}^{(v)}) = \alpha/2$ eller $P(T_v \geq t_{\alpha}^{(v)}) = \alpha, \quad v = n - 1$

t-test for to utvalg, uavhengige grupper

- Er forventningen i to utvalg like eller ikke? Normalfordelte data.
- Modellantagelser: $H_0 : \mu_1 = \mu_2, \quad \mu_1 \leq \mu_2, \quad \mu_1 \geq \mu_2$
- Alternativer: $H_A : \mu_1 \neq \mu_2, \quad \mu_1 > \mu_2, \quad \mu_1 < \mu_2$
- Testobservator: $T = \frac{\bar{X} - \bar{Y}}{S[\delta]}$
- $S^2 = \frac{1}{n_1 + n_2 - 2}\left(\sum_{i=1}^{n_1}(X_i - \bar{X})^2 + \sum_{i=1}^{n_2}(Y_i - \bar{Y})^2\right)$
- $S[\hat{\delta}] = S \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
- Grense: $P(T_v \geq t_{\alpha/2}^{(v)}) = \alpha/2$ eller $P(T_v \geq t_{\alpha}^{(v)}) = \alpha, \quad v = n_1 + n_2 - 2$

t-test for to utvalg, parede observasjoner

- Er forventningen i to utvalg like eller ikke? Normalfordelte data.
- Modellantagelser: $H_0 : \mu_1 = \mu_2, \quad \mu_1 \leq \mu_2, \quad \mu_1 \geq \mu_2$
- Alternativer: $H_A : \mu_1 \neq \mu_2, \quad \mu_1 > \mu_2, \quad \mu_1 < \mu_2$
- Testobservator: $T = \frac{\bar{X} - \bar{Y}}{S[\delta]}$
- $S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}(X_i - Y_i - \bar{X} + \bar{Y})^2\right)$
- $S[\hat{\delta}] = \frac{S}{\sqrt{n}}$
- Grense: $P(T_v \geq t_{\alpha/2}^{(v)}) = \alpha/2$ eller $P(T_v \geq t_{\alpha}^{(v)}) = \alpha, \quad v = n - 1$

Wilcoxons rangtest

- Er forventningen i to utvalg like eller ikke? Data med vilkårlig fordeling.
- Modellantagelser: $H_0 : \mu_1 = \mu_2, \quad \mu_1 \leq \mu_2, \quad \mu_1 \geq \mu_2$
- Alternativer: $H_A : \mu_1 \neq \mu_2, \quad \mu_1 > \mu_2, \quad \mu_1 < \mu_2$

U-testen

- Er sannsynligheten for suksess i to grupper like eller ikke?
- Modellantagelse: $H_0 : p_1 = p_2, \quad p_1 \leq p_2, \quad p_1 \geq p_2$
- Alternativ: $H_A : p_1 \neq p_2, \quad p_1 > p_2, \quad p_1 < p_2$
- Testobservator: $U = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{(\frac{1}{n_1} + \frac{1}{n_2})\hat{p}(1-\hat{p})}}$
- $\hat{p}_1 = \frac{X_1}{n_1}, \quad \hat{p}_2 = \frac{X_2}{n_2}, \quad \hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$
- Grense: $P(U \geq z_{\alpha/2}) = \alpha/2$

Kjikvadrattesten for sannsynligheter

- Er en oppgitt modell troverdig eller ikke?
- Modellantagelser: H_0: Utfallene $u_1, \ldots, u_m$ har sannsynlighetene $p_1, \ldots, p_m$.
- Alternativer: H_A: Ikke alle sannsynlighetene er lik $p_1, \ldots, p_m$.

- Testobservator: $Q = \frac{(X_1-np_1)^2}{np_1} + \frac{(X_2-np_2)^2}{np_2} + \ldots + \frac{(X_m-np_m)^2}{np_m}$
- Grense: $P(Q_v \geq q_\alpha^v) = \alpha$, $\quad v = m - 1$

Kjikvadrattesten for uavhengighet

- Er det sammenheng mellom to variabler?
- Modellantagelser: H_0: Uavhengighet mellom variabler.
- Alternativer: H_A: Avhengighet mellom variabler.
- Testobservator: $Q = \sum_{j=1}^{J}\sum_{i=1}^{I} \frac{(X_{ij}-E_{ij})^2}{E_{ij}}$
- $E_{ij} = \frac{A_i B_j}{n}$, $\quad i = 1, \ldots, I \quad j = 1, \ldots, J$
- $A_i = \sum_{j=1}^{J} X_{ij}, i = 1, \ldots, I, \quad B_j = \sum_{i=1}^{I} X_{ij}, j = 1, \ldots, J$
- Grense: $P(Q_v \geq q_\alpha^v) = \alpha$, $\quad v = (i-1)(j-1)$

10 Regresjonsanalyse

10.1 Lineær regresjon

$$y = \alpha + \beta x + e \tag{59}$$

Utregning av de ukjente koeffisientene α og β ut fra n uavhengige observasjoner av (X, Y), kalt $(X_1, Y_1), (X_2, Y_2), \ldots (X_n, Y_n)$:

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n) \tag{60}$$

$$\bar{Y} = \frac{1}{n}(Y_1 + Y_2 + \ldots + Y_n) \tag{61}$$

$$M = (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \ldots + (X_n - \bar{X})^2 \tag{62}$$

$$\hat{\beta} = \frac{1}{M}\Big((X_1 - \bar{X})(Y_1 - \bar{Y}) + (X_2 - \bar{X})(Y_2 - \bar{Y}) + \ldots + (X_n - \bar{X})(Y_n - \bar{Y})\Big) \tag{63}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} \tag{64}$$

Feilleddet e:

$$E[e] = 0, \qquad Var[e] = \sigma^2 \tag{65}$$

Forventningsrett estimator for variansen til e:

$$S^2 = \frac{1}{n-2}\Big((Y_1 - \hat{Y}_1)^2 + (Y_2 - \hat{Y}_2)^2 + \ldots + (Y_n - \hat{Y}_n)^2\Big) \tag{66}$$

10.2 Residualer og forklaringskraft

Residualer (avvikene mellom Y_i og $\hat{Y}_i$):

$$R_i = Y_i - \hat{Y}_i, \qquad i = 1, 2, \ldots, n \tag{67}$$

SSE (Sum of Squared Errors):

$$SSE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{68}$$

SST (Sum Squares Total):

$$SST = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 \tag{69}$$

Modellens forklaringskraft uttrykkes med R^2:

$$R^2 = 1 - \frac{SSE}{SST} \tag{70}$$

10.3 Hypotesetesting i regresjonsmodellen

Grensene til et 95% konfidensintervall for β er gitt ved

$$\hat{\beta} \pm t^{(v)}_{0.025} \cdot S[\hat{\beta}] \tag{71}$$

Hypotesetestene forutsetter enten mange uavhengige observasjoner eller at residualene er uavhengige og normalfordelte.

10.3.1 Hypotesetest for stigningstallet i regresjonsmodellen

1. Modeller:
   (a) $H_0 : \beta = 0, \quad H_A : \beta \neq 0$
   (b) $H_0 : \beta \leq 0, \quad H_A : \beta > 0$
   (c) $H_0 : \beta \geq 0, \quad H_A : \beta < 0$

---

1. Finn verdiene til $\hat{\beta}$, S og M.
2. Regn ut $S[\hat{\beta}]$ ved

$$S[\hat{\beta}] = \frac{S}{\sqrt{M}} \tag{72}$$

3. Regn ut testestimatoren T gitt ved

$$T = \frac{\hat{\beta}}{S[\hat{\beta}]} \tag{73}$$

4. Bruk t-tabellen med parameter $v = n - 2$, og finn $t^{(v)}_{0.025}$ (tosidig) eller $t^{(v)}_{0.05}$ (ensidig) slik at

$$P\left(T_v \geq t^{(v)}_{0.025}\right) = 2,5\% \quad \text{eller} \quad P\left(T_v \geq t^{(v)}_{0.05}\right) = 5\% \tag{74}$$

5. (a) Tosidig: Forkast H_0 hvis $T_v \geq t^{(v)}_{0.025}$ eller $T_v \leq -t^{(v)}_{0.025}$. I motsatt fall beholder vi H_0
   (b) Høyresidig: Forkast H_0 hvis $T_v \geq t^{(v)}_{0.05}$. I motsatt fall beholder vi H_0, eller
   (c) Venstresidig: H_0 hvis $T_v \leq -t^{(v)}_{0.05}$. I motsatt fall beholder vi H_0

11 Normalfordelingen

	Normalkurven (arealtabell)									
Z	*0*	*0,01*	*0,02*	*0,03*	*0,04*	*0,05*	*0,06*	*0,07*	*0,08*	*0,09*
0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
4	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Figur 1: Normalfordelingen areal

12 Normalkurven (fraktiltabell)

Normalkurven (Fraktiltabell)

α	z
0,100	1,2816
0,090	1,3408
0,080	1,4051
0,070	1,4758
0,060	1,5548
0,050	1,6449
0,040	1,7507
0,030	1,8808
0,025	1,9600
0,020	2,0537
0,010	2,3263
0,005	2,5758

Figur 2: Normalfordelingen fraktiler

13 t-kurver

Frihets-grader	Sikkerhetsnivå							
	0,25	*0,1*	*0,05*	*0,025*	*0,01*	*0,005*	*0,001*	*0,0005*
1	1,000	3,078	6,314	12,706	31,821	63,657	318,309	636,619
2	0,816	1,886	2,920	4,303	6,965	9,925	22,327	31,599
3	0,765	1,638	2,353	3,182	4,541	5,841	10,215	12,924
4	0,741	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	5,893	6,869
6	0,718	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	0,706	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,688	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	0,688	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,687	1,325	1,725	2,086	2,528	2,845	3,552	3,850
25	0,684	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,684	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,684	1,314	1,703	2,052	2,473	2,771	3,421	3,690
30	0,683	1,310	1,697	2,042	2,457	2,750	3,385	3,646
40	0,681	1,303	1,684	2,021	2,423	2,704	3,307	3,551
50	0,679	1,299	1,676	2,009	2,403	2,678	3,261	3,496
100	0,677	1,290	1,660	1,984	2,364	2,626	3,174	3,390
200	0,676	1,286	1,653	1,972	2,345	2,601	3,131	3,340
500	0,675	1,283	1,648	1,965	2,334	2,586	3,107	3,310
1000	0,675	1,282	1,646	1,962	2,330	2,581	3,098	3,300

Figur 3: t-kurver (fraktiltabell)

14 Kjikvadratkurver

Frihets-grader (v)	Sikkerhetsnivå og signifikans					
	70 %	*80 %*	*90 %*	***95 %***	*98 %*	***99 %***
	0,3	*0,2*	*0,1*	*0,05*	*0,02*	*0,01*
1	1,074	1,642	2,706	3,841	5,412	6,635
2	2,408	3,219	4,605	5,991	7,824	9,210
3	3,665	4,642	6,251	7,815	9,837	11,345
4	4,878	5,989	7,779	9,488	11,668	13,277
5	6,064	7,289	9,236	11,070	13,388	15,086
6	7,231	8,558	10,645	12,592	15,033	16,812
7	8,383	9,803	12,017	14,067	16,622	18,475
8	9,524	11,030	13,362	15,507	18,168	20,090
9	10,656	12,242	14,684	16,919	19,679	21,666
10	11,781	13,442	15,987	18,307	21,161	23,209
11	12,899	14,631	17,275	19,675	22,618	24,725
12	14,011	15,812	18,549	21,026	24,054	26,217
13	15,119	16,985	19,812	22,362	25,472	27,688
14	16,222	18,151	21,064	23,685	26,873	29,141
15	17,322	19,311	22,307	24,996	28,259	30,578
16	18,418	20,465	23,542	26,296	29,633	32,000
17	19,511	21,615	24,769	27,587	30,995	33,409
18	20,601	22,760	25,989	28,869	32,346	34,805
19	21,689	23,900	27,204	30,144	33,687	36,191
20	22,775	25,038	28,412	31,410	35,020	37,566
25	28,172	30,675	34,382	37,652	41,566	44,314
30	33,530	36,250	40,256	43,773	47,962	50,892
40	44,165	47,269	51,805	55,758	60,436	63,691
50	54,723	58,164	63,167	67,505	72,613	76,154

Figur 4: Kjikvadratkurver (fraktiltabell)

15 Ekstra

15.1 To-punktsformel for rett linje

Ettpunktsformel:

$$y - y_1 = \alpha(x - x_1) \tag{75}$$

Topunktsformelen (kombineres med ettpunktsformelen):

$$\alpha = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1} \tag{76}$$