

Inlämningsuppgift 1 matematik statistik KF

Anton Lindbro

7 mars 2024

Innehåll

1	Uppgift 1	3
1.1	a)	3
1.2	b)	3
1.3	c)	3
1.4	d)	4
2	Uppgift 2	5
2.1	a)	5
2.2	b)	6
2.3	c)	6
2.4	d)	7
2.5	e)	9
2.6	f)	9

1 Uppgift 1

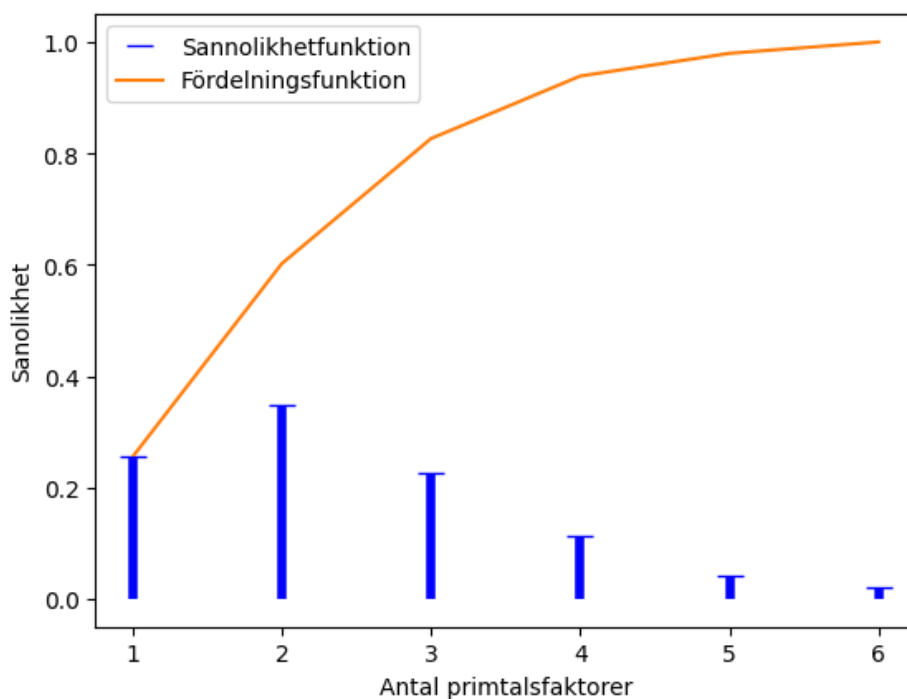
1.1 a)

För att beräkna väntevärdet för $E(X)$ behövs för det första värdet på för alla tal i det i uppgiften givna intervallet. För att ta fram dessa värden så användes den i uppgiften givna kodsnutten på alla tal i intervallet. Några ytterligare rader kod användes för att skapa en lista med alla värden för X för talen i intervallet. Dessa summerades sedan och dividerades sedan med antalet försök för att få fram $E(X) = 2,4$.

1.2 b)

För att definiera en fördelning för $E(X)$ behövs först sannolikheterna för samtliga möjliga utfall. Dessa beräknas genom att räkna hur många gånger varje utfall förekommer och sedan dividera det med totala antalet försök. Från mean kommandot får vi också samma svar $E(X) = 2,4$

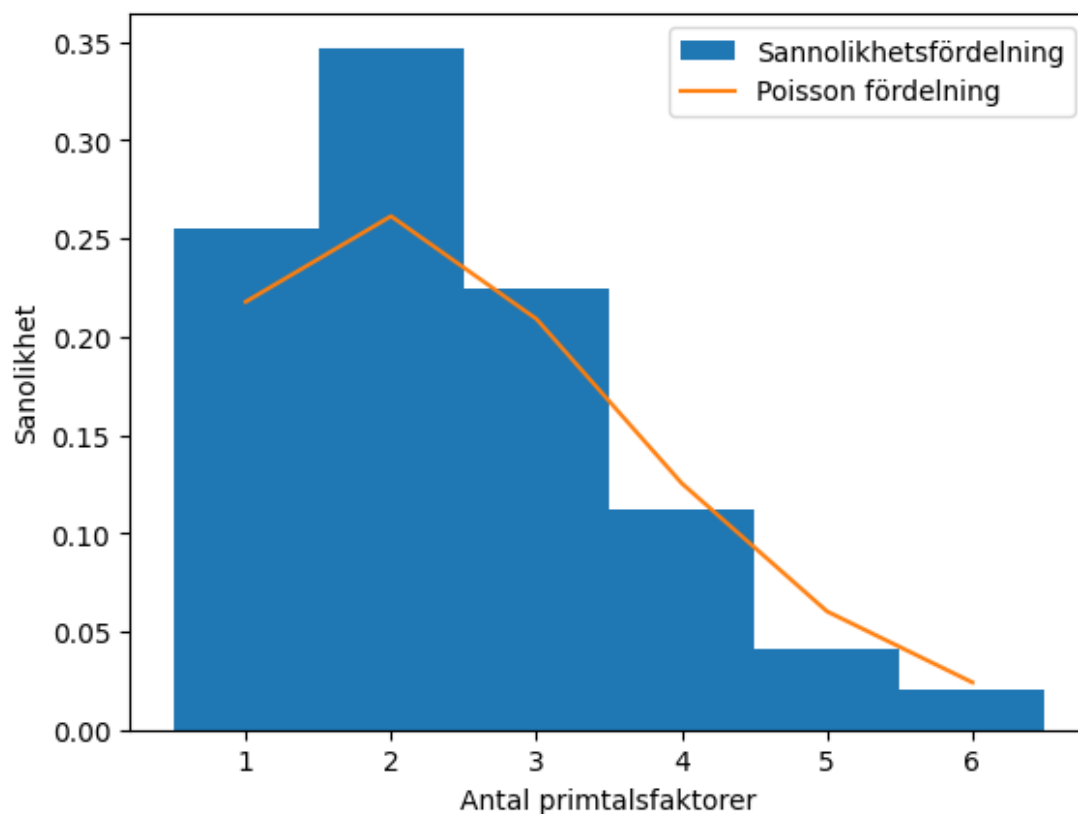
1.3 c)



Figur 1: Graf över den diskreta sannolikhetsfördelningen för X samt dess fördelningsfunktion

Med hjälp av de inbyggda funktionerna i `scipy.stats` och fördelningen som definierades i föregående uppgift erhöles övan graf.

1.4 d)



Figur 2: Graf som visar hur en poisson fördelning förhåller sig till den faktiska fördelningen

Poisson fördelningen verkar vara en bra approximation. I figuren syns det att den följer fördelningen ganska bra. verkar underskatta för det mesta men ju fler primtalfaktorer desto mer överskatta poisson fördelningen.

2 Uppgift 2

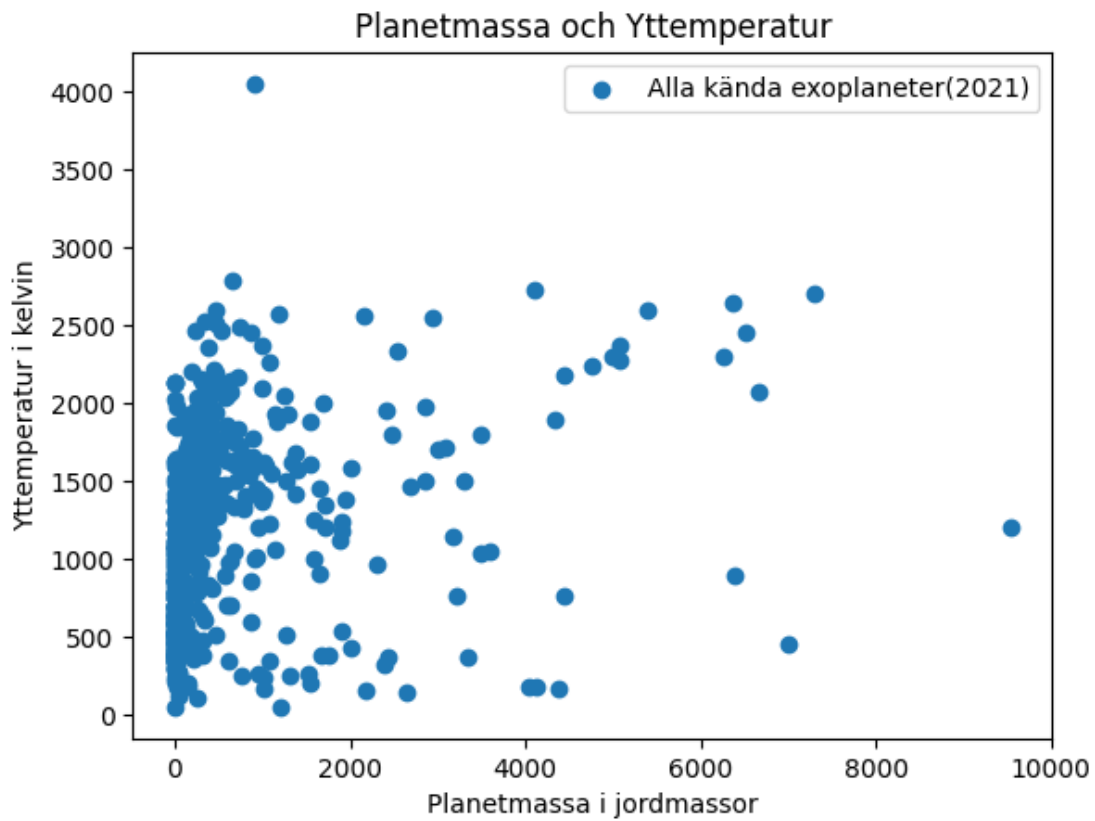
2.1 a)

Tabell 1: Statistiska grund data för alla upptäckta planeter fram till och med 2021

	Planetmassa[Jordmassor]	Yttretemperatur[Kelvin]	Stjärn radie[Solradier]	Stjärnmassa[Solmassor]
Medelvärde	725	1060	1,54	0,973
Standardavvikelse	1340	561	3,82	0,474
Minsta värde	0,02	50	0,01	0,01
Största värde	17700	4050	83,8	10,9
Median	221	961	0,960	0,960
Variationsbredd	1340	511	3,81	0,473

Med hjälp av den inbyggda funktionen `describe` i `pandas` samt lite huvudräkning erhöles övan tabell med statistiska grunddata för det givna mängden data

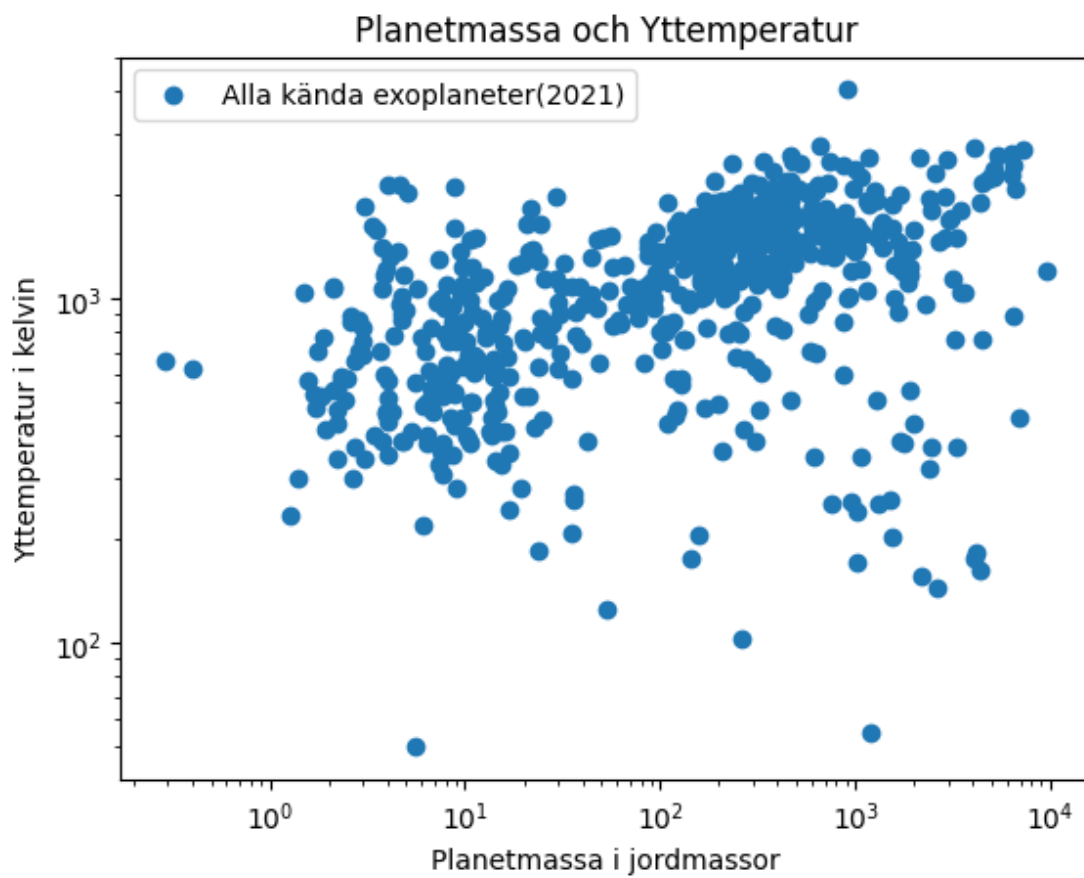
2.2 b)



Figur 3: Scatter plot över planetmassa kontra yttemperatur

2.3 c)

För att kunna få en överblick av variabler som har stor variationsbredd är användandet av logaritmerade axlar att föredra. Om axlarna på ovan graf logarimeras erhålls följande plot.



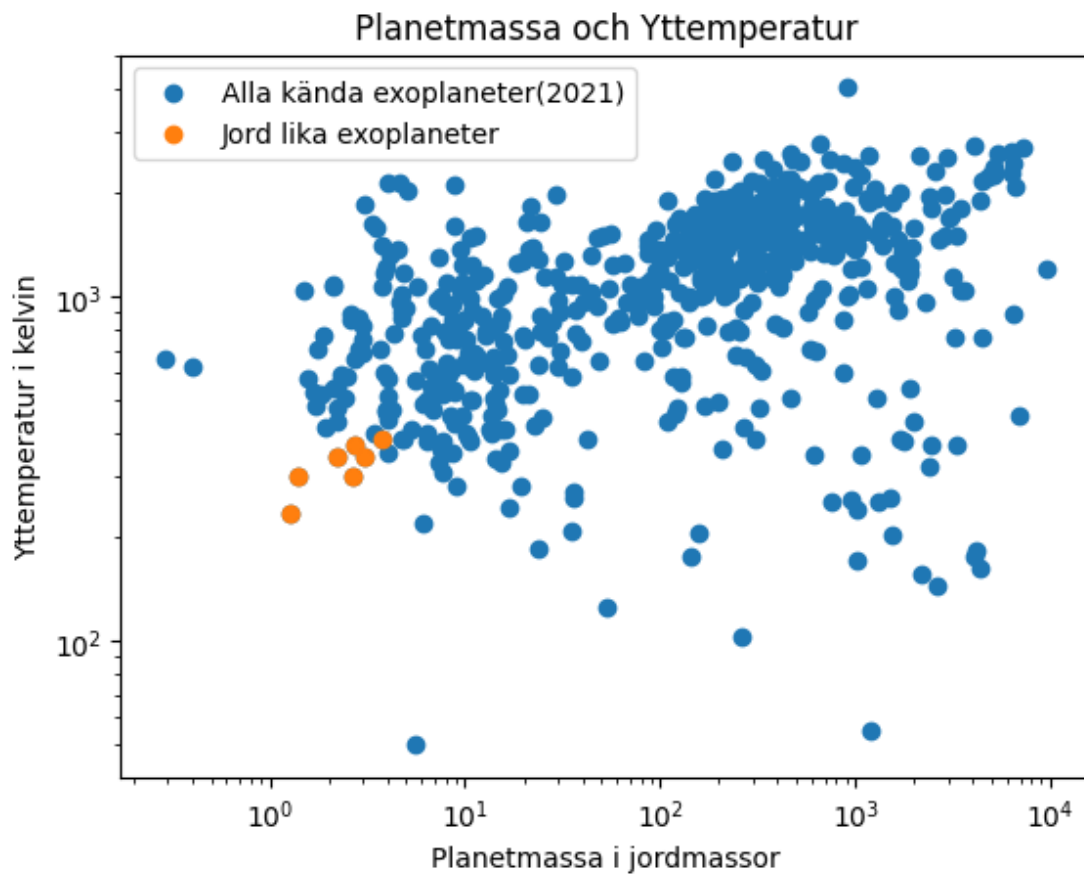
Figur 4: Log log scatter plot över planetmassa kontra yttemperatur

För att kunna få log axlar i matplotlib finns det för figurer en egenskap som heter axis scale där man med en sträng kan skala om axlarna till log skalor.

2.4 d)

Tabell 2

Planetmassa[Jordmassor]	Yttremperatur[Kelvin]	Stjärn radie[Solradier]	Stjärnmassa[Solmassor]
3,75	387	0,30	0,30
2,64	300	0,21	0,18
2,69	370	0,39	0,39
3,06	342	0,30	0,27
1,27	234	0,14	0,12
1,40	301	0,20	0,17
2,20	344	0,42	0,45



Figur 5

I tabellen ovan kan läsaren hitta intressanta data om de mest jordlika planeterna i datamängden. Dessa togs fram genom att använda den inbyggda filterfunktionen för pandas dataframes. Filtret som användes va:

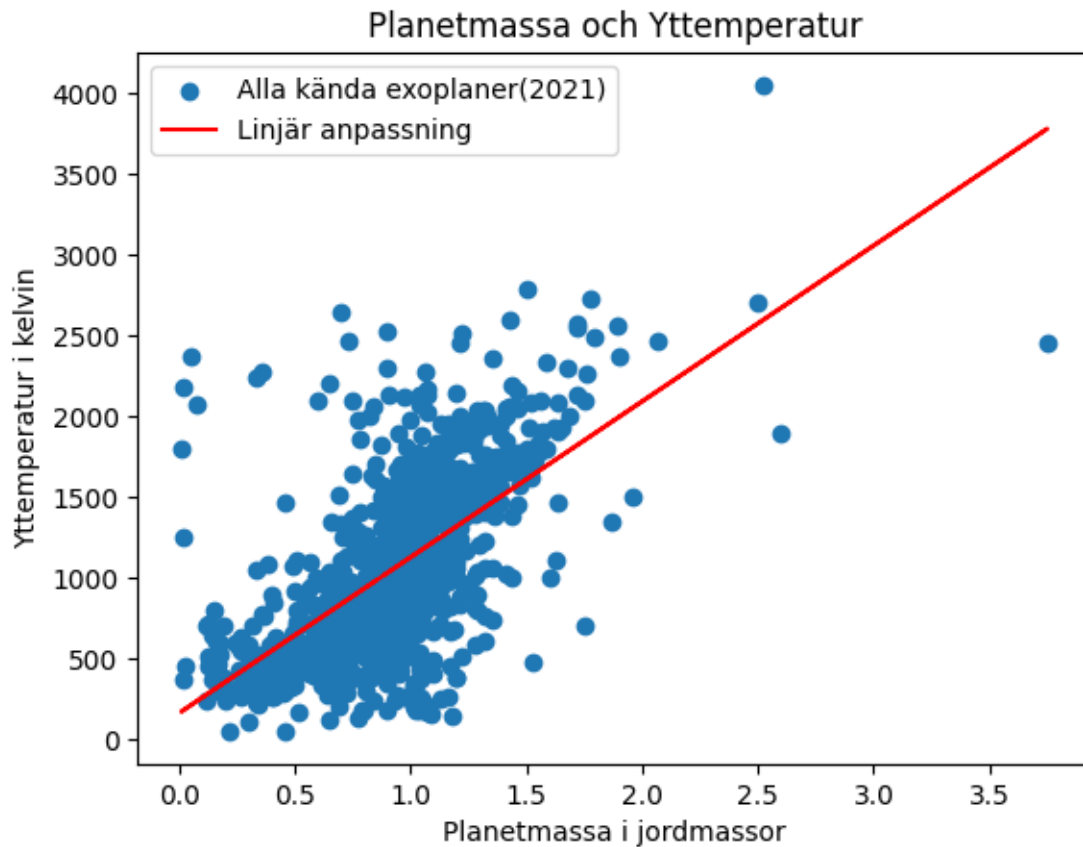
$$0.5 < Mass < 4 \quad (1)$$

$$200 < EquilibriumTemperature < 400 \quad (2)$$

Sedan för att se hur dom ligger i fördelningen med alla planeter färglades dom orange som läsaren kan se i figur 5

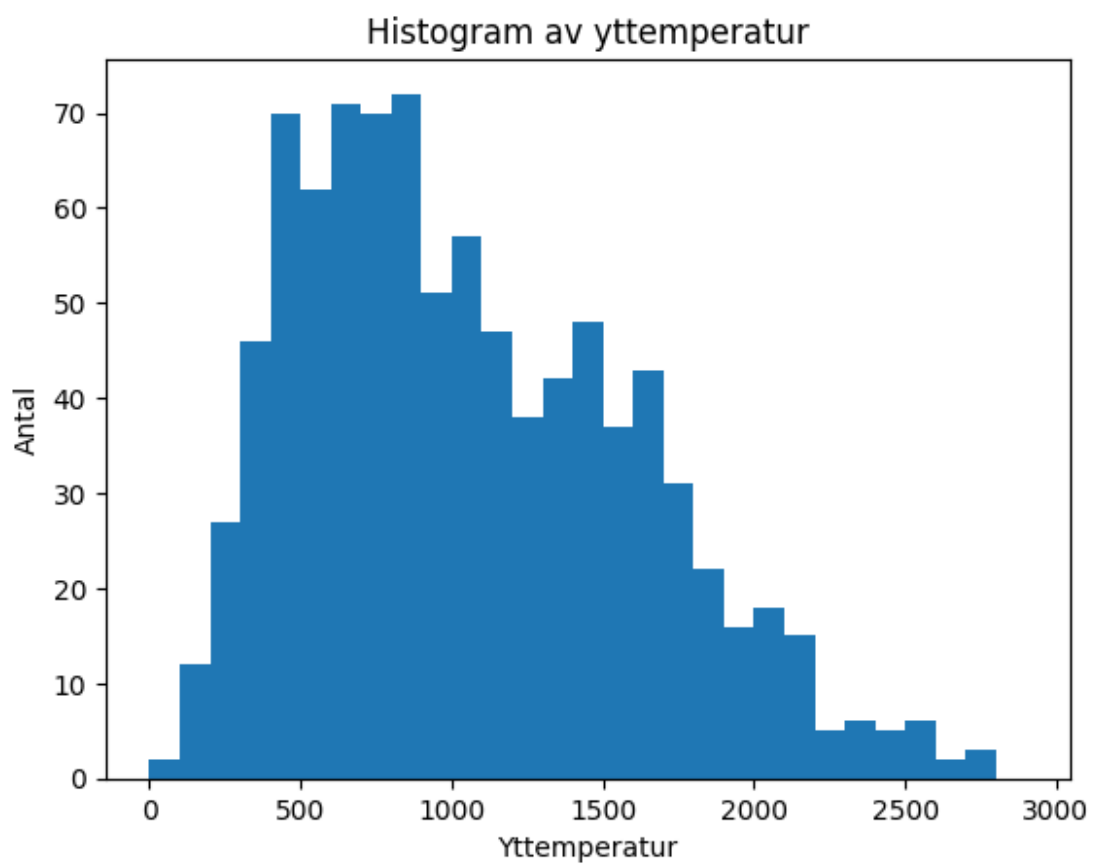
2.5 e)

Korrelationskoefficienter togs fram genom att använda .corr metoden som finns inbyggd i pandas. De variabler som har störst korrelation är yttemperatur och stjärn massa med en Korrelationskoefficient på 0,62. Sambandet illustreras i följande figur.

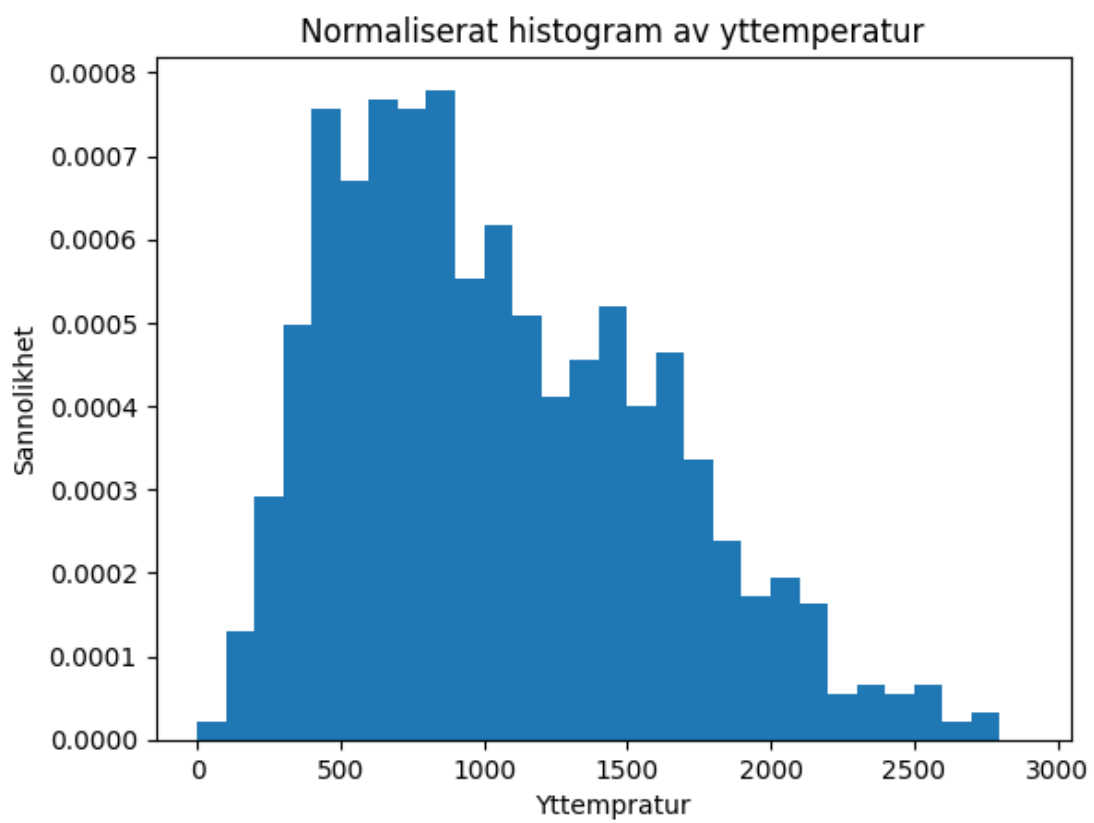


Figur 6: Scatter plot över planetmassa och yttemperatur med en linjäranpassning för att förtydliga korrelationen

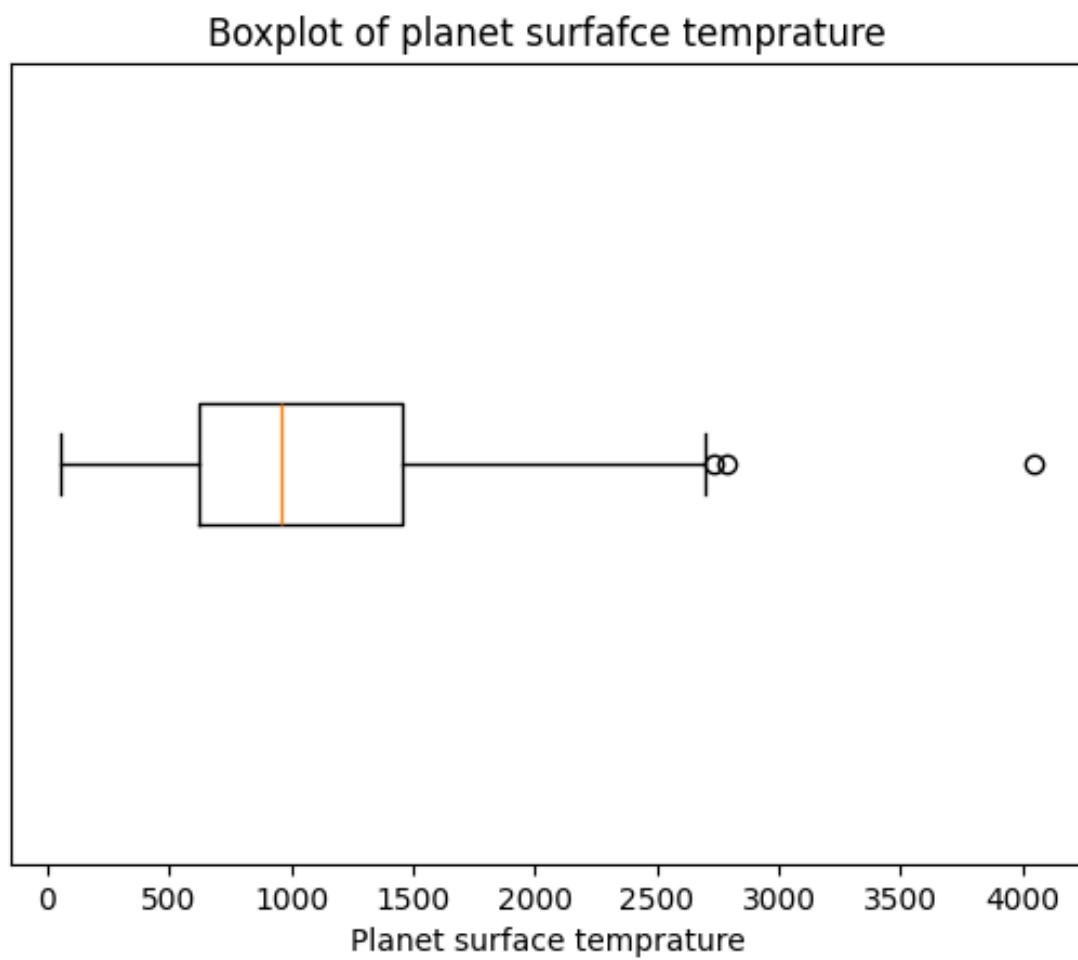
2.6 f)



Figur 7: Histogram över yttemperaturen, med en bin bredd på 100 kelvin



Figur 8: Normaliserat histogram över yttemperaturen, med en bin bredd på 100 kelvin



Figur 9: Boxplot av planeternas yttemperatur