Høgskolen i Østfold

5. april 2017

EKSAMEN – løsningsforslag

Emnekode:	**Emnenavn:**
ITD20106	Statistikk og økonomi
Dato:	**Eksamenstid:**
2. mai 2016	09.00 – 13.00
Hjelpemidler:	**Faglærer:**
- Alle trykte og skrevne. - Kalkulator.	Christian F Heide

Om eksamensoppgaven og poengberegning:

Oppgavesettet består av 12 sider inklusiv denne forsiden og seks sider vedlegg. Kontroller at oppgavesettet er komplett.

Oppgavesettet består av 7 oppgaver med i alt 23 deloppgaver. Ved sensur vil alle deloppgaver telle omtrent like mye.

Der det er mulig skal du:

- vise utregninger og hvordan du kommer fram til svarene
- begrunne dine svar, selv om dette ikke er eksplisitt sagt i hvert spørsmål

Om noe er uklart eller mangelfullt i oppgaven, gjør selv de nødvendige forutsetninger.

Sensurfrist: 26. mai 2016

Karakterene er tilgjengelige for studenter på Studentweb senest 2 virkedager etter oppgitt sensurfrist. www.hiof.no/studentweb

Oppgave 1

Se bort fra MVA i denne oppgaven.

En bedrift har på bakgrunn av tidligere erfaringer utarbeidet tilleggssatser for de indirekte variable kostnadene i en produksjonslinje:

- *Materialavdeling:* *25 % av direkte material*
- *Tilvirkningsavdeling:* *60 % av direkte lønn*
- *Salgs-/administrasjonskostnader:* *10 % av tilvirkningsmerkost*

a) *Sett opp en bidragskalkyle for en ny bestilling som ventes å forbruke kr 30 000 i direkte materialkostnad og kr 40 000 i direkte lønnskostnad.*

	Tilleggssats	Kalkyle
Direkte material		30 000
Direkte lønn		40 000
Indirekte variable:		
Mat avd.	25 %	7 500
Tilv. avd	60 %	24 000
Tilvirkningsmerkost		101 500
S/A avd. variable	10 %	10 150
Merkost/Sum variable		111 650

Bedriften har fått en forespørsel fra en potensiell kunde om å levere fire bestillinger som beskrevet i a) pr. år i 3 år fremover. Prisen/inntekten er satt til kr 170 000 pr. bestilling. Denne produksjonen vil i så fall medføre behov investeringer i en maskin til kr 300 000. Maskinen forventes å ha en utrangeringsverdi på kr 20 000 etter tre år.

Disse bestillingene vil også medføre økte faste, betalbare kostnader på kr 100 000 pr. år. De direkte og variable kostnadene fra a) er også betalbare. Man har dessuten beregnet at omløpsmidlene knyttet til varebeholdninger vil måtte øke med kr 50 000 dersom investeringen gjennomføres og forespørselen fra kunden aksepteres.

b) *Sett opp kontantstrømmene for dette prosjektet fra og med år 0 til og med år 3.*

Årlig innbetalingsoverskudd fra driften:				
	Ordrer	Per ordre	Totalt	
Inntekter	4	170 000	680 000	
Merkost/Utbetaling drift	4	111 650	446 600	
Kontantstrøm:				
	0	1	2	3
Investeringsbeløp	-300 000			
Utrangeringsverdi				20 000
Innbetalinger		680 000	680 000	680 000
Utbetalinger drift		-446 600	-446 600	-446 600
Faste betalbare		-100 000	-100 000	-100 000
Økning OM	-50 000			50 000
Årlig KS	**-350 000**	**133 400**	**133 400**	**203 400**

Bedriften har et helt annet investeringsprosjekt, «Super», hvor det er budsjettert følgende netto kontantstrømmer over 3 år:

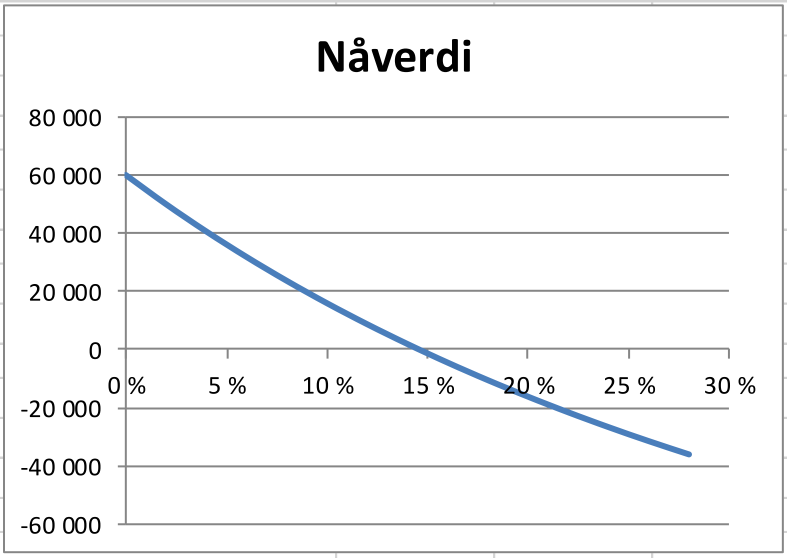
År 0	**År 1**	**År 2**	**År 3**
-175 000	55 000	75 000	105 000

I resten av denne oppgaven skal du bruke kontantstrømmen for «Super» gjengitt over her når du gjør dine beregninger.

c) *Beregn nåverdien av prosjekt «Super» og kommenter lønnsomheten når avkastningskravet/kalkulasjonsrenten er 10 %.*

	Prosjekt "Super"			
	År 0	**År 1**	**År 2**	**År 3**
	-175 000	55 000	75 000	105 000
Avkastningskrav:	10 %			
Nåverdi prosjekt:	15 872	=> NV > 0, lønnsom		

d) *Lag en nåverdiprofil for prosjekt «Super» og estimer investeringens internrente. Marker internrenten i figuren.*



Rente	Nåverdi
0 %	60 000
2 %	49 953
4 %	40 571
6 %	31 797
8 %	23 579
10 %	15 872
12 %	8 634
14 %	1 828
16 %	-4 580
18 %	-10 620
20 %	-16 319
22 %	-21 704
24 %	-26 797
26 %	-31 618
28 %	-36 187

Internrenten er	14,56 %

Oppgave 2

Bedriften CapsLock AS produserer en spesiell, delvis håndsydd baseballcap i modellene Hipster, FK (Felleskjøpet) og Army. Tabellen under viser en oversikt over aktuelle tall for produksjon og salg av cap'ene i et kvartal. Kronebeløpene i tabellen er pr. stk. av aktuell modell.

Produkt	**Hipster**	**FK**	**Army**
Direkte materialkostnader	Kr 330	Kr 200	Kr 225
Direkte lønn	Kr 140	Kr 110	Kr 120
Indirekte variable kostnader	Kr 80	Kr 60	Kr 60
Salgspris uten mva.	Kr 800	Kr 550	Kr 600
Tidsforbruk systue pr. enhet	2 t	1 t	1.5 t
Antall produserte og solgte enheter	300	400	300

Materialene som brukes i de tre modellene er av samme type.

De faste kostnadene er anslått til kr. 160 000 pr kvartal, og de direkte kostnadene er som vanlig å betrakte som variable.

a) *Beregn det totale dekningsbidraget for produksjonen i kvartalet gjengitt i tabellen.*

Produkt	Hipster	FK	Army	
Direkte materialkostnader	330	200	225	
Direkte lønn	140	110	120	
Indirekte variable kostnader	80	60	60	
Salgspris uten mva	800	550	600	
DB	250	180	195	
Tidsforbruk «syavdeling» pr. enhet (timer)	2,0	1,0	1,5	
Antall solgte enheter forrige periode	300	400	300	
Inntekt	240 000	220 000	180 000	
Direkte materialkostnader	99 000	80 000	67 500	
Direkte lønn	42 000	44 000	36 000	
Indirekte variable kostnader	24 000	24 000	18 000	
DB	75 000	72 000	58 500	**205 500**

b) *Hvilket resultat oppnådde bedriften dette kvartalet?*
Resultatet er dekningsbidraget som vi fant i a) minus 160 000 i faste kostnader som er oppgitt i oppgaven.
Resultatet er følgelig 205 500 – 160 000 = 45 500

c) *Hvilken cap bør CapsLock AS satse på hvis de har problemer med å skaffe nok materialer? Hvilken bør de satse på hvis de mangler kapasitet i systua?*

		Material	
DB per enhet	250	180	195
Forbruk knapp faktor (Material)	330	200	225
DB per knapp faktor	0,76	0,90	0,87
Rangerer etter DB per knapp faktor	**3**	**1**	**2**

De bør satse på FK-modellen dersom de har problemer med å skaffe nok materialer.

		Tid i systua	
DB per enhet	250	180	195
Forbruk knapp faktor (TID)	2,0	1,0	1,5
DB per knapp faktor	125	180	130
Rangerer etter DB per knapp faktor	**3**	**1**	**2**

De bør satse på FK-modellen dersom de mangler kapasitet i systua.

d) *Anta nå at bedriften kun produserer og selger Army. Hva er dekningsgraden, dekningspunkt mengde og dekningspunktomsetning?*

Dekningsgrad ARMY		33 %
Dekningspunkt mengde(FTK/DB per enhet)		821
Dekningspunkt omsetning (FTK/DG)		492 308
Dekningspunkt omsetning (DP mengde * pris)		492 308

e) *Anta at du i dag produserer og selger 1500 stk. av Army pr kvartal. Hva er sikkerhetsmarginen i prosent og kroner?*

Antall solgte ARMY i dag:		1 500
Omsetning i dag:		900 000
Sikkerhetsmargin (Omst - Dekningspunkt oms)		407 692
Sikkerhetsmargin %		45 %

Oppgave 3

Gitt følgende sannsynlighetsfordeling (simultanfordeling) for de stokastiske variablene X og Y.

x	y	
	0	1
0	0.1	0.3
1	0.3	0.1
2	0.1	0.1

a) *Beregn forventningsverdi og varians for X og Y.*
Her kan det være en fordel å regne ut de marginale sannsynlighetene først:

x	y		
	0	1	$P(X=x)$
0	0.1	0.3	0.4
1	0.3	0.1	0.4
2	0.1	0.1	0.2
$P(Y=y)$	0.5	0.5	

Vi får da

$$\mu_X = E(X) = \sum_{i=0}^{2} x_i \cdot P(X = x_i) = 0 \cdot 0.4 + 1 \cdot 0.4 + 2 \cdot 0.2 = 0 + 0.4 + 0.4 = \underline{\underline{0.8}}$$

$$\mu_Y = E(Y) = \sum_{i=0}^{1} y_i \cdot P(Y = y_i) = 0 \cdot 0.5 + 1 \cdot 0.5 = 0 + 0.5 = \underline{\underline{0.5}}$$

$$\sigma_X^{\ 2} = Var(X) = E(X^2) - \mu_X^{\ 2} =$$

$$\sum_{i=0}^{2} x_i^{\,2} \cdot P(X = x_i) - \mu_X^{\,2} =$$

$$0^2 \cdot 0.4 + 1^2 \cdot 0.4 + 2^2 \cdot 0.2 - 0.8^2 = 0 + 0.4 + 0.8 - 0.64 = \underline{\underline{0.56}}$$

$$\sigma_Y^{\,2} = Var(Y) = E(Y^2) - \mu_Y^{\,2} =$$

$$0^2 \cdot 0.5 + 1^2 \cdot 0.5 - 0.5^2 = 0 + 0.5 - 0.25 = \underline{\underline{0.25}}$$

b) *Undersøk om X og Y er uavhengige variable.*
Vi må her undersøke om $P(X = x_i \wedge Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$ for alle mulige par av *x* og *y*.
Her ser vi for eksempel at

$$P(X = 0 \wedge Y = 0) = 0.1$$

mens

$$P(X = 0) \cdot P(Y = 0) = 0.4 \cdot 0.5 = 0.2 \neq P(X = 0 \wedge Y = 0)$$

Følgelig er *X* og *Y* ikke uavhengige.

Dette kan også begrunnes av korrelasjonen vi regner ut i spørsmål c). Dersom to variabler er uavhengige, så er korrelasjonen 0. Siden korrelasjonen vi finner i spørsmål c) ikke er 0, så er de heller ikke uavhengige.

c) *Finn korrelasjonen til X og Y.*
For å finne korrelasjonen, må vi først finne kovariansen. Denne er definert som

$$Cov(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y)$$

Vi må først finne $E(X \cdot Y)$:

$$E(X \cdot Y) = \sum_{i=0}^{2} \sum_{j=0}^{1} x_i \cdot y_j \cdot P(X = x_i \wedge Y = y_j)$$

Her trenger vi ikke å ta med ledd der minst en av x_i, y_i eller $P(X = x_i \wedge Y = y_j)$ er null (fordi da blir produktet 0). Vi får derfor bare med leddene der *x* = 1 og *y* = 1, og der *x* = 2 og *y* = 1:

$$E(X \cdot Y) = 1 \cdot 1 \cdot 0.1 + 2 \cdot 1 \cdot 0.1 = 0.1 + 0.2 = \underline{0.3}$$

Kovariansen blir derfor

$$Cov(X,Y) = 0.3 - 0.8 \cdot 0.5 = 0.3 - 0.4 = \underline{-0.1}$$

Korrelasjonen er følgelig

$$\rho(X,Y) = Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{-0.1}{\sqrt{0.56} \cdot \sqrt{0.25}} = \underline{\underline{-0.267}}$$

Oppgave 4

En dyreart er i ferd med å dø ut. Det antas å være bare 25 dyr igjen av denne arten i et område. Forskere ønsker å studere arten og fanger derfor fem eksemplarer og merker disse. En måned etter kommer forskerne tilbake og fanger ti dyr.

a) *Dersom det virkelig er 25 dyr av denne arten i området, hva er sannsynligheten for at det blant de ti fangede dyrene er to som er merket?*
Siden dette er en «trekning» uten tilbakelegging fra en relativt liten populasjon, blir dette beskrevet av en hypergeometrisk fordeling. Sannsynligheten er derfor gitt ved

$$P(X = x) = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}}$$

Her er x = 2, M = 5, N = 25, n = 10. Vi får derfor

$$P(X = 2) = \frac{\binom{5}{2} \cdot \binom{20}{8}}{\binom{25}{10}} = \frac{10 \cdot 125\ 970}{3\ 268\ 760} = \underline{\underline{0.385}}$$

b) *Hva ville resultatet i a) blitt dersom vi hadde benyttet binomisk fordeling som en approksimasjon ved beregningen? Forklar hvorfor dette er en god eller dårlig approksimasjon.*
Å approksimere med en binomisk fordeling innebærer at vi ikke tar hensyn til at sannsynligheten for å fange et merket dyr endrer seg underveis, men at den i alle «trekningene» er

$$p = \frac{5\ merkede\ dyr}{25\ dyr\ totalt} = 0.2$$

Sannsynligheten i binomisk fordeling er gitt ved

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Vi får derfor

$$P(X = 2) = \binom{10}{2} \cdot 0.2^2 \cdot 0.8^8 = \underline{\underline{0.302}}$$

Vi ser at approksimasjonen gir oss et tall som er noe i nærheten av det korrekte. Det er allikevel et stykke unna fordi populasjonen vi trekker fra er såpass liten at sannsynligheten endrer noe seg fra trekning til trekning (den er altså 0.2 kun i den første trekningen og ikke konstant 0.2 i hver trekning).

Regelen for tilnærming til binomisk fordeling, er at populasjonen *N* må være større enn 10 ganger utvalget vi trekker. Her ville dette kravet vært at populasjonen skulle vært større enn 100 (siden vi fanger 10 dyr). Vi ser at dette kravet ikke er oppfylt, siden populasjonen er på kun 25 dyr. En tilnærming til binomisk fordeling kan derfor ikke anbefales i dette tilfellet, men dette avhenger selvsagt også av hva man skal bruke resultatet til.

Oppgave 5

Et kaffebrenneri benytter en maskin som fyller kaffe i poser. Hver pose skal inneholde en kvart kilo (250 g) kaffe. Maskinen innstilles på vekt μ . Mengden kaffe som fylles i hver pose kan da oppfattes som uavhengig og normalfordelt.

Maskinen innstilles på μ = 250 g og standardavviket er oppgitt til $\sigma = 12$ g.

a) *Hva er sannsynligheten for at en kaffepose inneholder mindre enn 245 g kaffe?*
Vi kaller mengden kaffe som fylles i en tilfeldig pose for *X*. Siden *X* er normalfordelt, finner vi denne sannsynligheten slik (siden vi har en kontinuerlig fordeling, trenger vi ikke å skille mellom < og $\leq$):

$$P(X < 245) = P\left(\frac{X - \mu}{\sigma} \leq \frac{245 - \mu}{\sigma}\right) = P\left(Z \leq \frac{245 - 250}{12}\right) =$$

$$P(Z < -0.4167) = G(-0.4167) \approx G(-0.42) = \underline{\underline{0.3372}}$$

b) *Kaffeposer som inneholder mindre enn 245 g kaffe blir ansett som undervektige, og produsenten ønsker ikke at for mange av disse skal bli sendt ut på markedet. Produsenten ønsker derfor å stille inn fyllingsgraden μ på maskinen på mer enn 250 g slik at andelen av undervektige kaffeposer blir mindre enn 1 %. Bestem μ slik at dette kravet blir oppfylt.*
Her ønsker vi altså å finne en normalfordeling for *X* med en μ slik at

$$P(X < 245) < 0.01$$

Dette tilsvarer at

$$P(Z < \frac{245 - \mu}{\sigma}) < 0.01$$

Dette oppfylles dersom Z er mindre enn venstre 0.01-kvantil, altså dersom

$$Z < -z_{0.01}$$

Denne finner vi i kvantiltabellen for standardnormalfordelingen:

$$z_{0.01} = 2.326$$

Vi må altså kreve at

$$Z < -2.326$$

Dersom vi skal oppfylle dette, må altså

$$\frac{245-\mu}{\sigma} = -2.326$$

Snur vi på denne, finner vi at

$$\mu = 245 + 2.326 \cdot 12 = 272.9$$

Vi må altså stille fyllingsgraden på mer enn 272.9 g for at andelen undervektige kaffeposer skal være mindre enn 1 %.

Oppgave 6

Anta at det årlige forbruket av nitrogengjødsel i Norge er normalfordelt med forventning μ og kjent standardavvik σ = 5000 tonn. Gjennomsnittlig forbruk av nitrogengjødsel i norsk landbruk for årene 2001–2008 er 105 700 tonn (kilde: SSB).

a) *Finn et 95 % konfidensintervall for nitrogenforbruket μ.*
Her er standardavviket kjent, og vi lager derfor det som kalles et Z-intervall for forventningsverdien, μ. Vi tar da utgangspunkt i gjennomsnittet av målingene som punktestimat for μ, og lar intervallet strekke seg så langt til sidene som konfidensnivået $1-\alpha$ angir. Et slikt intervall er gitt ved

$$\overline{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

I vårt tilfelle er følgende oppgitt i oppgaven: $\overline{X} = 105\,700$ og $\sigma = 5000$. Antall målinger er n = 8, siden vi har gjennomsnittet for åtte år.

Siden vi skal ha et 95 % konfidensintervall, betyr at konfidensnivået er 0.95. Følgelig er $\alpha = 0.05$, noe som igjen betyr at $\alpha/2 = 0.025$. Fra standardnormalfordelingens kvantiltabell finner vi at

$$z_{\alpha/2} = z_{0.025} = 1.960$$

Følgelig blir

$$z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 1.960 \cdot \frac{5000}{\sqrt{8}} = 3464.8 \approx \underline{3465}$$

Konfidensintervallet blir følgelig

$$\left[\overline{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}},\ \overline{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right] = \left[105\,700 - 3465, 105\,700 + 3465\ \right] =$$

$$\underline{\underline{[102\,235, 109\,165]}}$$

b) *Hvor mange år (n) må vi måle nitrogenforbruket (X) for at lengden på konfidensintervallet ikke skal overstige 4000 tonn*?
Lengden av konfidensintervallet er

$$2z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Krever vi nå at dette skal være mindre enn eller lik 4000, får vi

$$2z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq 4000$$

Løser vi denne med hensyn på $\sqrt{n}$ får vi

$$\sqrt{n} \geq \frac{2\sigma}{4000} \cdot z_{\alpha/2}$$

Setter vi inn tall blir dette

$$\sqrt{n} \geq \frac{2 \cdot 5000}{4000} \cdot 1.960 = 4.90$$

Kvadrerer vi begge sider, får vi

$$n \geq 4.90^2 = 24.01 \approx \underline{\underline{24}}$$

Siden det nøyaktige svaret på regnestykket er så nær 24, godtar vi både 24 og 25 som svar (i et praktisk tilfelle ville man benyttet 24, selv om det matematisk sett er 25 som er korrekt).

c) *Utfør en test for å undersøke om det er grunnlag for å hevde at forventet nitrogenforbruk er over 100 000 tonn. Velg signifikansnivå α = 0.05*.
Her er det naturlig å velge følgende hypoteser:

$H_0: \ \mu \leq 100\,000$ (Forventet nitrogenforbruk er mindre enn eller lik 100 000 tonn.)

$H_1: \ \mu > 100\,000$ (Forventet nitrogenforbruk er større enn 100 000 tonn.)

Siden standardavviket er kjent, kan vi bruke en *Z*-test. Et signifikansnivå på $\alpha = 0.05$ betyr at $z_\alpha = z_{0.05} = 1.645$. Vi forkaster nullhypotesen dersom

$$Z > z_\alpha$$

Her er *Z* gitt ved

$$Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$$

hvor $\overline{X} = 105\ 700$, $\sigma = 5000$ og $n = 8$ som tidligere. Videre er $\mu_0 = 100\ 000$, altså grensen som er angitt i hypotesene. Setter vi inn disse tallene får vi

$$z = \frac{105\ 700 - 100\ 000}{5000 / \sqrt{8}} = \underline{\underline{3.22}}$$

Vi ser av dette at

$$z = 3.22 > z_\alpha = 1.645$$

Vi kan følgelig forkaste nullhypotesen. Konklusjon:

<u>Forventet årlig nitrogenforbruk er større enn 100 000 tonn</u>.

d) *Beregn testens p-verdi.*
Dette angir sannsynligheten for at vi forkaster nullhypotesen gitt at den er sann. Altså: hva er sannsynligheten for at vi kunne fått de måleresultatene vi fikk dersom forventningsverdien faktisk ikke *var* større enn 100 000. Denne er

$$p = P(\overline{X} \geq 105\ 700 \mid \mu = 100\ 000) =$$

$$P\left(Z \geq \frac{105\ 700 - 100\ 000}{5000 / \sqrt{8}} \right) = P(Z \geq 3.22) =$$

$$1 - P(Z < 3.22) = 1 - G(3.22)$$

Vår tabell går ikke så langt som til 3.22. Den høyeste verdien er 3.09 med verdi $G(3.09) = 0.9990$. Vi vet derfor at $G(3.22)$ er større enn dette. Vi kan derfor konkludere med at

$p < 1 - 0.9990 = \underline{\underline{0.001}}$

(En mer nøyaktig verdi er $p = 1 - 0.9936 = 0.00064$).

e) *I årene 2009 og 2010 var nitrogenforbruket henholdsvis 82 550 og 83 080 tonn. Er det på bakgrunn av disse tallene grunnlag for å hevde at forventningsverdien til nitrogenforbruket i disse siste årene (2009 og 2010) har sunket og nå er lavere enn 100 000 tonn? Velg signifikansnivå $\alpha = 0.025$. Anta at standardavviket er ukjent for disse årene.*

Siden standardavviket er ukjent, må vi bruke en *T*-test. Hypotesene blir nå:

$H_0: \ \mu \geq 100\,000$ (Forventet nitrogenforbruk er nå større enn eller lik 100 000 tonn.)

$H_1: \ \mu < 100\,000$ (Forventet nitrogenforbruk er mindre enn 100 000 tonn.)

Gjennomsnittet for disse to årene er

$$\overline{X} = \frac{1}{2}(82\ \ 550 + 83\ \ 080) = 82\ \ 815$$

Vårt estimat for standardavviket, *S*, er gitt ved

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 =$$

$$\frac{1}{2-1}\sum_{i=1}^{2}\left(x_i - \overline{x}\right)^2 = (82\ \ 550 - 82\ \ 815)^2 + (83\ \ 080 - 82\ \ 815)^2 =$$

$$70\ \ 225 + 70\ \ 225 = 140\ \ 450$$

og følgelig er

$s = 374.8$

Siden antall frihetsgrader her er $n - 1 = 1$ (vi har bare målinger for to år), må vi finne *t*-fordelingens kvantil $t_{\alpha,n-1} = t_{0.025,1}$ som er

$$t_{0.025,1} = 12.706$$

Vi forkaster nullhypotesen dersom den observerte verdien av *T* er mindre enn –12.706 (fordi vi benytter venstre hale av sannsynlighetsfordelingen).

Vår observerte *T*-verdi er

$$t = \frac{\overline{x} - \mu_0}{s\big/\sqrt{n}} = \frac{82\ 815 - 100\ 000}{374.8\big/\sqrt{2}} = \underline{-64.8}$$

Vi ser at vår verdi er mindre enn –12.706. <u>Vi kan derfor forkaste nullhypotesen</u>.

Konklusjon: Forventet årlig nitrogenforbruk er mindre enn 100 000 tonn.

Oppgave 7

En italiensk kjede som selger pizza, har etablert seg i nærheten av store universiteter i USA. Kjeden består av 10 restauranter, og sammenhengen mellom antall tusen studenter (x) pr universitet og årlig salg (Y) oppgitt i 1000 dollar, er gitt i tabellen nedenfor:

Antall studenter (*x*)	2	6	8	8	12	16	20	20	22	26
Årlig salg (*Y*)	60	105	88	120	118	136	158	170	152	202

Videre er det oppgitt at:

$$\bar{x} = 14$$

$$\bar{y} = 130.9$$

$$\sum_{i=1}^{10}(x_i - \bar{x})^2 = 568$$

$$\sum_{i=1}^{10}(y_i - \bar{y})^2 = 15\,612.9$$

$$\sum_{i=1}^{10}(x_i - \bar{x})(y_i - \bar{y}) = 2836$$

a) *Finn den empiriske korrelasjonskoeffisienten mellom x og Y, og forklar hva resultatet betyr*.
Den empiriske korrelasjonskoeffisienten er

$$r = \frac{\sum_{i=1}^{10}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{10}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{10}(y_i - \bar{y})^2}} = \frac{2836}{\sqrt{568} \cdot \sqrt{15612.9}} = \underline{\underline{0.952}}$$

En verdi nær 1 betyr at det er tilnærmet lineær sammenheng mellom variablene. En verdi på 0.95 er ganske nær 1 og betyr at det er sterk sammenheng mellom antall studenter og årlig pizzasalg.

b) *Bestem koeffisientene for den estimerte regresjonslinjen* $\hat{\alpha} + \hat{\beta}x$.
Regresjonslinjens koeffisienter finner vi ved uttrykkene

$$\hat{\alpha} = \bar{y} - \hat{\beta} \cdot \bar{x} \qquad \hat{\beta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{r \cdot S_Y}{S_X}$$

Vi får

$$\hat{\beta} = \frac{2836}{568} = \underline{\underline{4.99}}$$

og

$$\hat{\alpha} = 130.9 - 4.99 \cdot 14 = \underline{\underline{61.0}}$$

Den estimerte regresjonslinjen (som det ikke er eksplisitt spurt etter) er derfor

$$\underline{Y = 61.0 + 4.99x}$$