

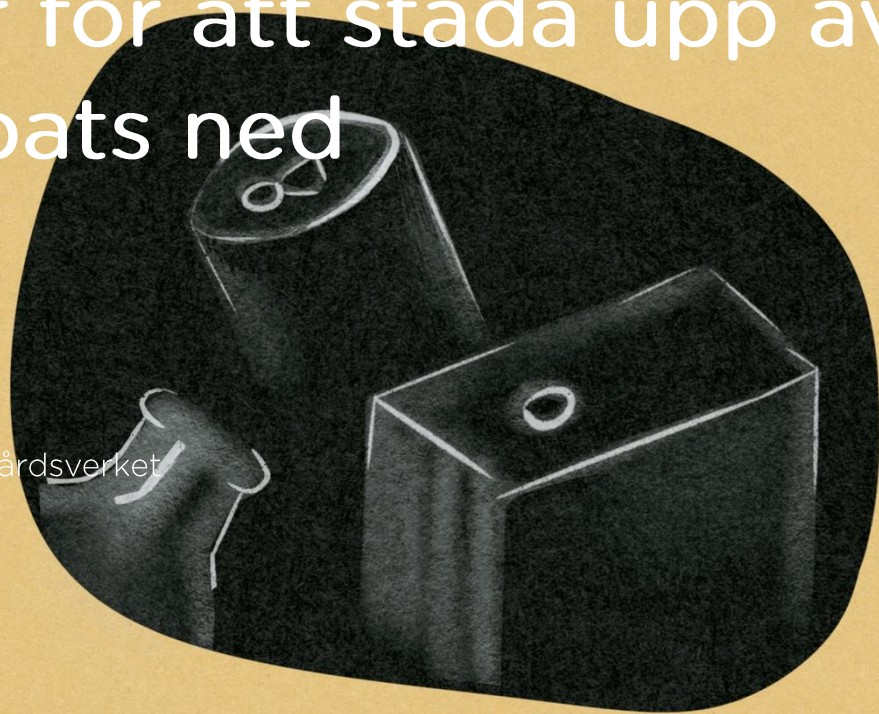
Maj 2025

Utveckling och uppdatering av bedömningsmodell för kommuners kostnader för att städa upp avfall som skräpats ned

Rapport beställd av Naturvårdsverket

RAMBOLL

Bright ideas.
Sustainable change.



Bedömningsmodell för kommuners nedskräpningskostnader – en översikt

För att kunna göra en rättvis och jämförbar bedömning av kommunernas kostnader för att städa upp avfall som skräpats ned har Ramboll utvecklat en prediktionsmodell. Modellen hjälper till att förstå och jämföra kostnader för nedskräpning mellan olika kommuner, baserat på deras specifika förutsättningar, exempelvis demografiska och geografiska faktorer.

Resultatet av modellen är ett förväntat kostnadsvärde (prediktionsvärde) per person (capita) för varje kommun, som sedan jämförs med kommunens inrapporterade kostnader för att städa upp avfall som skräpats ned. På så sätt ger modellen en tydlig referenspunkt för att bedöma om en kommuns rapporterade kostnader är för höga med hänsyn till olika faktorer som kan påverka kommunens kostnader för nedskräpning.

Vad prediktionen visar

- Prediktionsvärdet ger en **rimlig uppskattning** av vad en kommuns kostnader kan förväntas vara, givet dess förutsättningar.
- **Avvikelser** från detta värde kan signalera att de rapporterade kostnaderna är högre än förväntat. Naturvårdsverket använder därför prediktionsvärdet som stöd i bedömningen av om kostnaderna är för höga, med hänsyn till olika relevanta faktorer.

Så fungerar modellen

Modellen baseras på **inrapporterade kostnader** från 281 kommuner, där kostnaderna har räknats om till kostnad per capita för att möjliggöra jämförelser.

Genom en **statistisk modell** (multipel linjär regression) beräknas ett prediktionsvärde för varje kommun utifrån samtliga kommuners inrapporterade kostnader. Prediktionen baseras på ett antal bakgrundsfaktorer som syftar till att fånga strukturella egenskaper som kan förklara avvikande kostnader mellan kommuner.

Modellens bakgrundvariabler:



Demografiska faktorer:

Befolkningsstorlek, befolkningstäthet, antal människor som arbetar i kommunen men bor utanför den, antal fritidshus.



Geografiska faktorer: Kommunens yta, Kommungrupp enligt SKR:s indelning*.



Skräprelaterade faktorer: HSR kommunindex**, invånarnas upplevelse av nedskräpning (medborgarundersökning).

* Indelning utifrån vissa kriterier som t.ex. tätortsstorlek och pendlingsmönster.

** HSR kommunindex är ett verktyg framtaget av Håll Sverige Rent (HSR) som mäter hur väl Sveriges kommuner arbetar med att förebygga nedskräpning.

Bakgrund och syfte

Bakgrund

I november 2021 utfärdade regeringen förordningen (2021:1002) om nedskräpningsavgifter. Syftet med denna förordning är att minska negativ påverkan på miljön från nedskräpning av engångsprodukter i utomhusmiljöer genom att producenterna tar det ekonomiska ansvaret för hantering av avfall som skräpats ned och för spridning av information om nedskräpning.

Enligt förordningen ska företag som tillverkar, importerar eller säljer vissa produkter som riskerar att orsaka nedskräpning, betala en särskild avgift. Avgiften varierar beroende på vilken typ av produkt det rör sig om, och syftar till att återspegla produktens miljöpåverkan.

Den avgift som producenterna ska betala inom ramen för det utökade producentansvaret sätts utifrån inrapporterade kostnader från kommunerna för att städa upp avfall som skräpats ned, för föregående år. Nationella skräpmätningar genomförs för att bedöma den andel av de engångsplastprodukter som skräpats ned som berörs av förordningen, vilket sedan ligger till grund för kommunernas ersättning.

Rapporteringen utgår från fyra kostnadsposter som anges i 10 §

1. Kostnader som kommunen har haft föregående år för att städa upp avfall som skräpats ned på de platser som kommunen har renhållningsansvar för enligt lagen (1998:814) med särskilda bestämmelser om gaturenhållning och skyltning
2. Administration och planering av verksamhet
3. Insamling och rapportering av uppgifter till Naturvårdsverket
4. Informationsspridning kopplat till nedskräpning av engångsplastprodukter

Syfte med bedömningsmodellen

Ramboll har på uppdrag av Naturvårdsverket utvecklat en kvantitativ metod för att bedöma rimligheten i kommunernas inrapporterade kostnader för att städa upp avfall som skräpats ned. Metoden baseras på de fyra kostnadsposter som specificeras i 10 § i förordningen.

Genom att analysera variationen i kommunernas rapporterade kostnader och modellera dessa mot ett antal bakgrundsfaktorer, kan modellen identifiera avvikande värden. Resultatet används för att justera kostnader som väsentligt avviker från vad som kan förväntas givet kommunens förutsättningar. Detta ligger sedan till grund för Naturvårdsverkets förslag på fastställande av kostnader.

Modellen bidrar därmed till en mer systematisk och likvärdig bedömning av rimligheten i kommunernas inrapporterade kostnader. Den fastställer förväntade intervall för kostnader per capita och lyfter fram faktorer som kan förklara skillnader mellan kommuner.

Ramboll tog fram den första versionen av bedömningsmodellen under 2023

I denna metodrapport beskrivs en vidareutveckling av den modell som användes då. Syftet och utgångspunkten för modellerna är desamma, men metoden har justerats något mellan de två versionerna.

Bedömningsmodellen

Om bedömningsmodellen

I analysen beaktas ett antal bakgrundsfaktorer som kan påverka kommunernas kostnader, såsom demografiska och geografiska förhållanden. Dessa faktorer representeras i modellen av ett antal bakgrundsvariabler – så kallade påverkansfaktorer – exempelvis kommunens yta, befolkningsmängd och befolkningstäthet.

Genom att analysera sambanden mellan dessa variabler och kommunernas inrapporterade kostnader beräknas ett prediktionsvärde för varje kommun. Detta värde utgör en uppskattning av vad kommunens kostnad per capita rimligen kan förväntas vara, givet dess strukturella förutsättningar och mönstren i samtliga kommuners data.

Modellen syftar inte till att fastställa en absolut sanning om kostnadsnivåer, utan till att skapa en konsekvent referensram där extrema avvikelser (outliers) dämpas. Prediktionen fungerar som ett verktyg för att bedöma om en kommuns inrapporterade kostnader är rimliga i förhållande till dess specifika förutsättningar, och anger därmed ett övre riktvärde för vad som är rimligt att låta ligga till grund för ersättning.



Osäkerheter med bedömningsmodellen

En central metodologisk utmaning är att det saknas ett oberoende och verifierat utfall att jämföra modellens prediktioner med. Det innebär att träffsäkerheten inte kan valideras på traditionellt sätt – det går inte att avgöra om en enskild prediktion är ”rätt” eller ”fel”. I stället får modellens värde bedömas utifrån dess förmåga att fånga mönster och förklaringskraft jämfört med de rådata som rapporterats in.

Det är också viktigt att understryka att förklaringsgraden hos de använda variablerna är begränsad. Strukturella skillnader mellan kommuner som inte fångas i tillgänglig kvantitativ data kan påverka resultaten. Prediktionerna bör därför tolkas som en indikation på hur kostnadsskillnader mellan kommuner kan förstås eller motiveras utifrån de faktorer som modellen fångar.



Modellutveckling

Nedan beskrivs en kort sammanfattning av processen. Mer detaljerad beskrivning av data samt beskrivning av regressionsanalysen presenteras i följande avsnitt.

Kortfattad beskrivning av processen:

- Kartläggning över relevanta offentliga data, det vill säga kvantitativa bakgrundsvariabler som förväntas påverka kostnader för att städa upp avfall som skräpats ned
- Test av samtliga identifierade variabler mot inrapporterade kostnadsdata
- Utveckling av bedömningsmodell med utgångspunkt i inrapporterade kostnadsdata. Detta inkluderar:
 - Regressionskörningar av inrapporterade kostnadsdata mot bakgrundsvariabler
 - Sammanställning av regressionsresultat
 - Framtagande av nyckeltal utifrån bedömningsmodellen



Ramboll tog fram en första version av bedömningsmodellen under 2023

Skillnad mellan den gamla och nya bedömningsmodellen

Tidigare modell

- Körde en regression per oberoende variabel (förutom kommungrupp)
- Använde tre metoder för estimering:
 - Individuella parameterestimat som sammanvägdes enligt två olika metoder
 - Genomsnittliga kostnader för de olika kommungrupperna
- Alla tre metoder applicerades för att få fram tre estimat
- Dessa estimat gavs samma vikt, och producerade en samlad bedömning

Utvecklad modell

- Kör en stor regression med alla oberoende variabler samtidigt
 - Även kommungrupp används som oberoende variabel
- Parameterestimatet appliceras samtidigt, enligt modellens resultat
- En enda prediktion tas fram och används
- Onödiga variabler elimineras

Data

Parametrar som driver kostnader för skräphantering

Kostnadsnivån för skräphantering beror på tre huvudsakliga parametrar:

1. **Nivå av nedskräpning** (beror t.ex. på befolkningsstorlek, dagbefolkning, andel unga i befolkningen, turism)
2. **Effektivitet i arbetet** (beror t.ex. på geografiska förutsättningar, stordriftsfördelar, organisation)
3. **Ambition i arbetet** (beror t.ex. på besöksnäring, kommunekonomi) – lägre tolerans för skräp och en ökad ambition att hålla skräpet borta påverkar kostnaderna.

Dessa tre parametrar har varit vägledande i valet av bakgrundsvariabler.



Identifiering av bakgrundsfaktorer och -variabler

I arbetet med att identifiera bakgrundsvariabler som kan förklara kostnader för skräphantering har vi utgått från de parametrar som bedöms driva dessa kostnader. Vi har kategoriserat variablerna i tre huvudgrupper: **demografiska faktorer**, **geografiska faktorer** samt **skräprelaterade faktorer**.

De demografiska och geografiska faktorerna syftar till att fånga strukturella och omvärldsrelaterade förutsättningar som påverkar både nivån av nedskräpning och effektiviteten i skräphanteringsarbetet. De skräprelaterade faktorerna avser att spegla den faktiska omfattningen av nedskräpning samt, i viss mån, ambitionsnivån i kommunernas arbete mot nedskräpning. Tillsammans utgör dessa variabler en grund för att förklara variationer i kostnader för skräphantering mellan kommuner. Det finns dock inneboende komplexiteter i samspelet mellan nedskräpningsnivå, arbetseffektivitet och ambitionsnivå.

En ytterligare utmaning är att samma variabel kan påverka kostnader i olika riktningar beroende på kontext. Till exempel kan en hög ambitionsnivå initialt innebära ökade kostnader, samtidigt som den på sikt kan bidra till minskad nedskräpning och därmed lägre kostnader. Detta skapar en komplexitet vid tolkningen av variablernas samband med kostnaderna, även om de underliggande parametrarna skulle vara möjliga att observera direkt.

Val av bakgrundsvariabler

För att modellera kostnader för skräphantering har vi identifierat och kategoriserat relevanta bakgrundsvariabler. Dessa delas in i tre huvudgrupper: **demografiska faktorer**, **geografiska faktorer** samt **skräprelaterade faktorer**.

Demografiska faktorer

- **Invånarantal:** Fler invånare kan leda till ökad nedskräpning och därmed ökat behov av skräphanteringstjänster.
- **Befolkningstäthet:** Tätare bebyggelse kan innebära både högre och mer koncentrerad nedskräpning. Det kan möjliggöra effektivare skräphantering per insats, men också ställa högre krav på städfrekvens.
- **Turismvolym (antal fritidshus per capita):** Tillfälliga befolkningsökningar kopplade till turism kan driva upp nedskräpningsnivåer och därmed öka kostnaderna under vissa perioder.
- **Antal människor som arbetar i kommunen men bor utanför den:** Fångar ytterligare dimensioner av daglig rörelse och vistelse i kommunen, vilket påverkar behovet av skräphantering.

Geografiska faktorer

- **Kommunens yta:** Påverkar omfattningen av områden som behöver städas. Större geografisk yta kan innebära högre kostnader, särskilt i glesbebyggda kommuner där transportavstånd och städfrekvens är kostnadsdrivande.
- **Kommungrupp:** Fångar kommunens övergripande struktur och typ (exempelvis storstad, förort eller glesbygd). Olika kommungrupper kan ha olika nedskräpningsmönster och olika kostnadsnivåer för skräphantering.

Skräprelaterade faktorer

- **HSR kommunindex:** Ett mått på kommunens samlade arbete med skräphantering och renhållning, inklusive förebyggande åtgärder och ambitionsnivå.
- **Andel invånare som upplever problem med nedskräpning:** Fångar invånarnas subjektiva upplevelse av nedskräpning, vilket kan indikera faktisk nedskräpningsnivå eller spegla kommunens ambitionsnivå i arbetet.

Vi har testat en stor mängd möjliga variabler

I arbetet med att ta fram det slutgiltiga valet av variabler har vi testat en stor mängd möjliga bakgrundsvariabler. Dessa presenteras nedan. Vi har valt att inte inkludera dessa eftersom vi inte ser att de bidrar med någon extra information, utöver de variabler vi har valt.

- **Andel av befolkningen i åldersgruppen 15–29 år**
Hypotes: Yngre personer skräpar ned mer än äldre, vilket innebär att en större andel unga i kommunen kan kopplas till högre kostnader för skräphantering.
- **NUTS1 och NUTS2-region**
Hypotes: Större geografiska gruppstillhörigheter kan påverka nedskräpningsnivåerna på grund av skillnader i mikroklimat mellan regioner.
- **Norrland (indikatorvariabel)**
Hypotes: Kommuner i Norrland kan ha särskilda nedskräpningsmönster på grund av sitt specifika mikroklimat.
- **Landsbygd**
Hypotes: Mindre nedskräpning förväntas på landsbygden eftersom rörligheten till fots i offentliga miljöer är lägre än i tätorter.
- **Tätortsgrad**
Variabeln fångar hur urbaniserad kommunen är, vilket kan påverka nedskräpningsnivåerna.
- **Kvot Dag-/Nattbefolkning**
Hypotes: Kommuner med många inpendlare har fler personer som vistas där dagtid, vilket kan öka nedskräpningen i förhållande till nattbefolkningen.
- **Vissa turismrelaterade variabler**
(Antal gästnätter, Logiintäkter, Beläggningsgrad, Ankomster, Gästnätter per invånare)
Hypotes: En hög volym av turism kan leda till ökad nedskräpning, eftersom fler människor rör sig i kommunen och kan generera mer avfall.
- **Möjlighet till återvinning utan bil (Medborgarundersökning)**
Hypotes: Om det är svårt att återvinna avfall på ett smidigt sätt kan risken för nedskräpning i offentliga miljöer öka.
- **Kommunal information om miljöpåverkan (Medborgarundersökning)**
Hypotes: Bristande information från kommunen om hur invånare kan minska sin miljöpåverkan kan leda till högre nivåer av nedskräpning.
- **Problem med skadegörelse och klotter (Medborgarundersökning)**
Hypotes: Förekomst av skadegörelse och klotter kan vara ett tecken på bristande respekt för den offentliga miljön och samvariera med nedskräpning.
- **Upplevd trygghet utomhus (Medborgarundersökning)**
Hypotes: Lägre trygghet kan spegla en allmänt sämre offentlig miljö, vilket i sin tur kan korrelera med högre nedskräpningsnivåer.
- **Anmälda brott om nedskräpning (per 100 000 invånare, enligt Brå)**
Ett direkt mått på rapporterade fall av nedskräpning.

Om datan som används

För en av variablerna i datan behöver vissa justeringar göras. Detta beskrivs nedan:

- **Andel invånare som upplever problem med nedskräpning:** För vissa kommuner och år saknas data från Medborgarundersökningen. För att maximera antalet kommuner med relevant data använder vi det senast rapporterade icke-tomma värdet för varje kommun. Om en kommun saknar rapporterade värden helt, även efter denna process, ersätts det saknade värdet med genomsnittet av alla tillgängliga kommunvärden. För 2025 saknar 49 kommuner data efter dessa justeringar, men detta antal förväntas minska i takt med att fler observationer tillkommer i kommande undersökningar.

Regressionsanalys

Användning av regressionsanalys för prediktion

Vi använder en multipel linjär regressionsmodell för att beräkna prediktionsvärden. Modellen analyserar hur variationen i kommunernas inrapporterade kostnader för att städa upp avfall som skräpats ned kan förklaras av ett antal bakgrundsvariabler – så kallade påverkansfaktorer – som bedöms ha ett relevant samband med kostnadsnivåerna.

Linjär regression är en etablerad statistisk metod och en typ av maskininlärningsmodell som estimerar linjära samband mellan en beroende variabel och flera oberoende variabler. I denna modell estimeras en koefficient för varje bakgrundsvariabel, som beskriver hur förändringar i variabeln är associerade med förändringar i kostnader. Koefficienterna tas fram genom Ordinary Least Squares (OLS), en metod som syftar till att minimera summan av kvadrerade fel mellan observerade och predikterade värden.

Modellen tränas på existerande data och används därefter för att prediktera rimliga kostnader per capita för varje kommun, baserat på dess strukturella förutsättningar.

Eftersom syftet är att skapa en konsistent referensram för bedömning – inte att fastställa kausala samband – läggs större vikt vid modellens prediktiva förmåga än vid enskilda koefficienters statistiska signifikans. Vi eftersträvar en modell som fångar de övergripande mönstren i datan utan att överanpassa till enskilda variationer, eftersom en alltför hög anpassning skulle kunna introducera bias och minska modellens generaliserbarhet.

Prediktionen används som ett verktyg för att uppskatta vad kommunens kostnad per capita rimligen kan förväntas vara, givet dess strukturella förutsättningar och mönstren i samtliga kommuners data.

Modellspecifikation

En multipel linjär regressionsmodell används alltså för att beräkna sambandet mellan kostnader per capita och ett antal förklarande variabler. Med kostnader per capita avses den totala summan av kostnader för uppstädning, administration och planering samt informationsspridning, dividerad med folkmängden i kommunen. Kostnader för insamling och rapportering av uppgifter ingår inte i regressionsanalysen, men används för den slutgiltiga beräkningen av berättigade kostnadsersättningar.

Formellt uttrycks modellen som:

$$\begin{aligned} & \text{Kostnad.per.capita} \\ &= \text{Landareal} + \text{Befolkning} + \text{Befolkningstäthet} + \text{Nedskräpning} + \text{HSR Kommundindex} + \text{Fritidshus per capita} + \text{Inpendlarkvot} + \text{Kommungrupp} + \varepsilon \end{aligned}$$

där ε representerar modellens felterm.

För att minska risken att resultaten påverkas av kommuner med orimliga värden, sorteras vissa extremvärden bort i träningsdatan. Specifikt exkluderas kommuner som rapporterat kostnader under 1 krona per capita, eftersom mycket låga kostnader ofta beror på felaktig eller ofullständig rapportering.

Efter att modellen har skattats används de framtagna koefficienterna på hela datasättet (inklusive de bortfiltrerade kommunerna) för att generera prediktioner. På så sätt tilldelas varje kommun en datadriven uppskattning av vilka kostnader som kan förväntas, givet dess strukturella och geografiska förutsättningar.

Vilka poster ingår var?

Olika kostnadsposter behandlas delvis olika i analysen. I regressionsmodellen ingår följande kostnadstyper:

- Uppstädning
- Administration och planering
- Informationsspridning

Summan av dessa kostnader utgör den beroende variabel som modellen är utformad för att estimeras.

När regressionsmodellen har körts och preliminära kostnadsförslag per capita tagits fram för varje kommun, delas de tre kostnadstyperna upp igen. De två första – uppstädning samt administration och planering – justeras därefter med en skräpmättningsfaktor, för att säkerställa att endast kostnader som är ersättningsgrundande inkluderas.

Efter justeringen summeras samtliga tre kostnadstyper, och kompletteras med kostnader för insamling och rapportering av uppgifter.

Det resulterande värdet per capita multipliceras med respektive kommuns folkmängd, vilket ger den totala kostnaden per kommun. Denna utgör modellens slutgiltiga förslag till ersättningsgrundande kostnad.

Mer detaljerade steg, inklusive tillämpade matematiska formler, presenteras på nästa sida.

Beräkningar av prediktionsvärde

Nedan beskrivs hur vi, baserat på regressionsmodellen, beräknar prediktionsvärden för kommunernas kostnader för att städa upp avfall som skräpats ned. Arbetssättet säkerställer att det predikterade värdet speglar kommunens förväntade behov utifrån dess egenskaper, samtidigt som den inte överstiger de kostnader som kommunen själv har rapporterat. Beräkningarna görs både på totalnivå och nedbrutet per kostnadstyp.

1. Vi skapar prediktionsvärden för enskilda kostnadstyper (uppstädning, administration och planering samt informationsspridning) baserat på modellens samlade prediktionsvärde. Detta gör vi genom att multiplicera prediktionsvärdet med de rapporterade andelarna av de totala kostnaderna per kostnadstyp.

$$Kostnad.per.capita_{Typ}^{Pred} = Kostnad.per.capita_{Total}^{Pred} \times \left(\frac{Kostnad_{Typ}^{Rapport}}{Kostnad_{Total}^{Rapport}} \right)$$

2. Vi omvandlar prediktionsvärden för kostnader per capita per kostnadstyp till totala kostnader genom att multiplicera dem med kommunens folkmängd.

$$Kostnad_{Typ}^{Pred} = Kostnad.per.capita_{Typ}^{Pred} \times Befolkning$$

3. Vi jämför rapporterade kostnader med modellens föreslagna kostnader. Det lägre av dessa två väljs för vardera kostnadstyp. Detta för att modellen inte ska ge en föreslagen ersättningsgrundande kostnad som är högre än den kommunen rapporterat.

$$Kostnad_{Typ}^{Berättigad} = \min(Kostnad_{Typ}^{Pred}, Kostnad_{Typ}^{Rapport})$$

4. För vissa kostnadstyper (se nästa sida för detaljer) multipliceras den föreslagna ersättningsgrundande kostnaden med en skräpmättningsfaktor, för att säkerställa att endast kostnader för relevanta skräptyper omfattas.

$$Ersättning_{Typ}^{Berättigad} = Kostnad_{Typ}^{Berättigad} \times Skräpmättningsfaktor$$

5. Det slutgiltiga förslaget på fastställd kostnad per kommun beräknas som summan av föreslagna kostnader per kostnadstyp, tillsammans med de inrapporterade kostnaderna för insamling och rapportering av uppgifter (dessa justeras inte).

$$Ersättning_{Total}^{Berättigad} = \sum Ersättning_{Typ}^{Berättigad} + \text{Kostnader för insamling och rapportering av uppgifter}$$

Om skräpmätningsfaktorn

Skräpmätningsfaktorn anger vilken andel av det skräp som uppmätts i Naturvårdsverkets nationella skräpmätning som består av produkter som omfattas av förordningen.

För år 2025 inkluderas följande skräpkategorier:

- Engångsplastlock till muggar
- Engångsplastlock till matlådor
- Flexibla omslag
- Matlådor som är engångsplastprodukter
- Muggar som är engångsplastprodukter
- Andra dryckesbehållare än plastflaskor som rymmer mindre än 0,6 liter
- Fimpar
- Ballonger
- Våtserverter

Enligt den nationella skräpmätningen från 2023 utgör dessa kategorier cirka 58 procent av allt skräp. Detta innebär att endast cirka 58 procent av de rapporterade eller beräknade kostnaderna ska anses vara ersättningsgrundande. Notera dock att kommunernas kostnader för insamling och rapportering av uppgifter till Naturvårdsverket samt informationsspridning kopplat till nedskräpning av engångsplastprodukter inte justeras av skräpmätningsfaktorn.

Om betydelsen av statistisk signifikans vid prediktiv regressionsanalys

Vid estimering av en linjär regressionsmodell i syfte att enbart generera prediktioner ligger fokus huvudsakligen på att uppnå en hög prediktionsnoggrannhet, snarare än på att tolka koefficienter eller bedöma deras statistiska signifikans. Några viktiga överväganden i en sådan kontext är:

- **Variabelurval:** Även om statistisk signifikans inte är huvudsaklig fokus kan den ge viss vägledning vid urval av variabler. Variabler med icke-signifikanta koefficienter kan ibland bidra med brus eller redundans som försämrar modellens prediktiva förmåga.
- **Robusthet:** Att inkludera variabler som inte är statistiskt signifikanta behöver inte alltid vara negativt om de tillför viss prediktiv information. Samtidigt kan en modell som exkluderar överflödiga variabler bli mer robust och lättare att tolka.
- **Överanpassning:** En överkomplex modell riskerar att fånga upp brus i stället för underliggande samband. Att begränsa modellen till betydelsefulla variabler kan bidra till att minska risken för överanpassning och förbättra dess generaliserbarhet.

I denna modellering har vi därför inte fokuserat på statistiska signifikansnivåer, och de enskilda koefficienterna bör inte tolkas som kausala effekter. De är i stället en del av en prediktionsalgorithm som syftar till att skapa en rimlig uppskattning av kommunernas kostnader utifrån deras strukturella förutsättningar. Dock har signifikans delvis vägts in i bedömningar om specifikation, eftersom ett parameterestimat som varken är statistiskt signifikant eller leder till någon nämnvärd förbättring i modellens förklaringsvärde oftast inte kan anses vara tillräckligt viktig för att inkludera i modellen.

Automatisering av modell

Beräkningarna sker automatiskt i R

Ett R-skript används för att genomföra alla de nödvändiga beräkningarna som beskrivits i modellen. Skriptet är strukturerat i flera övergripande steg, och varje steg är noggrant kommenterat för att underlätta användarupplevelsen och säkerställa att processen kan följas och förstås enkelt. De centrala stegen som genomförs i skriptet är följande:

1. **Läsa in data:** Skriptet importerar nödvändiga data om kommunernas rapporterade kostnader samt bakgrundsvariablerna och annan data (t.ex. befolkning, landareal, och nationella skräpmättningsresultat) från den specificerade datakällan.
2. **Transformation av datapunkter:** Vissa datapunkter transformeras för att säkerställa att data är i rätt format för analys. Ett exempel på detta är konvertering av kostnader till kostnader per capita, vilket görs för att kunna jämföra kommunernas kostnader på ett rättvist sätt.
3. **Filtrering av extremvärden:** För att säkerställa att resultaten inte påverkas av felaktiga eller extrema värden, filtreras data för att ta bort observationer med ovanligt låga värden, såsom kommuner med rapporterade kostnader under 1 kr per capita.
4. **Genomförande av regressionsanalys:** Skriptet genomför en multipel linjär regressionsanalys baserat på filtrerade och transformerade data. Här beräknas sambandet mellan kostnader och de förklarande variablerna enligt den specificerade modellen.
5. **Applicering av resultat och beräkning av slutgiltiga kostnader:** Efter att regressionsmodellen har körts appliceras de beräknade koefficienterna på hela datamängden för att generera prediktioner för kostnader per kommun. Dessa prediktioner används för att beräkna slutgiltiga föreslagen ersättningsgrundande kostnad.
6. **Export av resultat i Excelformat:** Slutligen exporteras resultaten i ett strukturerat Excelformat.

Automatiserat datainsamlingssystem via SCB:s API

Ett särskilt R-skript har även tagits fram för att automatisera uttaget av data från SCB:s API*. Detta gör det möjligt att på ett effektivt sätt samla in en mängd nyckeldata, bland annat:

- Befolkningsmängder, landarealer och befolkningstätheter
- Dagbefolkning (inpendlare)
- Resultat från medborgarundersökningar
- Antal fritidshus

Utöver dessa variabler kan skriptet även hämta andra typer av data så länge dessa är tillgängliga via SCB:s API eller någon annan relevant datakälla.

Förutom att automatiskt samla in och extrahera data, ansvarar skriptet även för att transformera dessa uppgifter till ett format som överensstämmer med de krav som regressionsmodellen ställer. Detta säkerställer att data är korrekt definierade och formaterade för vidare analys.

Det är dock fortfarande nödvändigt att hämta vissa specifika dataset manuellt. Detta beskrivs på nästa sida.

* API (Application Programming Interface) är en uppsättning verktyg och regler som gör det möjligt för olika program att kommunicera med varandra och utbyta data på ett strukturerat sätt.

Manuella moment inför modellkörning

Trots att det har gjorts ett stort arbete för att automatisera processen återstår fortfarande vissa manuella steg. Dessa beskrivs nedan.

Följande data hämtas inte automatiskt, och behöver därför tas fram och sparas i en dedikerad mapp för rådata:

- HSR Kommunindex
- SKR:s kommungruppsindelning (om den har förändrats sedan föregående år)
- Kommunernas inrapporterade kostnader för att städa upp avfall som skräpats ned
- Resultat från den senaste nationella skräpmätningen

Därutöver behöver JSON-filerna för frågor mot SCB:s API uppdateras:

- I varje JSON-fil (med undantag för Medborgarundersökningen) ska det aktuella året anges under "values" i det block som har **"code": "Tid"** nära filens slut.

När datagrunden är på plats behöver två R-skript köras i rätt ordning. Skripten tar cirka en sekund vardera att köra.

1.hämta_data.R

2.beräkningsmodell.R

Efter körningen ska en Excelfil med resultat automatiskt ha genererats och sparats i samma mapp där skripten ligger.

Resultat/output

Export och sammanställning av resultat

När resultaten har genererats i R exporteras de till ett Excelformat för vidare hantering och analys. Resultaten sammanställs i en Excelfil där varje rad motsvarar en kommun och där följande kolumner ingår:

Kolumn	Förklaring
Kommun	Namn på kommunen i fråga
Total rapporterad kostnad	Summan av alla rapporterade kostnader
Prediktionsvärde kostnader per capita	Regressionsmodellens förslag på kostnader per capita*
Befolkning	Senast observerade folkmängdsstatistiken från SCB
Total prediktionskostnad	Prediktionsvärdet för kostnader per capita multiplicerat med folkmängden*
Rapporterad kostnad för insamling och rapportering av uppgifter	Denna kostnadspost har exkluderats från analysen och inkluderas nu för att adderas tillbaka till den totala predikterade kostnaden
Bedömd total kostnad	Summan av total prediktionskostnad och rapporterad kostnad för insamling och rapportering av uppgifter (som inte ingår i modellen)
Förslag på fastställd kostnad	Summan av fastställda typkostnader, som tagits fram genom jämförelse av predikterade och rapporterade typkostnader och multiplicerats med skräpmätningsskalfaktorn

*Dessa prediktionsvärden omfattar kostnader för uppstädning, administration och planering samt informationsspridning.

Excelfilen placeras automatiskt i samma mapp som där R-skripten kördes.

Exempelresultat

Strukturen på de sammanställda resultaten är enligt tabellen till höger, fast i en Excelfil.

OBS! De faktiska siffrorna här är endast exempelsiffror som inte utgör någon slutgiltig bedömning.

Kommun	Totala rapporterade kostnader (kr)	Kostnader per capita (kr), prediktion	Folkmängd	Totala kostnader (kr), prediktion (exkl. insamling och rapportering av utgifter)	Rapporterade kostnader (kr) för insamling och rapportering av uppgifter	Totala kostnader (kr), samlad prediktion	Förslag på fastställd kostnad (kr)
0114 Upplands Väsby	1 159 150	38,13	50 110	1 910 783	2 560	1 913 343	676 210
0115 Vallentuna	741 000	33,41	35 076	1 171 896	16 000	1 187 896	438 564
0117 Österåker	882 000	36,77	49 282	1 811 928	1 300	1 813 228	512 062
...	...	...	...	...	...	...	...
2582 Boden	391 260	24,02	27 943	671 092	25 000	696 092	237 412
2583 Haparanda	58 000	38,04	9 177	349 081	1 000	350 081	36 157
2584 Kiruna	514 000	10,85	22 433	243 314	14 000	257 314	155 110

Utökat resultat

I Excelfilen med resultat exporteras även utökade resultat till en annan flik. Dessa rapporterar mer detaljerade siffror, vilka använts i de olika beräkningsstegen. De utökade resultaten kan exempelvis användas för att kvalitetssäkra beräkningar eller för att analysera specifika avvikelser mellan rapporterade och beräknade kostnader.

Tabellen innehåller 33 kolumner som för varje kommun rapporterar:

- Kommunens namn och folkmängd
- Rapporterade kostnader per kostnadstyp (totalt och per capita)
- Prediktionsvärden för kostnader per kostnadstyp (totalt och per capita)
- Avvikelse mellan rapporterade kostnader och prediktionsvärden (i kronor per capita och i procent)
- Förslag på fastställd kostnad per kostnadstyp efter jämförelse mellan rapporterade och beräknade kostnader, samt efter eventuell justering för skräpmätning
- Slutgiltigt totalt förslag på fastställd kostnad

Bright
ideas.
Sustainable
change.

RAMBOLL