*Spørgsmål 1.1 (25 %)*

*Opstil en model der kan forklare kunders placering i rating-systemet. Alle forklarende variable skal være signifikante på 5 % niveau.*

*Fortolk dine resultater og vurdér i hvilken grad modellen er brugbar for Nykøbing Bank.*

Alder målt som kundens alder i år

Bruttoindkomst pr. år målt i kr.

Kundeanciennitet målt som antal år kunden har været kunde i banken

Om kunden har haft overtræk (0 = ikke overtræk og 1 = overtræk)

Om kunden bor i ejerbolig (0 = ikke ejerbolig og 1 = ejerbolig)

Startmodellen bliver:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Summary of fit** |  |  |  |
| R-kvadreret | 0.662038 |  |  |  |
| R-kvadreret justeret | 0.585229 | 2.10092204 |  |  |
| Standardfejl | 2.22 | 7.323989705 |  |  |
| Middel af respons | 7.571428571 | 16.65262877 |  |  |
| Observationer | 28 |  |  |  |
| **Analyse af varians** |  |  |  |  |
| Source | DF | Sum af kvadrater | Mean square | F-teststørrelse |
| Model | 5 | 212.420 | 42 | 8.6192 |
| Error SSE/SSR | 22 | 108.437 | 5 | p-værdi F-test fuld model |
| Total SST | 27 | 320.857 |  | 0.01% |
|  |  |  |  |  |
| **Parameter estimater** | Estimater | Std. Fejl | t-teststørrelse | p-værdi koefficienter |
| Skæring | 12.02 | 2.0216 | 5.9435 | 0.00056% |
| Alder | 0.00 | 0.0247 | -0.1087 | 91.44066% |
| Bruttoindkomst | 0.000 | 0.0000 | -1.8842 | 7.28229% |
| Kundeanciennitet | -0.14 | 0.0470 | -2.8764 | 0.87672% |
| Overtræk | 2.7350 | 1.1473 | 2.3838 | 2.62% |
| Boligejer | -2.1080 | 0.9767 | -2.1583 | 4.21% |

Vi ser af F-testet at mindst en forklarende variabel har altså effekt på responsvariablen.

Vi reducerer nu trinvist i modellen ved at fjerne forklarende variable der ikke er signifikante.

Da p-værdien er 91.44% har den forklarende variabel ikke effekt og fjernes derfor fra modellen.

Vi fjerner insignifikante variabel fra modellen, og når frem til følgende slutmodel

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Summary of fit** |  |  |  |
| R-kvadreret | 0.576532 |  |  |  |
| R-kvadreret justeret | 0.542654 | 2.10092204 |  |  |
| Standardfejl | 2.33 | 3.146732835 |  |  |
| Middel af respons | 7.571428571 | 12.9424498 |  |  |
| Observationer | 28 |  |  |  |
| **Analyse af varians** |  |  |  |  |
| Source | DF | Sum af kvadrater | Mean square | F-teststørrelse |
| Model | 2 | 184.984 | 92 | 17.0181 |
| Error SSE/SSR | 25 | 135.873 | 5 | p-værdi F-test fuld model |
| Total SST | 27 | 320.857 |  | 0.00% |
|  |  |  |  |  |
| **Parameter estimater** | Estimater | Std. Fejl | t-teststørrelse | p-værdi koefficienter |
| Skæring | 9.61 | 1.1165 | 8.6025 | 0.00000% |
| Kundeanciennitet | -0.16 | 0.0459 | -3.3980 | 0.22784% |
| Overtræk | 2.421 | 1.1713 | 2.0668 | 4.92605% |

Slutmodellen bliver:

Vi ser at når kundeanciennitet vokser med 1 år, falder ratingen med 0.16. Dette giver god mening jo ældre kunder jo mere stabile og sikre.

Når en kunde har haft overtræk vokser ratingen med 2.421. Hvilket giver god mening da kunder med overtræk må forventes at være mere risikable for banken.

En helt nu kunde, uden overtræk vil ifølge modellen have en rating på 9.61.

54% af variationen i responsvariablen rating, forklares ved variationen i de 2 forklarende variable.

46% af variationen forklares altså ikke af modellen. Dog er begge variable signifikante så vi kan bruge modellen, men den ville forbedres hvis man tilføjede yderligere forklarende variable.

*Spørgsmål 1.2 (5 %)*

*Beregn ved hjælp af din signifikante model fra spørgsmål 1.1 hvilken rating nedenstående kunde vil opnå:*

*Alder 34 år Bruttoindkomst 580.000 kr. Har været kunde i banken i 3 år Har ikke haft overtræk Bor i ejerbolig.*

Kundens rating vil ifølge modellen være 9.13.

Opgave 2 (30 %)

Nykøbing Bank er bekymret for kundernes overtræk. I 2015 foretog banken en analyse, hvor man i en stikprøve på 75 kunder registrerede 26 kunder med overtræk.

Spørgsmål 2.1 (10 %)

Du bedes beregne et punktestimat og et 95 % konfidensinterval for andelen af bankens kunder, der havde overtræk i 2015.

Vi benytter test af en andel, forudsætningen for approximation med normalfordelingen er opfyldt.

|  |  |  |  |
| --- | --- | --- | --- |
| Succeser | Stikprøve str. | Stikprøve andel | n·p·(1-p)>9 Forudsætningen er opfyldt |
| 26 | 75 | 0.3467 | 16.9867 |
| Nedre grænse 95 KI andelen p | |  | Fejlmargin |
| 0.2390 |  |  | 10.77% |
| Øvre grænse 95 KI andelen p | |  |  |
| 0.4544 |  |  |  |

Punktestimatet bliver  . Dette er stikprøveandelen hvilket er vores estimat på andelen i populationen.

Vi kan med 95% sikkerhed sige at andelen af kunder med overtræk i populationen ligger mellem 23.9% og 45.44%.

Spørgsmål 2.2 (10 %)

Ved hjælp at en test på 5 % testniveau bedes du undersøge, om andelen af kunder med overtræk var større end 25 % i 2015.

Da vi forkaster nulhypotesen. Kan vi konkludere at andelen af kunder med overtræk var større end 25% i 2015.

|  |  |  |  |
| --- | --- | --- | --- |
| Et-sidet alternativ hypotese | |  |  |
| p-værdi | 2.66% |  |  |
| H0 | p ≤ 0.25 |  |  |
| HA: | p > 0.25 |  |  |
| Da p værdien/signifikanssandsynligheden 0.0266 < 0.05 signifikansniveauet, afviser vi nulhypotesen | | | |
|

Spørgsmål 2.3 (10 %)

Banken har en formodning om, at andelen af kunder med overtræk er faldet siden 2015. Anvend resultaterne i den vedlagte stikprøve fra maj 2016 til at teste på 5 % testniveau, om banken har ret i denne formodning.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Stikprøve 1 | |  |  |  |  |
|  | Succeser | Stikprøve str. | Stikprøve andel punktestimat | | n·p·(1-p)>9 Forudsætningen er opfyldt |
|  | 26 | 75 | 0.346666667 |  | 16.9867 |
| Stikprøve 2 |  |  |  |  |  |
|  | Succeser | Stikprøve str. | Stikprøve andel | | (1-p)>9 Forudsætningen er ikke opfyldt |
|  | 8 | 28 | 0.285714286 |  | 5.7143 |

Bemærk forudsætningen for approximation med normalfordelingen er ikke opfyldt for 2016 stikprøven.

Da p-værdien 27.92% er større end 5% signifikansniveauet kan vi ikke afvise nulhypotesen. Bankens formodning er ikke underbygget, da andelen af kunder med overtræk ikke er faldet.

|  |  |  |  |
| --- | --- | --- | --- |
| Et-sidet alternativ hypotese | | |  |
| Prob>z = | 27.918370% |  |  |
| H0 | p1 - p2 ≤ D |  |  |
| HA: | p1 - p2 > D |  |  |
| Da p værdien/signifikanssandsynligheden 0.2792 > 0.05 signifikansniveauet, kan vi ikke afvise nulhypotesen. | | | |
|

*Opgave 3 (30 %)*

*En række klagesager i Nykøbing Bank har fået en del kunder til at forlade banken. Det er vigtigt for Nykøbing Bank at fastholde sine kunder i lokalområdet i mange år. Banken ønsker derfor en nærmere analyse af kundeancienniteten.*

*Spørgsmål 3.1 (5 %)*

*Undersøg om kundeancienniteten i banken er normalfordelt.*

Det er ud fra normalfraktildiagrammet rimeligt at antage at kundeancienniteten i populationen er normalfordelt.



*Spørgsmål 3.2 (10 %)*

*Test på 5 % testniveau om den gennemsnitlige kundeanciennitet er under 20 år.*

Da p-værdien 13.14% er større end signifikansniveauet på 5% kan vi ikke forkaste nulhypotesen. Kundeancienniteten er altså ikke mindre en 20 år.

|  |  |
| --- | --- |
| Et-sidet alternativ hypotese |  |
| p-værdi | 13.14% |
| H0 | μ ≥ 20 |
| HA: | μ < 20 |
| Da p værdien/signifikanssandsynligheden 0.1314 > 0.1 signifikansniveauet, kan vi ikke afvise nulhypotesen. | |
|

*I undersøgelsen fra 2015 var kundeancienniteten på 20 år med en standardafvigelse på 10 år.*

*Spørgsmål 3.3 (5 %)*

*Test på 5 % testniveau om standardafvigelserne er ens i de 2 undersøgelser.*

|  |  |  |  |
| --- | --- | --- | --- |
| Test af varianshomogenitet | |  |  |
| Teststatisic | 1.3767 |  |  |
| p-værdi | 28.34% |  |  |
| H0 | σ21 = σ22 |  |  |
| HA: | σ21 ≠ σ22 |  |  |
| Da p værdien/signifikanssandsynligheden 0.2834 > 0.05 signifikansniveauet, kan vi ikke afvise nulhypotesen, varianserne er ens | | | |
|
|
| Forudsætning for test for varianshomogenitet er at begge populationer er normalfordelte | | | |

Vi ved 2016 kundeancienniteten kan antages at være normalfordelt. Vi har ikke rådata fra 2015 undersøgelsen, men forudsætter populationen er normalfordelt.

Da p-værdien 28.34% hvilket er større end signifikansniveauet, kan vi ikke forkaste nulhypotesen. Standardafvigelserne er altså ens.

*Spørgsmål 3.4 (10 %)*

*Test på 5 % testniveau om den gennemsnitlige anciennitet er faldet fra 2015 til 2016.*

|  |  |  |
| --- | --- | --- |
| Et-sidet alternativ hypotese | |  |
| p-værdi | 13.89% |  |
| H0 | μ1 -μ2 ≥ D |  |
| HA: | μ1 -μ2 < D |  |
| Da p værdien/signifikanssandsynligheden 0.1389 > 0.05 signifikansniveauet, kan vi ikke afvise nulhypotesen. | | |
|

Vi har en signifikanssandsynlighed på 13.89% hvilket er større end 5% signifikansniveauet vi kan altså ikke forkaste nulhypotesen.

Den gennemsnitlige anciennitet er altså ikke faldet fra 2015 til 2016.

*Opgave 4 (10 %)*

*Det er mange år siden bankens markedschef har beskæftiget sig med statistik. Han beder dig derfor forklare nogle begreber.*

*Spørgsmål 4.1 (5 %)*

*Forklar forskellen på et punktestimat og et konfidensinterval.*

*Forklar endvidere formålet med at beregne et konfidensinterval.*

Et punktestimat er et tal der angiver vort bedste gæt på den sande parameter i populationen.

Et konfidensinterval er en nedre og øvre grænse indenfor hvilket den sande parameter i populationen befinder sig med en bestemt sandsynlighed (ofte 95% konfidensniveau).

Formålet med konfidensintervallet er at angive præcisionen for punktestimatet.

*Spørgsmål 4.2 (5 %)*

*Hvilken betydning har testniveauet for konklusionen i en hypotesetest?*

*Hvilken betydning har stikprøvens størrelse for p-værdien, når der udarbejdes en hypotesetest?*

Testniveauet angiver hvornår vi forkaster nulhypotesen, et 1% signifikansniveau betyder at vi skal have stærkere beviser for at forkaste nulhypotesen, i forhold til et 5% signifikansniveau.

Jo større stikprøve jo mere ekstrem p-værdi dvs. tydeligere konklusion.