**HT 2020 Inlämningsuppgift i Undersökningsmetodik, dagtid**

|  |
| --- |
|  |

Övningslärare: Maria Anna Di Lucca

Övningsgrupp: Raum B2

Inlämningsdatum: 7 januari, kl. 17:00

Gruppmedlemmar:

|  |  |  |  |
| --- | --- | --- | --- |
| **För- och efternamn** | **Pnr** | **Mobilnr** | **E–mail** |
| Leo Söderberg | 980114–0717 | 0793101331 | leo.soderberg@gmail.com |
| Michaela Machacny | 950322–6947 | 0700596078 | michaela247@hotmail.com |
| Moa Dahlin | 840521–3904 | 0702703207 | moadahlin@gmail.com |
| Sara Standert | 790330–0205 | 0701088429 | sara.standert@gmail.com |

*Övningslärarens anteckningar*:

**Resultat efter första rättningen**:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| □ | Godkänt | □ | Komplettering | Lärarens signatur: ……………………… |

*OBS! Vid eventuell komplettering skall ursprungsuppgiften med försättsblad bifogas*

Kommentarer (vid komplettering):

**Inledning och ämnesbeskrivning**

Valt ämne: Nummer 10 – Restauranger, caféer, hotell. Vi har begränsat vår frågeställning till att gälla hotell för att minska omfånget av undersökningen.

**Uppdragsgivare**

Hotell AB franchisekedja ska öppna ett nytt hotell och är intresserade av att veta vad de kan ta betalt för ett hotellrum på semesterort A, beroende på avståndet till diverse objekt.

Vi antar att samtliga hotell som medverkar i undersökningen har likvärdig standard och likvärdiga rum.

**Undersökningsfrågor**

För att kunna undersöka vilket pris på hotellrummen som är lämpligt för det nya hotellet har vi valt att undersöka priset av hotellrum för likvärdiga hotell/hotellrum inom Franchise-koncernen i relation till avståndet för olika relevanta objekt.

Våra undersökningsfrågor är: “Vad kan man ta betalt för ett hotellrum per natt vid semesterort A?” samt “Vilka faktorer påverkar priset”.

För att få en uppfattning om hur priset skiljer sig beroende på placering har vi valt att undersöka olika hotellpriser beroende på följande faktorer: avstånd till kollektivtrafik, sevärdheter, restauranger samt huruvida hotellen ligger centralt eller icke-centralt. Detta är några faktorer som tros kunna påverka efterfrågan och därmed priset på hotellrum.

För att illustrera resultatet planerar vi att göra konfidensintervall, histogram för priset, scatterplots för variablerna i intervallskala samt parvisa stapeldiagram för variablerna i nominalskala. Punktskattning av medelvärdet och standardavvikelsen för beroendevariabeln, samt frekvenstabeller för samtliga oberoende variabler.

* Kollektivtrafik definieras som offentliga transportmedel, däribland buss, tåg, spårvagn och metro.
* Med sevärdheter menas aktiviteter som lockar turister.
* Med restaurangområde menas ett område med fem eller fler restauranger.
* Centralt och icke-centralt definieras enligt semesterorternas stadsgränser.

**Metod**

Målpopulationen är samtliga hotell likvärdiga det påtänkta hotellet, men då det kan vara svårt att få tillförlitliga svar på våra frågor från konkurrenter utgörs ramen, och därmed urvalet för undersökningen, av hotell inom samma franchisekedja. Ramen innehåller 150 hotell på semesterorter likvärdiga semesterort A och urvalet kommer att bestå av 50 hotell för att skapa en översiktlig prisbild.   
  
**Rampopulationen** innefattar samtliga hotell inom franchisekedjan belägna på semesterorter.

**Målpopulationen** är samtliga hotell på semesterorter likvärdiga semesterort A.  
  
Urvalet sker genom gruppurval då vi enbart undersöker hotell inom franchisekedjan. Och därmed har övriga hotell på semesterorterna ingen sannolikhet att komma med i urvalet. Problematiken med ett urval som enbart ingår i en enda franchisekedja är att resultatet blir svårt att generalisera samt att resultatet kan vi skevt.

Valet av stratifierat OSU grundar sig på att det tros vara en skillnad på prissättningen beroende på om hotellet ligger i centrum eller utanför. Därav kommer det att ses till att urvalet är proportionerligt utifrån andelen hotell som ligger centralt/icke-centralt.

Kvotskattning är inte lämpligt i det här fallet då det är orimligt att priset för ett hotellrum går genom origo. Om det mot förmodan skulle gå att få information om ett faktiskt genomsnittligt pris på ett hotellrum i exempelvis ett specifikt område, skulle en regressionsskattning vara en möjlighet (Nyquist, 2020).

Undersökningen kommer ske via webbenkät då det anses vara kostnadseffektivt och smidigt att skicka ut samt för mottagaren att besvara. Enkäten kommer adresseras till hotellchefen över e-post då denne anses kunna besvara undersökningen korrekt. Då mejlet kommer att skickas ut via det interna systemet förväntas svarsfrekvensen vara hög (Dahmström, 2011).

Frågeformuläret kommer inledas med en introduktion där bakgrunden till undersökningen redovisas. Frågorna följer en ordning där det först kommer frågor direkt kopplade till priset på hotellrum som följs av frågor rörande faktorer som indirekt kan påverka hotellrummets pris. Svarsalternativen består antingen av kryssfrågor eller frågor där mottagaren själv får fylla i pris och avstånd i kilometer. Vi har valt att inte ha de frågorna i färdiga klasser i enkäten, och istället klassindela efteråt, för att undvika att kategorisera i klasser som inte hade uppkommit naturligt (Dahmström, 2011).  
  
Variablerna “Placering” och “Frukost” är dummyvariabler och därmed kodade antingen 1 eller 0 beroende på om hotellet har egenskapen eller inte. “Centralt” och “Frukost ingår” innebär att hotellet har egenskapen (1) och “Icke-centralt” och “Frukost ingår ej” innebär att hotellet inte har egenskapen (0).

**Metodkritik**

Partiellt bortfall beräknas kunna undvikas på frågorna gällande “ange kilometer/avstånd” då hotellen har möjlighet att via en karta undersöka avståndet. Därav finns inget “Vet ej”-alternativ på de frågorna. Dock kan mätfel uppstå om respondenter uppger avståndet i exempelvis meter istället för kilometer. Det kan även förekomma att respondenter väljer att inte kolla upp det korrekta avståndet och istället gissar eller lämnar blankt. Det finns risk för att respondenter avstår från att svara på hela enkäten och att objektsbortfall uppstår.

Det kommer att uppstå undertäckning då enkäten enbart skickas ut till hotell inom franchise-organisationen som kan hända inte representerar alla hotell i målpopulationen. Detta kan medföra en snedvridning av resultatet och det “sanna rimliga priset” kan komma att skilja sig från resterande hotell på semesterorterna. Det kan även finnas en risk för övertäckning då hotell som inte har likvärdiga förhållanden av misstag inkluderas i urvalet.

Då undersökningen sker via webbenkät minskas risken för bearbetningsfel eftersom svaren automatiskt kan överföras till exempelvis en Excel fil. Det innebär att risken för felregistrering från undersökarens sida elimineras. Dock kan exempelvis “buggar” uppstå vid överföringen som kan medföra fel i Excel filen. Det kan även uppstå fel vid bearbetningen av datan vid exempelvis bearbetning via exempel SAS, där felkodning kan uppstå.

**Fråge- och variabelöversikt**

|  |  |  |  |
| --- | --- | --- | --- |
| ***Undersökningsfrågor*** | ***Frågor i frågeformuläret*** | ***Motsvarande variabel per fråga i frågeformuläret*** | ***Plan för “svar per undersökningsfråga”*** |
| Vad kan man ta betalt för ett hotellrum per natt vid semesterort A? | Fråga 1:  Vad tar ni betalt för ett standard dubbelrum? (Inklusive moms) | Pris:   * Undersökningsvariabel * Kvotskala * Modell: kvantitativ kontinuerlig variabel * Data: kontinuerlig med kronor som enhet * Utfallsrum för data:  > 0 kr | Klassindela i efterhand med kronor som enhet. Histogram med avseende på variabeln “Pris”.  Punktskattning av medelvärdet samt standardavvikelsen.   Konfidensintervall (intervallskattning) med avseende på variabeln “Pris”. |
| Vilka faktorer påverkar priset  (i beskrivande syfte)? | Fråga 2:  Ingår frukost i ovan angivet rumspris? | Frukost:   * Bakgrundsvariabel * Nominalskala * Kategorisk * Kodad: Ingår (1)   Ingår ej (0) | Stapeldiagram för variabeln “Frukost”.  Histogram där variabeln “Pris” är betingad på variabeln “Frukost”.  Punktskattning: andel “Ingår” respektive “Ingår ej”.   Frekvenstabell för “Pris” betingat på “Frukost”. |
|  | Fråga 3:  Ligger hotellet centralt eller inte centralt? (Enligt ortens gränser.) | Placering av hotellet:   * Bakgrundsvariabel * Nominalskala * Kategorisk * Kodad: Centralt (1) Icke-central (0) | Stapeldiagram för variabeln “Placering”.  Histogram där variabeln “Pris” är betingad på variabeln “Placering”.  Punktskattning: andel “Centralt” respektive “Icke-centralt”.  Frekvenstabell för “Pris” betingat på “Placering”. |
|  | Fråga 4:  Inom vilket avstånd finns kollektivtrafik? Ange i kilometer. | Hotellets närhet till kollektivtrafik:   * Bakgrundsvariabel * Kvotskala * Modell: kvantitativ kontinuerlig variabel * Utfallsrum för data:  > 0 kilometer | Scatterplott med “Pris” som beroende variabel och “Avstånd till kollektivtrafik” som förklaringsvariabel.  Klassindela i efterhand med meter som enhet.  Frekvenstabell för “Pris” betingat på “Avstånd till kollektivtrafik”. |
|  | Fråga 5:  Inom vilket avstånd finns närmaste sevärdheter? Ange i kilometer. | Hotellets avstånd till sevärdheter:   * Bakgrundsvariabel * Kvotskala * Modell: Kvantitativ kontinuerlig variabel * Utfallsrum för data:  > 0 kilometer | Scatterplott med pris som beroende variabel och “Avstånd till kollektivtrafik” som förklaringsvariabel.  Klassindela i efterhand med meter som enhet.  Frekvenstabell för “Pris” betingat på “Avstånd till sevärdheter”. |
|  | Fråga 6:  Inom vilket avstånd finns närmaste restaurangområde? Ange i kilometer. | Hotellets avstånd till restaurangområde:   * Bakgrundsvariabel * Kvotskala * Modell: Kvantitativ kontinuerlig variabel * Utfallsrum för data:  > 0 kilometer | Scatterplott med “Pris” som beroende variabel och “Avstånd till restaurangområde” som förklaringsvariabel.  Klassindela i efterhand med meter som enhet.  Frekvenstabell för “Pris” betingat på “Avstånd till restaurangområde”. |

**Avslutning**

Undersökningen syftar till att skapa en bättre förståelse för vad som är prissättningen beroende på diverse förutsättningar för ett hotell på semesterorten. Målet med undersökningen är att skapa bra förutsättningar för prissättningen av rummen hos det nya hotellet.

Målet förväntas nås genom att vi skickar ut en webbenkät till likvärdiga hotell, i franchise-organisationen, som får svara på frågor rörande prissättningen och externa faktorer som kan påverka priset. Detta analyseras sedan för att ge en bild av vad ett rimligt pris för ett hotellrum är givet olika faktorer.

Den största fördelen med undersökningen är att då enkäterna enbart skickas inom en befintlig koncern kommer bortfallet troligen vara litet. Den största nackdelen med undersökningen är att den enbart kommer att innehålla svar från hotell inom franchisekedjan. Detta då vi tror att chansen att få in tillförlitliga svar från konkurrenter, inte är troligt.

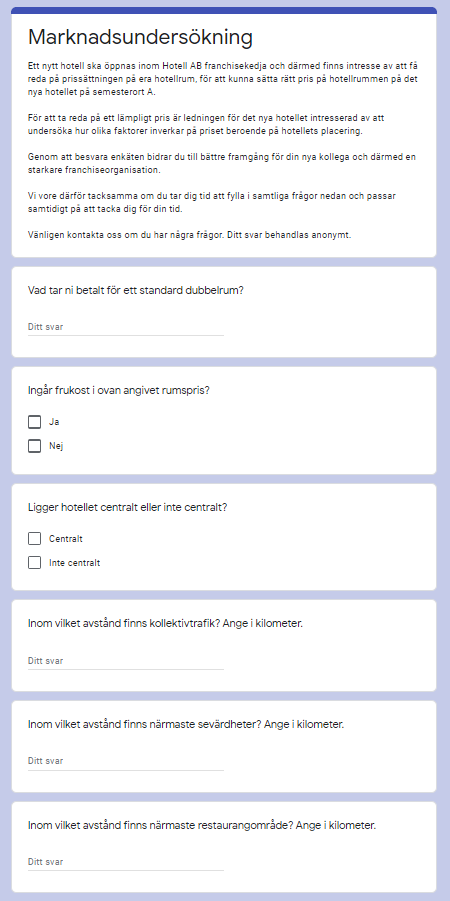
**Referensförteckning**

Dahmström, K. (2011). *Från Datainsamling till rapport – att göra en statistisk undersökning.* (Femte upplagan). Studentlitteratur.

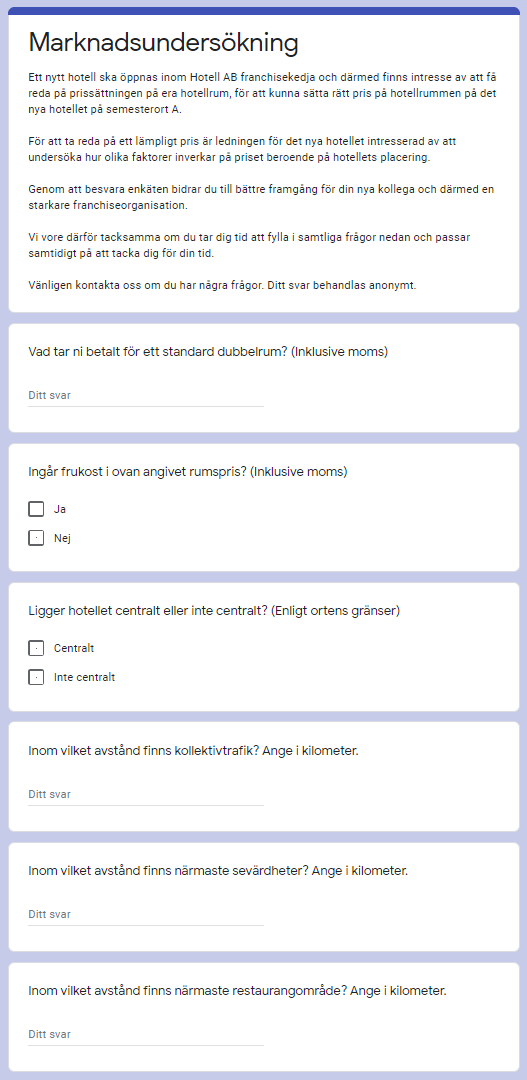
Nyquist, H. (2017). *Statistikens grunder – vetenskap, empiriska undersökningar och statistisk analys.* (Första upplagan). Studentlitteratur.

**Bilagor**

**Länk till enkäten:** [Marknadsundersökning Enkät](https://docs.google.com/forms/d/e/1FAIpQLSeCdvmp7IU41vKI6pEgJAfGlg9QzUh8Qhq-EUS3MazOFvOTwQ/viewform?usp=pp_url)  
Pilotstudieenkät före pilotundersökning:



Efter revidering:



Del B

**Obundet slumpmässigt urval, utan återläggning (OSU u.å.)**

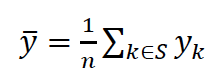
***Tabell 1: Slumpmässigt urval utan återläggning, frö=22***

|  |  |  |  |
| --- | --- | --- | --- |
| **Obs** | **Kommun** | **Kommunal Skattesats** | **Grupp** |
| **1** | Järfälla | 31.23 | 2 |
| **2** | Tyresö | 31.58 | 3 |
| **3** | Vaxholm | 31.86 | 3 |
| **4** | Vingåker | 33.5 | 1 |
| **5** | Katrineholm | 32.95 | 2 |
| **6** | Eskilstuna | 32.85 | 2 |
| **7** | Vadstena | 33.5 | 2 |
| **8** | Mjölby | 32.6 | 2 |
| **9** | Habo | 33.43 | 2 |
| **10** | Nässjö | 33.45 | 2 |
| **11** | Lessebo | 33.81 | 1 |
| **12** | Markaryd | 33.31 | 1 |
| **13** | Hultsfred | 33.28 | 1 |
| **14** | Vimmerby | 33.73 | 2 |
| **15** | Sölvesborg | 33.46 | 2 |
| **16** | Malmö | 31.93 | 1 |
| **17** | Hylte | 32.27 | 2 |
| **18** | Karlsborg | 32.8 | 2 |
| **19** | Vara | 33.25 | 2 |
| **20** | Lysekil | 33.94 | 2 |
| **21** | Ulricehamn | 32.53 | 2 |
| **22** | Hammarö | 33.3 | 3 |
| **23** | Forshaga | 33.55 | 2 |
| **24** | Sunne | 33.35 | 1 |
| **25** | Lindesberg | 33.85 | 2 |
| **26** | Hallstahammar | 32.69 | 2 |
| **27** | Smedjebacken | 34.09 | 2 |
| **28** | Borlänge | 34.05 | 2 |
| **29** | Bräcke | 34.59 | 1 |
| **30** | Robertsfors | 34.45 | 1 |

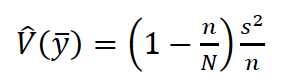
***Tabell 2: Punktskattning av genomsnittlig kommunalskatt i hela landet, varians, standardavvikelse och 95%-igt konfidensintervall (undre och övre gräns)***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Obs** | **Medelvärde Kommunal Skattesats** | **Varians** | **Standardavvikelse** | **Undre gräns** | **Övre gräns** |
| **1** | 33.172666667 | 0.019932 | 0.14118 | 32.8960 | 33.4494 |

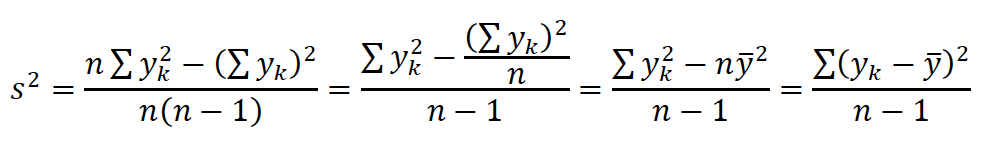
De olika värdena i tabellen beräknas enligt följande formler:  
Medelvärdet för kommunalskatt:



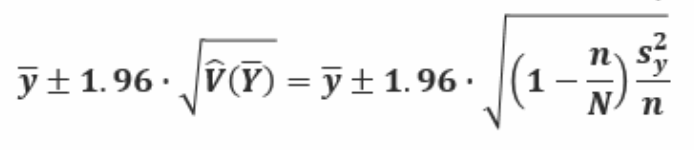
Skattad varians för medelvärdet:



Skattad varians för y:



Konfidensintervall 95%:



**Stratifierat urval**

Baserat på stratifieringsvariabeln inkomst definieras vilket stratum som kommunen tillhör. Det finns tre stratum:  
1 = låg medianinkomst (≤ 220,0 tkr)   
2 = mellan medianinkomst (220,0–260,0 tkr)   
3 = hög medianinkomst (≥ 260,0 tkr)

***Tabell 3: Stratifierat urvalgrupp 1, frö=22***

|  |  |  |
| --- | --- | --- |
| **Obs** | **Kommun** | **Kommunal Skattesats** |
| **1** | Flen | 33.1 |
| **2** | Östra Göinge | 31.68 |
| **3** | Bjuv | 31.68 |
| **4** | Perstorp | 31.5 |
| **5** | Eda | 33.45 |
| **6** | Skinnskatteberg | 33.39 |
| **7** | Rättvik | 33.45 |
| **8** | Nordmaling | 34.6 |
| **9** | Arjeplog | 34.84 |
| **10** | Överkalix | 34.14 |

***Tabell 4: Stratifierat urvalgrupp 2, frö=22***

|  |  |  |
| --- | --- | --- |
| **Obs** | **Kommun** | **Kommunal Skattesats** |
| **11** | Upplands-Bro | 31.68 |
| **12** | Norrtälje | 31.8 |
| **13** | Östhammar | 33.4 |
| **14** | Boxholm | 32.52 |
| **15** | Gislaved | 33.39 |
| **16** | Skurup | 31.11 |
| **17** | Hörby | 31.77 |
| **18** | Höör | 32.57 |
| **19** | Hjo | 33.05 |
| **20** | Falköping | 33.08 |
| **21** | Askersund | 33.5 |
| **22** | Nora | 33.8 |
| **23** | Kungsör | 32.91 |
| **24** | Hallstahammar | 32.69 |
| **25** | Mora | 33.97 |
| **26** | Umeå | 34.15 |

***Tabell 5: Stratifierat urvalgrupp 3, frö=22***

|  |  |  |
| --- | --- | --- |
| **Obs** | **Kommun** | **Kommunal Skattesats** |
| **27** | Tyresö | 31.58 |
| **28** | Trosa | 32.43 |
| **29** | Staffanstorp | 29.83 |
| **30** | Lomma | 30.33 |

***Tabell 6: Punktskattning av genomsnittlig kommunalskatt för grupp 1, variansstandardavvikelse och 95%-igt konfidensintervall (undre och övre gräns)***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Analysis Variable: Kommunal Skattesats** | | | | |
| **Medelvärde** | **Varians** | **Standardavvikelse** | **Undre gräns** | **Övre gräns** |
| 33.1830000 | 1.4680233 | 1.2116201 | 32.3162592 | 34.0497408 |

***Tabell 7: Punktskattning av genomsnittlig kommunalskatt för grupp 2, variansstandardavvikelse och 95%-igt konfidensintervall (undre och övre gräns)***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Analysis Variable: Kommunal Skattesats** | | | | |
| **Medelvärde** | **Varians** | **Standardavvikelse** | **Undre gräns** | **Övre gräns** |
| 32.8368750 | 0.7890362 | 0.8882771 | 32.3635455 | 33.3102045 |

***Tabell 8: Punktskattning av genomsnittlig kommunalskatt för grupp 3, variansstandardavvikelse och 95%-igt konfidensintervall (undre och övre gräns)***

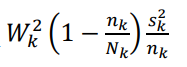
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Analysis Variable: Kommunal Skattesats** | | | | |
| **Medelvärde** | **Varians** | **Standardavvikelse** | **Undre gräns** | **Övre gräns** |
| 31.0425000 | 1.3972917 | 1.1820709 | 29.1615614 | 32.9234386 |

De olika värdena i tabellerna beräknas enligt följande formler:

Stickprovsmedelvärde för varje stratum =



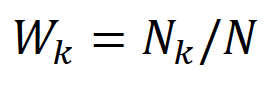
Stickprovsvarians för varje stratum =

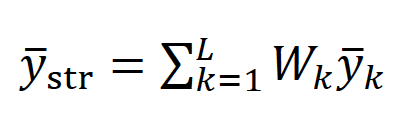


***Tabell 9: Sammanvägd skattning av genomsnittlig kommunalskatt, variansstandardavvikelse och 95%-igt konfidensintervall (undre och övre gräns) för samtliga strata.***

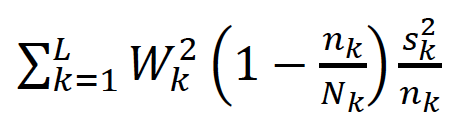
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Obs** | **Medelvärde** | **Varians** | **Standardavvikelse OSU u.å.** | **Undre gräns** | **Övre gräns** |
| **1** | 32.7435 | 0.032447 | 0.18013 | 32.3904 | 33.0966 |

Sammanvägd skattning av genomsnittliga kommunalskatten:

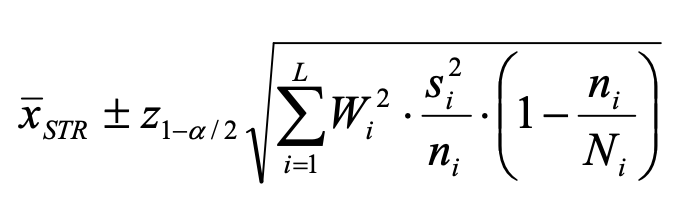




Sammanvägd variansskattning=



Konfidensintervall för sammanvägda strata:



**För- och nackdelar med de två olika urvalsmetoderna**

**OSU u.å.**

OSU är okomplicerat och teoretiskt sett enkelt att genomföra, vilket är en fördel vid stora populationer då ett generaliserbart resultat är eftersträvat. Men det krävs en ram över alla elementen, vilket inte alltid finns, och därmed blir precisionen sämre. Precisionen ökar inte heller proportionerligt när stickprovsstorleken ökar.  
  
Med ett OSU har varje element samma sannolikhet att bli valt, dock tar urvalsmetoden inte hänsyn till om (hur) populationen är heterogen och grupper med avvikande variabelvärden missas lättare.

**Stratifierat urval:**

Fördelar med stratifierat urval är att alla strata blir representerade och att precisionen oftast blir bättre än med OSU. Variansen blir mindre då precisionen blir bättre när strata är mer homogena vilket medför mindre stickprovsvarianser i varje stratum. Stratifierat urval kan därför vara att föredra vid snedfördelningar.

Det finns risk för att stratifieringen inte är entydig utan kan innebära olika resultat beroende på val av stratifieringsvariabel. Det är därför viktigt att ha stor kännedom om sin data vid användningen. Problem i planeringen kan vara att bedöma vilken stratifieringsvariabel som ska användas, hur många strata som behövs och vilka gränser som ska sättas mellan dem. Det vill säga efter vilka grunder indelningen i strata ska göras och hur allokeringen mellan strata ska göras. Det kan krävas subjektiva bedömningar för att kunna skapa homogena strata.

**Val av metod:**  
Differensen mellan den undre och övre gränsen för de två konfidensintervallen är för OSU: 0,5534 och för Stratifierat: 0,7062.  
Variansen för OSU u.å. (0,019932) är mindre än variansen för det sammanvägda stratifierade urvalet (0,032447) och för de enskilda strata (1.4680233, 0.7890362, 1.3972917). Detta kan bero på att urvalen inte innefattar samma kommuner (se tabell 1 och 3/4/5).

Utifrån de här siffrorna kan den bästa modellen tyckas vara OSU u.å. dock missas mönstret av att det är olika genomsnittlig kommunalskatt i förhållande till medianinkomst i de olika strata.

Kollar man enskilda medelvärden för varje stratum går det att se att det är högst medelskatt (33.1830) där inkomsten är lägst (≤ 220,0 tkr) och lägst medelskatt (31.0425) där inkomsten är högst (≥ 260,0 tkr).

Vårt val av metod faller därmed på stratifierat urval.

**Sammanfattning av resultaten**

Det har framkommit i tidigare text/delar att det är viktigt att se till helhetsperspektivet och inte enbart enskilda värden. Då vi sett att även om variansen för OSU var lägre och kunde anses vara det bästa valet, gav STR mer information som var relevant. I STR sågs skillnader mellan medianinkomst och skatt i kommunerna. Detta är förväntat då kommuner med hög andel låginkomsttagare behöver en högre skattesats för att finansiera samma offentliga utgifter som kommuner med många höginkomsttagare. Allmänt brukar STR ge en lägre varians, men att vår data visade motsatta siffror tros bero på att urvalet i OSU och STR inte består av samma kommuner. Det går att se att medelvärdet för kommunalskatten via OSU (33.172) och via det sammanvägda STR (32.743) visar på liknande resultat, vilket antyder att det sanna populationsmedelvärdet troligen ligger däromkring.

**SAS kod:**

