

Indhold

Udregning	9	
<i>2019_feb - Opgave 1.1 - bionormalfordeling</i>	9	
I gennem længere tid har et flyselskab erfaret, at 10 % af passagererne på en given rute ikke mødte op. Flyet på denne rute har 200 sæder, og flyselskabets politik har været at sælge op til et maksimum på 210 billetter til hver afgang. a) Hvad er sandsynligheden for, at mere end 200 passagerer møder op, når der er solgt 210 billetter? b) Diskuter om forudsætningerne kan antages at være opfyldte.....		9
<i>Feb_2019 - Opgave 1.5 - Poissionfordeling</i>	10	
Nogle passagerer medtager kæledyr og for at imødekomme diverse krav for transport af kæledyr, har flyselskabet 2 luksusbokse samt adskillige andre bokse til denne rute. Erfaring viser, at antal kæledyr per afgang er poissonfordelt med $\mu=1,5$. a) Hvad er sandsynligheden for, at mere end 2 kæledyr møder op til en afgang? b) Hvad er det forventede antal kæledyr, der bliver transporteret i luksusbokse per afgang, når der kun er 2 luksusbokse til rådighed? (Passagerne har ikke mulighed for at vælge boks på forhånd) c) Hvor mange luksusbokse skulle der være til rådighed, hvis flyselskabet ville være 90 % sikker på ikke at løbe tør for luksusbokse?		10
<i>Feb_2020 - Opgave 1.1 - Poissionfordeling</i>	12	
En rutineret korrekturlæser har erfaring for, at fejl (det registrerede antal) per side er Poissonfordelt med $\mu=0,3$. a) Hvad er sandsynligheden for, at der på en side ingen fejl er? b) Hvad er det forventede antal sider uden fejl i en bog på 200 sider? c) Hvad er sandsynligheden for at have 5 sider i træk uden fejl?		12
<i>Jan_2019 - Opgave 1.6 - bionomialfordeling</i>	12	
Hvad er sandsynligheden for, at 4 ud af 25 tilfældigt udvalgte studerende anvender mindre end 20 timer i ugen på at studere?		12
<i>Jan_2020 - opgave 3.1</i>	13	
a) Hvad er sandsynligheden for at svare rigtigt på 7 ud af 10 spørgsmål, hvis man for hvert spørgsmål har 80 % chance for at svare rigtigt? b) Hvad er sandsynligheden for, at der på en dag ankommer mere end 7 kunder, hvis man antager, at antal kunder, der ankommer, er Poissonfordelt med $\mu=10$? c) Hvad er sandsynligheden for, at den stokastiske variabel X antager en værdi på under 7, når det oplyses, at X er normalfordelt med $\mu=10$ og $\sigma=2$?		13
<i>Jan_2018 - Opgave 2.1 (poissionfordeling - mest sandsynlige udfald, sandsynligheden for et udfald 2 dage i træk, sandsynligheden for x antal på x dage)</i>	14	
En virksomhed har erfaring for, at antallet af indkomne klager per dag følger en poissonfordeling med et gennemsnit på 3. a) Hvad er det mest sandsynlige antal klager per dag? b) Hvad er sandsynligheden for, at der to dage i træk ikke kommer klager? c) Hvad er sandsynligheden for, at der på en uge (5 arbejdsdage) kommer mere end 10 klager?		14
<i>Feb_2018 - Opgave 1.1 - Hypergeometrisk - Diskussion af fordeling, forventet antal og sandsynlighed...</i>	15	
På et øvelseshold bestående af 15 mænd og 21 kvinder beslutter man sig for at deltage i en case-competition. Man udvælger derfor tilfældigt en gruppe på 4 studerende. a) Diskuter, hvilken fordeling antal kvinder i en gruppe følger. b) Hvad er det forventede antal kvinder i en gruppe? c) Hvad er sandsynligheden for, at alle 4 i en gruppe er kvinder?		15

udtale sig om, hvad sandsynligheden er for, at der på en dag i alt for begge forretninger kommer mere end 800 kunder?	44
<i>Jan_2018 - Opgave 2.3 - Bionominalfordeling - Forventede antal, standardafvigelse og sandsynlighed ...</i>	45
Antag, at en ud af fem klager imødekommes. a) Hvad er det forventede antal imødekommer ud af 250 klager, og hvad er standardafvigelsen? b) Hvad er sandsynligheden for, at mere end 70 ud af 250 klager bliver imødekommet?	45
<i>Feb_2018 - Opgave 3.1 - Udregn fælles forventet værdi og fælles varians</i>	46
Antal børn i tusinder, der er tilskuere til det lokale fodboldholds kampe, har en forventet værdi på 4 og en varians på 0,8. Antal voksne i tusinder, der er tilskuere til det lokale fodboldholds kampe, har en forventet værdi på 14 og en varians på 2,8. Billetprisen er hhv. 50 DKK for børn og 100 DKK for voksne. a) Hvad er den forventede omsætning (i tusinder) ved den næste kamp? b) Hvad er variansen, når man antager, at antal børn og antal voksne tilskuere er uafhængige af hinanden? c) Er det realistisk, at antal børn og antal voksne tilskuere er uafhængige af hinanden?	46
<i>Prøveeksamen - Opgave 1.1</i>	47
De danske Folkeskoler har på det seneste været i fokus både med hensyn til elevernes præstationer i faget matematik og med hensyn til sammenligninger af skolernes karakterniveau. Karaktererne ved Folkeskolens afgangsprøve i matematik var undervisningsårene 2018/19 og 2019/20 (<i>se opgaven</i>). a) Gennemfør en test for, om der er sket en ændring fra 2018/19 til 2019/20? (Antag, at varianserne er ens)	47
Hypoteseprøvning vedrørende p (test for andele)	49
<i>2019_Feb - Opgave 1.4</i>	49
En passager, der det seneste år har rejst regelmæssigt på en anden rute, har oplevet væsentlige forsinkelser på 13 ud af 47 afgang. Flyselskabet har for denne rute i den samme periode registreret, at 167 afgang af i alt 945 afgang havde væsentlige forsinkelser. a) Diskuter hvilken fordeling passagerens antal forsinkelser kunne følge. b) Gennemfør en test for at afdække om passagerens oplevelse afviger fra flyselskabets opgørelse. c) Anfør mindst tre grunde til, at det er muligt at have en signifikant forskel mellem estimatet og parameterværdien i nul-hypotesen, når man tester.	49
<i>Feb_2020 - Opgave 1.6</i>	51
Forlaget har set på, om der tre år efter udgivelsen er udsolgt af første oplag. Af 50 bøger der var trykt i et første oplag på 1 tusinde eksemplarer, var 25 udsolgt og af 25 bøger der var trykt i et første oplag på 2 tusinde eksemplarer, var 20 udsolgt. a) Gennemfør en test for om man på grundlag af ovenstående kan konkludere, at der er forskel i den andel af bøger, der er udsolgt, afhængigt af om bogen er trykt i 1 eller i 2 tusinde eksemplarer.	51
<i>Jan_2019 - Opgave 1.5</i>	53
Erfaring har vist, at 20 % af studerende bruger mindre end 20 timer om ugen på at studere. Gennemfør en test for at undersøge, om man kan konkludere, at andelen af studerende, der anvender mindre end 20 timer om ugen på at studere, er mindsket. Det oplyses, at 8 af de 61 adspurgte svarede, at de brugte mindre end 20 timer om ugen.	53
<i>Jan_2020 - Opgave 1.1</i>	54
Ved en opinionsmåling har man tilfældigt udvalgt og kontaktet 1.500 personer i alderen 18-67 år. Af de 1.500 personer har 900 besvaret et spørgeskema, der blandt andet omhandler køn, alder, indkomst	

igennem længere tid og derfor, medmindre der er omstændigheder vi ikke kender til, må de 10% antages at være et gennemsnit tæt på flyselskabets generelle "no-show" procent.

- e. $n/N < 5\%$ (hypergeometrisk) eller proces (binomial) + konstanthed
 - i. Her argumenterer jeg for en proces betragtning, da der konstant letter fly og det samme antages der at gøres på denne route. Derfor vil passagertallet og antallet af "No-shows" stige konstant. Der er gennemsnitligt 90% sandsynlighed for at passager møder op, jævnfør opgaveteksten og diskussionen af bionominalfordelingen.

2) Beregning af sandsynlighed

$$P(X > 200 | n = 210, p = 0.90) = P(X < 10 | n = 210, p = 0.10) = P(X \leq 9 | n = 210, p = 0.10)$$

For at udregne sandsynligheden approksimerer jeg til normalfordelingen, således:

$$Z \sim \frac{X + \frac{1}{2} - n * p}{\sqrt{n * p * (1 - p)}}$$

- HUSK approximationsled for overgangen mellem diskret og kontinuert fordeling. Det er den halve der er lagt til i formelen ovenfor.

$$P\left(Z < \frac{9 + \frac{1}{2} - 210 * 0.10}{\sqrt{210 * 0.10 * (1 - 0.10)}} = \frac{9.5 - 21}{\sqrt{210 * 0.10 * 0.90}} = -2.65\right) = 0.0040$$

Dermed er sandsynligheden for at mere end 200 passagerer møder op, når man har solgt 210 billetter, 0.4% under antagelsen af at X er bionominalfordelingen samt de øvrige nævnte problemstillinger under forudsætningsdiskussionen.

Feb_2019 - Opgave 1.5 - Poissionfordeling

Nogle passagerer medtager kæledyr og for at imødekomme diverse krav for transport af kæledyr, har flyselskabet 2 luksusbokse samt adskillige andre bokse til denne rute. Erfaring viser, at antal kæledyr per afgang er poissonfordelt med $\mu=1,5$. **a)** Hvad er sandsynligheden for, at mere end 2 kæledyr møder op til en afgang? **b)** Hvad er det forventede antal kæledyr, der bliver transporteret i luksusbokse per afgang, når der kun er 2 luksusbokse til rådighed? (Passagerne har ikke mulighed for at vælge boks på forhånd) **c)** Hvor mange luksusbokse skulle der være til rådighed, hvis flyselskabet ville være 90 % sikker på ikke at løbe tør for luksusbokse?

Opgave 1.5

$X = \text{Antal kæledyr per flyafgang}$

Antallet af kæledyr per afgang antages at følge en poissonfordeling med $\mu = 1.5$, jævnfør opgaveteksten.

1) Forudsætninger

- a. Troværdigt svar

$$2 * P(Z \leq -2.58) = 2 * 0.0049 = 0.0098$$

8) Konklusion

På det foreliggende grundlag forkastes H_0 -hypotesen da Z_{obs} ligger udenfor de kritiske grænser på ± 1.96 . Det betyder at den nye forfatter har signifikant færre fejl, per side, end det korrekturlæseren har erfaring for under den antagelse at han ikke har misset nogle fejl. Ligeledes antages det at han har talt fejlene som han plejer. Det vil sige hvis han normalt har talt alle fejl, også hvis det er den samme fejl der er gået igen, har han forsat gjort det i rettelsen af den nye forfatter. Dette er en sikker konklusion da P-værdien på 0.0098 er skarpt mindre end signifikansniveauet på 0.05.

Hypoteseprøvning vedrørende én μ med ukendt varians

Jan_2019 - Opgave 1.1

Antag, at man i et auditorium tilfældigt har udvalgt 10 studerende og spurgt dem, hvor mange timer de studerer per uge. Der er 250 studerende tilmeldt på holdet, og der er 220 studerende til stede i lokalet. Af de 10 studerendes svar fremgik det, at de i gennemsnit brugte 29 timer i ugen med en standardafvigelse på 5 timer. **a)** Diskuter om stikprøveresultatet kan generaliseres til de tilmeldte. Anfør mindst tre ting, der meget vel kunne betyde, at man ikke kan generalisere. **b)** Gennemfør en hypotesetest for at teste om $\mu=37$. Se bort fra ovenstående anfægtelser

Spørgsmål 1.1 b

$X = \text{antal timer en studerende studerer per uge}$

$$n = 10, \bar{x} = 29, s = 5, s^2 = 25, \mu = 37$$

$$H_0: \mu = 37$$

$$H_1: \mu \neq 37$$

Jeg vælger en tosidet test da der ikke er noget der ligger til grund for at antage at gennemsnittet hverken er over eller under 37.

Jeg vælger et signifikansniveau på 0.05, da der ikke er opgivet noget i opgaveteksten.

Valg af observator

Da variansen er ukendt bruger jeg nedenstående observator:

$$\frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \sim T_{n-1}$$

Forudsætninger

Der ses bort fra de allerede diskutererede emner i opgave 1.1 a og disse antages opfyldte, jævnfør opgaveteksten.

- H_0 antages at være sand.
- $\bar{X} \sim NF$

Fordelingen antages at være forholdsvis symmetrisk omkring middelværdien. En smule højre skæv da det formodentlig kun er relativt få studerende der bruger mere end 29 timer

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}} = T_v$$

Da antallet af frihedsgrader ($n_1 + n_2 - 2$) ikke er over 30, kan der ikke approksimeres til normalfordelingen.

5) Udregning af observator værdien

$$t_{obs} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}} = \frac{(20 - 19)}{\sqrt{\left(\frac{3.3^2}{20} + \frac{1.7^2}{10}\right)}} = \frac{1}{\sqrt{\left(\frac{3.3^2}{20} + \frac{1.7^2}{10}\right)}} = 1.095$$

6) Kritiske grænser

$$\pm t_{v;a} = \pm t_{30+10-2;0.05} = \pm t_{28;0.05} = \pm 1.701$$

7) Beregning af P-værdi

$$P(T_v > |t_{obs}|)$$

$$P(T_{28;0.05} > 1.095) > 10\%$$

8) Konklusion

På foreliggende grundlag fastholdes H_0 -hypotesen da observationsværdien på 1.095 ligger inden for de kritiske grænser på ± 1.701 . Det betyder med andre ord at den nye bording procedure, på nuværende tidspunkt, ikke kan konkluderes til at være signifikant anderledes end den gamle, hvad angår gennemsnitlig boarding tid. Dette kommer af at stikprøverne har været meget små og det er forventeligt at der vil kunne påvises en forskel, med en større stikprøve, da variansen er faldet markant. En mindre varians betyder at sandsynligheden for lange boarding sessioner er betydeligt mindre og derfor vil jeg anbefale at bruge denne, hvis ikke den hverken er for dyr eller for besværlig.

Feb_2020 - Opgave 1.3

Det forlag, hvor korrekturlæseren er ansat, har sammenholdt denne korrekturlæseres registrerede antal fejl per bog med de øvrige korrekturlæseres registrerede antal fejl per bog (se opgaven).

Antag at de korrekturlæste bøger af denne og af de øvrige korrekturlæsere er sammenlignelige. a) Test om man på grundlag af ovenstående tabel kan konkludere, at der er forskel i det antal fejl, der registreres.

Opgave 1.3

1) Definer de stokastiske variable og opskriv de givne oplysninger

X_2 = Gennemsnitlige antal fejl denne korrekturlæser har fundet

X_1 = Gennemsnitlige antal fejl øvrige korrekturlæsere har fundet

$X_2 = 60.5, X_1 = 54.7, n_2 = 31, n_1 = 91, S_2 = 7.6, S_1 = 10.5$

Altså kan de to varianser ikke antages at være ens da observationsværdien på 1.92 ligger over de 1.32.

2) Formel

Da vi nu ved at de to varianser både er ukendte og ikke kan antages at være ens, bruger jeg nedenstående formel til at beregne konfidensintervallet:

$$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm Z_{\frac{\alpha}{2}} * \sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}$$

$$\mu_m - \mu_k = (370 - 305) \pm Z_{\frac{0.05}{2}} * \sqrt{\left(\frac{32400}{420} + \frac{16900}{480}\right)}$$

$$= 65 \pm 1.96 * \sqrt{(77.14 + 35.20)} = 65 \pm 1.96 * 10.599 = 65 \pm 20.77$$

$$\Rightarrow \mu_m - \mu_k \in [44.23 ; 85.77]$$

3) Konklusion

Altså kan jeg med 95% sikker konkludere at differencen mellem mænd og kvinders gennemsnitlige indkomst er indeholdt i intervallet [44.23 ; 85.77]. Da nul ikke er indeholdt i intervallet betyder det at der ikke er noget overlap imellem kvinder og mænds gennemsnitlige indkomst, mænds gennemsnitlige indkomst er højest. Ovenstående er selvfølgelig med de forbehold at forudsætningerne holder samt at stikprøven er repræsentativ. Hvis mænd med lav løn har en større tendens til ikke at svare end kvinder med lav løn, vil det betyde at lønforskellen vil se større ud end den i virkeligheden er.

Prøveeksamen - Opgave 1.3 - Forklaring af standardafvigelse og konfidensinterval

I en kommune med to meget små skoler har man i undervisningsåret 2019/20 sammenlignet disse to skoler, med hensyn til afgangsprøven i matematik, og har fået følgende resultat (*se opgaven*). **a)** Forklar med egne ord mulige årsager til forskellen mellem de to gennemsnit, samt årsager til at standardafvigelserne er noget mindre end 3,36, som ses i opgave 1.1. (Der ønskes ingen beregninger). **b)** Opstil et 95 % konfidensinterval for karakterniveauet i matematik for Skole 2. Kan man på grundlag heraf konkludere, at der er signifikant forskel på karakterniveauet for Skole 1 og Skole 2?

Spørgsmål 1.3

- a) En af grundene til forskellen i gennemsnittene og standardafvigelser kan være at de elever der går på de mindre skoler, kommer fra mere ressourcer stærke familier og de dermed har råd til at lade dem gå på, hvad jeg antager, er en privatskole. Det betyder både at eleverne kan forvente at få bedre undervisning, da der er mere tid til den enkelte elev, derudover har familien nok bedre mulighed (rent forståelsesmæssigt) for at hjælpe deres børn med lektierne. Til sidst kan der nævnes at de børn der går på en privatskole med alt sandsynlighed også er mere ens end de børn der går på en normal folkeskole, fordi der - i