Appendix 4

Metodbeskrivning för The Definitive Ranking (TDR)

Detta appendix beskriver den beräkningsmässiga konstruktionen av TDR från riktningskontrollerade indikatorer till slutligt totalindex och rangordning. Beskrivningen avser den slutliga analysmodellen med 290 kommuner och 13 delindex. Validering, alternativa normativa viktmodeller och jämförelser med externa rankingar ingår inte.

1. Analysenhet, datamängd och struktur

Analysenheten är kommun. Samtliga 290 svenska kommuner ingår i den slutliga analysfilen. TDR är hierarkiskt uppbyggt i två nivåer: indikatorer sammanförs först till kategoriindex (delindex), varefter 13 delindex sammanvägs till TDR Total.

De 13 delindexen är: Befolkning; Arbetsmarknad; Inkomster; Företagande; Bostäder; Tillgänglighet/Infrastruktur; Utbildning; Hälsa; Trygghet; Miljö; Kultur/Fritid; Välfärd/Robusthet – Ekonomiblock; samt Välfärd/Robusthet – Välfärdsblock.

Varje kategori har konstruerats enligt samma övergripande analyskedja. Kategorispecifika uppgifter om indikatorval, transformationer och eventuell hantering av saknade värden dokumenteras i respektive kategorimaterial.

2. Förbehandling av indikatorer

2.1 Bruttourval och slutligt indikatorurval

Utgångspunkten var ett bruttourval om cirka 180 kandidatindikatorer. Genom en successiv avgränsning reducerades detta till 121 indikatorer, vilka utgör TDR:s slutliga indikatoruppsättning. Urvalssteget föregick den statistiska sammanvägningen och fastställde därmed den datamängd som användes i de efterföljande korrelations- och PCA-analyserna.

Efter att den slutliga indikatoruppsättningen fastställts kontrollerades varje indikator med avseende på riktning, fördelning och datatäckning. Syftet var att skapa variabler som är jämförbara i riktning och lämpliga för korrelations- och PCA-analys.

2.2 Riktningskontroll

Samtliga indikatorer orienteras så att ett högre numeriskt värde representerar ett bättre utfall enligt den riktning som fastställts för indikatorn. Om en indikator ursprungligen är orienterad så att lägre värde är bättre, vänds den före fortsatt analys. Den enklaste vändningen är

$$x' = -x$$

En sådan teckenvändning ändrar inte indikatorns korrelationsstruktur annat än att tecknet på korrelationerna vänds. Rangordningen blir omvänd och den gemensamma indexriktningen kan därefter upprätthållas konsekvent.

2.3 Transformationer

Transformationer används när indikatorns fördelning motiverar detta. I utbildningskategorin logtransformeras exempelvis andelen med eftergymnasial utbildning enligt

$$x' = \ln(x)$$

Transformationen genomförs före standardisering och PCA. Den exakta transformationsregeln för varje indikator framgår av indikator- och kategoridokumentationen.

2.4 Saknade värden

Datatäckningen kontrolleras före den multivariata analysen. Hantering av saknade värden är kategorispecifik och dokumenteras i respektive kategorimaterial. Utbildningskategorin har full täckning för samtliga 10 indikatorer och kräver därför ingen imputering. Den slutliga filen med 13 delindex innehåller ett fullständigt kategoriindex för samtliga 290 kommuner.

3. Deskriptiv statistik och standardisering

För varje indikator beräknas deskriptiva fördelningsmått. Dessa omfattar n, antal och andel saknade värden, medelvärde, stickprovsstandardavvikelse, minimum, kvartiler, median, maximum, korrigerad stickprovsskevhet samt korrigerad excess kurtosis. Kvartiler beräknas med linjär interpolation på position (n - 1)p.

PCA:n genomförs på standardiserade indikatorer. För indikator i och kommun j beräknas z-värdet som

$$z_{ij} = (x_{ij} - \bar{x}_i) / s_i$$

där $\bar{x}_i$ är indikatorns medelvärde och s_i dess stickprovsstandardavvikelse.

Standardiseringen innebär att varje indikator går in i PCA:n med medelvärde 0 och standardavvikelse 1. PCA på de standardiserade variablerna är ekvivalent med PCA på deras korrelationsmatris.

4. Korrelationsanalys

Inom varje kategori beräknas Pearsons produktmomentkorrelation mellan de riktningskontrollerade och transformerade indikatorerna. Korrelationsmatrisen utgör underlag för PCA. För två variabler x och y beräknas korrelationen som

$$r_{xy} = \Sigma[(x_j - \bar{x})(y_j - \bar{y})] / \sqrt{\{\Sigma(x_j - \bar{x})^2 \cdot \Sigma(y_j - \bar{y})^2\}}$$

Korrelationsmatrisen används både som diagnostiskt underlag och som direkt indata till PCA. Inga rotationer används i den slutliga TDR-modellen.

5. PCA inom kategori

För varje kategori genomförs principal component analysis (PCA) på korrelationsmatrisen. Om en kategori innehåller p indikatorer ger analysen p ortogonala komponenter. För komponent k gäller eigenvärdesproblemet

$$R\, a_k = \lambda_k\, a_k$$

där R är korrelationsmatrisen, λ_k komponentens eigenvärde och a_k dess eigenvektor.

Komponentpoängen för kommun j kan skrivas

$$PC_{kj} = \Sigma_i\, a_{ik}\, z_{ij}$$

Faktorladdningen l_{ik} redovisas som korrelationen mellan indikator i och komponent k. För PCA på standardiserade variabler gäller sambandet

$$l_{ik} = a_{ik} \sqrt{\lambda_k}$$

Komponenternas tecken är matematiskt godtyckliga. Tecknen orienteras därför konsekvent för tolkning och beräkning. En samtidig teckenvändning av en komponent och dess koefficienter förändrar inte lösningens informationsinnehåll.

5.1 Urval av komponenter

Bärande komponenter väljs med Kaisers kriterium: komponenter med eigenvärde större än 1 behålls. Antalet behållna komponenter kan därför variera mellan kategorier. I utbildningskategorin behålls PC1 och PC2; de förklarar tillsammans 67,93 procent av total varians.

6. PCA-baserad viktning inom kategori

De behållna komponenterna viktas proportionellt mot sin förklarade varians inom den behållna lösningen. Om K komponenter behålls och λ_k är eigenvärdet för komponent k definieras komponentvikten som

$$\alpha_k = \lambda_k \,/\, \Sigma_{m=1}{}^{K}\, \lambda_m$$

Därmed gäller $\Sigma\alpha_k = 1$. Ett rått kategoriindex kan skrivas

$$I_j = \Sigma_{k=1}{}^{K}\, \alpha_k\, PC_{kj}$$

Eftersom varje komponent är en linjär kombination av de standardiserade indikatorerna kan samma index skrivas direkt som en indikatorform

$$I_j = \Sigma_i\, b_i\, z_{ij}, \quad \text{där} \quad b_i = \Sigma_{k=1}{}^{K}\, \alpha_k\, a_{ik}$$

Koefficienterna b_i kan därefter normaliseras till en explicit indikatorviktmodell med summan 1. Om $c = \Sigma_i\, b_i$ och $c > 0$ sätts

$$w_i = b_i \,/\, c, \quad \Sigma_i\, w_i = 1$$

Den normaliserade indikatorformen $I^* = \Sigma w_i z_i$ är en positiv linjär omskalning av I. Eftersom kategoriindexet i nästa steg z-standardiseras ger I och I^* samma standardiserade indexvärden och samma rangordning.

6.1 Exempel: utbildningskategorin

Utbildningskategorin omfattar 10 indikatorer. PC1 har eigenvärdet 5,6445 och PC2 eigenvärdet 1,1482. Komponentvikterna blir 0,8310 respektive 0,1690. Den behållna PC-modellen är därför

$$I = 0{,}8310 \cdot PC1 + 0{,}1690 \cdot PC2$$

Återräkning till de standardiserade indikatorerna ger en explicit indikatorviktmodell där vikterna summerar till 1. Denna procedur är den principiella modell som används för kategoriindexen; indikatorer, antal komponenter och erhållna vikter är kategorispecifika.

7. Skalning och rangordning av kategoriindex

Det råa kategoriindexet standardiseras över de 290 kommunerna och transformeras till TDR-skalan med medelvärde 50 och standardavvikelse 10:

$$TDR_category,j = 50 + 10 \cdot (I_j - \bar{I}) \,/\, s^I$$

Här är $\bar{I}$ medelvärdet och s^I stickprovsstandardavvikelsen för det råa kategoriindexet. En kommun med indexvärdet 50 ligger vid kommunmedelvärdet. En differens på 10 indexenheter motsvarar en standardavvikelse.

Kommunerna rangordnas fallande efter kategoriindexet. Rang 1 avser högsta indexvärde. Den slutliga gemensamma analysfilen innehåller kategoriindex och rang för samtliga 13 delindex och samtliga 290 kommuner.

8. Sammanvägning av 13 delindex till TDR Total

De 13 färdiga kategoriindexen sammanförs på kommunnamn. Därefter genomförs en ny PCA med kommun som analysenhet och kategoriindexen som variabler. PCA:n bygger på korrelationsmatrisen, vilket innebär att kategoriindexen behandlas som standardiserade variabler i den övergripande komponentanalysen.

I den slutliga modellen är PC1 och PC2 bärande enligt Kaisers kriterium. PC1 har eigenvärdet 5,5759 och förklarar 42,89 procent av total varians. PC2 har eigenvärdet 2,6359 och förklarar 20,28 procent. Tillsammans förklarar PC1 och PC2 63,17 procent av variationen i de 13 delindexen. PC3 har eigenvärdet 0,9581 och behålls därför inte.

8.1 Kategorivikter

Vid sammanvägningen till TDR Total används kvadrerade faktorladdningar. Syftet är att vikten ska uttrycka hur stor del av en kategoris standardiserade varians som fångas av de bärande komponenterna, oberoende av komponenternas godtyckliga teckenorientering.

Komponentvikten inom den behållna tvåkomponentslösningen definieras som

$$\alpha_k = \lambda_k / (\lambda_1 + \lambda_2), \quad k = 1, 2$$

För kategori j med laddning l_{jk} på komponent k definieras kategorivikten som

$$w_j = \Sigma_{k=1}{}^{2}\, \alpha_k \cdot (l_{jk}{}^{2} / \lambda_k)$$

Eftersom $\alpha_k = \lambda_k/(\lambda_1+\lambda_2)$ kan uttrycket förenklas till

$$w_j = (l_{j1}{}^{2} + l_{j2}{}^{2}) / (\lambda_1 + \lambda_2) = h_j{}^{2} / (\lambda_1 + \lambda_2)$$

där h_j^2 är kommunaliteten för kategori j i den behållna tvåkomponentslösningen. Kategorivikterna är icke-negativa och summerar till 1.

Delindex	Vikt	Vikt (%)	Rang vikt
1 Befolkning	0.0710	7.10	10
2 Arbetsmarknad	0.0846	8.46	6
3 Inkomster	0.1050	10.50	1
4 Företagande	0.0719	7.19	9
5 Bostäder	0.0780	7.80	8
6 Tillgänglighet/Infrastruktur	0.0927	9.27	4
7 Utbildning	0.0929	9.29	3
8 Hälsa	0.0932	9.32	2
9 Trygghet	0.0877	8.77	5
10 Miljö	0.0804	8.04	7
11 Kultur/Fritid	0.0462	4.62	12
12a Välfärd/Robusthet – Ekonomiblock	0.0645	6.45	11
12b Välfärd/Robusthet – Välfärdsblock	0.0321	3.21	13

9. Beräkning av TDR Total

Det råa totalindexet beräknas som det viktade medelvärdet av de 13 kategoriindexen. Om I_{jc} betecknar kategoriindexet för kommun j och kategori c gäller

$$\mathrm{TDR_raw,j} = \Sigma c_{=1}{}^{13}\, \mathrm{w_c}\ I_{jc}$$

Det råa totalindexet standardiseras därefter över samtliga 290 kommuner och transformeras till samma indexskala som kategoriindexen:

$$\mathrm{TDR_Total,j} = 50 + 10 \cdot (\mathrm{TDR_raw,j} - \bar{\mathrm{TDR}}_\mathrm{raw}) / \mathrm{s_raw}$$

TDR Total har således medelvärde 50 och standardavvikelse 10 i kommunpopulationen. Kommunerna rangordnas fallande efter TDR Total. Rang 1 avser högsta TDR Total.

10. PC1 och PC2 som kompletterande mått

Den övergripande PCA:n ger även separata PC1- och PC2-poäng. Poängen beräknas från standardiserade kategoriindex och PCA:ns komponentkoefficienter:

$$PC_{kj} = \Sigma c \; a_c_k \; z_c_j$$

PC1 och PC2 är ortogonala och därmed okorrelerade i den använda PCA-lösningen. De redovisas separat med värde och rang som kompletterande beskrivningar av variationsstrukturen. De används inte som direkta additiva komponenter i TDR Total; totalindexets kategorivikter härleds i stället från de kvadrerade laddningarna på PC1 och PC2 enligt avsnitt 8.1.

11. Formell sammanfattning av beräkningskedjan

Steg	Definition
1. Riktningskontroll/transformation	x → x' enligt indikatorns fastställda riktning och vid behov fördelning.
2. Standardisering	$z_{ij} = (x_{ij} - \bar{x}_i)/s_i$.
3. PCA inom kategori	R $a_k = \lambda_k a_k$; komponenter med $\lambda_k > 1$ behålls.
4. PC-viktning inom kategori	$\alpha_k = \lambda_k / \Sigma\lambda_m$ för behållna komponenter.
5. Kategoriindex	$I_j = \Sigma\alpha_k PC_{kj} = \Sigma b_i z_{ij}$.
6. Kategoriskala	TDR_category = 50 + 10·z(I).
7. PCA av 13 delindex	PC1 och PC2 behålls; $\lambda_1 = 5{,}5759$ och $\lambda_2 = 2{,}6359$.
8. Kategorivikt i totalindex	$w_c = h_c^2/(\lambda_1+\lambda_2)$, $\Sigma w_c = 1$.
9. Rått totalindex	TDR_raw = Σw_cI_c.
10. Slutskala	TDR_Total = 50 + 10·z(TDR_raw).
11. Rang	Fallande rangordning; rang 1 = högsta indexvärde.

12. Metodologiska egenskaper och avgränsningar

TDR:s viktning är empiriskt bestämd av samvariationen i den valda indikatoruppsättningen. Normativa ställningstaganden finns i urvalet av indikatorer, deras riktning och de transformationer som görs före analysen. Själva PCA-viktningen följer därefter den observerade korrelationsstrukturen.

PCA-vikter är datamängdsberoende. Om indikatoruppsättningen, tidsperioden eller kommunernas observerade värden ändras kan komponentstruktur, laddningar och vikter förändras. TDR 2026 ska därför tolkas som en definierad modell för den använda datamängden, inte som en tidsinvariant viktfunktion.

Kaiser-kriteriet används konsekvent som regel för komponenturval inom kategori och på totalnivå. Ingen rotationslösning används. På totalnivå används kvadrerade laddningar för att skapa icke-negativa kategorivikter som är invarianta mot komponenternas teckenorientering.

Indexskalan 50 + 10z är en linjär standardisering. Den påverkar inte kommunernas ordningsföljd. Detsamma gäller positiv linjär omskalning av ett råindex före slutlig z-standardisering.

13. Dokumentation och reproducerbarhet

Metoden är dokumenterad på tre nivåer: den slutliga analysmodellen för TDR Total, det gemensamma underlaget med 13 delindex samt kategorispecifik dokumentation av beräkningarna inom respektive delindex. Denna struktur gör det möjligt att följa beräkningskedjan från enskilda indikatorer till kategoriindex och vidare till TDR Total.

- Detaljer om enskilda indikatorers definitioner, datakällor, riktning och transformationer redovisas i indikatorförteckningen och i den kategorispecifika dokumentationen.
- Beräkningsstegen är definierade så att analysen kan reproduceras från den fastställda indikatoruppsättningen, givna transformationsregler, PCA-lösningar och viktformler.