OM386 Marketing Analytics II
Assignment 3
Due: March 13th, 11:59pm

Censored Regression

In this exercise, we will apply censored regression to the dataset "CreditCard_SOW_Data2.csv". The dataset has the following variables.

ConsumerID	ID's of the sampled consumers
SOW	The card's share of wallet in the consumer's total monthly spending
Promotion	Index of monthly promotion activity –higher index indicates more promotions
Balance	The customer's unpaid balance at the beginning of the month

1). In this data set, the share of wallet (SOW) can be 0% or 100%. Therefore, the share of wallet is considered a truncated variable at 0% (censored at 0) or 100% (censored at 1). We would like fit the following regression model

$SOW_{ij}^* = \beta_0 + \beta_1 \times Balance_{ij} + \beta_2 \times Promotion_{ij} + \varepsilon_{ij}$
$SOW_{ij} = SOW_{ij}^*$ if $0 < SOW_{ij}^* < 1$
$SOW_{ij} = 0$ if $SOW_{ij}^* \leq 0$
$SOW_{ij} = 1$ if $SOW_{ij}^* \geq 1$
$\varepsilon_{ij} \sim N(0, \sigma^2)$ (We can reparametrize $\tau = 1/\sigma^2$, which is often named the precision)

Please use the R function censReg() in library(censReg) to fit this model by MLE. Copy and paste the summary of the results here. Interpret the parameters $\beta_{1\,2}$ in the model.

```
Call:
censReg(formula = SOW ~ Promotion + Balance, left = 0, right = 1,
    data = sow)

Observations:
         Total  Left-censored     Uncensored Right-censored
          3600            831           2436            333

Coefficients:
              Estimate Std. error t value Pr(> t)
(Intercept)  6.016e-01  8.357e-03   71.98  <2e-16 ***
Promotion    5.072e-01  1.174e-02   43.21  <2e-16 ***
Balance     -5.107e-04  6.135e-06  -83.23  <2e-16 ***
logSigma    -1.604e+00  1.465e-02 -109.43  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Newton-Raphson maximisation, 9 iterations
Return code 2: successive function values within tolerance limit
Log-likelihood: -65.88155 on 4 Df
```

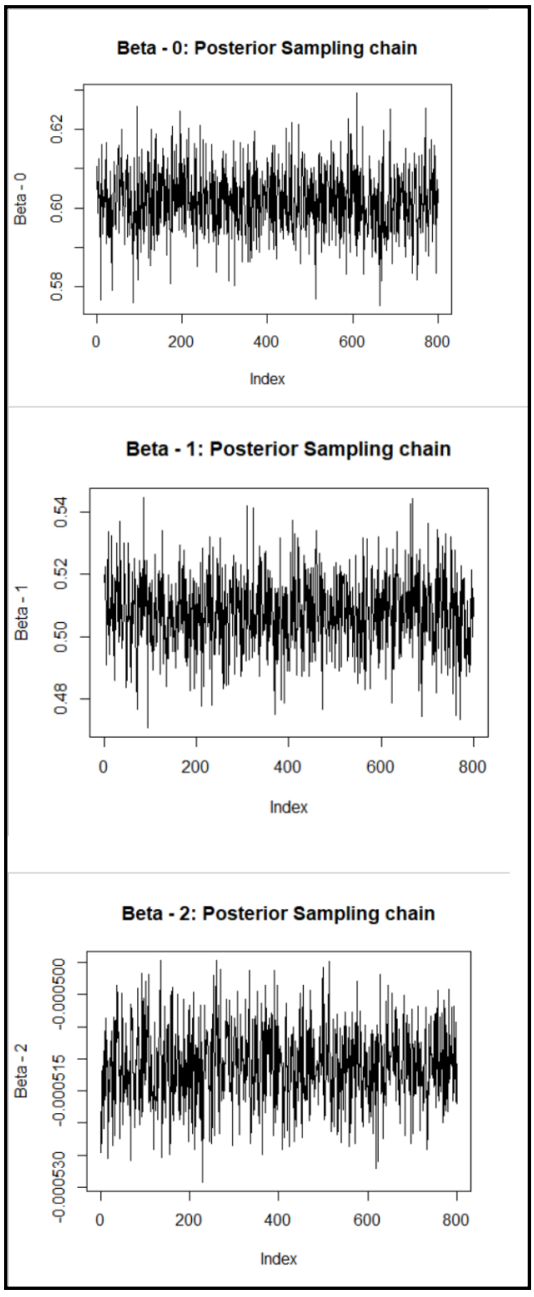
β_1 – **coefficient for balance:** Here, the negative coefficient for Balance suggests that:

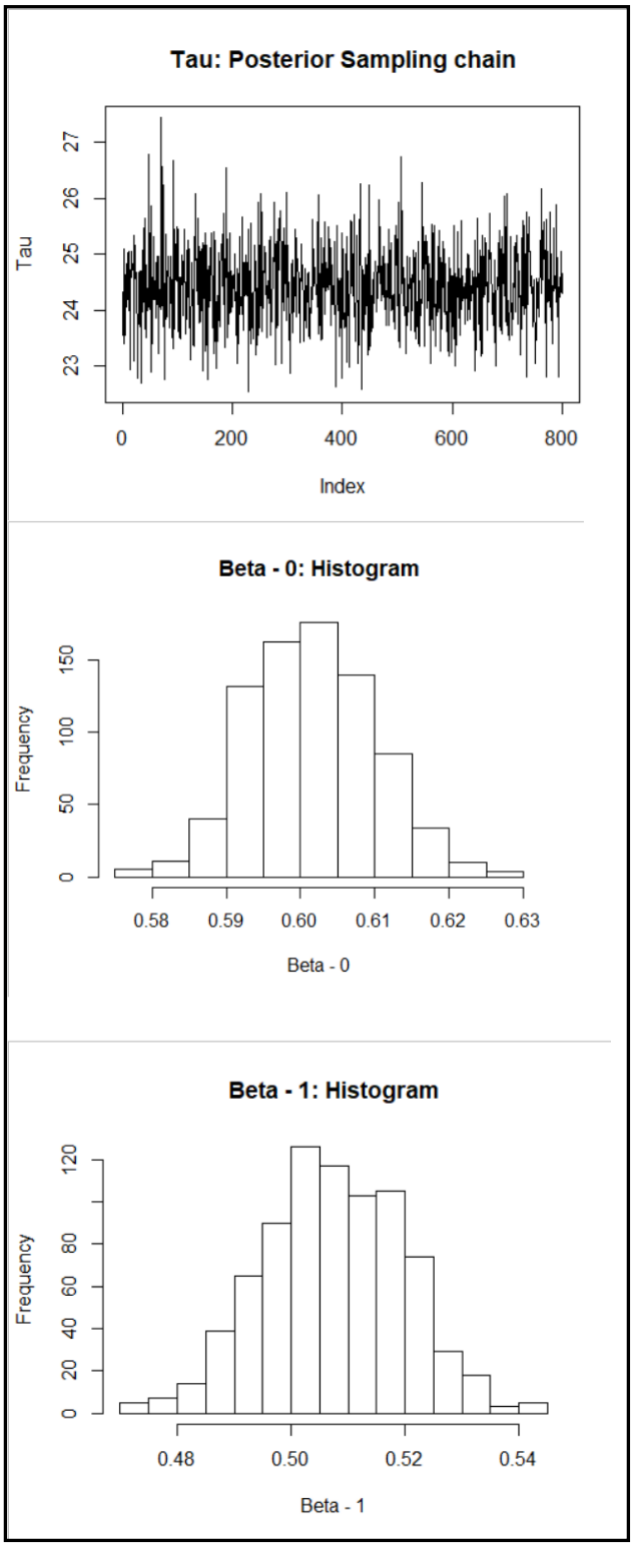
- as the customers unpaid balance at the beginning of the month increases, the share of wallet decreases.
- But the beta value is almost close to zero, this suggests that balance has a practically insignificant impact on the share of wallet.

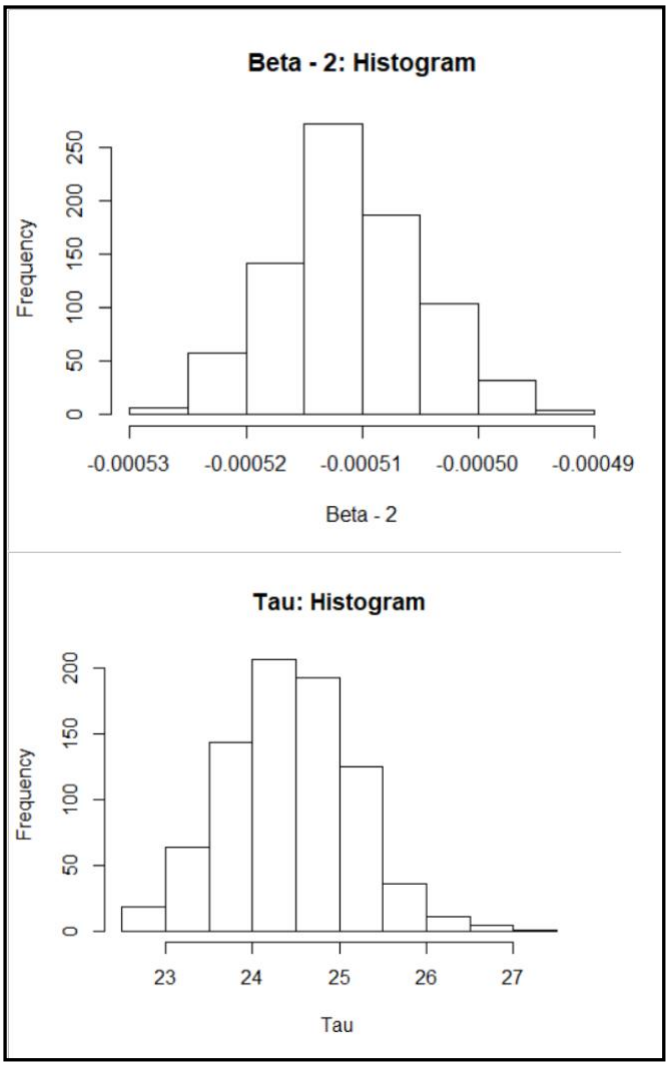
B_2 – **coefficient for promotion:** The positive coefficient for Promotion shows that the promotional activity and customers share of wallet are directly proportional.

2). Next, we will fit the model above using Bayesian estimation, which involves sampling the latent SOW_{ij}^{*} when the observed SOW_{ij} =0 or SOW_{ij} =1. The R code using MCMC (Gibbs sampling) for this inference problem is in "Assignment-4_blankcode.r". There are 2 blanks in the "step 1.a" and "step 1.b" of the "main loop" in the code. Please read the code carefully and fill in the code for sampling the latent SOW_{ij}^{*} in the 2 blanks. You may use the rtruncnorm( ) function in the library(truncnorm) to sample from truncated normal distributions. This method is called data augmentation, which is widely applied in Bayesian statistics for missing data.

Please run the completed code. Use the plot() function to plot the posterior sampling chains and hist() to plot posterior histograms for β_0, β_1, $_2$ and . Copy and paste the results here. Please also calculate the 95% posterior intervals for β_0, β_1, $_2$ and . Copy and paste the results here.




Tau: Posterior Sampling chain
Tau
23
24
25
26
27
0
200
400
600
800
Index
Beta - 0: Histogram
Frequency
0
50
100
150
0.58
0.59
0.60
0.61
0.62
0.63
Beta - 0
Beta - 1: Histogram
Frequency
0
20
40
60
80
120
0.48
0.50
0.52
0.54
Beta - 1



Posterior Intervals β0:

2.5%	50%	97.5%
0.5869358	0.6023205	0.6189392

Posterior Intervals β1:

2.5%	50%	97.5%
0.4856172	0.5078789	0.5297551

Posterior Intervals β2:

2.5%	50%	97.5%
-0.0005240876	-0.0005118647	-0.0004999240

Posterior Intervals:

2.5%	50%	97.5%
22.94419	24.43893	25.87934