

Projekt MA 1487

Temperaturförändringar i Sverige senaste 70 åren

Rapport gjord utifrån årsmedeltemperaturen som uppmätts vid 3 olika mätstationer mellan åren 1952 - 2022

1. Beskriv data

Jag har valt att titta på data som visar lufttemperaturen vid tre olika mätstationer i Sverige, från norr till söder. För att få fram årsmedeltemperaturen vid varje station har jag adderat varje månads medeltemperatur och dividerat med tolv. Detta enligt hur SMHI gör sina beräkningar av årsmedeltemperaturen¹.

Anledningen till valet av min data är att man nuförtiden ofta hör om hur klimatet har blivit varmare med åren. Därför tycker jag det är intressant att se ifall dessa påståenden också stämmer överens med mina undersökningar.

Under arbetets gång har det för mig framkommit en del data som blivit extra intressant att titta på. Dels ville jag titta på om variablerna från de tre olika stationerna följer samma mönster. Min hypotes är att det vid alla mätstationer sker en temperaturökning över tid, alltså att årsmedeltemperaturen ökar med tiden.

En observation jag gjort är att det finns en stark korrelation mellan februaris medeltemperatur och årsmedeltemperaturen. Man skulle alltså kunna anta att om det bli en hög medeltemperatur i februari så blir det även en hög årstemperatur. Detta samband tittar jag närmare på i uppgift 4.

Min data är tagen från SMHI Open API. Datan hämtas dynamiskt från SMHI URL och sparas som CSV-fil. Varje månad/variabel extraheras från CSV filen till en egen Data Frame som jag sedan kan jobba utefter.

Så häng med och ta del av lite roliga klimatstatistik och antaganden.

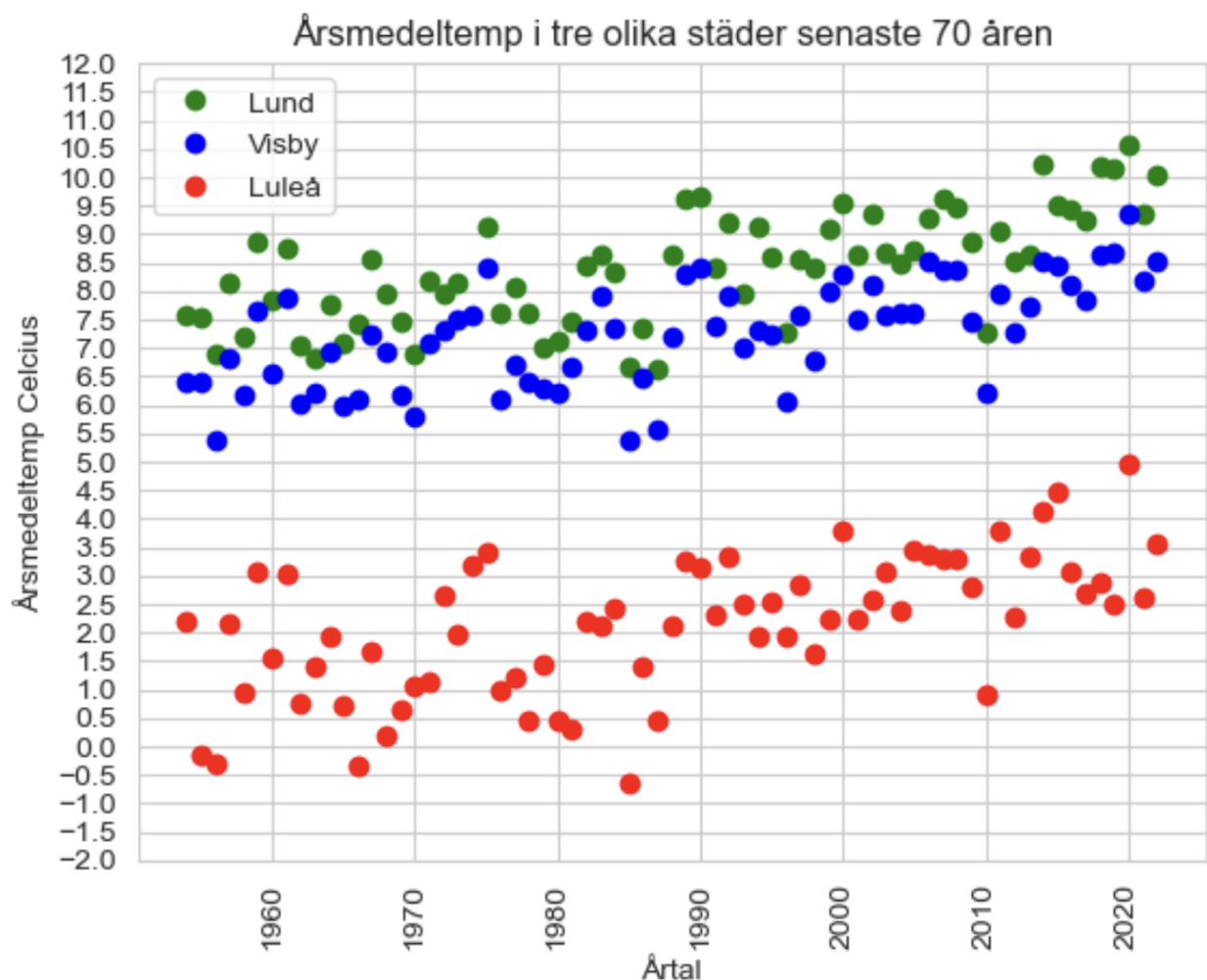
¹

<https://www.smhi.se/kunskapsbanken/klimat/normaler/hur-beraknas-medeltemperatur-1.392>

Jag har valt en mätstation i norr, Luleå, en på Gotland, Visby och en i söder , Lund, för att få spridning över hela landet. Det finns drygt 60 datapunkter i varje variabel. Enheten på x axeln är årtal och enheten på y-axeln är årsmedeltemperaturen i grader celsius. Här ett utdrag från Lunds data frame. Total = Årsmedeltemperaturen, Representativ månad = Årtal

Representativ månad	Temp jan	Temp feb	Temp march	Temp april	Temp may	Temp june	Temp july	Temp aug	Temp sep	Temp okt	Temp nov	Temp dec	total
2020	5.0	4.9	4.6	8.6	11.4	17.9	16.8	19.9	15.2	10.5	7.6	4.2	10.550000
2021	0.5	-0.5	4.1	6.7	11.6	18.7	20.4	16.8	15.1	10.6	6.8	1.4	9.350000
2022	3.3	3.7	3.9	7.2	12.7	17.2	18.4	20.1	13.8	12.0	7.2	0.7	10.016667

Grafen här nedanför visar årsmedeltemperaturen i de tre olika städerna. Som väntat ligger Luleå betydligt lägre i temperaturer, medan Lund som ligger längst söderut har flest mätningar med högst temperaturer. Det finns inga direkta outliers och min hypotes om att årsmedeltemperaturen stiger med tiden i hela Sverige tycks stämma då vi ser en lätt positiv lutning på alla variabler, vilket tyder på att ju senare årtal, ju högre medeltemperatur.



2. Beskrivande Statistik

Nedan följer lite olika statistik från de tre olika mätstationerna.

Det man kan se är att standardavvikelsen för Lund och Visby ligger mellan 1.2 och 2.5 grader varje månad och medan den totala std är något lägre, ca. 0.9, vilket tyder på att årsmedeltemperaturen har mindre avvikelser i temperaturskillnader än enskilda månader. Man skulle kunna anta att vissa månader går medeltemperaturen lite mer crazy men att det rätar ut sig och blir mer jämnt, sett över ett helt år.

Om man tittar på Luleå så ligger standardavvikelsen på vissa månader betydligt högre, som högst i februari månad då standardavvikelsen ligger på hela 4.5 grader. Detta är något man också kan se på grafen från uppgift 1, där plottarna för Luleå har en större spridning än de andra två mätstationerna.

LUND - Beskrivande statistik

	Temp jan	Temp feb	Temp march	Temp april	Temp may	Temp june	
count	68.000000	68.000000	68.000000	68.000000	68.000000	68.000000	
mean	0.064706	0.001471	2.436765	6.695588	11.826471	15.670588	
std	2.572400	2.723447	2.151147	1.547114	1.388935	1.337636	
min	-6.800000	-6.400000	-2.200000	3.300000	9.000000	12.400000	
25%	-1.600000	-1.625000	0.900000	5.600000	10.900000	14.775000	
50%	0.100000	-0.350000	2.800000	6.650000	11.700000	15.500000	
75%	1.900000	1.850000	4.100000	7.825000	12.800000	16.475000	
max	5.000000	5.400000	6.300000	10.300000	16.400000	18.900000	

	Temp july	Temp aug	Temp sep	Temp okt	Temp nov	Temp dec	total
count	68.000000	68.000000	68.000000	68.000000	68.000000	68.000000	68.000000
mean	17.605882	17.160294	13.535294	9.251471	4.886765	1.673529	8.400735
std	1.664635	1.521894	1.248493	1.388061	1.477079	2.134690	0.989865
min	14.200000	14.100000	11.000000	5.900000	1.400000	-5.000000	6.608333
25%	16.350000	16.200000	12.850000	8.350000	4.100000	0.575000	7.550000
50%	17.550000	17.200000	13.750000	9.300000	5.000000	2.050000	8.500000
75%	18.700000	18.025000	14.300000	10.025000	5.800000	3.300000	9.152083
max	21.700000	20.900000	16.600000	12.000000	7.600000	6.200000	10.550000

VISBY - Beskrivande statistik

	Temp jan	Temp feb	Temp march	Temp april	Temp may	Temp june
count	69.000000	69.000000	69.000000	69.000000	69.000000	69.000000
mean	-0.466667	-1.179710	0.620290	4.497101	9.779710	14.391304
std	2.200312	2.707755	2.116294	1.434345	1.512262	1.407933
min	-6.500000	-7.400000	-3.700000	1.100000	6.600000	11.500000
25%	-1.900000	-2.900000	-1.100000	3.500000	8.800000	13.200000
50%	-0.100000	-1.200000	1.000000	4.400000	9.800000	14.300000
75%	1.200000	1.100000	2.000000	5.600000	10.700000	15.300000
max	4.200000	4.300000	4.400000	7.300000	14.400000	17.900000

	Temp july	Temp aug	Temp sep	Temp okt	Temp nov	Temp dec	total
count	69.000000	69.000000	69.000000	69.000000	69.000000	69.000000	69.000000
mean	16.908696	16.427536	12.523188	8.231884	4.149275	1.194203	7.256401
std	1.537065	1.339763	1.330651	1.382718	1.509833	1.890758	0.929101
min	14.000000	13.800000	9.500000	4.800000	0.300000	-3.600000	5.366667
25%	15.700000	15.400000	11.600000	7.400000	3.400000	0.100000	6.391667
50%	16.900000	16.400000	12.600000	8.300000	4.200000	1.100000	7.325000
75%	17.800000	17.300000	13.400000	8.900000	5.300000	2.600000	7.958333
max	20.400000	20.100000	16.100000	11.300000	7.500000	5.600000	9.366667

LULEÅ - Beskrivande statistik

	Temp jan	Temp feb	Temp march	Temp april	Temp may	Temp june
count	69.000000	69.000000	69.000000	69.000000	69.000000	69.000000
mean	-10.086957	-9.926087	-5.152174	0.336232	6.730435	12.910145
std	3.863818	4.543730	3.151779	1.523345	1.566215	1.504615
min	-18.500000	-21.400000	-14.300000	-3.600000	3.600000	9.900000
25%	-13.100000	-12.700000	-7.500000	-0.500000	5.700000	11.900000
50%	-10.000000	-9.900000	-4.900000	0.400000	6.700000	12.800000
75%	-7.000000	-6.300000	-2.400000	1.400000	7.800000	13.900000
max	-3.000000	-1.300000	-0.100000	3.800000	10.200000	16.600000

	Temp july	Temp aug	Temp sep	Temp okt	Temp nov	Temp dec	total
count	69.000000	69.000000	69.000000	69.000000	69.000000	69.000000	69.000000
mean	15.965217	14.097101	8.876812	2.789855	-3.336232	-7.686957	2.126449
std	1.514715	1.235501	1.377567	2.011302	2.876628	3.896979	1.222391
min	13.500000	11.100000	5.600000	-3.500000	-9.400000	-17.100000	-0.633333
25%	14.800000	13.400000	8.100000	1.600000	-5.300000	-10.200000	1.191667
50%	15.700000	14.000000	8.900000	3.000000	-3.400000	-7.300000	2.233333
75%	16.700000	14.700000	10.000000	3.800000	-1.400000	-5.000000	3.075000
max	20.300000	17.600000	11.500000	7.400000	2.100000	-0.600000	4.958333

Korrelation

Det finns några korrelationer som är extra intressanta att titta på. Det första man kan se är att det i samtliga månader, i samtliga städer, finns en viss positiv korrelation, detta gäller även för månader och den totala årsmedeltemperaturen. Det jag fann extra intressant var att det även i samtliga städer är högst korrelation mellan de fyra första månadernas månadstemperatur och årstemperaturen, detta kan man tydligt se på SNS heatmap för varje stad. Där jan, feb, mar och april är lysande orangea, i synnerhet högst korrelation är det mellan februari och årsmedeltemperaturen. Där ligger korrelationen i både Lund, Visby och Luleå på över 0.7.

Man ska nog vara mer insatt i meteorologiska samband än vad jag är för att kunna dra riktiga slutsatser. Men om vi enbart tittar på den data vi har, skulle man kunna säga att om medeltemperaturen i februari månad är hög, så blir också årstemperaturen hög. Därför kommer jag titta vidare på detta samband i senare uppgifter.

Däremot är korrelationen mellan temperaturen i februari och temperaturen i juni och juli väldigt låg. Vilket tyder på att vi inte kan se något samband dessa månader emellan. Man kan alltså inte säga "Vi fick en varm februari i år, då borde det bli en varm sommar", baserat på enbart korrelationen

Om man tittar på korrelationen mellan de olika mätstationerna så är det ett väldigt starkt samband. Korrelationen mellan två mätstationer kan ni se nedan. Detta tyder på att alla tre mätstationer följer varandra när det gäller årsmedeltemperaturen. Alltså blir det en hög årstemperatur i Lund så blir det även det i Visby, i alla fall om man ska tro korrelationen. Minst samband har Lund och Luleå, vilket man skulle kunna anta är också för att de ligger långt ifrån varandra.

Korrelation mellan Lund och Visby

Temp feb	0.946559
Temp aug	0.942949
Temp okt	0.918190
total	0.965012

Korrelation mellan Lund och Luleå

Temp feb	0.687949
Temp aug	0.589279
Temp okt	0.557373
total	0.839766

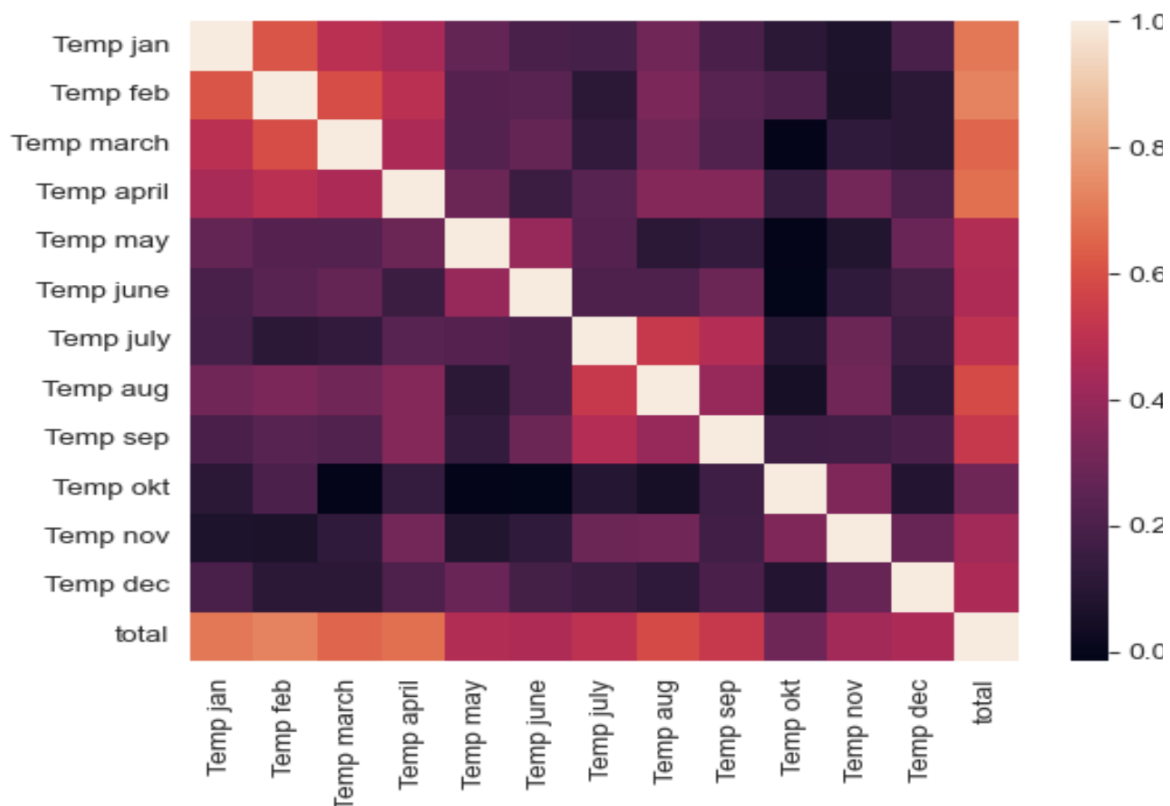
Korrelation mellan Luleå och Visby

Temp feb	0.820782
Temp aug	0.665300
Temp okt	0.633762
total	0.876557

Korrelationen mellan olika månads temperaturer och årsmedeltemperatur i Lunds stad illustrerad i en tabell

	Temp jan	Temp feb	Temp march	Temp april	Temp may	Temp june	Temp july	Temp aug	Temp sep	Temp okt	Temp nov	Temp dec	total
Temp jan	1.000000	0.614765	0.490728	0.441519	0.264109	0.194885	0.184921	0.300323	0.197393	0.108695	0.071485	0.193187	0.698167
Temp feb	0.614765	1.000000	0.594786	0.490183	0.231011	0.240384	0.111867	0.330333	0.234957	0.200192	0.068793	0.112171	0.722855
Temp march	0.490728	0.594786	1.000000	0.453048	0.224514	0.270106	0.129524	0.299159	0.222082	-0.008491	0.126842	0.108124	0.653453
Temp april	0.441519	0.490183	0.453048	1.000000	0.293028	0.158315	0.231421	0.351927	0.351819	0.137790	0.306618	0.204597	0.677886
Temp may	0.264109	0.231011	0.224514	0.293028	1.000000	0.401458	0.229745	0.109313	0.133638	-0.008846	0.084128	0.282393	0.471549
Temp june	0.194885	0.240384	0.270106	0.158315	0.401458	1.000000	0.204319	0.205511	0.291447	-0.014365	0.126407	0.181780	0.458713
Temp july	0.184921	0.111867	0.129524	0.231421	0.229745	0.204319	1.000000	0.532032	0.476109	0.093336	0.289398	0.152722	0.501853
Temp aug	0.300323	0.330333	0.299159	0.351927	0.109313	0.205511	0.532032	1.000000	0.398376	0.048037	0.301662	0.120728	0.586090
Temp sep	0.197393	0.234957	0.222082	0.351819	0.133638	0.291447	0.476109	0.398376	1.000000	0.166277	0.167954	0.195747	0.529469
Temp okt	0.108695	0.200192	-0.008491	0.137790	-0.008846	-0.014365	0.093336	0.048037	0.166277	1.000000	0.339935	0.087810	0.294815
Temp nov	0.071485	0.068793	0.126842	0.306618	0.084128	0.126407	0.289398	0.301662	0.167954	0.339935	1.000000	0.278504	0.429215
Temp dec	0.193187	0.112171	0.108124	0.204597	0.282393	0.181780	0.152722	0.120728	0.195747	0.087810	0.278504	1.000000	0.449324
total	0.698167	0.722855	0.653453	0.677886	0.471549	0.458713	0.501853	0.586090	0.529469	0.294815	0.429215	0.449324	1.000000

Korrelationen mellan olika månads temperaturer och årsmedeltemperatur i Lunds stad illustrerad av en heatmap



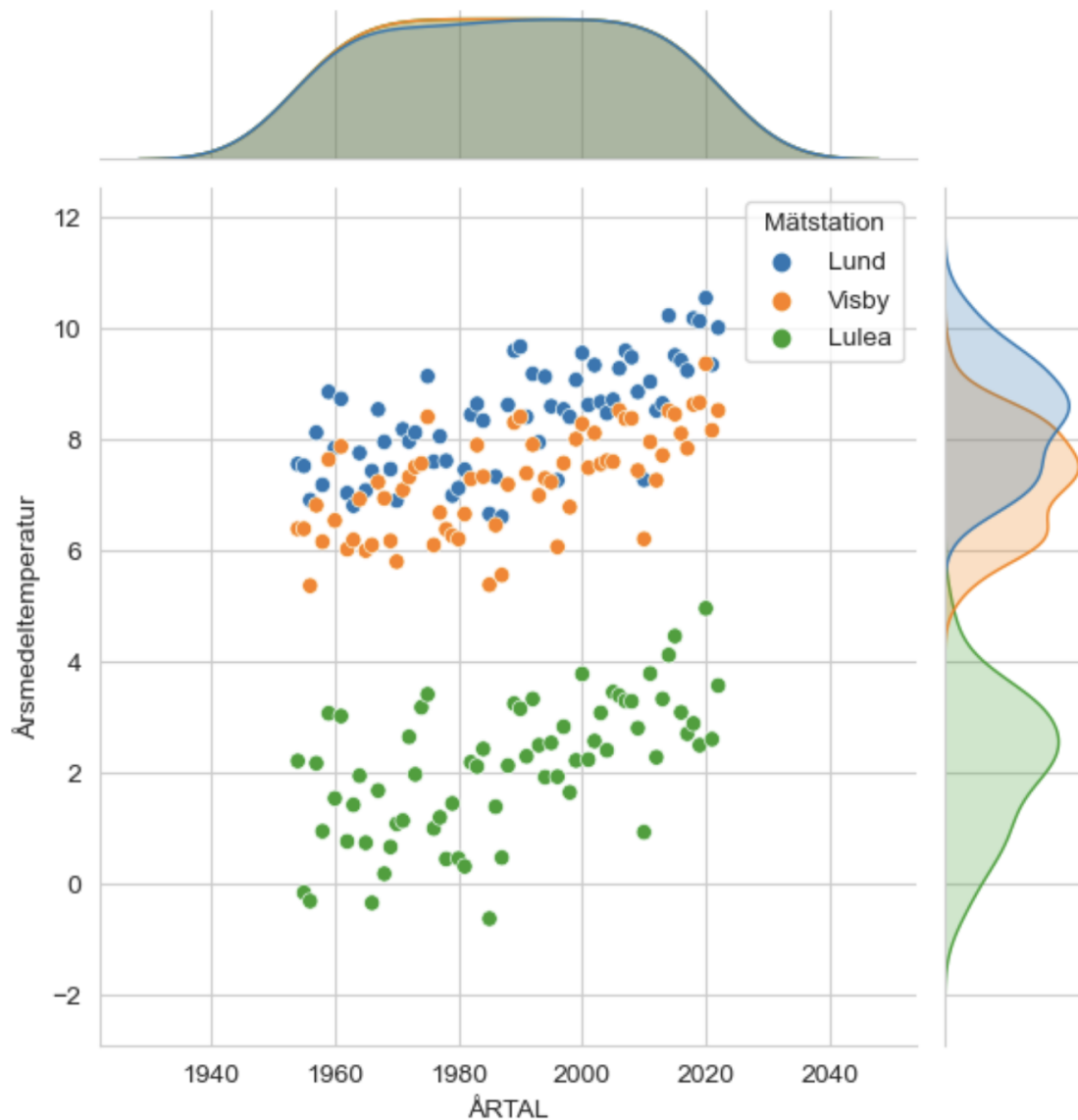
För fler illustrationer av de övriga två städerna och deras kolumners korrelation se ipynb.

3. Beskrivande Plottar

Här nedan följer lite fler grafer för att beskriva sambanden mellan min data.

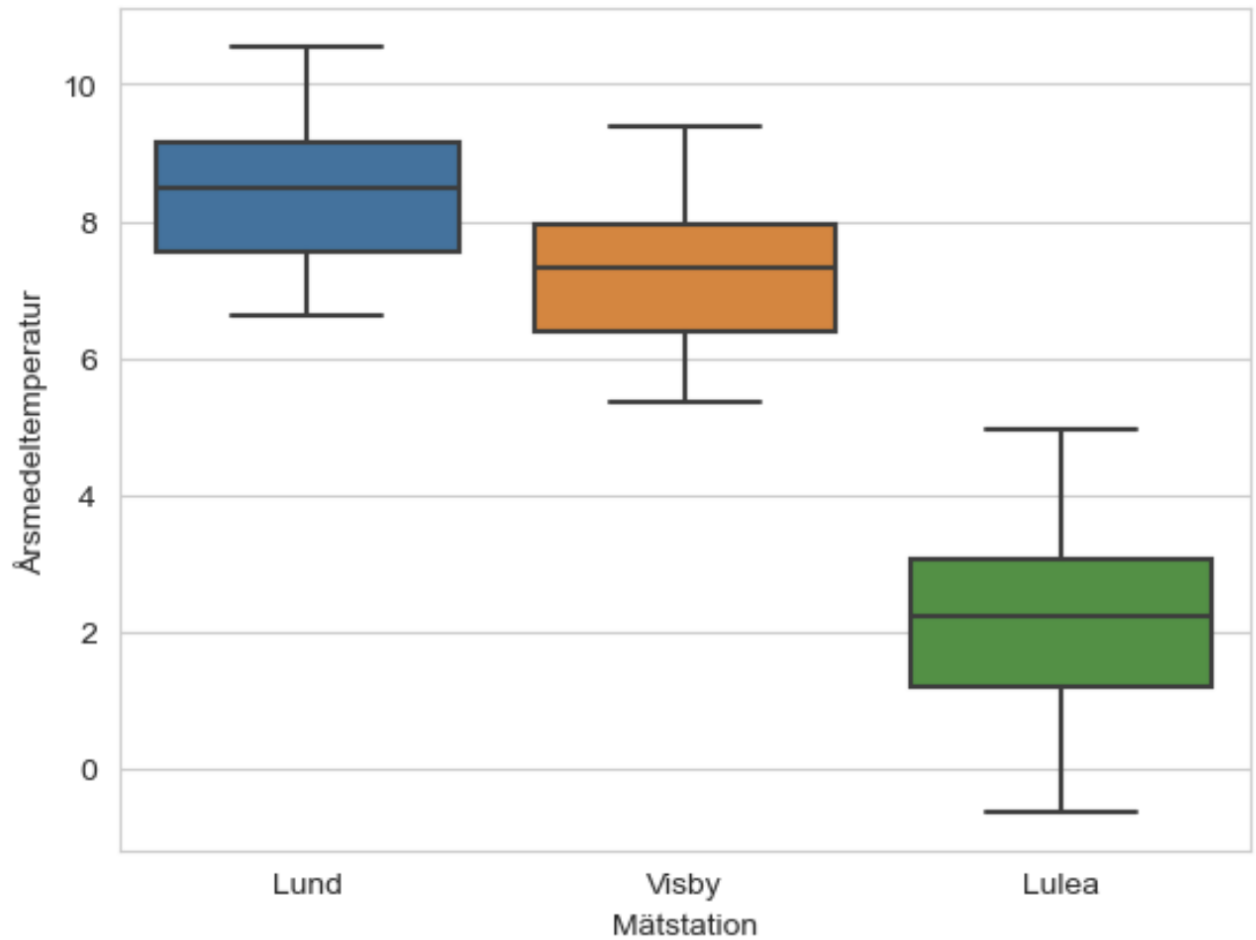
Den första grafen visar fördelningen av årsmedeltemperaturen i de olika städerna. Man kan se att alla variabler liknar en någorlunda normal fördelning på y axeln, alltså årsmedeltemperaturen. Jag kan tycka att Luleås normalfördelning är något flackare än Lund och Visby, vilket styrks av att Luleås std var större.

jointplot



Lådogram

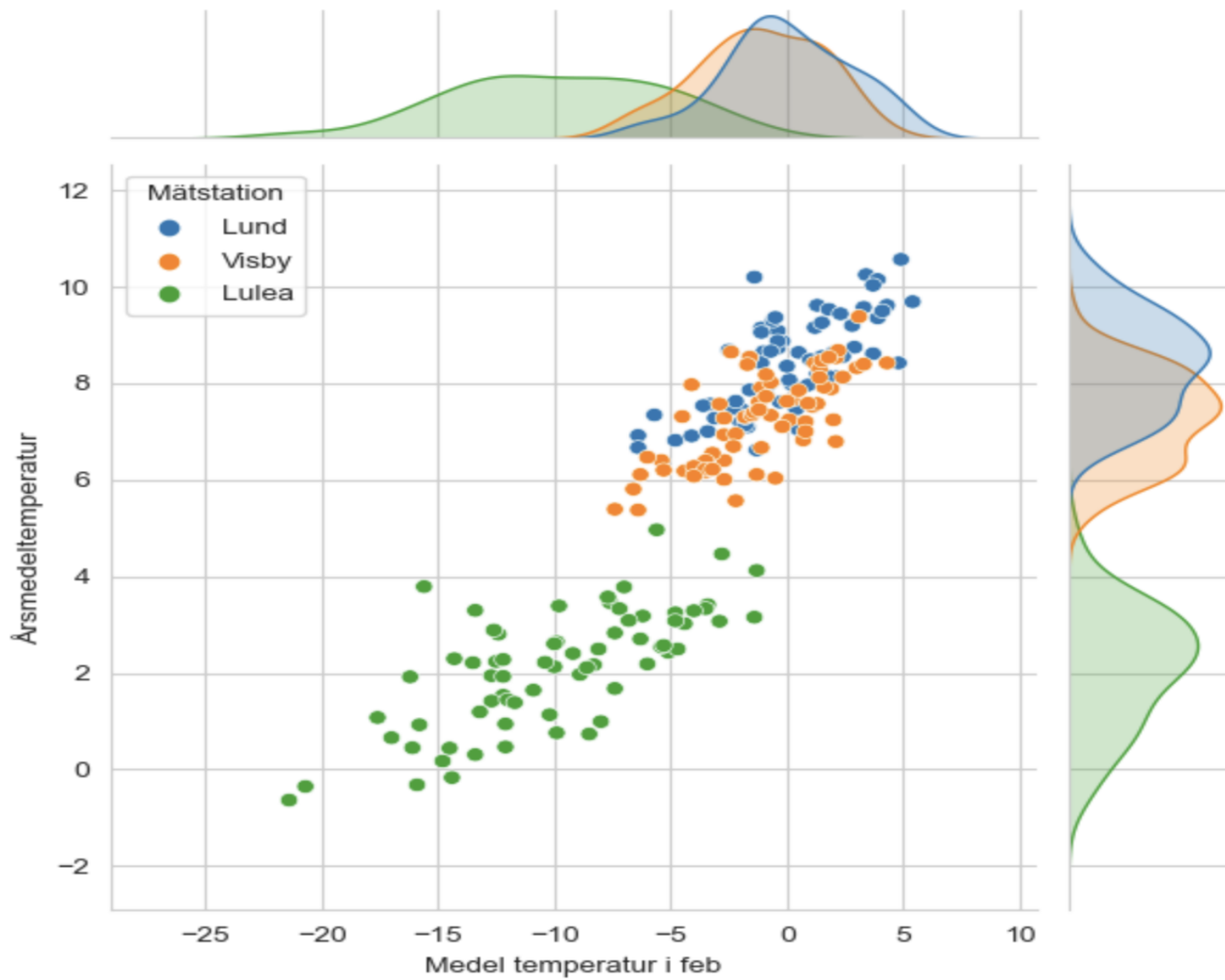
I mitt Lådogram kan man se att det inte finns några outliers och att det är rätt så normalt fördelat mellan data när det kommer till årstemperaturen. Som jag nämnde innan har Luleå det största spannet av årstemperaturer.



Joint plot

Den tredje grafen visar sambandet mellan årstemperaturen på y-axeln och februaris medeltemperatur på x-axeln.

Även här ser man att Luleås normalfördelning är flackast och att det finns ett positivt samband mellan y och x axeln.



4. Linjär regression

Till denna uppgift har jag valt att arbeta med en av mina variabler, nämligen Lunds mätstation. Den linjära regressionen jag kommer att titta på är mellan en hög temperatur i februari och en hög årsmedeltemperatur. Jag har valt temp i februari till x, alltså den variabeln som används för att prediktera. Den variabel som ska predikteras blir den totala års temperaturen som då blir på y axeln. Anledningen till att jag valt dessa två variabler är för att de är dom två kolumner som har högst korrelation till varandra, 0.7.

Efter att ha räknat ut variablerna *a* och *b* och dess punk skattnings konfidensintervall, där alpha är satt till 0.05 (95 % intervall) får jag fram dessa värden. Dess R-squared värde, vilket är värdet som visar hur väl regressionslinjen approximerar datapunkterna, där 1 är perfekt korrelation så ligger denna mätning på 0.5 vilket jag ändå kan tycka är värt en linjär regression.

$a = 8.400348927580474$

$b = 0.2627292452773423$

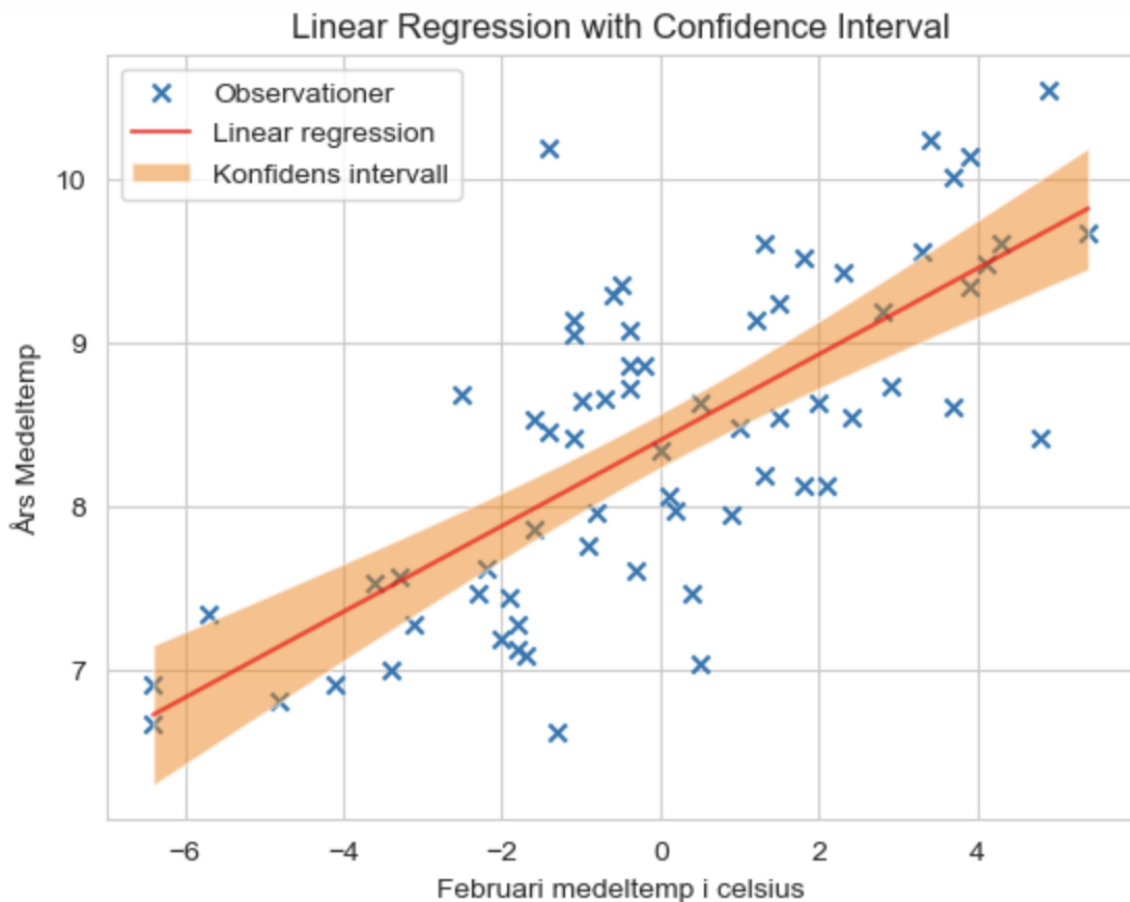
Confidence Intervals:

	0	1
Intercept	8.233490	8.567207
feb	0.201006	0.324452

samma data framtagen med stats models api

OLS Regression Results

Dep. Variable:	total		R-squared:	0.523		
Model:	OLS		Adj. R-squared:	0.515		
Method:	Least Squares		F-statistic:	72.23		
Date:	Fri, 30 Jun 2023		Prob (F-statistic):	3.40e-12		
Time:	10:31:13		Log-Likelihood:	-70.158		
No. Observations:	68		AIC:	144.3		
Df Residuals:	66		BIC:	148.8		
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	8.4003	0.084	100.515	0.000	8.233	8.567
feb	0.2627	0.031	8.499	0.000	0.201	0.324
Omnibus:	1.348	Durbin-Watson:	1.870			
Prob(Omnibus):	0.510	Jarque-Bera (JB):	0.804			
Skew:	0.241	Prob(JB):	0.669			
Kurtosis:	3.225	Cond. No.	2.70			



5. Transformerade data

Till denna jämförelse har jag valt att gå tillbaka till mätningen för hur årstemperaturen förändras över tid.

x = årtal

y = årsmedeltemperatur

variabel = Lund mätstation

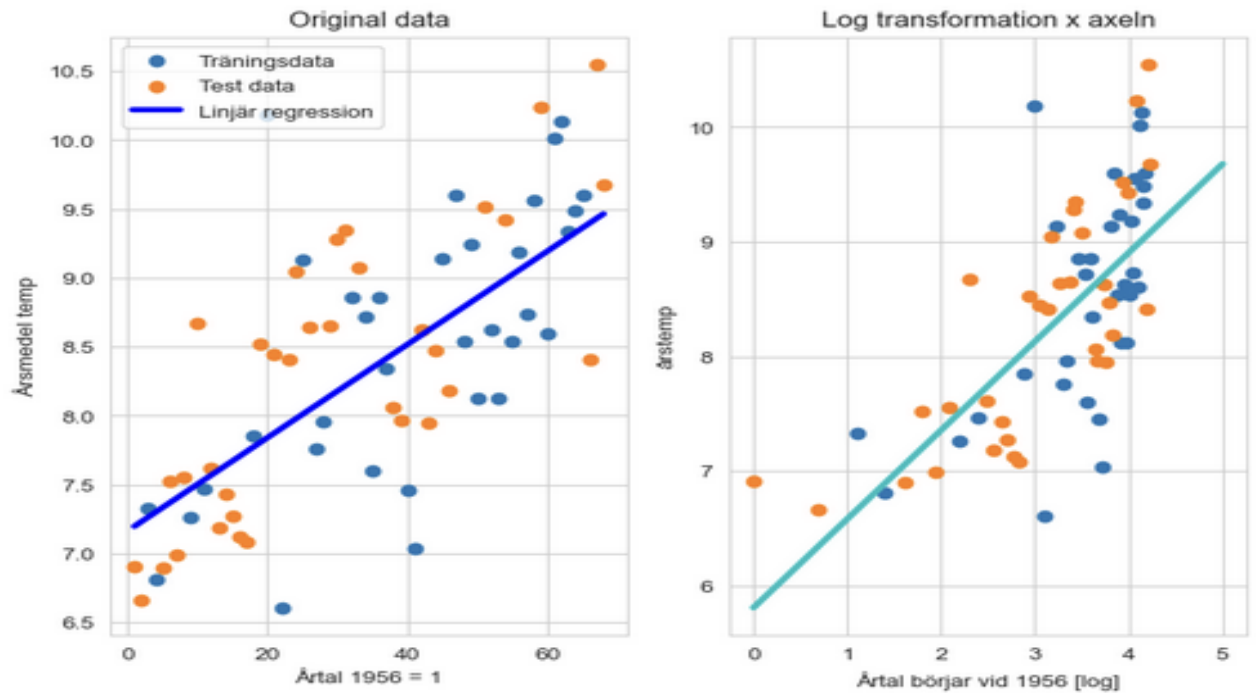
Här kan vi också se ett visst positivt linjärt samband även om det inte är lika starkt. Jag provar att transformera datan i tre ytterligare modeller. Logaritmera x-axeln, logaritmera enbart y-axeln och sen log-log modell som logaritmerar båda axlarna.

Vad min data visar är att det är väldigt marginell skillnad mellan de olika transformationerna. Vilket tyder på att en logaritmisk transformation kanske inte är det bästa valet. Eftersom det är lättast att göra en linjär modell så skulle jag i detta fallet inte använda logaritmisk transformerade data.

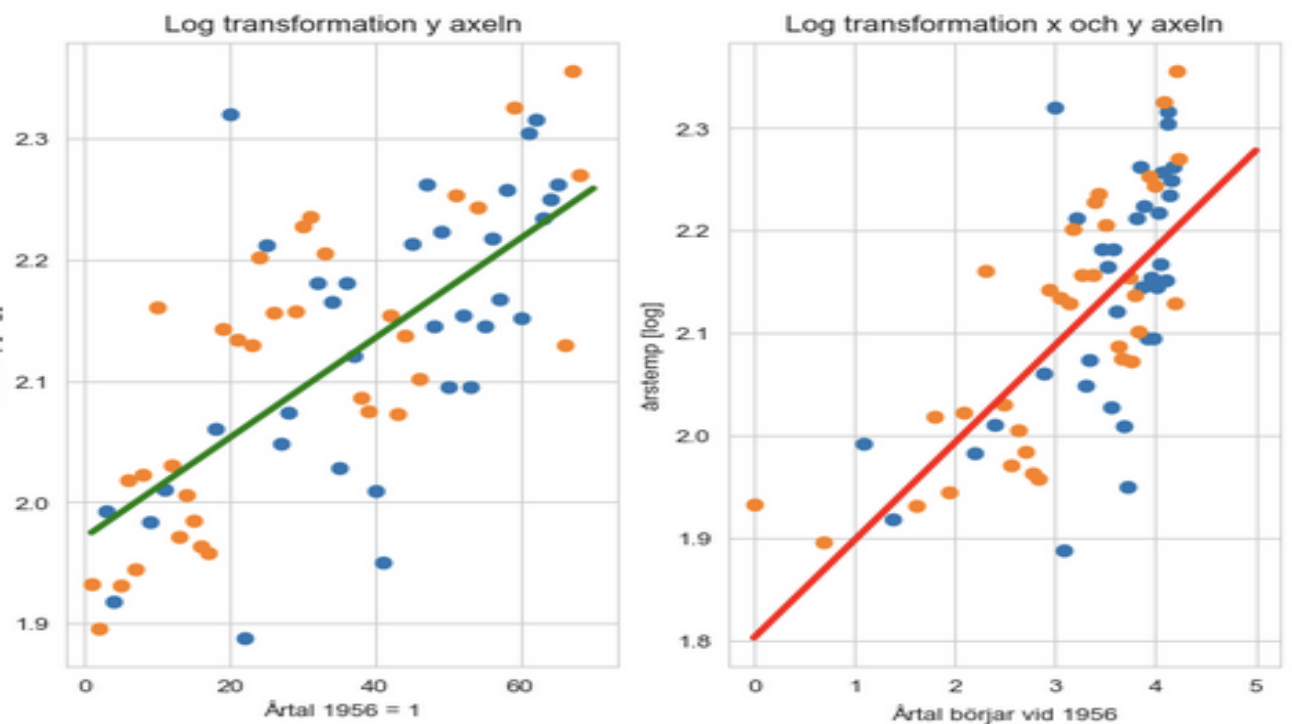
När det kommer till värdena på MSE så har den linjära och log-log modell något bättre värden:

```
Mean squared error linjär regression:    0.42592655218220393
Mean squared error log transformerad x:  0.4466438122431898
Mean squared error log transformerad y:  0.44160585613444486
Mean squared error log-log transformerad: 0.4357877657611927
```

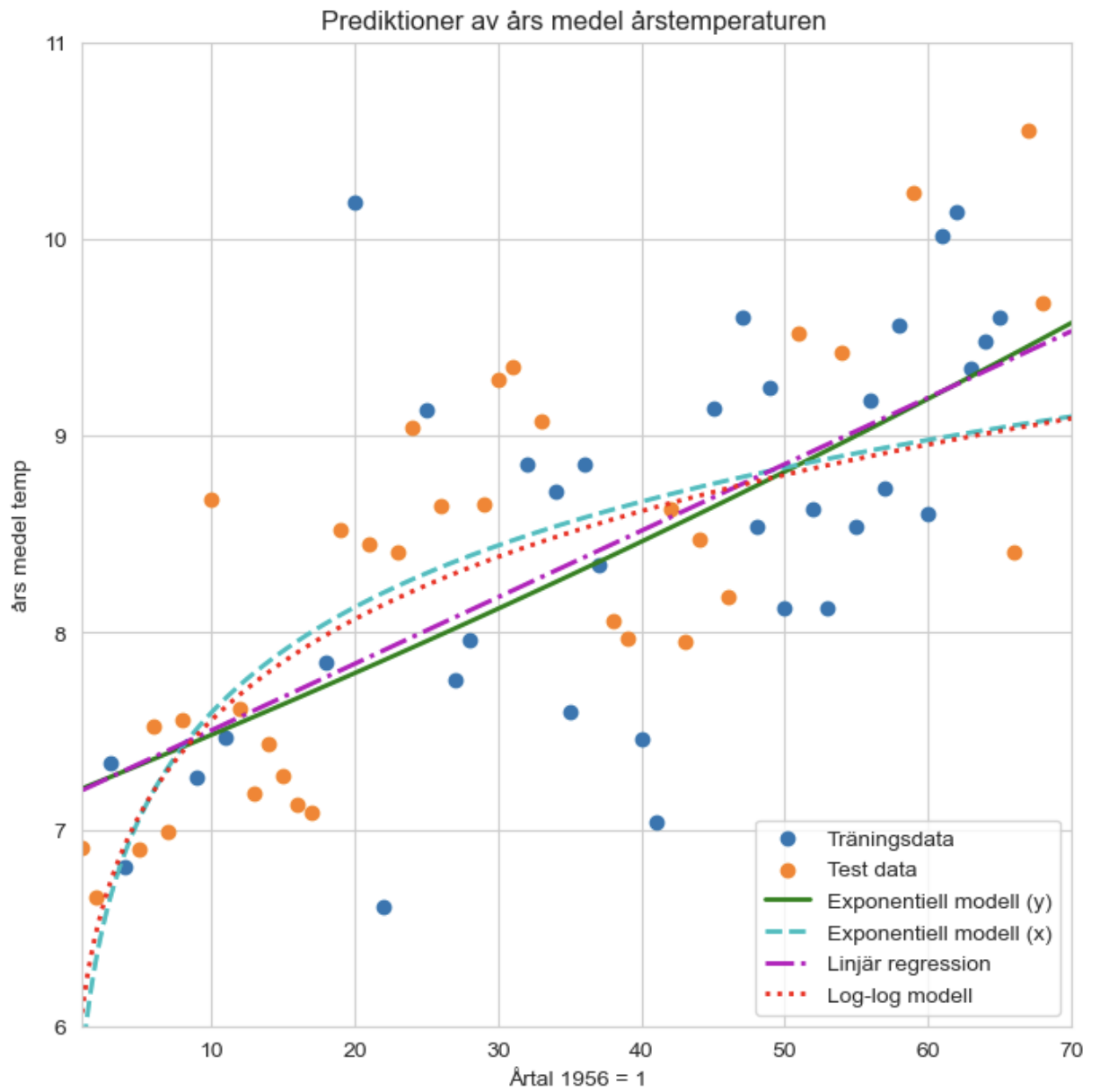
I graferna här nedanför ser vi de linjära regressionerna med sin egen data. De första två visar original data och sedan den logaritmerade x axeln:



Nästa graf visar den linjära regressionen där y axeln är logaritmerad och där både y och x är det



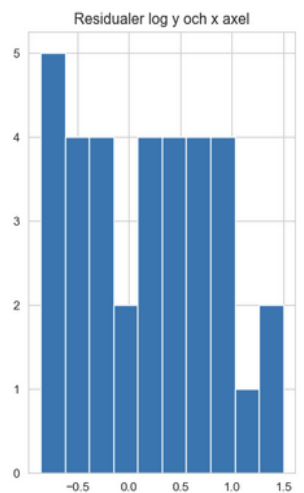
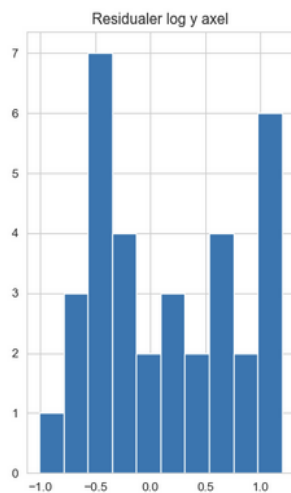
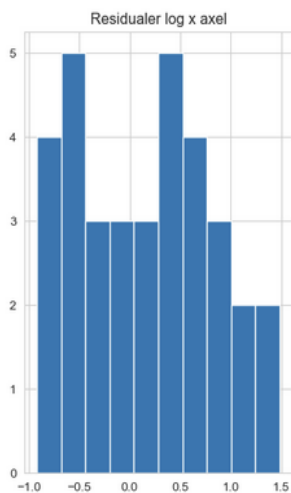
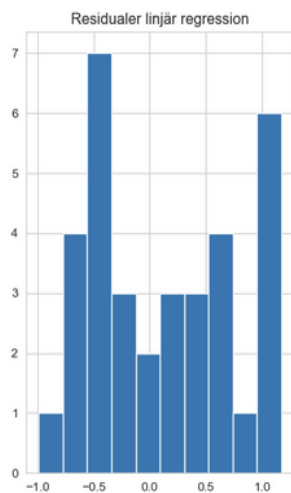
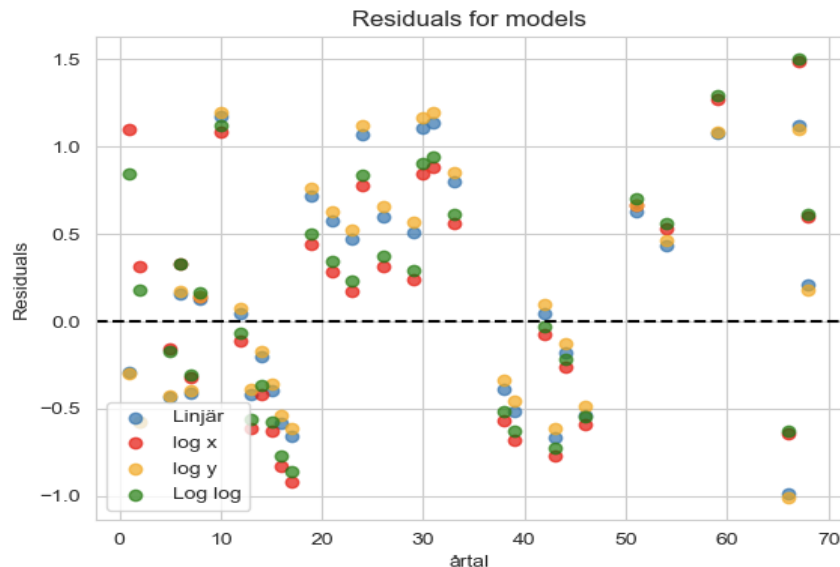
Denna graf visar de olika regressionerna tillsammans med original datat.



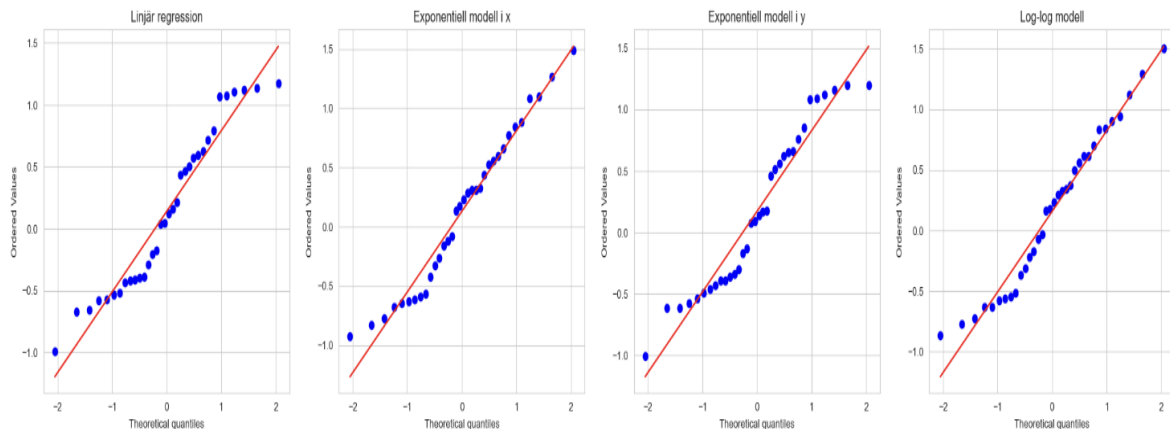
6. Residualer

När det kommer till residualerna som är skillnaden mellan de faktiska värdena och de predikterade så skulle jag nog inte tycka att några av mina modeller är optimala för att visa ett perfekt linjärt samband. Om vi ser på bilden nedan över hur residualerna är fördelade vill vi att det ska vara som en splash och inte följa några mönster.

Och en normalfördelning av residualerna tyder på att vi kan använda oss av statistiska tester för hur bra modellen fungerar. Ingen av dessa grafer skulle jag säga uppfyller dessa krav, detta tror jag främst beror på att mina test data är rätt utspridd och som nämnt innan kan det i sin tur ha att göra med att vädret är väldigt oförutsägbart och därför svårare att modellera



Q-Q plottarna ser ut som följer för de fyra modellerna:



När det kommer till R-squared värdet och residualvariansen, ser min tabell ut så här för de fyra modellerna

Var(R)

Residualvarians för log-log modell:	0.46302450112126725
Residualvarians för linjär regression:	0.45254696169359165
Residualvarians för exponentiell modell i y:	0.4692062221428477
Residualvarians för exponentiell modell i x:	0.47455905050838915

R²:

R ² för log-log modell:	0.5284936011322318
R ² för linjär regression:	0.5391630729907945
R ² för exponentiell modell i y:	0.5221986404754366
R ² för exponentiell modell i x:	0.516747756515124

Här kan vi se att det är lite bättre värden för log-log-modellen och för den linjära modellen när det kommer till både r-squared värdet och residualvariansen.

De slutsatser jag drar av dessa residuala grafer och värden är att både log-y och log-x går bort. Log-log modellen har något bättre värden men det är så marginellt att jag hade valt den linjära modellen framför det andra även efter att ha tittat på residualerna.

7. Sammanfattning

Sammanfattningsvis kan man säga att efter att ha tittat på statistik och samband mellan de olika variablerna och de olika kolumnerna i mina data frames så kan man se att all data och alla samband har en positiv lutning, som tyder på att min hypotes att det blir varmare med tiden stämmer. Vi kan dock också se att det inte finns en konstant ökning av årstemperaturen över tid och därför inte ett perfekt linjärt samband.

I SMHI egna rapporter jämför de istället medeltemperaturen av 30 års mätningar mellan 1950-1980 med medeltemperaturen åren 1990-2020, då fick de fram att medeltemperaturen ökat med 1.9 grader. Detta kanske är bättre metrologiska prediktioner, men mina grafer visar i stort sätt samma sak om än med lite spridning.

Min observation att korrelationen mellan medeltemperaturen i februari och årsmedeltemperaturen

är högst intressant och kanske till och med något att rapportera in till SMHI, ganska häftigt att man redan i februari månad kan prediktera att det kommer bli ett varmt år eller vad säger ni?