

```
In [1]: from cmdstanpy import cmdstan_path, CmdStanModel
import pandas as pd
import arviz as az
import numpy as np
from matplotlib import pyplot as plt

# First name: Maciej (6)
# Last name: Wojtyś (6)
```

Exercise 1 - generated quantities

```
In [2]: stan_file = r'C:\Studia\Data-Analytics\Lab-2\code_1.stan'
model = CmdStanModel(stan_file=stan_file)

data = {
    "N": 6
}
fit = model.sample(data = data)
```

INFO:cmdstanpy:found newer exe file, not recompiling

INFO:cmdstanpy:CmdStan start processing

chain 1 00:00 Sampling completed

chain 2 00:00 Sampling completed

chain 3 00:00 Sampling completed

chain 4 00:00 Sampling completed

INFO:cmdstanpy:CmdStan done processing.

```
In [3]: df = fit.draws_pd()
df
```

Out[3]:

	lp__	accept_stat__	lambda	y_sim[1]	y_sim[2]	y_sim[3]	y_sim[4]	y_sim[5]	y_sim[6]
0	0.0	0.0	67.1300	49.0	80.0	72.0	69.0	51.0	62.0
1	0.0	0.0	42.5498	47.0	44.0	41.0	40.0	34.0	38.0
2	0.0	0.0	245.0690	220.0	271.0	248.0	247.0	261.0	246.0
3	0.0	0.0	140.0270	141.0	143.0	126.0	136.0	153.0	117.0
4	0.0	0.0	281.4930	264.0	252.0	278.0	278.0	287.0	269.0
...
3995	0.0	0.0	241.6580	236.0	259.0	231.0	257.0	253.0	242.0
3996	0.0	0.0	81.0997	81.0	79.0	81.0	89.0	68.0	69.0
3997	0.0	0.0	42.7677	48.0	44.0	52.0	52.0	37.0	45.0
3998	0.0	0.0	25.6378	28.0	31.0	21.0	25.0	21.0	26.0
3999	0.0	0.0	140.9350	137.0	119.0	141.0	135.0	162.0	160.0

4000 rows × 9 columns

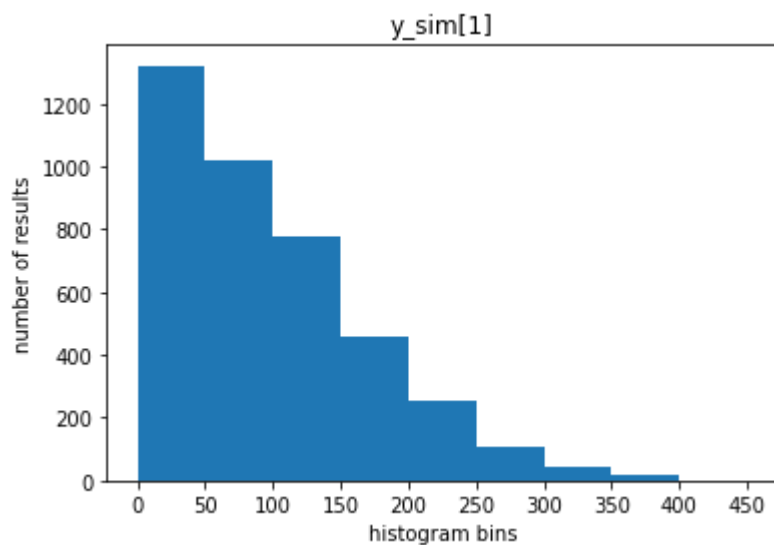
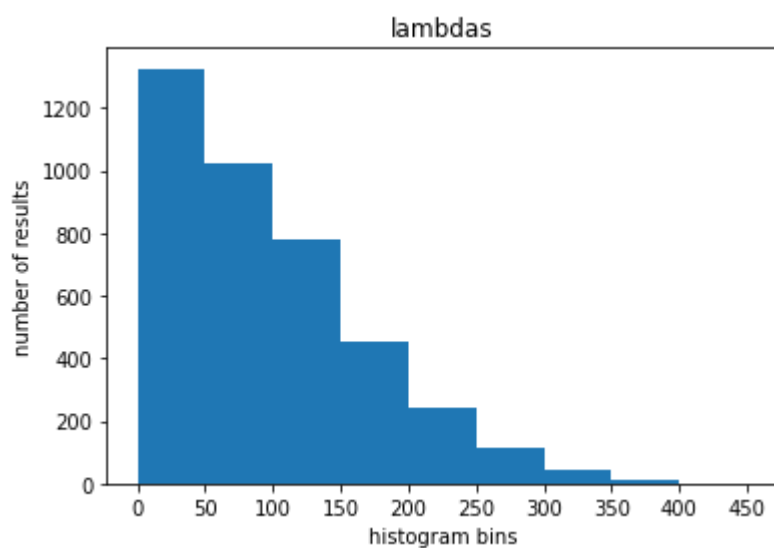
```
In [4]: df.iloc[:, 2:9]
```

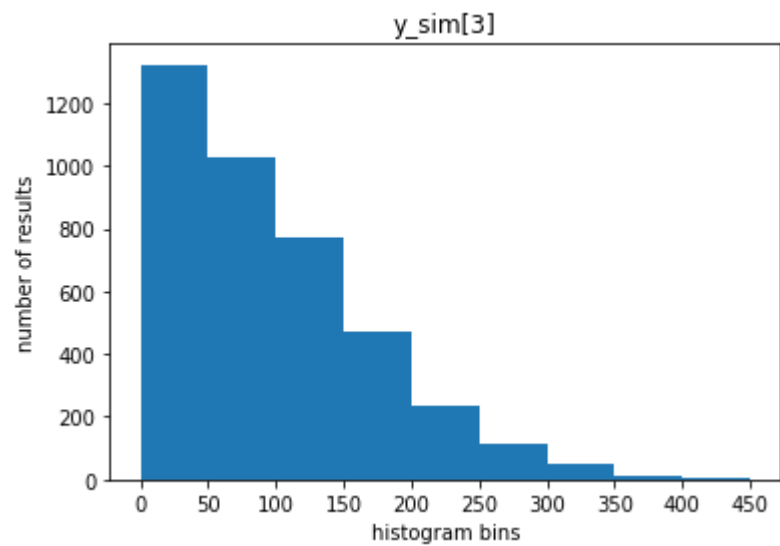
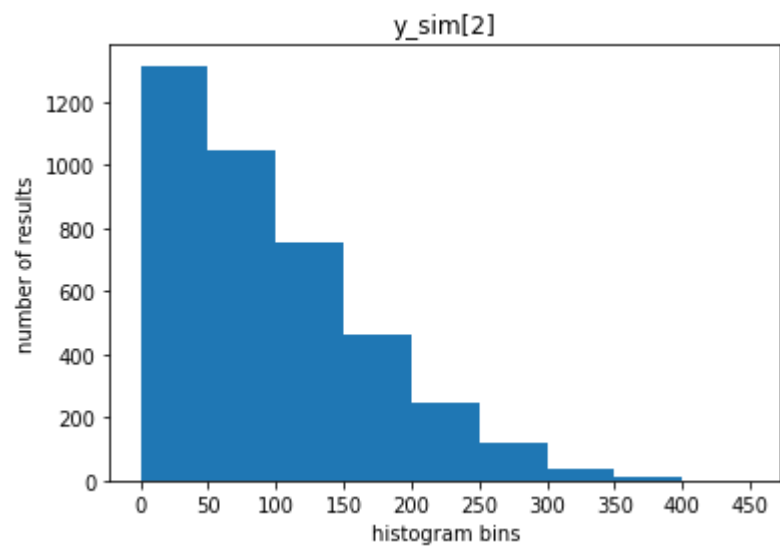
Out[4]:

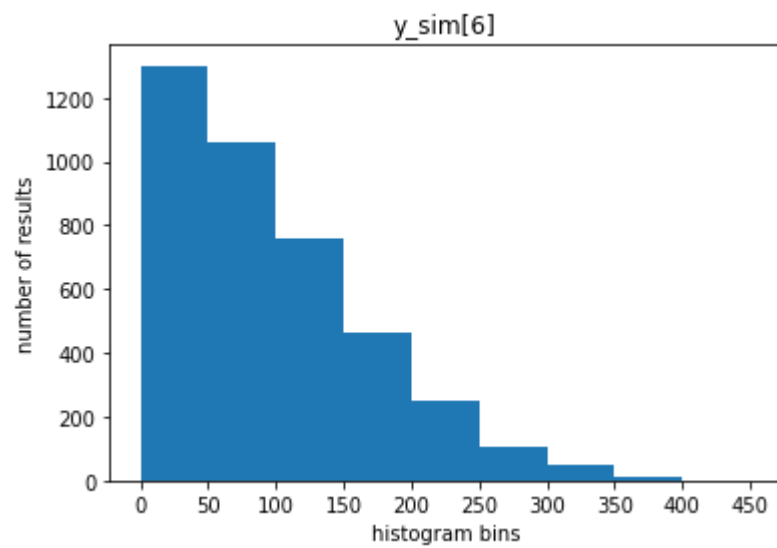
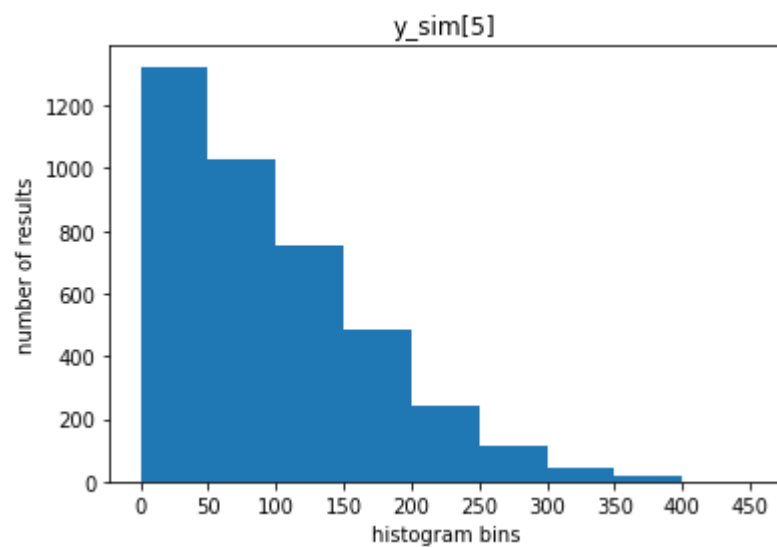
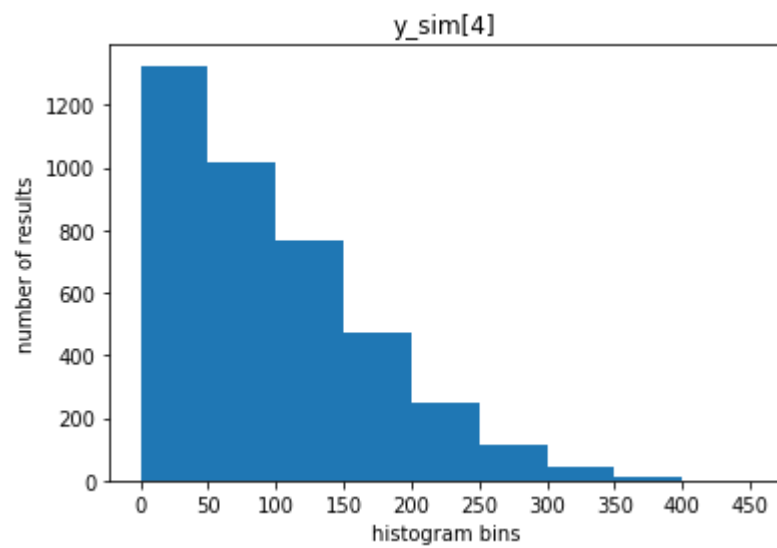
	lambda	y_sim[1]	y_sim[2]	y_sim[3]	y_sim[4]	y_sim[5]	y_sim[6]
0	67.1300	49.0	80.0	72.0	69.0	51.0	62.0
1	42.5498	47.0	44.0	41.0	40.0	34.0	38.0
2	245.0690	220.0	271.0	248.0	247.0	261.0	246.0
3	140.0270	141.0	143.0	126.0	136.0	153.0	117.0
4	281.4930	264.0	252.0	278.0	278.0	287.0	269.0
...
3995	241.6580	236.0	259.0	231.0	257.0	253.0	242.0
3996	81.0997	81.0	79.0	81.0	89.0	68.0	69.0
3997	42.7677	48.0	44.0	52.0	52.0	37.0	45.0
3998	25.6378	28.0	31.0	21.0	25.0	21.0	26.0
3999	140.9350	137.0	119.0	141.0	135.0	162.0	160.0

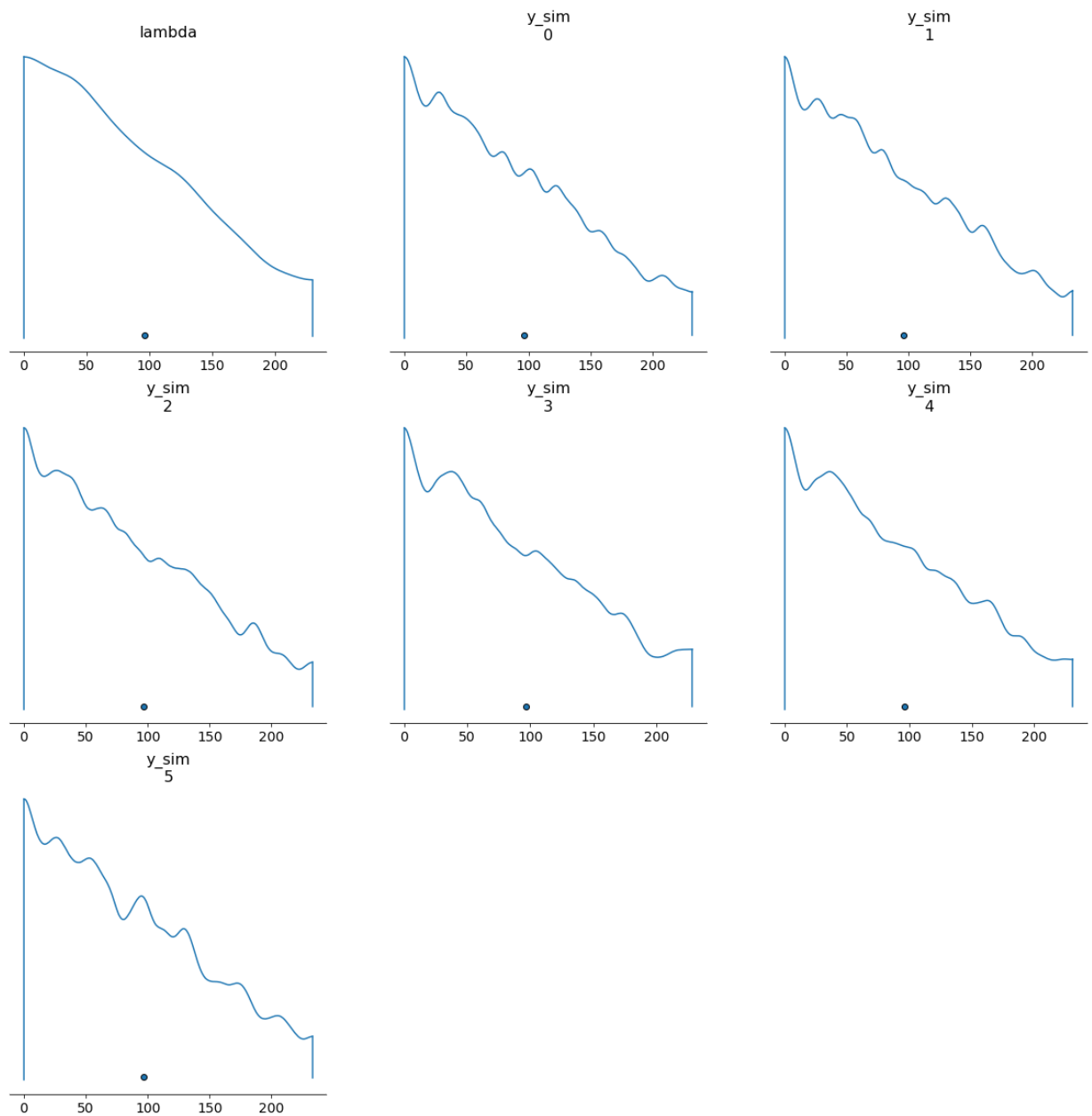
4000 rows × 7 columns

```
In [5]: titles = ["lambdas", "y_sim[1]", "y_sim[2]", "y_sim[3]", "y_sim[4]", "y_sim[5]",  
x = np.linspace(0, 450, 10)  
  
for i in range(7):  
    plt.figure(i)  
    plt.hist(df.iloc[:, i + 2], bins = x)  
    plt.title(titles[i])  
    plt.xticks(x)  
    plt.xlabel("histogram bins")  
    plt.ylabel("number of results")  
    plt.show()  
  
az.plot_density(fit, figsize=(20, 20))  
plt.show()
```









Excercise 2 - constraints on the data

```

In [6]: stan_file = r'C:\Studia\Data-Analytics\Lab-2\code_2.stan'
model1 = CmdStanModel(stan_file=stan_file)

data = {
    "N": 12,
    "y": [0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1]
}
fit1 = model1.sample(data = data)
df1 = fit1.draws_pd()

stan_file = r'C:\Studia\Data-Analytics\Lab-2\code_3.stan'
model2 = CmdStanModel(stan_file=stan_file)

data = {
    "N": 12,
    "y": [0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1]
}
fit2 = model2.sample(data = data)
df2 = fit2.draws_pd()

```

```

INFO:cmdstanpy:found newer exe file, not recompiling
INFO:cmdstanpy:CmdStan start processing

```

chain 1	00:00 Sampling completed
chain 2	00:00 Sampling completed
chain 3	00:00 Sampling completed
chain 4	00:00 Sampling completed

```

INFO:cmdstanpy:CmdStan done processing.

```

```

INFO:cmdstanpy:found newer exe file, not recompiling
INFO:cmdstanpy:CmdStan start processing

```

chain 1	00:00 Sampling completed
chain 2	00:00 Sampling completed
chain 3	00:00 Sampling completed
chain 4	00:00 Sampling completed

```

INFO:cmdstanpy:CmdStan done processing.

```

In [7]: df1

Out[7]:

	lp__	accept_stat__	stepsize__	treedepth__	n_leapfrog__	divergent__	energy__	tf
0	-9.70433	1.000000	1.01493	2.0	3.0	0.0	9.71908	0.496
1	-9.71328	0.927766	1.01493	2.0	3.0	0.0	10.17090	0.518
2	-10.29780	0.907313	1.01493	2.0	3.0	0.0	10.46820	0.642
3	-11.60830	0.868626	1.01493	2.0	7.0	0.0	11.64430	0.255
4	-9.88438	1.000000	1.01493	2.0	3.0	0.0	11.23500	0.579
...
3995	-10.14740	0.922916	1.11978	2.0	3.0	0.0	10.26910	0.623
3996	-9.77496	0.975577	1.11978	1.0	3.0	0.0	10.19360	0.449
3997	-9.74339	1.000000	1.11978	1.0	1.0	0.0	9.77363	0.462
3998	-9.71196	1.000000	1.11978	1.0	1.0	0.0	9.73318	0.483
3999	-9.97999	0.871880	1.11978	2.0	3.0	0.0	10.30230	0.598

4000 rows × 8 columns



In [8]: df2

Out[8]:

	lp__	accept_stat__	stepsize__	treedepth__	n_leapfrog__	divergent__	energy__	tf
0	-10.06900	1.000000	0.95837	1.0	1.0	0.0	10.28140	0.387
1	-10.40290	0.975959	0.95837	2.0	3.0	0.0	10.42820	0.345
2	-9.95943	1.000000	0.95837	2.0	3.0	0.0	10.30540	0.405
3	-9.91584	1.000000	0.95837	1.0	1.0	0.0	9.98896	0.413
4	-10.21790	0.932907	0.95837	1.0	1.0	0.0	10.22440	0.366
...
3995	-9.80266	1.000000	1.06310	1.0	1.0	0.0	9.85038	0.559
3996	-11.63790	0.548987	1.06310	1.0	3.0	0.0	12.52910	0.254
3997	-13.00450	0.762788	1.06310	1.0	1.0	0.0	13.18380	0.193
3998	-13.00450	0.881186	1.06310	1.0	1.0	0.0	14.15360	0.193
3999	-11.45820	1.000000	1.06310	1.0	1.0	0.0	12.81960	0.264

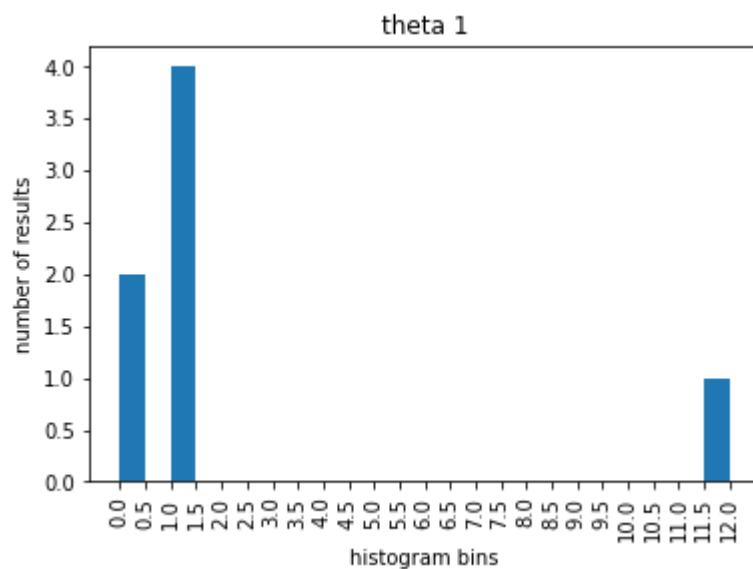
4000 rows × 8 columns

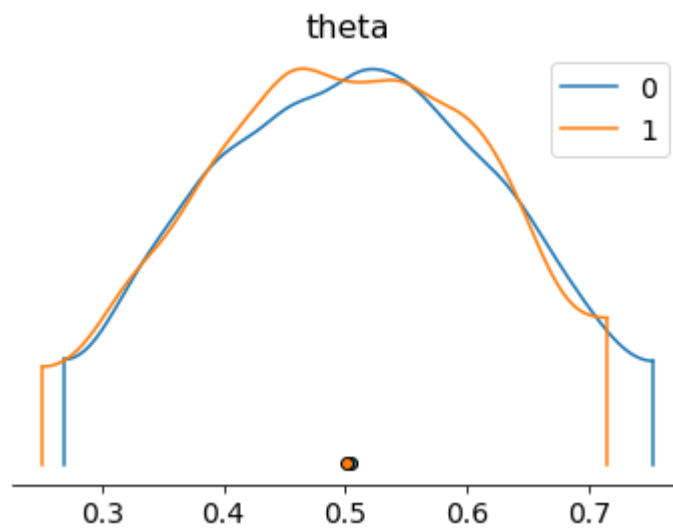
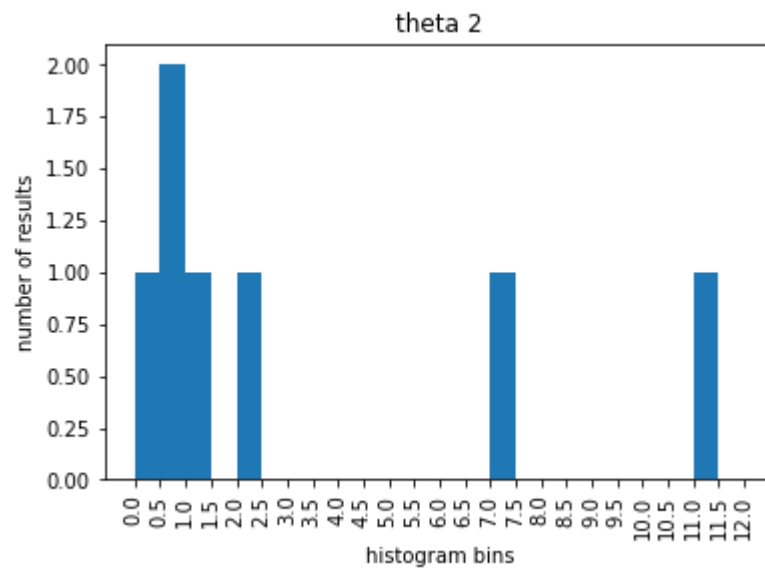



```
In [9]: x = np.linspace(0, 12, 25)
plt.figure(1)
plt.hist(df1.iloc[7], bins = x)
plt.title("theta 1")
plt.xticks(x, rotation = 90)
plt.xlabel("histogram bins")
plt.ylabel("number of results")
plt.show()

plt.figure(1)
plt.hist(df2.iloc[7], bins = x)
plt.title("theta 2")
plt.xticks(x, rotation = 90)
plt.xlabel("histogram bins")
plt.ylabel("number of results")
plt.show()

az.plot_density([fit1, fit2])
plt.show()
```





Excercise 3 - constraints on the parameters

```
In [10]: stan_file = r'C:\Studia\Data-Analytics\Lab-2\code_4.stan'
model1 = CmdStanModel(stan_file=stan_file)

data = {
    "N": 12,
}
fit1 = model1.sample(data = data)
df1 = fit1.draws_pd()

stan_file = r'C:\Studia\Data-Analytics\Lab-2\code_5.stan'
model2 = CmdStanModel(stan_file=stan_file)

data = {
    "N": 12,
}
fit2 = model2.sample(data = data)
df2 = fit2.draws_pd()
```

```
INFO:cmdstanpy:found newer exe file, not recompiling
INFO:cmdstanpy:CmdStan start processing
```

chain 1	00:00 Sampling completed
chain 2	00:00 Sampling completed
chain 3	00:00 Sampling completed
chain 4	00:00 Sampling completed

```
INFO:cmdstanpy:CmdStan done processing.
```

```
INFO:cmdstanpy:found newer exe file, not recompiling
INFO:cmdstanpy:CmdStan start processing
```

chain 1	00:00 Sampling completed
chain 2	00:00 Sampling completed
chain 3	00:00 Sampling completed
chain 4	00:00 Sampling completed

```
INFO:cmdstanpy:CmdStan done processing.
```

In [11]: df1

Out[11]:

	lp__	accept_stat__	stepsize__	treedepth__	n_leapfrog__	divergent__	energy__	
0	-0.964271	0.497646	0.302936	3.0	7.0	0.0	1.32109	0.0
1	-1.115450	0.997870	0.302936	1.0	3.0	0.0	1.19305	0.0
2	-1.652500	0.664324	0.302936	1.0	3.0	1.0	2.89419	1.0
3	-1.234190	0.666667	0.302936	1.0	3.0	1.0	2.33840	0.0
4	-1.036510	0.999526	0.302936	1.0	3.0	0.0	1.37153	0.0
...
3995	-0.722336	0.999266	0.184489	2.0	7.0	0.0	1.43907	0.0
3996	-1.629580	0.983470	0.184489	3.0	15.0	0.0	2.79399	1.0
3997	-1.621920	1.000000	0.184489	2.0	3.0	0.0	1.63119	1.0
3998	-1.681670	0.999948	0.184489	2.0	3.0	0.0	1.68245	1.0
3999	-1.118980	0.874983	0.184489	3.0	8.0	1.0	1.94268	0.0

4000 rows × 8 columns

In [12]: df2

Out[12]:

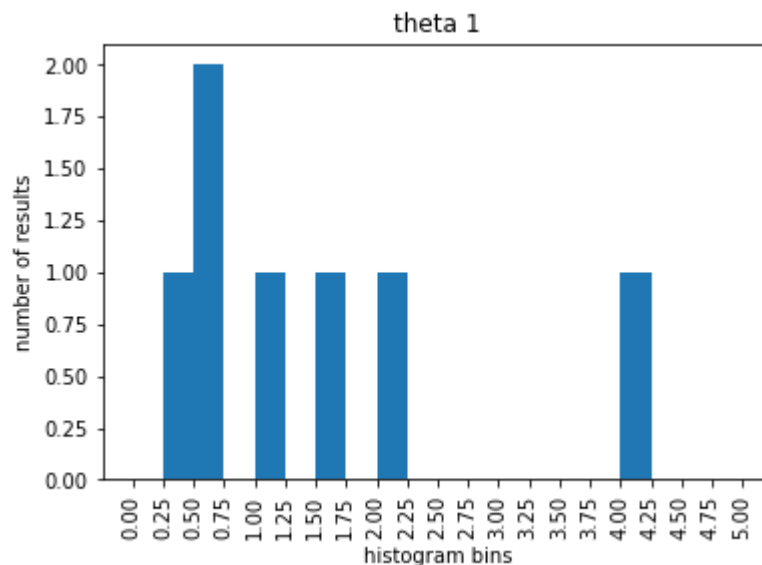
	lp__	accept_stat__	stepsize__	treedepth__	n_leapfrog__	divergent__	energy__	the
0	-2.45156	0.931896	0.606099	2.0	3.0	0.0	2.49924	3.0891
1	-2.36281	1.000000	0.606099	2.0	3.0	0.0	2.75897	2.9832
2	-2.40647	0.530039	0.606099	3.0	7.0	0.0	4.51078	0.1734
3	-1.54158	0.972309	0.606099	3.0	7.0	0.0	2.48376	0.4627
4	-1.58219	0.995924	0.606099	2.0	3.0	0.0	1.68737	0.4362
...
3995	-1.26784	0.990922	0.619901	2.0	3.0	0.0	1.33921	1.1785
3996	-2.70833	0.855028	0.619901	3.0	7.0	0.0	3.41965	0.1305
3997	-3.35559	0.983378	0.619901	2.0	3.0	0.0	3.83434	0.0734
3998	-1.74414	0.998624	0.619901	3.0	7.0	0.0	3.41914	2.1700
3999	-1.25571	0.982945	0.619901	2.0	3.0	0.0	1.90039	1.0986

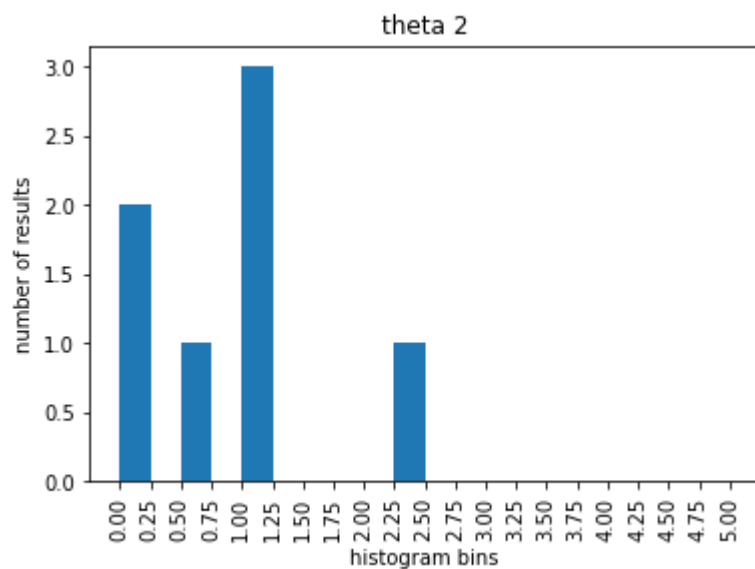
4000 rows × 8 columns

```
In [13]: x = np.linspace(0, 5, 21)
plt.figure(1)
plt.hist(df1.iloc[7], bins = x)
plt.title("theta 1")
plt.xticks(x, rotation = 90)
plt.xlabel("histogram bins")
plt.ylabel("number of results")
plt.show()

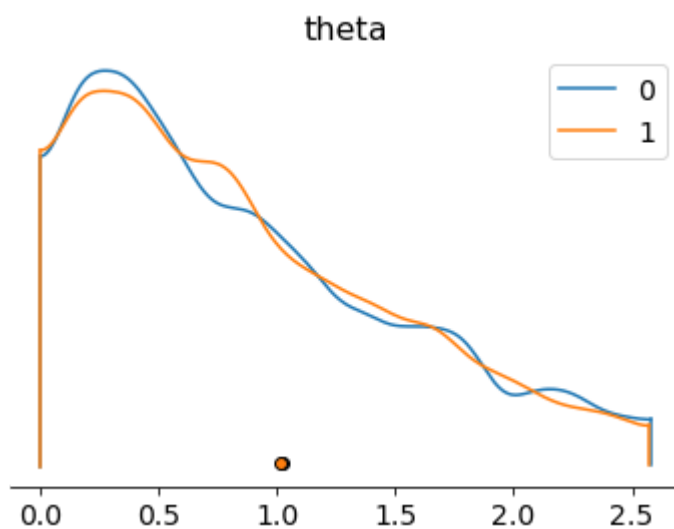
plt.figure(2)
plt.hist(df2.iloc[7], bins = x)
plt.title("theta 2")
plt.xticks(x, rotation = 90)
plt.xlabel("histogram bins")
plt.ylabel("number of results")
plt.show()

az.plot_density([fit1, fit2])
```





```
Out[13]: array([[<AxesSubplot:title={'center':'theta'}>]], dtype=object)
```



Excercise 4 - functions and different functionalities of stan

```
In [22]: stan_file = r'C:\Studia\Data-Analytics\Lab-2\code_6.stan'
model = CmdStanModel(stan_file=stan_file)

data = {
    "y_guess": [2],
    "theta": [6]
}

fit = model.sample(data = data, iter_sampling = 1, chains = 1)
fit.draws_pd()
```

INFO:cmdstanpy:found newer exe file, not recompiling
 INFO:cmdstanpy:CmdStan start processing

chain 1 00:00 Sampling completed

INFO:cmdstanpy:CmdStan done processing.

Out[22]:

	lp__	accept_stat__	sigma
0	0.0	0.0	2.32935

Excercise 5 - different methods of defining models

```
In [15]: data = {  
          "N": 6  
        }  
  
stan_file = r'C:\Studia\Data-Analytics\Lab-2\code_7.stan'  
model1 = CmdStanModel(stan_file=stan_file)  
  
fit1 = model1.sample(data = data)  
  
stan_file = r'C:\Studia\Data-Analytics\Lab-2\code_8.stan'  
model2 = CmdStanModel(stan_file=stan_file)  
  
fit2 = model2.sample(data = data)  
  
stan_file = r'C:\Studia\Data-Analytics\Lab-2\code_9.stan'  
model3 = CmdStanModel(stan_file=stan_file)  
  
fit3 = model3.sample(data = data)
```

```
INFO:cmdstanpy:found newer exe file, not recompiling  
INFO:cmdstanpy:CmdStan start processing
```

chain 1	00:00 Sampling completed
chain 2	00:00 Sampling completed
chain 3	00:00 Sampling completed
chain 4	00:00 Sampling completed

```
INFO:cmdstanpy:CmdStan done processing.
```

```
INFO:cmdstanpy:found newer exe file, not recompiling  
INFO:cmdstanpy:CmdStan start processing
```

chain 1	00:00 Sampling completed
chain 2	00:00 Sampling completed
chain 3	00:00 Sampling completed
chain 4	00:00 Sampling completed

```
INFO:cmdstanpy:CmdStan done processing.
```

```
INFO:cmdstanpy:found newer exe file, not recompiling
```


INFO:cmdstanpy:CmdStan start processing

chain 1 00:00 Sampling completed

chain 2 00:00 Sampling completed

chain 3 00:00 Sampling completed

chain 4 00:00 Sampling completed

INFO:cmdstanpy:CmdStan done processing.

In [16]: fit1.draws_pd()

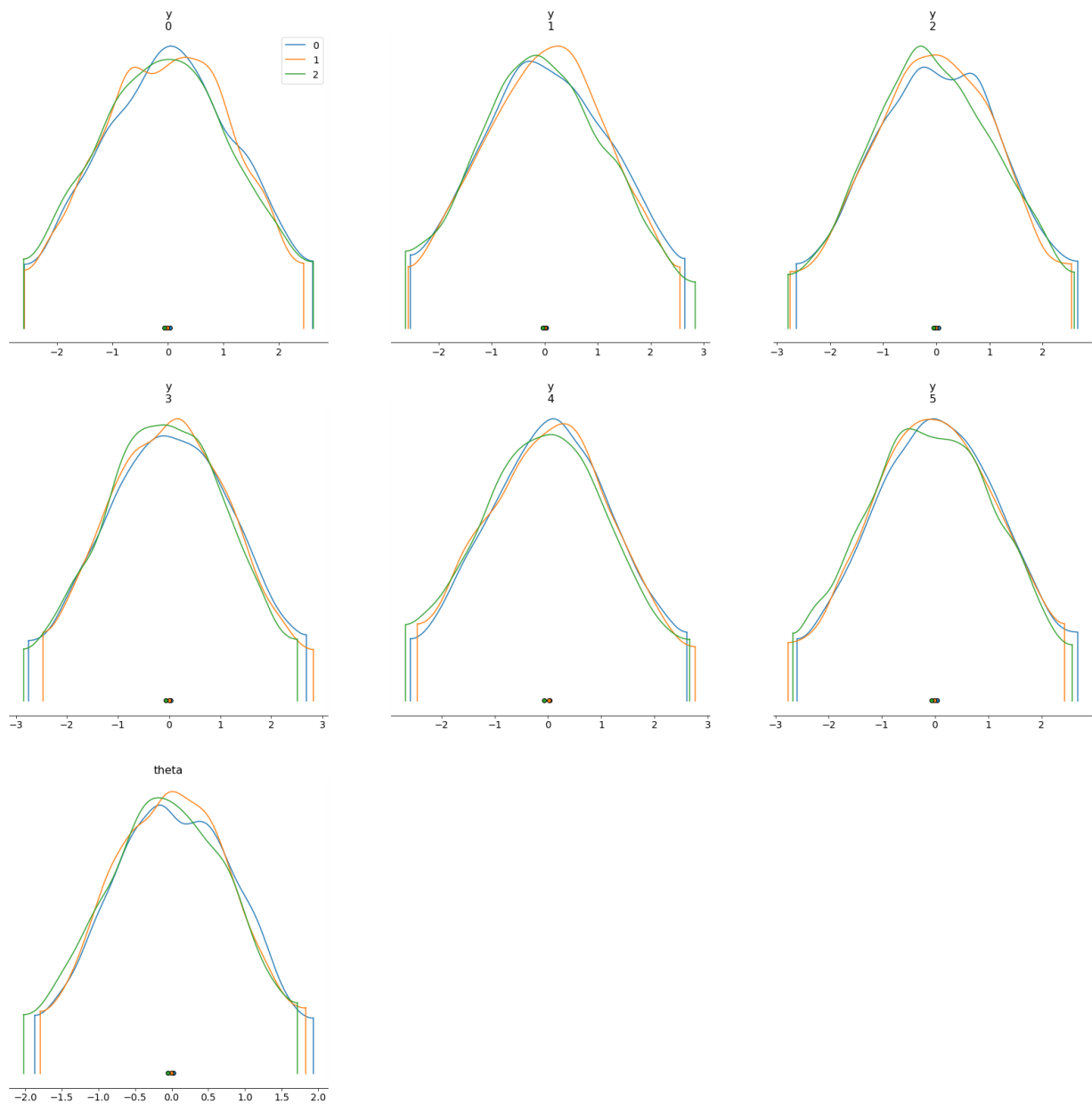
Out[16]:

	lp__	accept_stat__	stepsize__	treedepth__	n_leapfrog__	divergent__	energy__	
0	-6.576790	0.952756	0.309966	3.0	7.0	0.0	8.72844	2.961
1	-0.860473	0.954932	0.309966	3.0	15.0	0.0	8.05845	0.731
2	-3.782210	0.895443	0.309966	2.0	7.0	0.0	6.33435	0.13
3	-3.822720	0.987342	0.309966	3.0	15.0	0.0	5.54710	1.741
4	-6.257240	0.937748	0.309966	3.0	7.0	0.0	10.59870	0.301
...
3995	-3.901940	0.967352	0.343214	3.0	15.0	0.0	7.62473	-1.481
3996	-9.018920	0.879982	0.343214	3.0	7.0	0.0	11.23190	-0.791
3997	-4.775080	1.000000	0.343214	4.0	15.0	0.0	11.63140	-1.821
3998	-1.049780	0.633363	0.343214	3.0	7.0	0.0	8.42860	-0.751
3999	-1.720480	0.996535	0.343214	3.0	15.0	0.0	3.32515	-1.881

4000 rows × 14 columns

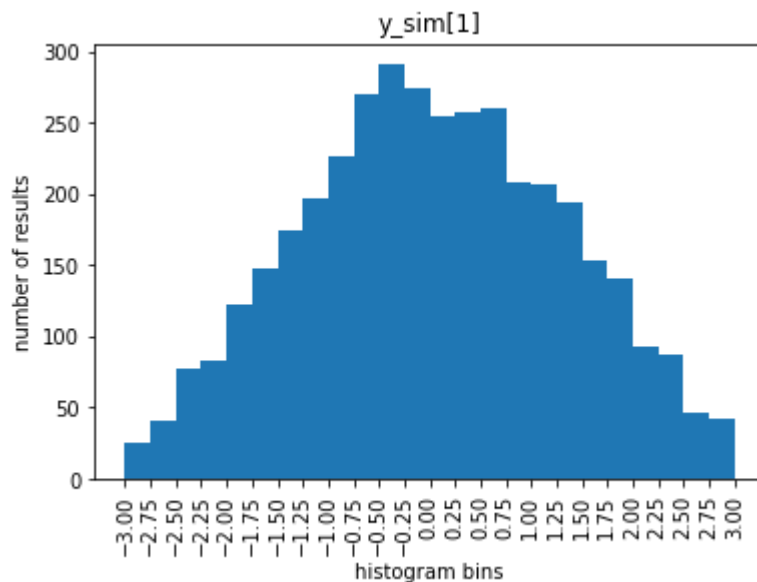
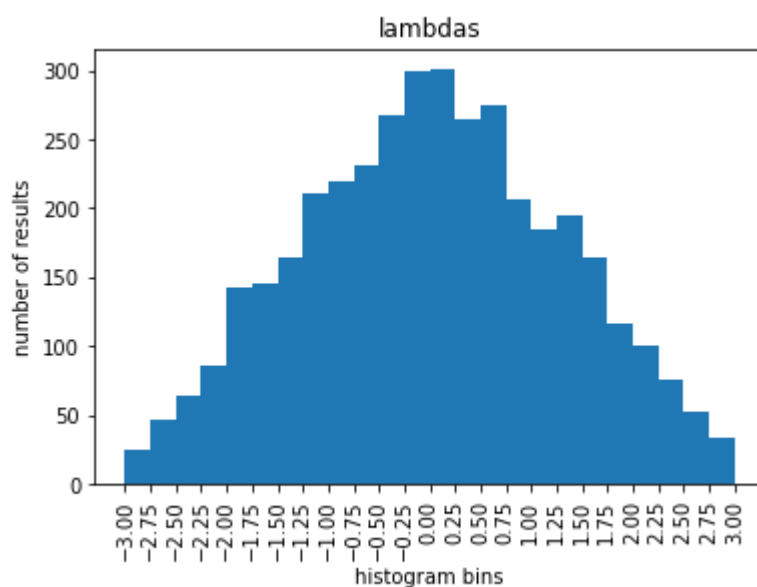


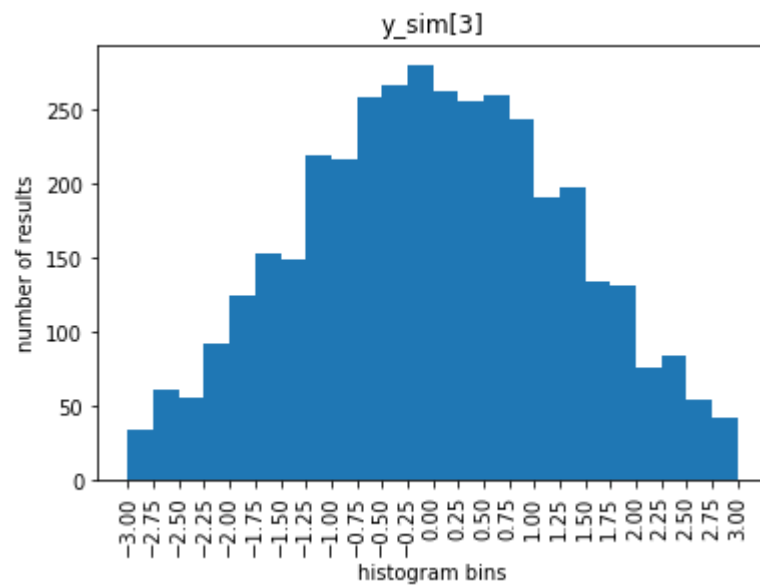
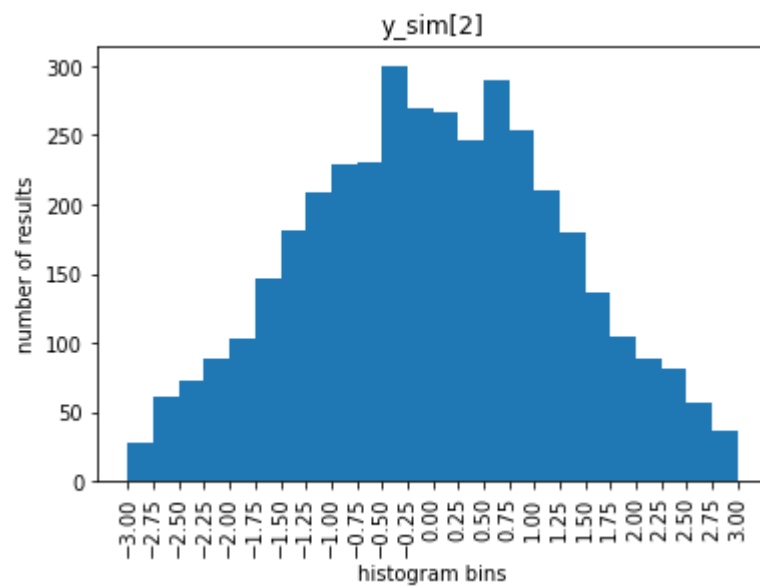
```
In [17]: az.plot_density([fit1, fit2, fit3], figsize = (28, 28))  
plt.show()
```

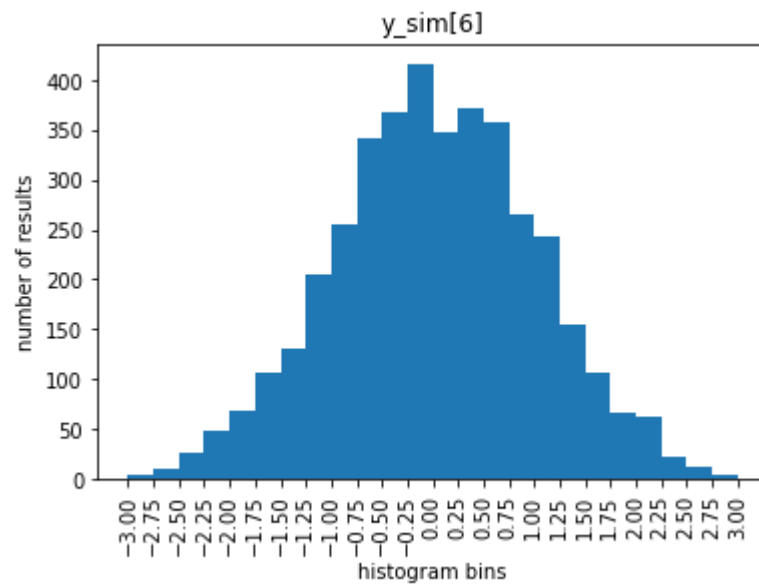
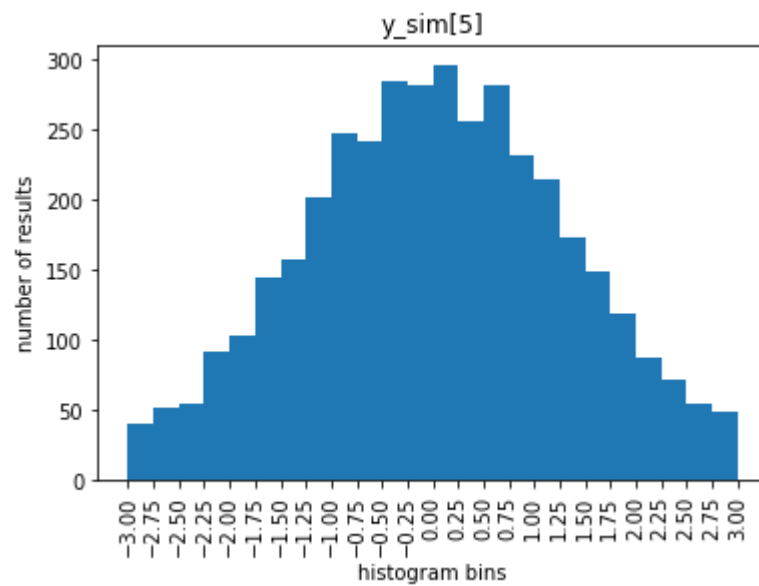
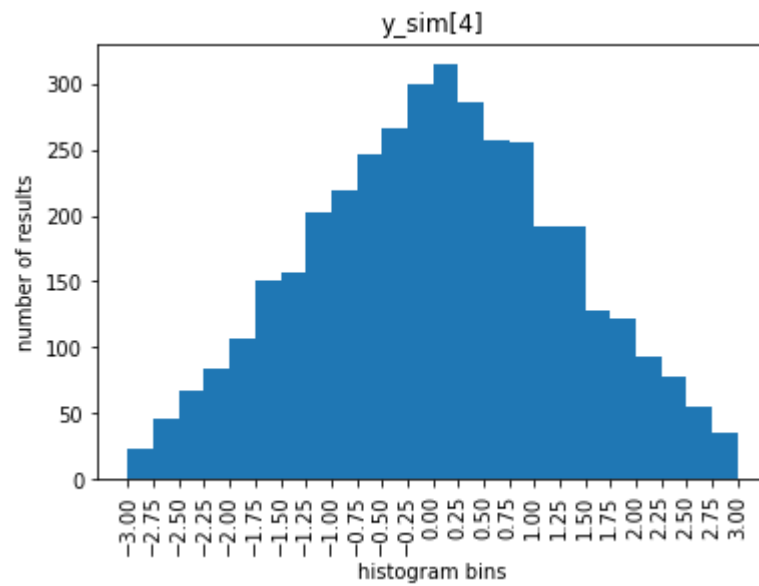


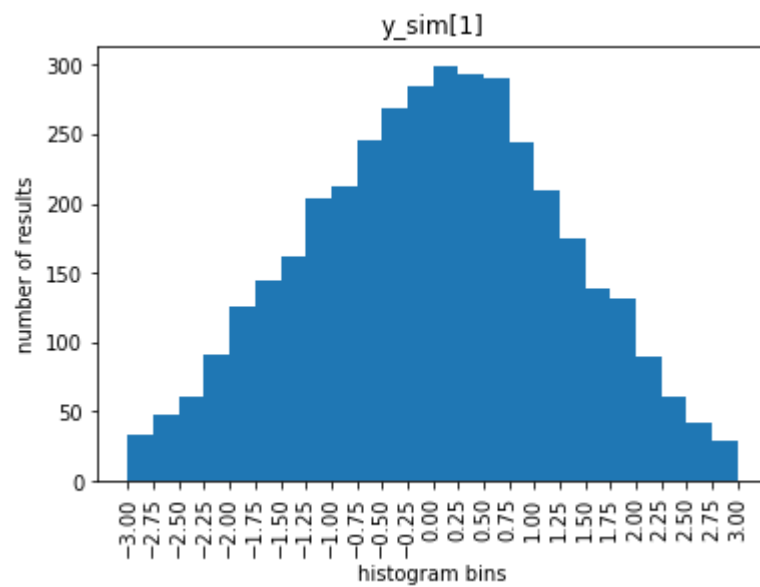
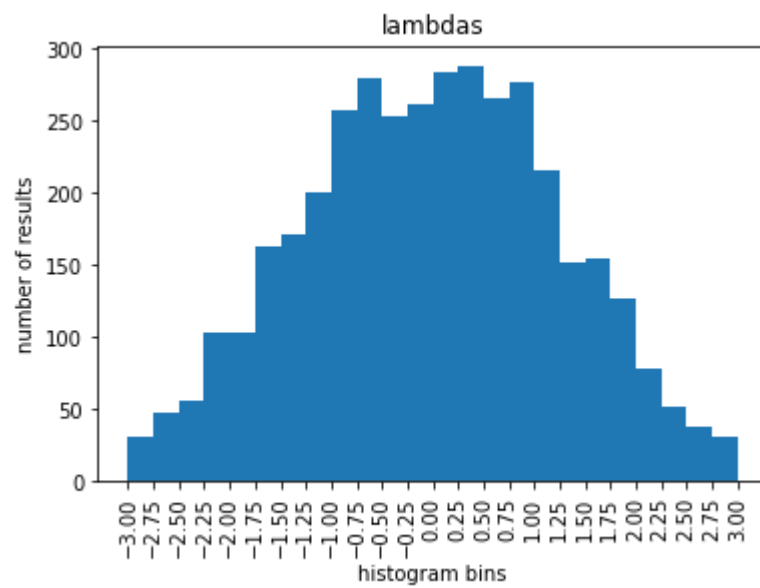
```
In [18]: titles = ["lambdas", "y_sim[1]", "y_sim[2]", "y_sim[3]", "y_sim[4]", "y_sim[5]",
fits = [fit1, fit2, fit3]
x = np.linspace(-3, 3, 25)
for j in range(3):
    df = fits[j].draws_pd()

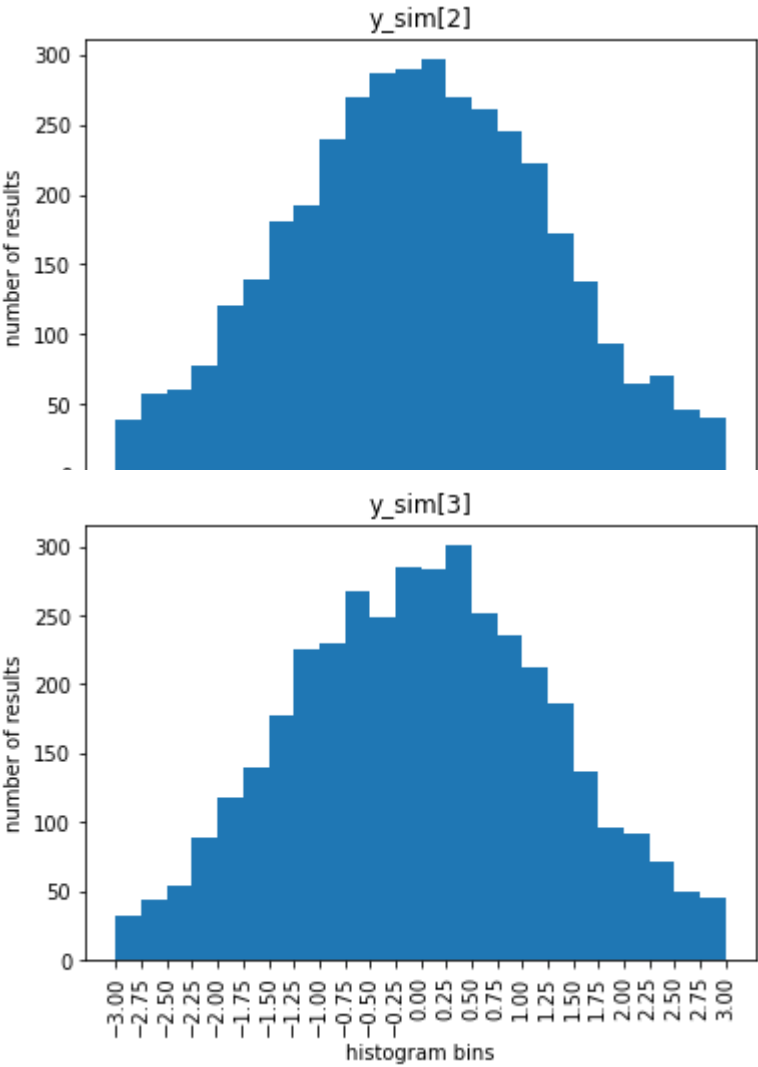
    for i in range(7):
        plt.figure(i)
        plt.hist(df.iloc[:, i + 7], bins = x)
        plt.title(titles[i])
        plt.xticks(x, rotation = 90)
        plt.xlabel("histogram bins")
        plt.ylabel("number of results")
        plt.show()
```

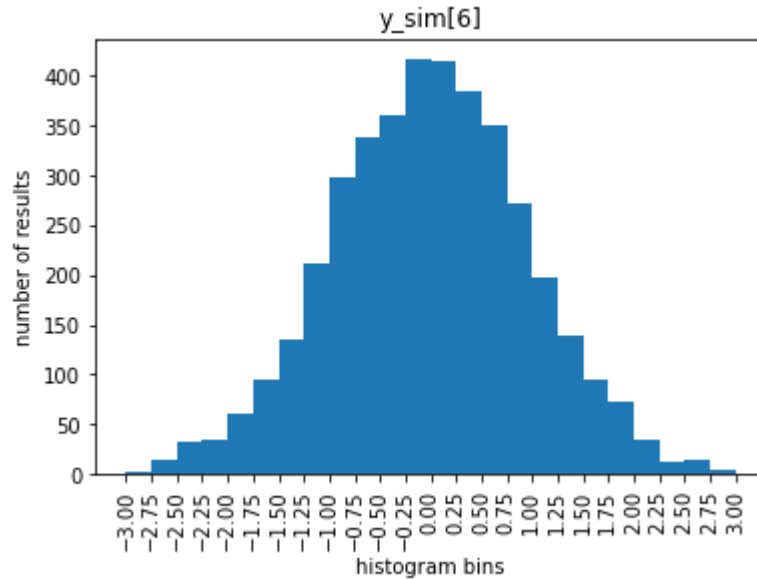
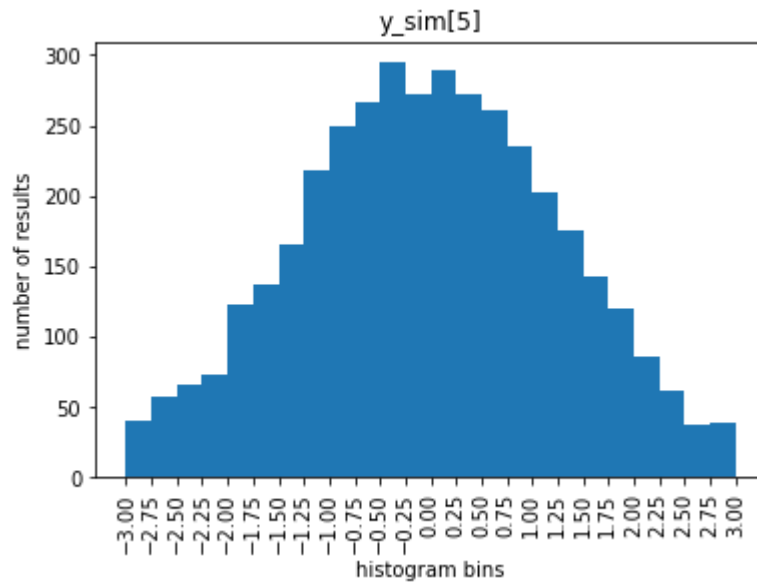
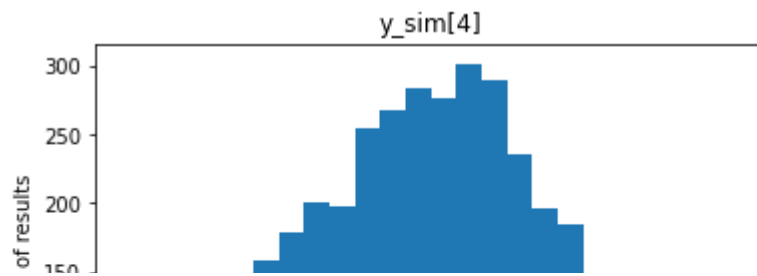


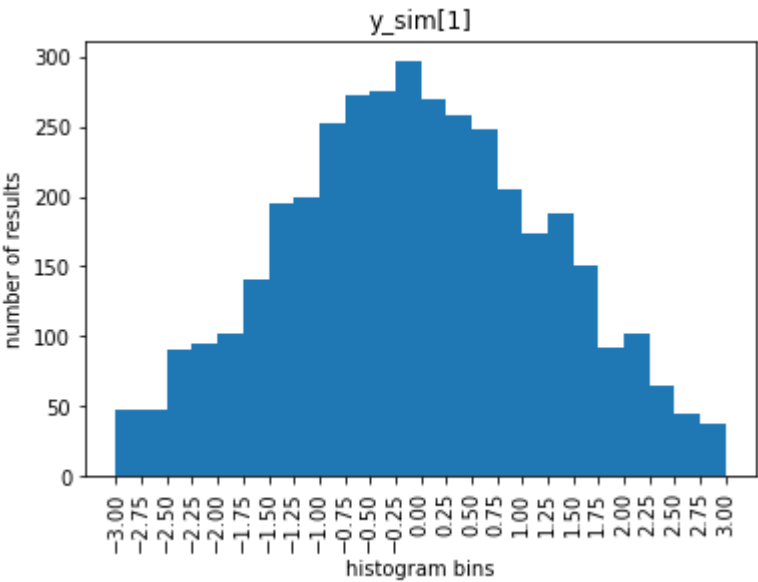
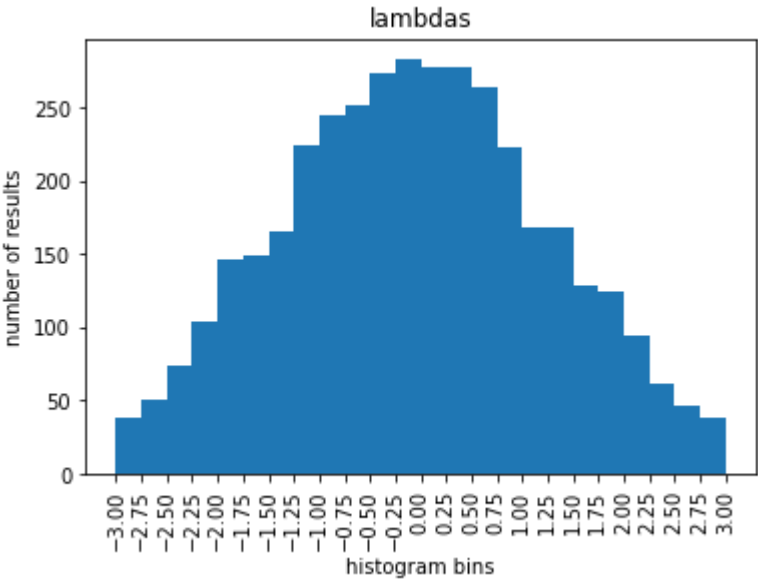


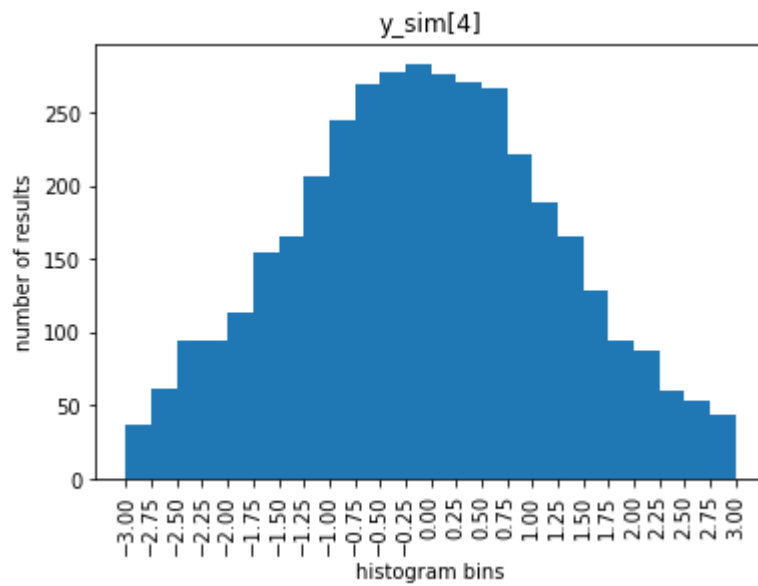
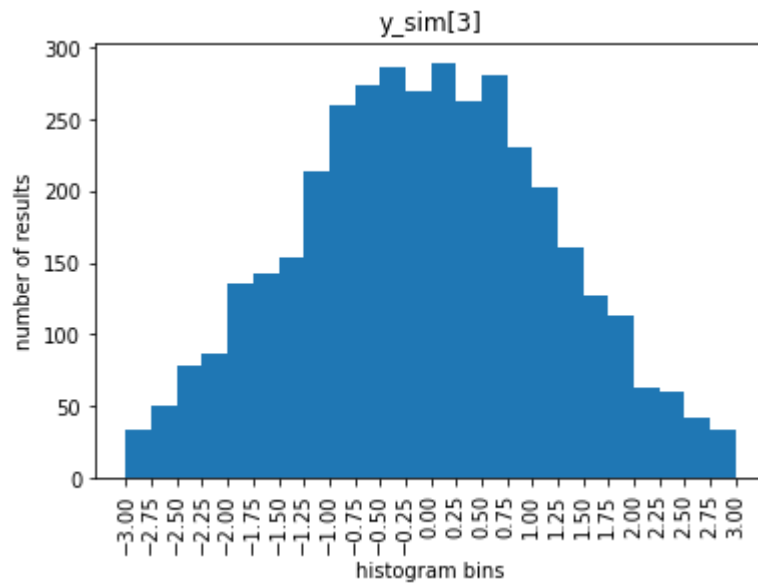
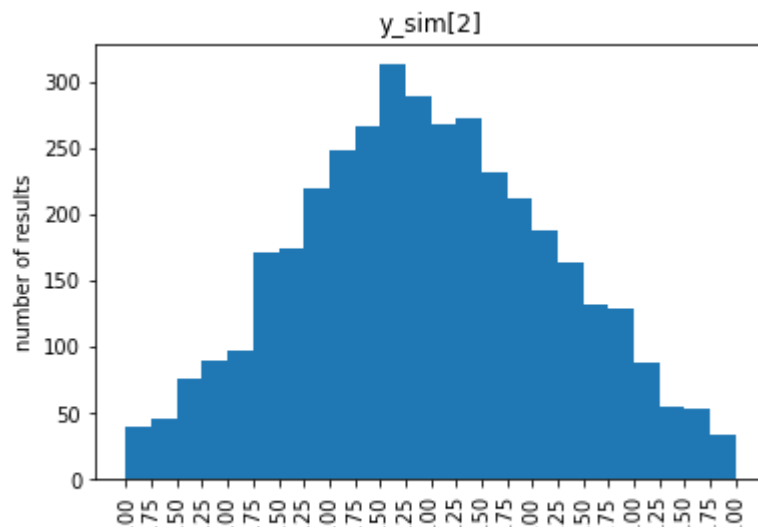


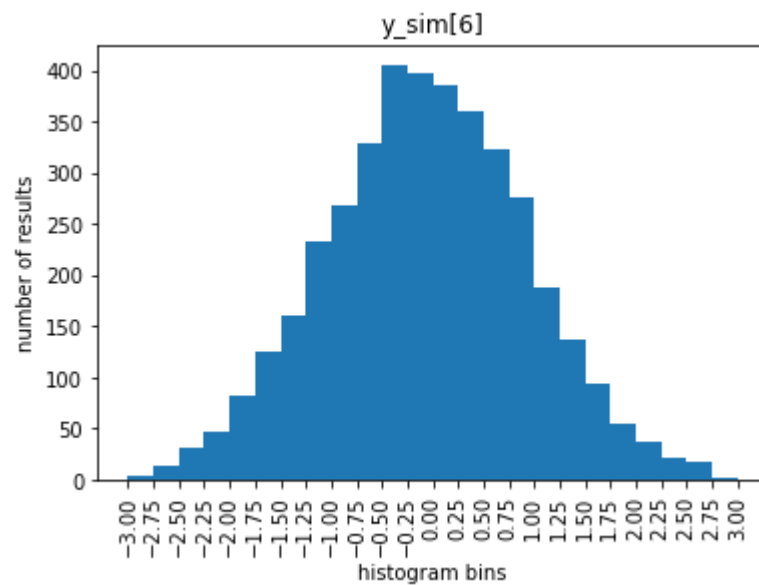
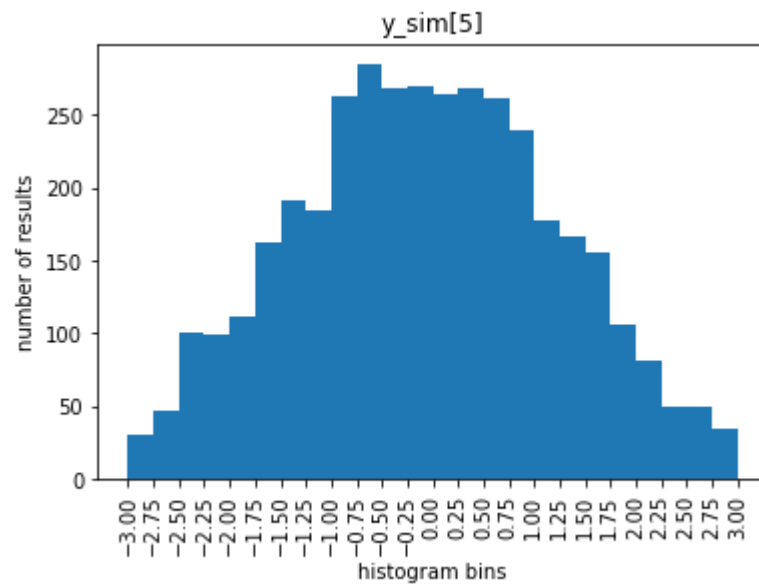












Excercise 6 - generated quantities post sampling

```
In [19]: stan_file = r'C:\Studia\Data-Analytics\Lab-2\code_10.stan'
model = CmdStanModel(stan_file=stan_file)
data = {
    "N": 6
}
fit4 = model.generate_quantities(data = data, mcmc_sample = fit3)
```

```
INFO:cmdstanpy:found newer exe file, not recompiling
INFO:cmdstanpy:Chain [1] start processing
INFO:cmdstanpy:Chain [2] start processing
INFO:cmdstanpy:Chain [3] start processing
INFO:cmdstanpy:Chain [4] start processing
INFO:cmdstanpy:Chain [4] done processing
INFO:cmdstanpy:Chain [2] done processing
INFO:cmdstanpy:Chain [1] done processing
INFO:cmdstanpy:Chain [3] done processing
```

```
In [20]: df = fit4.draws_pd()
df
```

Out[20]:

	mean_y
0	2.013150
1	-0.950739
2	-1.027340
3	-1.558180
4	2.176880
...	...
3995	-0.864170
3996	-1.150220
3997	0.887826
3998	0.031746
3999	0.099381

4000 rows × 1 columns

```
In [21]: plt.hist(df.iloc[:, 0], bins = 30)
plt.title("mean_y")
plt.show()
```

