1.Project Description

The purpose of the project is to compare the Bayesian Linear Regression and Bayesian Poisson Regression for the count data and compare the effectiveness. The data used for the project comes from "Bicycle Counts for East River Bridges" data from NYC open data (https://data.cityofnewyork.us/Transportation/Bicycle-Counts-for-East-River-Bridges/gua4-p9wg/data). The data contains the bicycle count on different east river bridges in Summer 2017. The scope of the project is to predict the number of bicycles in Brooklyn Bridge based on the temperature and precipitation on the given day. The project does not use any day related features like month or day of week, instead relies only on the weather features to predict the bicycle count.

The problem is analyzed with frequentist approach in the article, https://towardsdatascience.com/an-illustrated-guide-to-the-poisson-regression-model-50cccba15958. The project would implement a bayesian approach to the problem and compare Bayesian Linear and Bayesian Poisson models.

The data sample is shown below.

HIGH_T[]	LOW_T[]	PRECIP[]	BB_COUNT[]
46	37	0	606
62.1	41	0	2021
63	50	0.03	2470
51.1	46	1.18	723
63	46	0	2807

All three features, $HIGH\_T, LOW\_T, PRECIP$ are numeric features and BB_COUNT is the count of bicycles which is predicted variables.

The data set contains 214 records out of which 205 records are used for the building the model and another 10 records are used for prediction to evaluate the performance of the model.

The below sections examines Bayesian Linear Regression and Bayesian Poisson Regression for the given data set.

2.Bayesian Linear Regression

The Linear regression is formulated as below.

$$BB\_COUNT \sim \mathcal{N}(\mu, \tau) \tag{1}$$

$$\mu = \beta[1] + \beta[2] * HIGH\_T + \beta[3] * LOW\_T + \beta[4] * PRECIP \tag{2}$$

$$\beta = \mathcal{N}(0, 0.001) \tag{3}$$

$$\tau \sim \mathcal{G}(0.1, 0.1) \tag{4}$$

The below OpenBugs code implements Bayesian Linear regression for the given dataset. DATA2 has 204 records used to build the model and DATA3 has 10 records used for prediction.

```
model{
  for( i in 1:n) {
    BB_COUNT[i] ~ dnorm(mu[i] , tau)
    mu[i] <- beta[1]+ beta[2]*HIGH_T[i]  + beta[3]*LOW_T[i]   +beta[4] *
    ↪  PRECIP[i]
    }
  for( j in 1:4) {
    beta[j] ~ dnorm(0, 0.001)
    }
  for( k in 1:10) {
    mean_resp[k] <- beta[1]+ beta[2]*TEST_HIGH_T[k]  + beta[3]*TEST_LOW_T[k]
    ↪  +beta[4] * TEST_PRECIP[k]
  pred[k] ~ dnorm(mean_resp[k] , tau)
    }
  tau ~ dgamma(0.1, 0.1)
  sigma <- 1/sqrt(tau)
}
INITS
list(beta=c(1,0,0,0) ,tau=0,pred=c(0,0,0,0,0,0,0,0,0,0))
DATA
list(n =204)
->DATA2<-
->DATA3<-
```

Dataset and .odc files are included in the submission. The above OpenBugs code produces the below output.

Node statistics

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
beta[1]	0.2515	31.48	0.03903	-61.44	0.2246	62.01	1001	1000000
beta[2]	75.93	8.903	0.1571	58.58	75.99	93.35	1001	1000000
beta[3]	-47.3	10.6	0.1871	-68.05	-47.35	-26.63	1001	1000000
beta[4]	-59.22	31.14	0.03866	-120.3	-59.21	1.817	1001	1000000
pred[1]	3555.0	705.0	2.293	2169.0	3555.0	4937.0	1001	1000000
pred[2]	3348.0	697.7	1.397	1979.0	3348.0	4717.0	1001	1000000
pred[3]	2306.0	693.3	0.7086	941.7	2305.0	3665.0	1001	1000000
pred[4]	3102.0	702.1	2.114	1722.0	3103.0	4476.0	1001	1000000
pred[5]	2714.0	694.0	0.9877	1349.0	2715.0	4073.0	1001	1000000
pred[6]	2495.0	694.8	0.808	1131.0	2495.0	3859.0	1001	1000000
pred[7]	2361.0	694.8	0.7916	998.6	2361.0	3727.0	1001	1000000
pred[8]	3061.0	695.5	1.054	1695.0	3060.0	4424.0	1001	1000000
pred[9]	3068.0	695.4	0.7118	1704.0	3067.0	4432.0	1001	1000000
pred[10]	2353.0	694.9	0.9578	990.0	2353.0	3719.0	1001	1000000

Beta values and prediction for 10 records are available in the output. Results are analyzed in detail in Section 4, Result Analysis.

3.Bayesian Poisson Regression

The same model data and prediction data is used and Bayessian Poisson regression.The Poisson regression is formulated as below.

$$BB\_COUNT \sim \text{Pois}(\lambda) \tag{5}$$

$$\log(\lambda) = \beta[1] + \beta[2] * HIGH\_T + \beta[3] * LOW\_T + \beta[4] * PRECIP \tag{6}$$

$$\beta = \mathcal{N}(0, 0.001) \tag{7}$$

The below OpenBugs code implement Bayesian Poisson Regression and .odc file is included in the submission.

```
model{
  for( i in 1:n) {
    BB_COUNT[i] ~ dpois(lambda[i])
    log(lambda[i]) <- beta[1]+ beta[2]*HIGH_T[i]  + beta[3]*LOW_T[i]
     ↪  +beta[4] * PRECIP[i]
    }
  for( j in 1:4) {
    beta[j] ~ dnorm(0, 0.001)
    }
  for( k in 1:10) {
    log(mean_resp[k]) <- beta[1]+ beta[2]*TEST_HIGH_T[k]  +
     ↪  beta[3]*TEST_LOW_T[k]   +beta[4] * TEST_PRECIP[k]
  pred[k] ~ dpois(mean_resp[k])
    }
}
INIT
list(beta=c(1,0,0,0) ,pred=c(0,0,0,0,0,0,0,0,0,0))
DATA
list(n =204)
->DATA2<-
->DATA3<-
```

The OpenBugs code produces the below output

Node statistics

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
beta[1]	7.041	0.01048	3.639E-5	7.021	7.041	7.062	1001	100000
beta[2]	0.02302	3.102E-4	1.077E-6	0.02241	0.02302	0.02362	1001	100000
beta[3]	-0.01287	3.353E-4	1.127E-6	-0.01353	-0.01287	-0.01222	1001	100000
beta[4]	-0.7652	0.006896	2.255E-5	-0.7787	-0.7652	-0.7516	1001	100000
pred[1]	3654.0	62.05	0.1815	3533.0	3654.0	3777.0	1001	100000
pred[2]	3310.0	58.24	0.1851	3196.0	3310.0	3425.0	1001	100000
pred[3]	2487.0	50.28	0.1709	2389.0	2487.0	2587.0	1001	100000
pred[4]	3148.0	57.06	0.1885	3037.0	3147.0	3260.0	1001	100000
pred[5]	2819.0	53.24	0.1611	2716.0	2819.0	2924.0	1001	100000
pred[6]	1772.0	42.47	0.1315	1690.0	1772.0	1856.0	1001	100000
pred[7]	1468.0	38.91	0.1265	1392.0	1468.0	1544.0	1001	100000
pred[8]	3165.0	56.47	0.1774	3055.0	3164.0	3276.0	1001	100000
pred[9]	3183.0	56.64	0.1759	3073.0	3183.0	3294.0	1001	100000
pred[10]	1866.0	43.55	0.138	1781.0	1865.0	1952.0	1001	100000

Beta values and prediction for 10 records are available in the output. Results are analyzed in Section 4, Result Analysis.

4.Result Analysis

The below table shows output, mean error, mean absolute error for the predicted results.

HIGH_T	LOW_T	PRECIP	BB_COUNT			Error		Absolute Error	
			Actual	Linear	Poisson	Linear	Poisson	Linear	Poisson
82.9	57.9	0	3679	3666	3654	13	25	13	25
84	64	0.06	3315	3348	3310	-33	5	33	5
64	54	0	2225	2306	2487	-81	-262	81	262
72	50	0	3084	3102	3148	-18	-64	18	64
70	55	0	2945	2714	2819	231	126	231	126
66.9	54	0.53	1876	2485	1772	-609	104	609	104
69.1	60.1	0.74	1004	2361	1468	-1357	-464	1357	464
79	62.1	0	2866	3061	3165	-195	-299	195	299
84	70	0.01	3244	3068	3183	176	61	176	61
70	62.1	0.42	1232	2363	1866	-1131	-634	1131	634
					Mean Error	**-300.4**	**-140.2**	**384.4**	**204.4**

From the above results it is evident that the Bayesian Poisson Regression perform superior to Bayesian Linear regression on this count data. The data where the values are quite extreme(Rows 6,7,10), Poisson regression performs much better than the linear regression.

Both mean error and mean absolute error metrics are better for Bayesian Poisson regression compared to the Bayesian linear regression.

End.