

Copenhagen Business School
Dataanalyse, HD 1. del

ÆKSEMPEL PÅ EKSAMENSOPGAVE

4 timers skriftlig eksamen

Opgaven består af i alt **32** delspørgsmål fordelt på **10** opgaver. Ved bedømmelsen vægtes alle opgavens delspørgsmål ligeligt med undtagelse af delspørgsmål 10 a), der vægtes dobbelt. Vægtene skal alene opfattes som vejledende med henblik på disponering af tiden ved besvarelsen af eksamensopgaven.

NB:

- Hvis værktøjslinjen i JMP ikke fremgår af programmet, kan den genskabes v.hj.a. <Ctrl>-<K> -> “Windows Specific” -> “Auto-hide menu and toolbars” -> “never”
- Husk at medtage relevant JMP-output i opgavebesvarelsen (vedlæg ikke alt output men kun de dele, der refereres til i opgavebesvarelsen)
- Husk at JMP-output kun tæller med i bedømmelsen, såfremt det er eksplicit kommenteret/refereret.



REALKREDIT

BAGGRUND OG DATABESKRIVELSE

Danske realkreditinstitutter har siden indførelsen af realkreditlån med afdragsfrihed i 2003 oplevet en stigende interesse for optagelse af realkreditlån med afdragsfrihed til finansiering af køb af fast ejendom. Siden 2009 har realkreditlån med afdragsfrihed udgjort mere end halvdelen af alle realkreditlån.¹

Der har siden indførelsen i 2003 været meget debat om, hvilken betydning afdragsfrihed har for realkreditinstitutternes udlån, både i forhold til størrelsen af de lån der ydes, hvilke typer af ejendomme der ydes lån til, risikoen på de ydede lån osv. Denne opgave ser nærmere på forskellige aspekter af realkreditudlån i Danmark, herunder betydningen af afdragsfrihed, ved at analysere datamaterialet i JMP-filen *Realkredit.jmp*.

Datafilen indeholder information om et udsnit af de realkreditlån til privatpersoners køb af fast ejendom, som er ydet af et større dansk realkreditinstitut ved udgangen af 3. kvartal 2012. For hvert realkreditlån indeholder datamaterialet informationer om forskellige karakteristika vedrørende realkreditlånet, informationer om den ejendom lånet er knyttet til samt informationer om låntageren/låntagerne bag lånet.

Eksamensopgaven skal besvares på baggrund af datamaterialet i *Realkredit.jmp*, hvis indhold er nærmere beskrevet i tabellen på næste side.

¹Kilde: "Andelen af afdragsfrie realkreditlån er faldet svagt", Danmarks Nationalbank, 27. februar 2014.



VARIABELNAVN	FORKLARING
<i>Information om realkreditlån</i>	
<i>Ejendomsnr.</i>	Ejendomsnummer
<i>Restgæld</i>	Realkreditlånets restgæld i mio. kr.
<i>Belåningsgrad</i>	Forholdet mellem størrelsen af det optagne realkreditlån og værdien af den bagvedliggende ejendom. Belåningsgraden beregnes som $\frac{\text{Størrelse af restgæld på realkreditlån}}{\text{Værdi af belånt ejendom}}$.
<i>TekniskInsolvent</i>	Variabel der angiver, om belåningsgraden er større end 1, hvilket også kaldes teknisk insolvens (med de mulige værdier "Ja", "Nej")
<i>Lånetype</i>	Realkreditlånets type
<i>MaksimalLøbetid</i>	Realkreditlånets maksimale løbetid i år
<i>MulighedForAfdragsfrihed</i>	Variabel der angiver, om realkreditlånet giver mulighed for afdragsfrihed (med de mulige værdier "Ja", "Nej")
<i>Information om belånt ejendom</i>	
<i>Landsdel</i>	Hvilken landsdel den ejendom, realkreditlånet er knyttet til, er beliggende i (med de mulige værdier "Bornholm", "Fyn", "København by", "Københavns omegn", "Nordjylland", "Nordsjælland", "Sønderjylland", "Vest- og Sydsjælland", "Vestjylland", "Østjylland", "Østsjælland")
<i>Ejendomstype</i>	Ejendommens type (med de mulige værdier "Andet", "Ejerlejlighed", "Parcelhus", "Sommerhus")
<i>Information om låntagere</i>	
<i>AntalInteressenter</i>	Antal låntagere/interessenter på realkreditlånet
<i>GnsAlder</i>	Gennemsnitlig alder i år på låntagerne på realkreditlånet
<i>Anciennitet</i>	Antal måneder låntagerne har været kunder i realkreditinstituttet
<i>Andet</i>	
<i>TidligereIRestance</i>	Hvorvidt låntagerne tidligere har været i restance med betalinger til realkreditinstituttet
<i>Misligholdelse</i>	Hvorvidt realkreditlånet er blevet kategoriseret som misligholdt i løbet af 3. kvartal 2012. Misligholdelse indtræffer når låntager har været i restance med "et væsentligt beløb" i mindst 90 dage.



OPGAVE 1 (RESTGÆLD)

For at få overblik over realkreditinstituttets udlån ser vi på størrelsen af de enkelte udlån som angivet i variablen *Restgæld*.

- a). **Tegn** et histogram og et boxplot over fordelingen af variablen *Restgæld*.
Giv en kort beskrivelse af fordelingen af variablen *Restgæld* på baggrund af histogrammet, boxplottet og relevante nøgletal for centrum og variation.
- b). **Angiv** og **fortolk** medianen i fordelingen af variablen *Restgæld*.
Gør rede for om median eller middelværdi (eller begge) er retvisende mål for centrum i fordelingen af variablen *Restgæld*.
- c). **Tegn** den kumulerede fordelingsfunktion (CDF) for variablen *Restgæld*.
Bestem ud fra CDF-plottet omtrent hvor stor en andel af realkreditlånene, der har en restgæld på højst 5 mio. kr.
Bestem ud fra CDF-plottet omtrent hvor stor en andel af realkreditlånene, der har en restgæld på mellem 0,5 mio. kr. og 5 mio. kr.

OPGAVE 2 (BELÅNINGSGRAD)

Det, der tiltrækker sig størst opmærksomhed hos realkreditinstituttet, er ikke lånets restgæld, men derimod om værdien af den ejendom, realkreditlånet er optaget med sikkerhed i, er stor nok til at dække det udstedte lån. Forholdet mellem størrelsen på det udestående lån og værdien af den bagvedliggende ejendom - den såkaldte belåningsgrad - er angivet i variablen *Belåningsgrad*.

- a). **Tegn** et histogram over fordelingen af variablen *Belåningsgrad*.
Gør rede for hvorvidt det er muligt at bruge den empiriske regel til beskrivelse af variablen *Belåningsgrad*.
Bestem ved hjælp af den empiriske regel et interval som med cirka 95% sandsynlighed indeholder belåningsgraden for et tilfældigt valgt realkreditlån (idet vi her forudsætter, at den empiriske regel finder anvendelse).
- b). **Beregn** et 95%-konfidensinterval for den forventede værdi af *Belåningsgrad*. **Angiv** antagelser og **giv** en fortolkning af intervallet.
Gør rede for hvordan og hvorfor konfidensintervallet adskiller sig fra intervallet fundet i delspg. a).



OPGAVE 3 (TEKNISK INSOLVENS)

For de fleste typer af realkreditlån er den maksimale belåningsgrad på 80% af ejendommens værdi, når realkreditlånet optages.² Realkreditlån med en belåningsgrad på 80% eller derunder betragtes, alt andet lige, af realkreditinstituttet som forholdsvis sikre udlån, mens lån med en belåningsgrad over 80% anses for at rumme en ikke ubetydelig risiko for, at realkreditinstituttet kommer til at lide tab på lånet engang i fremtiden.

- a). **Gør rede for** om der er statistisk belæg for en nulhypotese om, at den forventede belåningsgrad på et af realkreditinstituttets udlån er lig 80%. **Angiv** i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.

Gør rede for om realkreditinstituttets udlån overordnet set kan karakteriseres som forholdsvis sikre udlån.

En værdi af *Belåningsgrad* større end 1 svarer til, at størrelsen af realkreditlånet overstiger værdien af ejendommen, og at låntager dermed er såkaldt "teknisk insolvent". Om låntager er teknisk involvent eller ej er angivet i variablen *TekniskInsolvent*.³

- b). **Beregn** et 95%-konfidensinterval for den forventede andel af realkreditlån hvis låntagere er teknisk insolvente. **Angiv** antagelser og **giv** en fortolkning af intervallet.

Gør rede for hvordan og hvorfor konfidensintervallet ændres, såfremt der i stedet beregnes et 99%-konfidensinterval.

Realkreditinstituttet yder lån til forskellige typer af ejendomme som angivet i variablen *Ejendoms-type*. Det er derfor relevant at vide, om der er særlig risiko for tab på lån knyttet til bestemte typer af ejendomme.

- c). **Estimér** for hver ejendomstype den forventede andel af teknisk insolvente realkreditlån.

Gør rede for om der er statistisk belæg for en nulhypotese om, at der er uafhængighed mellem ejendommens type og sandsynligheden for at være teknisk insolvent. **Angiv** i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.

²Belåningsgraden på et lån ændrer sig hele tiden, blandt andet afhængig af hvordan ejendommens værdi udvikler sig, og om og i givet fald hvor meget der afdrages på lånet.

³Når værdien af realkreditlånet overstiger værdien af ejendommen er låntager per definition teknisk insolvent. Teknisk insolvens fører imidlertid ikke til, at låntager reelt bliver insolvent og er nødt at afhænde ejendommen, så længe låntager blot er i stand til at overholde de på forhånd fastsatte ydelsesbetalinger på lånet. Teknisk insolvens er først og fremmest et problem for boligejere, der ønsker at sælge deres bolig, fordi de har udsigt til, at et eventuelt salg af boligen ikke vil indbringe nok til at kunne dække gælden til realkreditinstituttet.



OPGAVE 4 (LANDSDEL)

Realkreditinstituttet yder lån til ejendomme over hele landet som angivet i variablen *Landsdel*. Det er derfor relevant at undersøge, hvorvidt forskelle i ejendommenes geografiske placering har en sammenhæng med hvilke typer af ejendomme, realkreditinstituttet yder lån til.

- Tegn** en passende valgt figur af de to variable *Ejendomstype* og *Landsdel*, der viser eventuelle sammenhænge mellem de to variable.
Gør rede for de væsentligste tendenser, der kan udledes på baggrund af figuren.
- Opstil** en krydstabel for de to variable *Ejendomstype* og *Landsdel*.
Beregn sandsynligheden for at et tilfældigt valgt realkreditlån er knyttet til et sommerhus beliggende i Nordsjælland.
- Beregn** sandsynligheden for at et tilfældigt valgt realkreditlån er knyttet til et sommerhus, givet at vi kun ser på lån til ejendomme beliggende i Nordsjælland.
- Gør rede for** forskellen på de to sandsynligheder beregnet i delspg. b) og c).

OPGAVE 5 (AFDRAGSFRIHED)

Siden indførelsen i 2003 er afdragsfrie realkreditlån blevet populære blandt mange boligejere. Realkreditinstituttet udbyder en række forskellige typer af realkreditlån som angivet i variablen *Lånetype*, og flere af disse lån tilbydes både med og uden mulighed for afdragsfrihed som angivet i variablen *MulighedForAfdragsfrihed*.

Lån *med* afdragsfrihed anses af mange for at være *mere* risikable udlån end lån uden afdragsfrihed. For at mindske risikoen for tab på udlån er det derfor interessant for realkreditinstituttet at få klarlagt, om der er en sammenhæng mellem typen af realkreditlån der ydes, og hvorvidt lånet giver mulighed for afdragsfrihed eller ej

- Opstil** en krydstabel for de to variable *Lånetype* og *MulighedForAfdragsfrihed*.
Angiv de to hyppigst anvendte lånetyper.
Estimér for hver af de to hyppigst anvendte lånetyper den forventede andel af lån med mulighed for afdragsfrihed.
- Gør rede for** om der er statistisk belæg for en nulhypotese om, at der ikke er forskel på den forventede andel af lån med mulighed for afdragsfrihed for de to hyppigst anvendte lånetyper.
Angiv i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.



Én af antagelserne bag analysen i delspg. b) er antagelsen om observationernes indbyrdes uafhængighed.

- c). **Giv** et eksempel på, hvornår antagelsen om observationernes indbyrdes uafhængighed ikke er opfyldt i det givne datamateriale.

OPGAVE 6 (BELÅNINGSGRAD OG AFDRAGSFRIHED)

For et realkreditlån med afdragsfrihed vil lånets størrelse alt andet lige være større end for et tilsvarende lån uden afdragsfrihed. I forbindelse med realkreditinstituttets vurdering af, hvorvidt der skal ydes afdragsfrihed på fremtidige lån, er det derfor relevant at vide, om og i givet fald hvor meget det påvirker belåningsgraden, afhængig af om lånet rummer mulighed for afdragsfrihed eller ej.

- a). **Tegn** et histogram over fordelingen af *Belåningsgrad* og en tilhørende normalfordelingskurve for hver værdi af *MulighedForAfdragsfrihed*.
- b). **Beregn** et 95%-konfidensinterval for forskellen i den forventede værdi af *Belåningsgrad* for lån henholdsvis med og uden mulighed for afdragsfrihed. **Angiv** antagelser og **giv** en fortolkning af intervallet.
- c). **Gør rede for** om der er statistisk belæg for en nulhypotese om, at den forventede belåningsgrad er den samme for lån med og uden mulighed for afdragsfrihed. **Angiv** i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.
- d). **Gør rede for** sammenhængen mellem svarene i delspg. b) og c).

OPGAVE 7 (BELÅNINGSGRAD OG EJENDOMSTYPE)

Som tidligere nævnt har realkreditinstituttet fokus på, om der er en sammenhæng mellem risikoen for tab og hvilken type ejendom, realkreditlånet er knyttet til. I det omfang at belåningsgraden har en sammenhæng med risikoen for tab - større belåningsgrad vil alt andet lige øge risikoen for tab - er det væsentligt at få undersøgt, om der er en sammenhæng belåningsgraden og ejendommens type.

- a). **Tegn** et boxplot over fordelingen af variablen *Belåningsgrad* for hver værdi af variablen *Ejendomstype*.
Gør rede for hvad de tegnede boxplots indikerer om eventuelle forskelle i belåningsgrad på tværs af ejendomstyper.
- b). **Gør rede for** om der er statistisk belæg for en nulhypotese om, at den forventede belåningsgrad er den samme for alle ejendomstyper. **Angiv** i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.



OPGAVE 8 (LÅNTAGERNES ALDER)

Forud for udstedelsen af et realkreditlån indhenter realkreditinstituttet en række informationer om de kommende låntagere. Informationerne bruges til at vurdere, hvor kreditværdige låntagerne er. En af de oplysninger som indgår i vurderingen, er den gennemsnitlige alder blandt låntagerne som angivet i variablen *GnsAlder*.

- Estimér** middelværdi og standardafvigelse i den normalfordeling, der bedst beskriver fordelingen af variablen *GnsAlder*.
- Beregn** på baggrund af den i delspg. a) fundne normalfordeling sandsynligheden for, at låntagerne på et realkreditlån har en gennemsnitsalder på 40 år eller derunder.
- Beregn** på baggrund af den i delspg. a) fundne normalfordeling sandsynligheden for, at låntagerne på et realkreditlån har en gennemsnitsalder på mellem 40 og 70 år.

OPGAVE 9 (BELÅNINGSGRAD, SAMLET ANALYSE)

De ovenstående delanalyser har enkeltvis behandlet forskellige karakteristika vedrørende realkreditinstituttets udlån. For at få et samlet overblik over hvilke karakteristika, der har betydning for belåningsgraden på et realkreditlån foretager vi nu én samlet analyse.

- Tegn** et scatterplot af variablene *Anciennitet* og *GnsAlder* mod hinanden.
Beregn variablenes indbyrdes korrelation og **gør rede for** om variablene ser ud til, at indeholde stort set samme information om låntageren/låntagerne på et givet lån.

Vi ser nu på en lineær regressionsmodel med variablen *Belåningsgrad* som responsvariabel og variablene *MaksimalLøbetid*, *MulighedForAfdragsfrihed*, *Landsdel*, *Ejendomstype*, *AntalInteressenter*, *GnsAlder*, *Anciennitet*, *TidligereRestance* som forklarende variable.

- Angiv** estimater for regressionskoefficienter hørende til variablene *MaksimalLøbetid*, *AntalInteressenter*, *GnsAlder* og *Anciennitet* og **giv** en fortolkning af estimaterne.
- Angiv** estimater for regressionskoefficienter hørende til variablene *MulighedForAfdragsfrihed* og *TidligereRestance* og **giv** en fortolkning af estimaterne.
- Beregn** et 95%-konfidensinterval for betydningen af muligheden for afdragsfrihed og **giv** en fortolkning af intervallet.
- Beregn** et 95%-konfidensinterval for betydningen af 5 måneders ekstra anciennitet og **giv** en fortolkning af intervallet.
- Beregn** på baggrund af den estimerede regressionsmodel den forventede belåningsgrad for et realkreditlån med en maksimal løbetid på 30 år og mulighed for afdragsfrihed. Lånet er knyttet til et parcelhus beliggende i Nordjylland og er optaget af to personer med en gennemsnitsalder



på 55 år, der ikke tidligere har været kunder i realkreditinstituttet (dvs. ingen anciennitet har), og som følge heraf heller ikke tidligere har været i restance hos realkreditinstituttet.

- g). **Gør rede for** hvordan den forventede belåningsgrad i delspg. f) ændres, såfremt realkreditlånet ikke rummer mulighed for afdragsfrihed.

Gør rede for hvordan den forventede belåningsgrad i delspg. f) ændres, såfremt ejendommen i stedet er et sommerhus.

OPGAVE 10 (MISLIGHOLDELSE)

- a). **Gennemfør** en passende valgt analyse som undersøger, hvorvidt datamaterialet understøtter en påstand om, at mulighed for afdragsfrihed øger risikoen for, at et realkreditlån misligholdes som angivet i variablen *Misligholdelse*.

DATAKILDE

Datamaterialet i ovenstående opgaver er venligst stillet til rådighed af et større, unavngivent dansk realkreditinstitut.

DISCLAIMER:

Datamaterialet i nærværende opgave må udelukkende anvendes i forbindelse med faget Dataanalyse på CBS. Datamaterialet er fortroligt og må ikke videredistribueres eller offentliggøres andet end overfor studerende, der følger faget Dataanalyse på CBS, hverken skriftligt, mundtligt eller på anden vis, og ej heller de resultater og indsigter, der måtte komme ud af arbejdet med datamaterialet.



VEJLEDENDE LØSNING

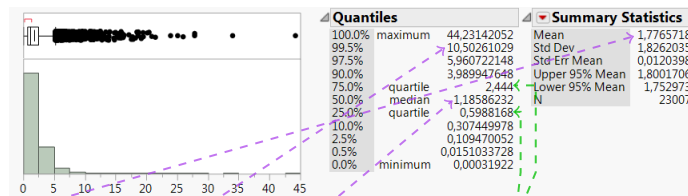
Bemærk at nedenstående løsning...

- udelukkende er **vejledende**, og at der i visse tilfælde kan være mere end én måde at besvare en given opgave på.
- visse steder har anført teoretiske formler sammen med det aflæste JMP-output for at fremme læsbarheden. Det er ikke et krav, at man gør noget tilsvarende i sin eksamensbesvarelse. Det er tilstrækkeligt at anføre de relevante tal direkte fra JMPs output samt inkludere det relevante output i besvarelsen.
- visse steder kommenterer på anførte antagelser. Det er ikke et krav, at man kommenterer på antagelser, medmindre det eksplicit efterspørges i opgaveteksten.
- til hver opgave i feltet ► *Løsning i JMP* indeholder et klikbart link til en kort video, der viser de nødvendige beregninger i JMP.



OPGAVE 1 (► Løsning i JMP)

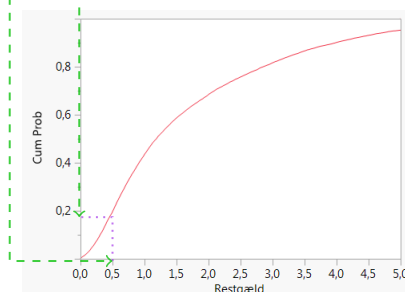
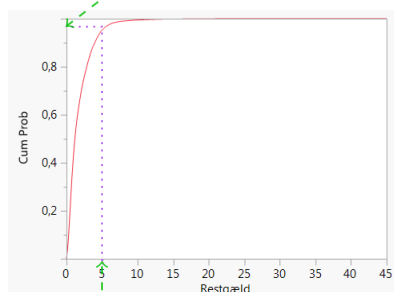
1A): Histogrammet over fordelingen af restgælden viser, at der er tale om en meget højreskæv fordeling, hvor enkelte realkreditlån har en meget stor restgæld på helt op til 44,2 mio. kr.. Sådanne lån må umiddelbart betragtes som outliers.



Højreskævheden af fordelingen fremgår også delvist af boxplottet, som blandt andet viser, at middelværdien på 1,78 mio. kr. er større end medianen på 1,19 mio. kr., hvilket alt andet lige er en indikation på højreskævhed. At der kun er tale om relativt få realkreditlån med meget stor restgæld ses eksempelvis af, at 99,5% af lånene har en restgæld på 10,50 mio. kr. eller mindre. Ligeledes ses det af, at standardafvigelsen, som er følsom overfor outliers, ikke er mærkbart påvirket af de få lån med stor restgæld, idet standardafvigelsen er af samme størrelsesorden (1,83 mio. kr.) som interkvartilbredden (på 2,44-0,60 = 1,84 mio. kr.), der ikke er følsom overfor outliers.

1B): Median restgælden er på 1,19 mio. kr., svarende til at 50% af realkreditlånene har en restgæld på 1,19 mio. kr. eller mindre. På grund af den konstaterede højreskævhed af fordelingen af restgælden er medianen et fint og retvisende mål for centrum af fordelingen, mens middelværdien på 1,78 mio. kr. i et vist omfang er påvirket af fordelings højreskævhed (idet den er noget større end medianen). Middelværdi og median er dog samlet set af nogenlunde samme størrelsesorden og kan dermed begge bruges som mål for centrum af fordelingen.

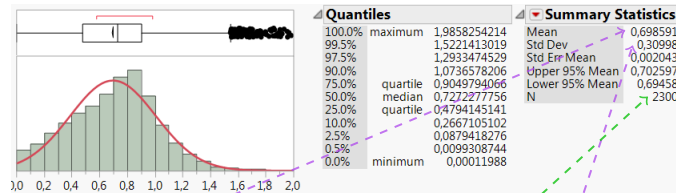
1C): Den kumulerede fordelingsfunktion for lånenes restgæld viser, at ca. 95% af lånene har restgæld på 5 mio. kr. eller derunder. Hvis der "zoomes ind" viser fordelingsfunktionen tillige, at ca. 20% af lånene har en restgæld på 0,5 mio. kr. eller derunder. Dermed har ca. $95\% - 20\% = 75\%$ af realkreditlånene en restgæld på mellem 0,5 mio. kr. og 5 mio. kr.





OPGAVE 2 (► Løsning i JMP)

2A): Histogrammet over fordelingen af belåningsgraden viser, at fordelingen med en vis tilnærmelse godt kan anses for at være klokkeformet, således at den empiriske regel kan anvendes.



Fordelingens gennemsnit er $\bar{x} = 0,699$ og standardafvigelsen er $s = 0,310$, hvilket ifølge den empiriske regel betyder, at belåningsgraden for et tilfældigt valgt realkreditlån med cirka 95% sandsynlighed vil ligge i intervallet

$$[\bar{x} - 2 \cdot s; \bar{x} + 2 \cdot s] = [0,699 - 2 \cdot 0,310; 0,699 + 2 \cdot 0,310] = [0,079; 1,319]$$

2B): Datamaterialet indeholder i alt $n = 23.007$ observationer, hvilket v.h.j.a tallene fra delspg. 2a) giver et 95%-konfidensinterval for den forventede værdi på⁴

$$\begin{aligned} & \left[\bar{x} - t_{0,025}(df = n - 1) \cdot \frac{s}{\sqrt{n}}; \bar{x} + t_{0,025}(df = n - 1) \cdot \frac{s}{\sqrt{n}} \right] \\ &= \left[0,699 - 1,96 \cdot \frac{0,310}{\sqrt{23.007}}; 0,699 + 1,96 \cdot \frac{0,310}{\sqrt{23.007}} \right] \\ &= [0,695; 0,703] \end{aligned}$$

Med 95% sandsynlighed vil den forventede belåningsgrad for realkreditinstituttets lån dermed ligge mellem 0,695 og 0,703.

Antagelserne bag beregningen af konfidensintervallet er, at observationerne af de enkelte låns belåningsgrader er indbyrdes uafhængige og approksimativt normalfordelte. Antagelsen om normalfordeling ser ud til at være opfyldt, da fordelingen er nogenlunde klokkeformet, jf. delspg. a).

Intervallet i delspg. a) viser, at belåningsgraden for ét tilfældigt valgt realkreditlån med 95% sandsynlighed vil ligge mellem 0,079 og 1,319. Omvendt viser intervallet her i delspg. b), at den gennemsnitlige belåningsgrad blandt en hel masse realkreditlån med 95% sandsynlighed vil ligge mellem 0,695 og 0,703. Sidstnævnte beregning ser altså på et gennemsnit af belåningsgrader blandt en hel masse (helt præcist 23.007) realkreditlån, og ikke kun på ét enkelt realkreditlån som i delspg. a). Det er forklaringen på, at vi får et meget mere præcist resultat (= smallere interval) her i delspg. b).

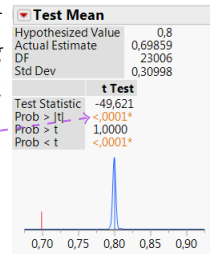
⁴ t -fordelingen har her så mange frihedsgrader ($n - 1 = 23.006$), at den i praksis er lig normalfordelingen. Dermed bliver $t_{0,025}(df = 23.006)$ lig med 97,5%-fraktilen i normalfordelingen, dvs. $t_{0,025} = 1,96$.



OPGAVE 3 (► Løsning i JMP)

3A): Test af nulhypotese om at den forventede belåningsgrad er lig 80%:

Antagelser:	Observationerne af de enkelte låns belåningsgrader er indbyrdes uafhængige og approksimativt normalfordelte. Antagelsen om normalfordeling ser ud til at være opfyldt, da fordelingen er nogenlunde klokkeformet, jf. delspg. 2a).
Hypoteser:	$H_0: \mu = 0,80$ $H_a: \mu \neq 0,80$
Signifikanssssh.:	$p < 0,0001$
Tekn. konklusion:	Signifikanssandsynligheden er mindre end 0,05 og dermed forkastes H_0 .
Let konklusion:	Der er ikke belæg for en påstand om, at den forventede belåningsgrad på realkreditinstituttets udlån er lig 80%. M.a.o. er der belæg for en påstand om, at den forventede belåningsgrad er forskellig fra 80%.



Eftersom datamaterialet giver belæg for en påstand om, at den forventede belåningsgrad på realkreditinstituttets udlån er forskellig fra 80% er spørgsmålet, om den forventeligt er større eller mindre end 80%. Den gennemsnitlige belåningsgrad i datamaterialet er 0,699 (jf. opgave 2), så umiddelbart må det være tegn på, at den forventede belåningsgrad er lavere end 80% og at realkreditinstituttets udlån derfor overordnet set kan karakteriseres som forholdsvis sikre.

3B): Et 95%-konfidensinterval for den forventede andel af teknisk insolvente realkreditlån er

$$\left[\hat{p} - \sqrt{\hat{p}(1-\hat{p})/n}; \hat{p} + \sqrt{\hat{p}(1-\hat{p})/n} \right]$$

$$= \left[0,150 - 1,96 \cdot \sqrt{0,150(1-0,150)/23.007}; 0,150 + 1,96 \cdot \sqrt{0,150(1-0,150)/23.007} \right]$$

$$= [0,146; 0,155]$$

Confidence Intervals					
Level	Count	Prob	Lower CI	Upper CI	1-Alpha
Ja	3457	0.15026	0.144291	0.156427	0.950
Nej	19550	0.84974	0.849066	0.8543	0.950
Total	23007				

Med 95% sandsynlighed vil den forventede andel af teknisk insolvente realkreditlån dermed ligge mellem 14,6% og 15,5%.

Antagelserne bag beregningen af konfidensintervallet er, at observationerne af om de enkelte lån er teknisk insolvente er indbyrdes uafhængige og at $n \cdot \hat{p} \geq 15$ og $n \cdot (1 - \hat{p}) \geq 15$. Sidstnævnte er opfyldt idet $n \cdot \hat{p} = 23.007 \cdot 0,150 = 3.457$ og $n \cdot (1 - \hat{p}) = 23.007 \cdot 0,850 = 19.550$.

Et 99%-konfidensinterval for den forventede andel af teknisk insolvente realkreditlån er

$$\left[\hat{p} - \sqrt{\hat{p}(1-\hat{p})/n}; \hat{p} + \sqrt{\hat{p}(1-\hat{p})/n} \right]$$

$$= \left[0,150 - 2,58 \cdot \sqrt{0,150(1-0,150)/23.007}; 0,150 + 2,58 \cdot \sqrt{0,150(1-0,150)/23.007} \right]$$

$$= [0,144; 0,156]$$

Confidence Intervals					
Level	Count	Prob	Lower CI	Upper CI	1-Alpha
Ja	3457	0.15026	0.144291	0.156427	0.990
Nej	19550	0.84974	0.849573	0.850427	0.990
Total	23007				

Med 99% sandsynlighed vil den forventede andel af teknisk insolvente realkreditlån dermed ligge mellem 14,4% og 15,6%.

99%-konfidensintervallet er brede end 95%-konfidensintervallet, fordi det med en større grad af sikkerhed indeholder den sande andel af teknisk insolvente realkreditlån.



3c): Den forventede andel af teknisk insolvente realkreditlån estimeres indenfor hver ejendomstype til

$$\hat{p}_{\text{andet}} = 15,91\% \quad \hat{p}_{\text{ejerlejlighed}} = 20,52\% \quad \hat{p}_{\text{parcelhus}} = 15,60\% \quad \hat{p}_{\text{sommerhus}} = 2,83\%$$

Test af nulhypotese om at der er uafhængighed mellem teknisk insolvens og ejendomstype:

Antagelser: Observationerne af om et lån er teknisk insolvent eller ej er indbyrdes uafhængige. Endvidere er det forventede antal observationer i hver af krydstabellens celler større end 5. Sidstnævnte antagelse er opfyldt, da alle celler har et forventet antal observationer på 89 eller mere.

Hypoteser: H_0 : Uafhængighed mellem TekniskInsolvens og Ejendomstype
 H_a : Afhængighed mellem TekniskInsolvens og Ejendomstype

Signifikanssssh.: $p < 0,0001$

Tekn. konklusion: Signifikanssandsynligheden er mindre end 0,05 og dermed forkastes H_0 .

Let konklusion: Der er ikke belæg for en påstand om, at der er uafhængighed mellem om et lån er teknisk insolvent og den bagvedliggende ejendoms type. M.a.o. er der belæg for en påstand om, at der er en sammenhæng (afhængighed) mellem sandsynligheden for at et lån er teknisk insolvent, og hvilken type ejendom lånet finansierer.

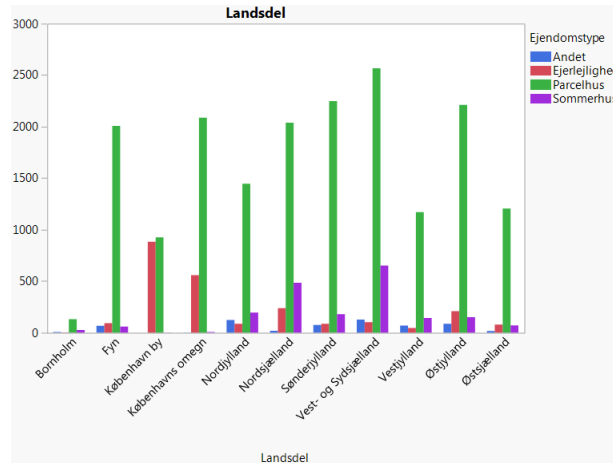
Contingency Table			
		TekniskInsolvent	
		Ja	Nej
Count			
Total			
Col %			
Row %			
Expected			
Andet		95	502
		0,41	2,18
		2,75	2,57
		15,91	84,09
		89,7044	507,296
Ejerlejlighed		491	1902
		2,13	8,27
		14,20	9,73
		20,52	79,48
		359,569	2033,43
Parcelhus		2815	15226
		12,24	66,18
		81,43	77,88
		15,60	84,40
		2710,82	15330,2
Sommerhus		56	1920
		0,24	8,35
		1,62	9,82
		2,83	97,17
		296,911	1679,09
Total		3457	19550
		15,03	84,97

Tests			
	N	DF	-LogLike
	23007	3	192,39832
			RSquare (U)
			0,0198
Test	ChiSquare	Prob>ChiSq	
Likelihood Ratio	384,797	<.0001*	
Pearson	291,655	<.0001*	

Af estimerne ovenfor fremgår det eksempelvis, at ca. 21% af samtlige låntagere af realkreditlån knyttet til ejerlejligheder er teknisk insolvente, mens det samme tal kun er ca. 3% for sommerhuse.

OPGAVE 4 (► [Løsning i JMP](#))

4A): Nedenstående figur viser for hver landsdel et søjlediagram af antallet af ejendomme af hver enkelt type:



Figuren viser at...

- parcelhuse udgør langt den største andel af de belånte ejendomme i alle landsdele med undtagelse af "København by".
- de belånte ejerlejligheder primært er beliggende i "København by" og "Københavns omegn".
- de belånte sommerhuse primært er beliggende i "Vest- og Sydsjælland" og "Nordsjælland".
- antallet af ejendomme er nogenlunde ligeligt fordelt på tværs af (de fleste) landsdele. Dog er antallet af ejendomme lavere i landsdelene "Bornholm", "København by", "Nordjylland", "Vestjylland" og "Østsjælland". Flere af disse områder er generelt tyndt befolket, hvilket delvist kan forklare det lavere antal i disse landsdele.



4b): Krydstabel mellem ejendommens type og hvilken landsdel, ejendommen er beliggende i:

Ejendoms- type	Landsdel											Total
	Bornhol m	Fyn	Københ avn by	Københ avns omegn	Nordjyll and	Nordsj ælland	Sønderj ylland	Vest- og Sydsjæll and	Vestjyll and	Østjylla nd	Østsjæll and	
Count	7	67	0	1	124	19	75	129	70	88	17	597
Total %	0,03	0,29	0,00	0,00	0,54	0,08	0,33	0,56	0,30	0,38	0,07	2,59
Col %	4,12	3,01	0,00	0,04	6,69	0,68	2,89	3,74	4,90	3,31	1,24	
Row %	1,17	11,22	0,00	0,17	20,77	3,18	12,56	21,61	11,73	14,74	2,85	
Andet	3	94	884	560	88	238	88	103	46	209	80	2393
Ejerlejlighed	0,01	0,41	3,84	2,43	0,38	1,03	0,38	0,45	0,20	0,91	0,35	10,40
	1,76	4,22	48,76	21,09	4,75	8,56	3,40	2,98	3,22	7,86	5,82	
	0,13	3,93	36,94	23,40	3,68	9,95	3,68	4,30	1,92	8,73	3,34	
Parcelhus	132	2007	927	2087	1446	2038	2249	2567	1171	2211	1206	18041
	0,57	8,72	4,03	9,07	6,29	8,86	9,78	11,16	5,09	9,61	5,24	78,42
	77,65	90,04	51,13	78,61	77,99	73,31	86,77	74,38	81,89	83,15	87,77	
	0,73	11,12	5,14	11,57	8,02	11,30	12,47	14,23	6,49	12,26	6,68	
Sommerhus	28	61	2	7	196	485	180	652	143	151	71	1976
	0,12	0,27	0,01	0,03	0,85	2,11	0,78	2,83	0,62	0,66	0,31	8,89
	16,47	2,74	0,11	0,26	10,57	17,45	6,94	18,89	10,00	5,68	5,17	
	1,42	3,09	0,10	0,35	9,92	24,54	9,11	33,00	7,24	7,64	3,59	
Total	170	2229	1813	2655	1854	2780	2592	3451	1430	2659	1374	23007
	0,74	9,69	7,88	11,54	8,06	12,08	11,27	15,00	6,22	11,56	5,97	

Sandsynligheden for at et tilfældigt valgt lån er knyttet til et sommerhus beliggende i Nordsjælland er:

$$P(\{\text{ejendom er sommerhus}\} \cap \{\text{ejendom ligger i Nordsjælland}\}) = \frac{485}{23.007} = 2,11\%$$

4c): Sandsynligheden for at et tilfældigt valgt lån er knyttet til et sommerhus, når vi udelukkende ser på ejendomme beliggende i Nordsjælland, er:

$$\begin{aligned} & P(\{\text{ejendom er sommerhus}\} | \{\text{ejendom ligger i Nordsjælland}\}) \\ &= \frac{P(\{\text{ejendom er sommerhus}\} \cap \{\text{ejendom ligger i Nordsjælland}\})}{P(\{\text{ejendom ligger i Nordsjælland}\})} \\ &= \frac{\frac{485}{23.007}}{\frac{2.780}{23.007}} = 17,45\% \end{aligned}$$

4d): Sandsynligheden i delspg. b) angiver, at 2,11% af samtlige realkreditinstituttets belånte ejendomme er sommerhuse beliggende i Nordsjælland (her ser vi på alle ejendomme). Sandsynligheden i delspg. c) angiver derimod, at 17,45% af de belånte ejendomme i Nordsjælland er sommerhuse (her ser vi kun på ejendomme i Nordsjælland). Sandsynligheden i delspg. c) vil altid være større end sandsynligheden i delspg. b), fordi vi i delspg. c) ser på et mindre udsnit af ejendomme (kun Nordsjælland) end i delspg. b) (ejendomme i hele landet).



OPGAVE 5 (► Løsning i JMP)

5A): Krydstabel mellem realkreditlånets type og hvorvidt lånet giver mulighed for afdragsfrihed:

	Count	Total %	Boligkredit	Boligkredit real	Boligkredit sparingslån	Garantilån	Indeks	Kontant	Kort	Obligationsslån	Obligationssparingslån	Obligationssparingslån	Rentefradragsslån	Rentetilpasningslån	Total
Mulighed for afdragsfrihed															
Ja	0	0	0	1586	0	0	0	0	364	1632	0	0	87	9938	13607
	0,00	0,00	0,00	6,89	0,00	0,00	0,00	0,00	1,58	7,09	0,00	0,00	0,38	43,20	59,14
	0,00	0,00	0,00	59,29	0,00	0,00	0,00	0,00	100,00	20,77	0,00	0,00	100,00	94,27	
	0,00	0,00	0,00	11,66	0,00	0,00	0,00	0,00	2,68	11,99	0,00	0,00	0,64	73,04	
Nej	847	20	1089	1	573	0	6227	38	1	0	604	9400			
	3,68	0,09	4,73	0,00	2,49	0,00	27,07	0,17	0,00	0,00	2,63	40,86			
	100,00	100,00	40,71	100,00	100,00	0,00	79,23	100,00	100,00	0,00	5,73				
	9,01	0,21	11,59	0,01	6,10	0,00	66,24	0,40	0,01	0,00	6,43				
Total	847	20	2675	1	573	364	7859	38	1	87	10542	23007			
	3,68	0,09	11,63	0,00	2,49	1,58	34,16	0,17	0,00	0,38	45,82				

De to hyppigst anvendte låntyper er "Obligationsslån" (i alt 7.859 lån) og "Rentetilpasningslån" (i alt 10.542 lån). For hver af disse låntyper estimeres den forventede andel af lån, som giver mulighed for afdragsfrihed, til

$$\hat{p}_{\text{obliga}} = \frac{1.632}{7.859} = 20,77\% \quad \hat{p}_{\text{tilpas}} = \frac{9.938}{10.542} = 94,27\%$$



5b): Test af nulhypotese om at den forventede andel af lån med mulighed for afdragsfrihed er den samme blandt obligations- og rentetilpasningslån:

Contingency Table			
		MulighedForAfdragsfrihed	
		Nej	Ja
Count			
Total %			
Col %			
Row %			
Expected			
Låntype	Obligationslån	6227	1632
		33,84	8,87
		91,16	14,11
		79,23	20,77
		2917,5	4941,5
Rentetilpasningslån		604	9938
		3,28	54,01
		8,84	85,89
		5,73	94,27
		3913,5	6628,5
Total		6831	11570
		37,12	62,88
		18401	

Two Sample Test for Proportions			
Description	Proportion Difference	Lower 95%	Upper 95%
P(Nej Obligationslån)-P(Nej Rentetilpasningslån)	0,735045	0,72488	0,744894
Adjusted Wald Test		Prob	
P(Nej Obligationslån)-P(Nej Rentetilpasningslån) ≥ 0		<,0001*	
P(Nej Obligationslån)-P(Nej Rentetilpasningslån) ≤ 0		1,0000	
P(Nej Obligationslån)-P(Nej Rentetilpasningslån) = 0		<,0001*	

Antagelser:

Observationerne af om et lån giver mulighed for afdragsfrihed eller ej er indbyrdes uafhængige. Endvidere er det forventede antal observationer i hver af krydstabelens celler større end 5. Sidstnævnte antagelse er opfyldt, da alle celler har et forventet antal observationer på 2.917 eller mere.

Hypoteser:

$$H_0 : p_{\text{obliga}} = p_{\text{tilpas}}$$

$$H_a : p_{\text{obliga}} \neq p_{\text{tilpas}}$$

Signifikanssssh.:

$$p < 0,0001$$

Tekn. konklusion:

Signifikanssandsynligheden er mindre end 0,05 og dermed forkastes H_0 .

Let konklusion:

Der er ikke belæg for en påstand om, at andelen af lån med mulighed for afdragsfrihed er den samme blandt henholdsvis obligations- og rentetilpasningslån. På baggrund af resultaterne i delspg. a) er den konklusion ikke overraskende, eftersom andelen af obligationslån med mulighed for afdragsfrihed estimeredes til 20,77%, mens den tilsvarende andel blandt rentetilpasningslån estimeredes til 94,27%. Muligheden for afdragsfrihed er således markant mere udtalt blandt rentetilpasningslån end blandt obligationslån.



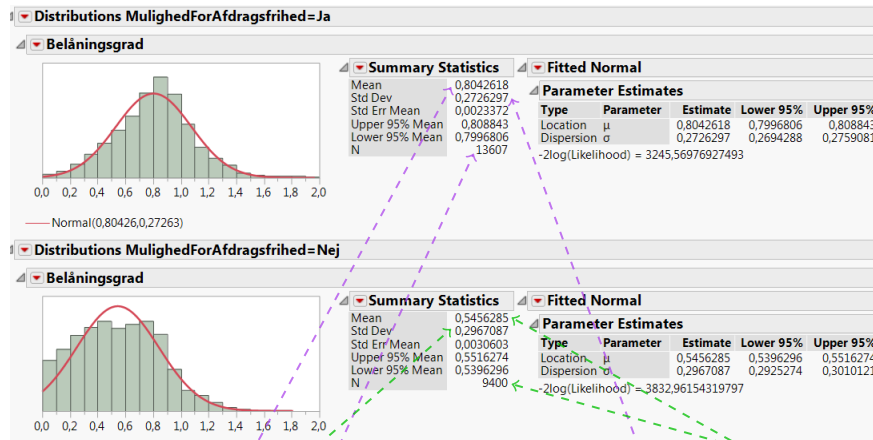
5c): To observationer i datamaterialet er ikke uafhængige, hvis der er tale om en eller anden form for afhængighed mellem dem, f.eks. som følge af...

- en låneomlægning. Hvis eksempelvis et obligationslån med afdragsfrihed omlægges til et rentetilpasningslån, så er det overvejende sandsynligt, at rentetilpasningslånet også vil rumme afdragsfrihed, fordi låntageren allerede i sit gamle lån har haft afdragsfrihed og derfor formentlig gerne vil fortsætte med det. Dermed vil der være to observationer i datamaterialet for den samme ejendom - ét for hver lånetype (obligationslån hhv. rentetilpasningslån) - og mellem disse to observationer vil der være en afhængighed.
- markedsføringskampagner fra realkreditinstituttet. Hvis eksempelvis en kampagne opfordrer til om-lægning af eksisterende lån til nye lån med rentetilpasning og afdragsfrihed, så vil det - i det omfang at kampagnerne rent faktisk virker - få låntagerne til at omlægge deres lån og dermed skabe en øget mængde nye rentetilpasningslån med afdragsfrihed. Det vil skabe en afhængighed mellem de nye lån der etableres, fordi de vil være ens både mht. type (rentetilpasningslån) og muligheden for afdragsfrihed.
- relationer mellem låntagere. Hvis eksempelvis én låntager, som har et obligationslån uden afdragsfrihed, og som er glad for sit lån, påvirker en anden låntager (kollega, ven, familie el.lign.) til at ændre sit lån til samme type, så vil det skabe en afhængighed mellem disse låntageres lån.



OPGAVE 6 (► Løsning i JMP)

6A): Histogram over fordelingen af belåningsgraden for lån henholdsvis med og uden mulighed for afdragsfrihed:



6B): Datamaterialet indeholder $n_{med} = 13.607$ observationer af lån med mulighed for afdragsfrihed med en gennemsnitlig værdi på $\bar{x}_{med} = 0,804$ og standardafvigelse på $s_{med} = 0,273$, og tilsvarende $n_{uden} = 9.400$ observationer af lån uden mulighed for afdragsfrihed med en gennemsnitlig værdi på $\bar{x}_{uden} = 0,546$ og standardafvigelse på $s_{uden} = 0,297$. Det giver et 95%-konfidensinterval for forskellen på de to forventede værdier på⁵

$$\left[\bar{x}_{med} - \bar{x}_{uden} - t_{0,025}(df = 19.077) \cdot \sqrt{\frac{s_{med}^2}{n_{med}} + \frac{s_{uden}^2}{n_{uden}}}; \bar{x}_{med} - \bar{x}_{uden} + t_{0,025}(df = 19.077) \cdot \sqrt{\frac{s_{med}^2}{n_{med}} + \frac{s_{uden}^2}{n_{uden}}} \right]$$

$$= \left[0,804 - 0,546 - 1,96 \cdot \sqrt{\frac{0,273^2}{13.607} + \frac{0,297^2}{9.400}}; 0,804 - 0,546 + 1,96 \cdot \sqrt{\frac{0,273^2}{13.607} + \frac{0,297^2}{9.400}} \right]$$

$$= [0,251; 0,266]$$

Kan aflæses i felterne "Upper CL Dif" og "Lower CL Dif" i JMP-outputtet øverst næste side

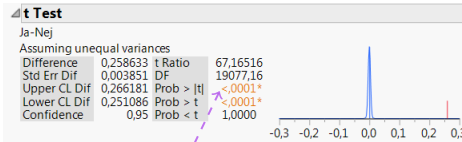
Med 95% sandsynlighed vil forskellen mellem den forventede belåningsgrad for lån henholdsvis med og uden mulighed for afdragsfrihed dermed ligge mellem 0,251 og 0,266.

Antagelserne bag beregningen af konfidensintervallet er, at observationerne af de enkelte låns belåningsgrader er indbyrdes uafhængige og approksimativt normalfordelte; normalfordelte indenfor hver gruppe af lån, i.e. med/uden mulighed for afdragsfrihed. Antagelsen om normalfordeling indenfor hver gruppe ser ud til at være opfyldt, da de to fordelinger begge er nogenlunde klokkeformede, jf. delspg. a).

⁵Se fodnote 5 s. 489 i lærebogen for udtrykket for antal frihedsgrader df i $t_{0,025}(df)$ -fraktilen.



6c): Test af nulhypotese om at den forventede belåningsgrad er den samme for lån med og uden mulighed for afdragsfrihed:

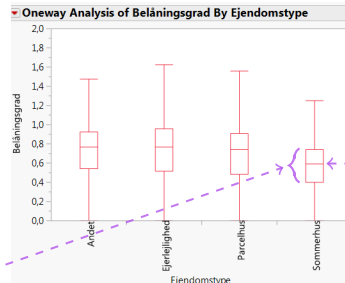


- Antagelser:** Observationerne af de enkelte låns belåningsgrader er indbyrdes uafhængige og approksimativt normalfordelte; normalfordelte indenfor hver gruppe af lån (med/uden mulighed for afdragsfrihed). Antagelsen om normalfordeling indenfor hver gruppe ser ud til at være opfyldt, da de to fordelinger begge er nogenlunde klokkeformede, jf. delspg. a).
- Hypoteser:** $H_0 : \mu_{med} = \mu_{uden}$
 $H_a : \mu_{med} \neq \mu_{uden}$
- Signifikansssh.:** $p < 0,0001$
- Tekn. konklusion:** Signifikanssandsynligheden er mindre end 0,05 og dermed forkastes H_0 .
- Let konklusion:** Der er ikke belæg for en påstand om, at den forventede belåningsgrad er den samme på lån med og uden mulighed for afdragsfrihed. På baggrund af resultaterne i delspg. a) er den konklusion ikke overraskende, eftersom den forventede belåningsgrad blandt lån med mulighed for afdragsfrihed estimeredes til 0,804, mens belåningsgraden blandt lån uden mulighed for afdragsfrihed estimeredes til 0,546. Den forventede belåningsgrad er således større blandt lån med mulighed for afdragsfrihed, hvilket naturligt kan forklares med, at afdragsfrihed betyder, at der i en periode ikke betales af på lånet, hvilket alt andet lige vil øge restgælden og dermed øge belåningsgraden.



OPGAVE 7 (► Løsning i JMP)

7A): Boxplot over fordelingen af belåningsgraden indenfor hver type af ejendom.



Figuren viser at niveauet af belåningsgrad, som angivet ved medianen, er lidt lavere (omkring 0,6) for sommerhuse sammenlignet med alle øvrige ejendomstyper. Variationen i belåningsgrad, som angivet ved interkvartilbredden, er ligeledes (en lille smule) mindre for sommerhuse end for de øvrige ejendomstyper.

7B): Test af nulhypotese om at den forventede belåningsgrad er den samme for de forskellige ejendomstyper:

Oneway Anova					
Summary of Fit					
Rsquare		0,017585			
Adj. Rsquare		0,017457			
Root Mean Square Error		0,307264			
Mean of Response		0,698592			
Observations (or Sum Wgts)		23007			
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Ejendomstype	3	36,8740	12,9580	137,2505	<.0001*
Error	23003	2171,7465	0,0944		
C. Total	23006	2210,6205			
Means for Oneway Anova					
Level	Number	Mean	Std Error	Lower 95%	Upper 95%
Andet	597	0,745393	0,01258	0,72094	0,77024
Ejendomslejlighed	2393	0,740497	0,00628	0,72819	0,75281
Parcelhus	18041	0,705522	0,00229	0,70104	0,71001
Sommerhus	1976	0,570371	0,00691	0,55682	0,58392

Antagelser:

Observationerne af de enkelte låns belåningsgrader er indbyrdes uafhængige og approksimativt normalfordelte; normalfordelte indenfor hver ejendomstype og med samme std.afvigelse for alle ejendomstyper. Hvorvidt antagelsen om normalfordeling indenfor hver ejendomstype er opfyldt kan ikke vurderes ud fra de optegnede box plots, men antagelsen om samme std.afvigelse indenfor hver ejendomstype ser ikke ud til helt at være opfyldt.

Hypoteser:

$$H_0 : \mu_{\text{andet}} = \mu_{\text{ejendomslejlighed}} = \mu_{\text{parcelhus}} = \mu_{\text{sommerhus}}$$

H_a : Mindst to af de forventede værdier μ_{andet} , $\mu_{\text{ejendomslejlighed}}$, $\mu_{\text{parcelhus}}$, $\mu_{\text{sommerhus}}$ er ikke ens

Signifikanssssh.:

$$p < 0,0001$$

Tekn. konklusion:

Signifikanssandsynligheden er mindre end 0,05 og dermed forkastes H_0 .

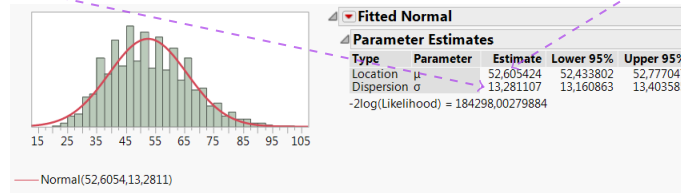
Let konklusion:

Der er ikke belæg for en påstand om, at den forventede belåningsgrad er den samme på tværs af ejendomstyper. M.a.o. er der belæg for en påstand om, at den forventede belåningsgrad er forskellig for de forskellige ejendomstyper, hvilket stemmer overens med de optegnede boxplots i delspg. a). Det er særligt sommerhuse som med en gennemsnitlig belåningsgrad på 0,57 skiller sig ud ved at have en markant lavere belåning end de øvrige ejendomstyper.



OPGAVE 8 (► Løsning i JMP)

8A): Låntagerne hos realkreditinstituttet har en estimeret forventet alder på $\hat{\mu} = 52,6$ år med en estimeret std.afvigelse på $\hat{\sigma} = 13,3$ år.



8B): Sandsynligheden for at låntagerne på et vilkårligt valgt realkreditlån har en gennemsnitsalder på 40 år eller derunder er $P(\text{GnsAlder} \leq 40) = 17,2\%$, hvor sandsynligheden er beregnet i en normalfordeling med forventet værdi og std.afvigelse som angivet i delspg. a).

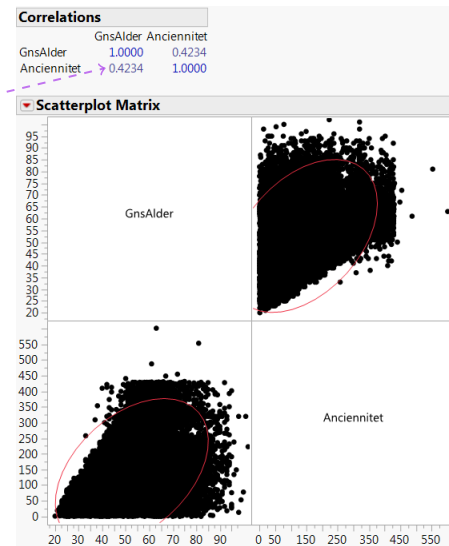
8C): På samme måde som i delspg. b) findes, at sandsynligheden for at låntagerne på et vilkårligt realkreditlån har en gennemsnitsalder på 70 år eller derunder er $P(\text{GnsAlder} \leq 70) = 90,5\%$. Dermed er sandsynligheden for at låntageres gennemsnitsalder er mellem 40 og 70 år

$$P(40 \leq \text{GnsAlder} \leq 70) = P(\text{GnsAlder} \leq 70) - P(\text{GnsAlder} \leq 40) = 90,5\% - 17,2\% = 73,3\%$$

OPGAVE 9 (► [Løsning i JMP](#))

9A): Scatterplottet af låntagernes anciennitet mod deres alder viser - ikke overraskende - at de to variable har en vis positiv sammenhæng (dvs. større anciennitet tenderer til at hænge sammen med højere alder).

Korrelationen mellem de to variable er 0,42, hvilket ligeledes indikerer en vis positiv sammenhæng. Sammenhængen er ganske intuitiv, idet man, alt andet lige, må forvente, at jo længere tid en låntager har været kunde i realkreditinstituttet desto ældre er vedkommende. For en given anciennitet er der dog fortsat en væsentlig spredning på kundernes alder, hvilket både fremgår af scatterplottet⁶, men også ses af at korrelationen ikke er tæt på 1, men "kun" er på 0,42.



⁶Det er ligegyldigt, hvilket af de to plots i JMP-outputtet, man refererer til.



9b): Parameterestimatet hørende til variablen...

- **MaksimalLøbetid** er $\hat{\beta}_{\text{MaksimalLøbetid}} = 0,0143$, således at den forventede belåningsgrad stiger med 0,0143, hvis lånets maksimale løbetid stiger med 1 år.
- **AntalInteressenter** er $\hat{\beta}_{\text{AntalInteressenter}} = 0,0125$, således at den forventede belåningsgrad stiger med 0,0125, hvis antallet af låntagere (= interessenter) på lånet stiger med 1.
- **GnsAlder** er $\hat{\beta}_{\text{GnsAlder}} = -0,0059$, således at den forventede belåningsgrad falder med 0,0059, hvis låntagernes gennemsnitsalder stiger med 1 år.
- **Anciennitet** er $\hat{\beta}_{\text{Anciennitet}} = -0,0005$, således at den forventede belåningsgrad falder med 0,0005, hvis låntagers anciennitet stiger med 1 måned.

Resultatet af den estimerede model:

Summary of Fit									
RSquare	0,422894								
RSquare Adj	0,422417								
Root Mean Square Error	0,235583								
Mean of Response	0,698592								
Observations (or Sum Wgts)	23007								

Analysis of Variance									
Source	DF	Sum of Squares	Mean Square	F Ratio					
Model	19	934,8571	49,2030	886,5511					
Error	22987	1275,7634	0,0555		Prob > F				
C. Total	23006	2210,6205			<,0001*				

Parameter Estimates									
Effect Tests									
Indicator Function Parameterization									
Term	Estimate	Std Error	t Ratio	Prob> t	Lower 95%	Upper 95%			
Intercept	0,5325393	0,014932	35,67	<,0001*	0,5032722	0,5618064			
MaksimalLøbetid	0,0142867	0,000325	43,91	<,0001*	0,0136489	0,0149245			
MulighedForAfdragsfrihed[Nej]	-0,170229	0,003405	-50,00	<,0001*	-0,176902	-0,163556			
Landsdel[Bornholm]	0,1596243	0,019182	8,32	<,0001*	0,1220272	0,1972214			
Landsdel[Fyn]	0,074416	0,008093	9,20	<,0001*	0,0585539	0,090278			
Landsdel[København by]	-0,097283	0,008762	-11,10	<,0001*	-0,114457	-0,080109			
Landsdel[Københavns omegn]	-0,03322	0,007886	-4,21	<,0001*	-0,048677	-0,017764			
Landsdel[Nordjylland]	0,004698	0,000843	0,55	0,5812	-0,011874	0,0211733			
Landsdel[Nordsjælland]	0,0547785	0,007806	7,02	<,0001*	0,0394786	0,0700784			
Landsdel[Sendjylland]	0,0144682	0,007874	1,84	0,0661	-0,000965	0,0299011			
Landsdel[Vest- og Sydjylland]	0,1607531	0,007573	21,23	<,0001*	0,1459092	0,1755971			
Landsdel[Vestjylland]	0,001123	0,008931	0,13	0,8999	-0,016383	0,018629			
Landsdel[Østjylland]	-0,037446	0,007835	-4,78	<,0001*	-0,052804	-0,022088			
Ejendomstype[Andet]	0,1555172	0,011141	13,96	<,0001*	0,1336796	0,1773548			
Ejendomstype[Ejerlejlighed]	0,2094777	0,00787	26,62	<,0001*	0,1940511	0,2249043			
Ejendomstype[Parcelhus]	0,1636124	0,005814	28,14	<,0001*	0,1522157	0,1750091			
AntalInteressenter	0,0124961	0,001547	8,08	<,0001*	0,0094639	0,0155282			
GnsAlder	-0,005909	0,000134	-43,97	<,0001*	-0,006173	-0,005646			
Anciennitet	-0,000471	1,876e-5	-25,11	<,0001*	-0,000508	-0,000434			
TidligereIRestance[Ja]	0,04287	0,003999	10,72	<,0001*	0,0350324	0,0507076			

9c): Parameterestimatene hørende til variablen...

- **MulighedForAfdragsfrihed** er $\hat{\beta}_{\text{MulighedForAfdragsfrihed,Nej}} = -0,1702$, hvis lånet ikke rummer mulighed for afdragsfrihed, og $\hat{\beta}_{\text{MulighedForAfdragsfrihed,Ja}} = 0$, hvis lånet rummer mulighed for afdragsfrihed. Den forventede belåningsgrad falder således med 0,1702, såfremt lånet ikke giver mulighed for afdragsfrihed.
- **TidligereIRestance** er $\hat{\beta}_{\text{TidligereIRestance,Ja}} = 0,0429$, hvis låntagerne tidligere har været i restance, og $\hat{\beta}_{\text{MulighedForAfdragsfrihed,Nej}} = 0$, hvis låntagerne ikke tidligere har været i restance. Den forventede belåningsgrad stiger således med 0,0429, såfremt låntagerne tidligere har været i restance (f.eks. som følge af at låntagerne pga. restancen har fået forlænget løbetiden på lånet, har fået udskudt et eller flere afdrag, at der er påløbet ekstra renter o.lign.).

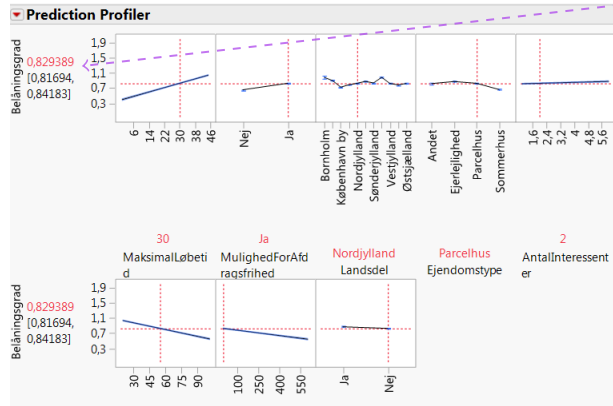
9d): Et 95%-konfidensinterval for betydningen af muligheden af afdragsfrihed er $[0,1636; 0,1769]$. Med 95% sandsynlighed vil den forventede belåningsgrad dermed stige med mellem 0,1636 og 0,1769, såfremt lånet giver mulighed for afdragsfrihed.

9e): Et 95%-konfidensinterval for betydningen af anciennitet er $[-0,0005; -0,0004]$. Med 95% sandsynlighed vil den forventede belåningsgrad dermed falde med mellem 0,0005 og 0,0004, såfremt låntagernes anciennitet stiger med 1 måned. Dermed vil den forventede belåningsgrad falde med mellem $0,0005 \cdot 5 = 0,0025$ og $0,0004 \cdot 5 = 0,0020$, såfremt låntagernes anciennitet stiger med 5 måneder.



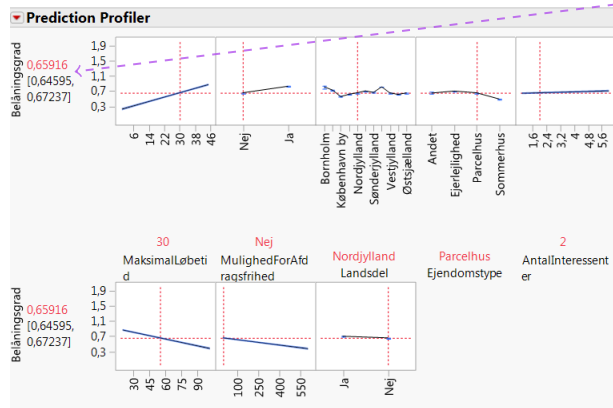
9f): Den forventede belåningsgrad på det i opgaveteksten angivne realkreditlån er⁷:

$$0,5325 + 0,0143 \cdot 30 + 0 + 0,0046 + 0,1636 + 0,0125 \cdot 2 - 0,0059 \cdot 55 - 0,0005 \cdot 0 = 0,8294$$



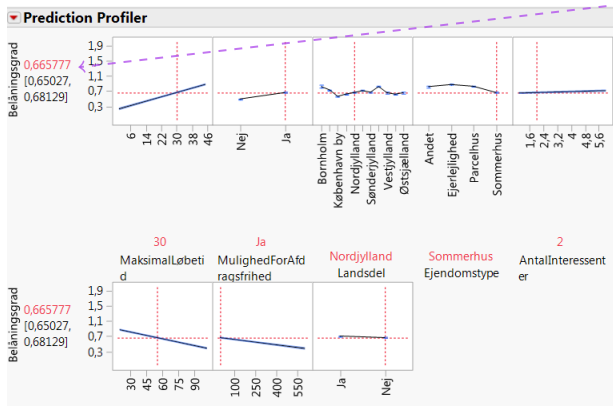
9g): Den forventede belåningsgrad, såfremt lånet i delspg. f) ikke rummer mulighed for afdragsfrihed, er:

$$0,5325 + 0,0143 \cdot 30 - 0,1702 + 0,0046 + 0,1636 + 0,0125 \cdot 2 - 0,0059 \cdot 55 - 0,0005 \cdot 0 = 0,6592$$



Den forventede belåningsgrad, såfremt lånet i delspg. f) i stedet er knyttet til et sommerhus, er:

$$0,5325 + 0,0143 \cdot 30 + 0 + 0,0046 + 0 + 0,0125 \cdot 2 - 0,0059 \cdot 55 - 0,0005 \cdot 0 = 0,6658$$



⁷Regressionsestimaterne er for læsbarhedens skyld afrundet til 4. decimal, hvilket - hvis man regner efter - giver en mindre numerisk afvigelse i forhold til det i JMP angivne output.



OPGAVE 10 (► Løsning i JMP)

Spørgsmålet om der er sammenhæng mellem muligheden for afdragsfrihed og hvorvidt et realkreditlån efterfølgende misligholdes eller ej kan eksempelvis undersøges ved at teste for, om andelen af misligholdte lån er den samme for lån hhv. med og uden mulighed for afdragsfrihed.

Test af nulhypotese om at den forventede andel af misligholdte lån er den samme blandt lån med og uden mulighed for afdragsfrihed:

By MulighedForAfdagsfrihed				
Contingency Table				
		Misligholdelse		
		Ja	Nej	Total
MulighedForAfdagsfrihed	Nej	34	9366	9400
	Expected	0,15	40,71	40,86
	Count	36,17	40,88	
	Col %	0,36	99,64	
Ja	Nej	38,4057	9361,59	
	Expected	60	13547	13607
	Count	0,26	58,88	59,14
	Col %	63,83	59,12	
Total	Nej	0,44	99,56	
	Expected	55,5943	13551,4	
	Count	94	22913	23007
	Col %	0,41	99,59	

Two Sample Test for Proportions				
Description	Proportion	Difference	Lower 95%	Upper 95%
P(Ja Nej)-P(Ja Ja)		-0,00079	-0,00243	0,000906
Adjusted Wald Test		Prob		
P(Ja Nej)-P(Ja Ja) ≥ 0		0,8143		
P(Ja Nej)-P(Ja Ja) ≤ 0		0,1857		
P(Ja Nej)-P(Ja Ja) = 0		0,3714		

Antagelser:

Observationerne af om et lån er misligholdt eller ej er indbyrdes uafhængige. Endvidere er det forventede antal observationer i hver af krydstabellens celler større end 5. Sidstnævnte antagelse er opfyldt, da alle celler har et forventet antal observationer på 38 eller mere.

Hypoteser:

$$H_0 : p_{med} = p_{uden}$$

$$H_a : p_{med} \neq p_{uden}$$

Signifikansssh.:

$$p = 0,3714$$

Tekn. konklusion:

Signifikanssandsynligheden er større end 0,05 og dermed kan vi ikke forkaste H_0 .

Let konklusion:

Vi kan ikke afvise hypotesen om, at andelen af misligholdte lån er den samme blandt lån med og uden mulighed for afdragsfrihed, og der er ikke belæg for en påstand om, at muligheden for afdragsfrihed påvirker sandsynligheden for efterfølgende at misligholde sit lån.