

Uppskattning av årskullar 2020/21, med exempel för region Stockholm

titlesec

Innehåll

1	Data	1
2	Metod 1: Linjär regression	1
3	Metod 2: ARIMA	2
4	Metod 3: ARIMAX	3
5	Sammanfattning	4

Data

Läsår	Region	Åk 1	Åk 2	Åk 3	Åk 2(es)	Åk 1(es)
2014/15	Stockholm	26470	21944	21712	22680	25857
2015/16	Stockholm	26863	22129	20884	21944	26470
2016/17	Stockholm	29622	22818	21197	22129	26863
2017/18	Stockholm	29853	23927	21430	22818	29622
2018/19	Stockholm	30780	24453	22378	23927	29853
2019/20	Stockholm	30436	25974	23155	24453	30780

Åk 2(es) är då samma elevkull som Åk 3, men är för samtliga läsår något mindre. Det kan bero på flera orsaker, ex. avhopp.

Metod 1: Linjär regression

Linjär regression. En linjär modell där antalet i årskurs 3 är beroende av eftersläpade variabeln för antal i årskurs 2.

$$Y_t = \beta_0 + \beta_1 \times X_{t-1} + \epsilon_t$$

Där Y, den variabel som ska uppskattas. ϵ_t termen kan ignoreras då det endast är en felterm som räknas in i modellen, om $t = 2020/21$ så är $t - 1 = 2019/20$

Uppskattningen behövs för Antal i Åk 2 och Åk 3 för varje region, med detta kan vi räkna ut årskullen som är summan av Antal i Åk 2 och Antal i Åk 3 som sedan delas på två.

Det vi har för att räkna ut detta är den eftersläpade variabeln för Årskull 1:

$$\text{Antal Åk 2}(t) = \beta_0 + \beta_1 \times \text{Antal Åk 1}(t-1)$$

Exempel modell för Stockholm. Data som används är Antal i Åk 1 (eftersläpad) och Åk 2 för läsåren 2013/14 till 2019/20.

$$\begin{aligned} \text{Antal i Åk 2}(t) &= 3361.64 + 0.7145 \times \text{Antal i Åk 1}(t-1) = \\ 33.61.64 + 0.7145 \times 30436 &= 25109.37 \end{aligned}$$

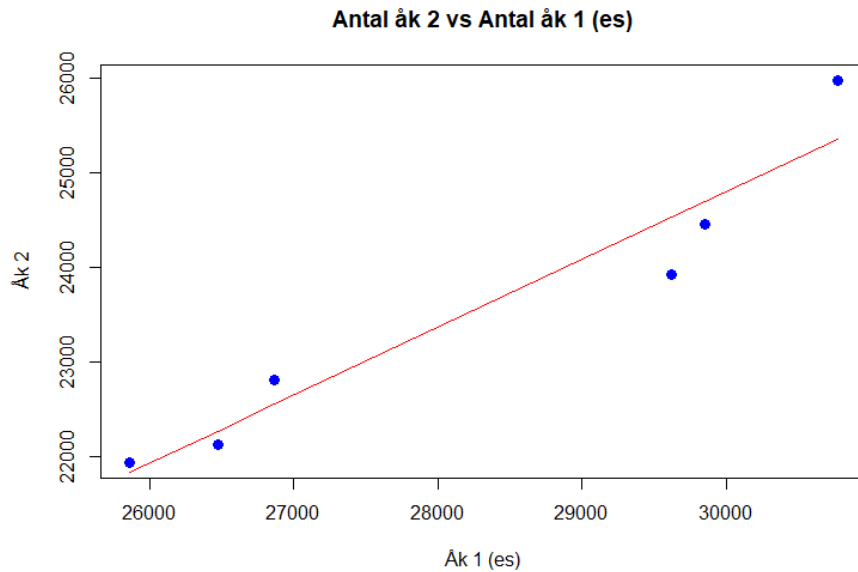
På ett liknande skapar vi en modell för Åk 3, med hjälp av eftersläpade värden för Åk 2.

$$\text{Antal Åk 3}(t) = \beta_0 + \beta_1 \times \text{Antal Åk 2}(t-1) = 2897.20 + 0.82183 \times 25974 = 24244$$

Årskullen för Stockholm 2020/21 blir då $\frac{(25109.37+24244)}{2} = 24676$. Något högre än 2019/20 då årskullen låg på 24565. Vilket beror på att antalet i Åk 2 i Stockholm

konstant ökat de senaste åren, men även att den ökningen inte förklarar hela scenariot då antalet i Åk 1 minskat senaste läsåret från tidigare läsår.

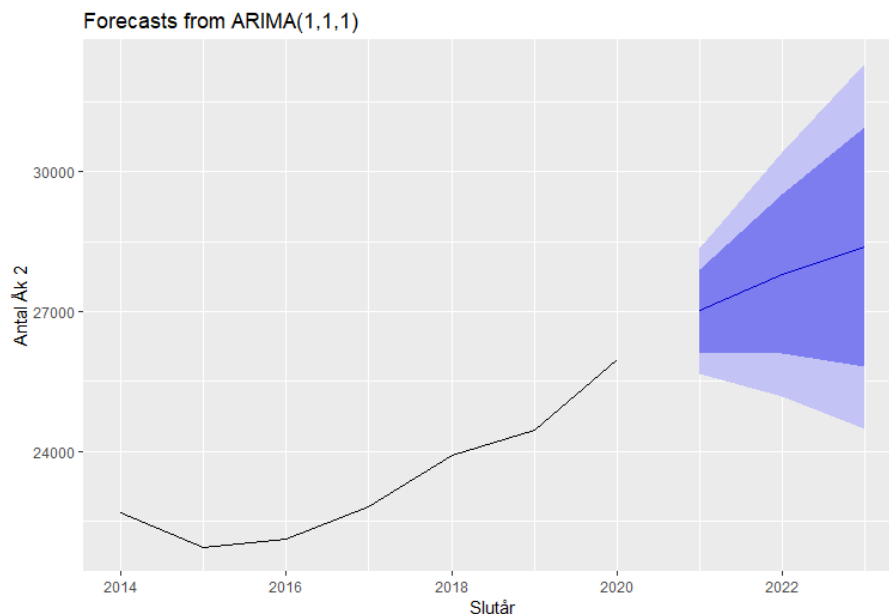
Modellen kan enkelt illustreras genom den röda linjen i grafen nedan, som kan ses som en centrerad linje mellan observationerna (i blått). Vet vi Åk 1(es) kan vi föra in värdet i linjen och estimerar Åk 2.



Metod 2: ARIMA

Tidsserier: modeller som är skapade för att förklara en tidsserie som också är en variant av Linjär regression (Metod 1). Förklarar en variabel och hur den rör sig med variabeln tid (dag, månad, kvartal etc.)

En simpel modell är ARIMA-modeller: $Y_t = Y_{t-1} + \epsilon_t$, skillnaden mellan en sån här modell och den i Metod 1 är att Y_t (variabeln som vi vill estimerar) endast beror på tidigare observationer av samma variabel. Dessa modeller tar upp trenden i en variabel och kan användas för att estimerar hur framtida utveckling kommer se ut, med viss säkerhet (se blåa området i grafen).



Slutår	Prognos (Åk 2)	Lo 80	Hi 80	Lo 95	Hi 95
2021	27021.22	26139.96	27902.48	25673.46	28368.99
2022	27806.38	26094.42	29518.34	25188.16	30424.60
2023	28395.06	25832.04	30958.08	24475.25	32314.86

Tabellen ovan visar modellens prognos för Åk2 3 år i framtiden, osäkerheten kring framtida prognoser visas av att den blåa ytan i grafen blir större.

Slutår	Prognos (Åk 3)	Lo 80	Hi 80	Lo 95	Hi 95
2021	23536.95	22824.15	24249.75	22446.82	24627.08
2022	23732.54	22468.80	24996.29	21799.81	25665.28
2023	23832.71	22081.08	25584.34	21153.82	26511.60

Utifrån denna modell har vi en prognos för Stockholms årskull 2020/21 som är lika med $\frac{27021.22+23536.95}{2} = 25279$

Metod 3: ARIMAX

ARIMAX modeller tillåter att man lägger till förklarande variabler i en ARIMA modell.

$$Y_t = \beta_0 \times X_{t-1} + Y_{t-1} + \epsilon_t$$

Dessa modeller kan ses som en kombination av Metod 1 och 2. Exempelvis för att estimerar en prognos för Åk 3 så använder vi tidigare värden för Åk 3 och värden för

Åk 2(es).

Resultatet från denna modell är prognos för Åk2 = 25354 samt Åk3 = 24984, vilket resulterar Årskull_{Sthlm}(2020/21) = 24984

Sammanfattning

Metod	Prognos (Åk 2)	Prognos (Åk 3)	Prognos Årskull
Linjär regression	25109	24244	24676
ARIMA	27021	23537	25279
ARIMAX	25353	24984	24984

En enkel uträkning som man kan göra är hur många av Åk 1 (2018/19) gick vidare till Åk 2 (2019/20). $\frac{25974}{30780} = 0.84$ och multiplicera den andelen med antal Åk 1(2019/20) för att ge oss en idé om hur många som kommer att gå i Åk 2 (2020/21). givet att andelen är densamma. $0.84 \times 30436 = 25684$. Vi kan göra detsamma för åk 3. $\frac{23155}{24453} = 0.947$. och multiplicera andelen med antalet i Åk 2(2019/20). $0.947 \times 25976 = 24597$. Detta ger oss en årskull på $\frac{(25684+24597)}{2} = 25141$.

Dessa andelar varierar en hel del däremot och tas upp någorlunda av modellerna. främst i ARIMAX-modellen. Men resultatet från den enklare uträkningen visar på att vi ligger nånstans mellan ARIMAX och ARIMA modellen gällande prognosen för årskull. En rimligt förväntning är att ARIMA modellen överskattar årskullen. då den endast tar upp trenden men inte räknar på förklarande variabler. Den tar alltså inte upp minskningen i Åk 1 och ökningen i Åk 2 (i uträkningen för antalet i Åk 3). ARIMAX känns mer rimlig och resultatet är väldigt nära den enklare uträkningen. Exempelvis har andelen från $\frac{ANTAL\ddot{A}k3}{Antal\ddot{A}k2(es)}$ varierat mellan 0.94 – 0.96 och liknande har $\frac{ANTAL\ddot{A}k2}{Antal\ddot{A}k1(es)}$ varierat mellan 0.81 – 0.85. ARIMAX modellen tar upp denna variation och borde ge det mest trovärdiga resultatet för en prognos av Årskull_{Sthlm}(2020/21).

Justering efter Skolverkets officiella statistik i Mars 2021. Utifrån resultat från Skolverket och den officiella statistiken för antalet elever i Sverige visar sig ARIMAX-modellen underskatta prognosen för totala antalet. Vilket förklaras av att modellen inte hittat någon kraftig trend på regional nivå. Modellen lyckas betydligt bättre i att uppskatta siffror på programnivå.

Program	Prognos årskull ARIMAX	Årskull Skolverket (2020/21)	Felberäkning
Barn- och Fritid	3422	3593.5	5%
Bygg- och Anläggning	4371	4447.5	2%
Ekonomi	16161	16268	1%
El- och Energi	5016	5071	1%
Estet	6489	6858	5%
Fordon- och Transport	4103	4081	1%
Handels- och Administration	3361	3319	1%
Hantverk	1994	2080.5	4%
Hotell- och Turism	854	869	2%
Humanistiska	701	689.5	2%
Industri teknik	1561	1546	1%
Introduktion	8903	8647.5	3%
Naturbruk	3064	3096	1%
Naturvetenskap	14865	15059	1%
Restaurang- och Livsmedel	1442	1475	2%
Samhäll	20160	20157.5	0%
Teknik	9986	10110	1%
VVS- och Fastighet	1200	1207.5	1%
Vård- och Omsorg	3417	3483	2%
Övriga program	1269	1396	9%
Alla program	112340	113454.5	1%

Resultat från uppskattning av årskullar genom ARIMAX metoden. kontra egentliga siffran från Skolverket. Felberäkningen av totalsiffran är betydligt lägre än när metoden räknar på regionala siffror. Vilket förklaras av att modellen tar upp en tydlig positiv trend i Samhällsprogrammet samt Ekonomiprogrammet

Utifrån dessa resultat finns en rimlig grund till att anta att resultatet från ARIMAX metoden är användbar för att räkna ut hur andelarna av årskullen ska fördelas per region. Med en justering för den officiella årskullen i Sverige 2020/21, 113454.4.

Region	Årskull (2020/21) prognos ARIMAX	Fördelning Årskull (2020/21) prognos ARIMAX	Årskull ARIMAX (justerad)
Blekinge	1807	1.65%	1870
Dalarna	3147	2.87%	3256
Fyrbodal	2917	2.66%	3018
Gotland	536	0.49%	554
Gävleborg	3060	2.79%	3166
Göteborg	9380	8.55%	9703
Halland	3768	3.44%	3898
Jämtland	1406	1.28%	1455
Jönköping	4071	3.71%	4211
Kalmar	2741	2.50%	2835
Kronoberg	2349	2.14%	2431
Norrbottn	2592	2.36%	2681
Skaraborg	2893	2.64%	2993
Skåne	14216	12.96%	14707
Stockholm	24984	22.78%	25847
Södermanland	3246	2.96%	3358
Uppsala	3871	3.53%	4004
Värmland	2935	2.68%	3036
Västerbotten	2813	2.56%	2910
Västernorrland	2688	2.45%	2781
Västmanland	3002	2.74%	3106
Älvsborg	2698	2.46%	2792
Örebro	3392	3.09%	3509
Östergötland	5154	4.70%	5332
Sverige	109667	100%	113454.5

Resultat för årskullar 2020/21, andelar uträknade från ARIMAX-modellen och har sedan justerats med den officiella Årskullen beräknad från Skolverkets nationella statistik

För frågor kontakta: edvard.aberg@ungforetagsamhet.se