PART 1: EDA

We first perform some exploratory analysis. Below, we can observe the histogram of the given data, which includes both when super discounts are present and are not.

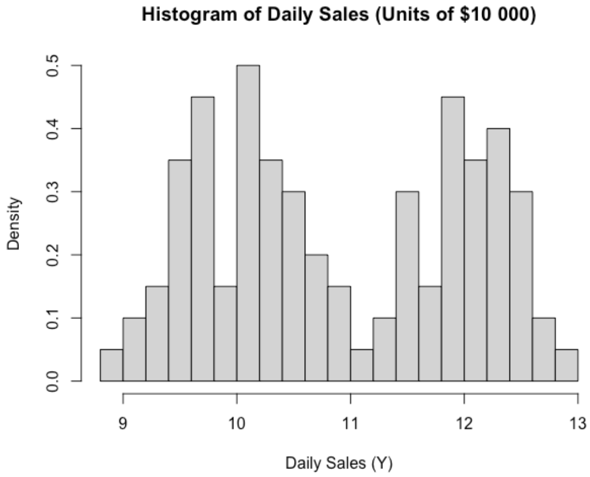

Figure 1.0. Histogram of Daily Sales

We observe that there seems to be 2 separate normal distributions here, with the first being when $y \leq 11$, and the second when $y > 11$. We first assume that the latter is when super discounts are present (since logically, super discounts should increase earnings, otherwise it wouldn't be implemented). Further, this also complies with the given information. First, cases when $y > 11$ accounts for 45% of all observations in our data, which is close to the departmental store's belief at 40%. Second, both cases seem to follow a normal distribution.

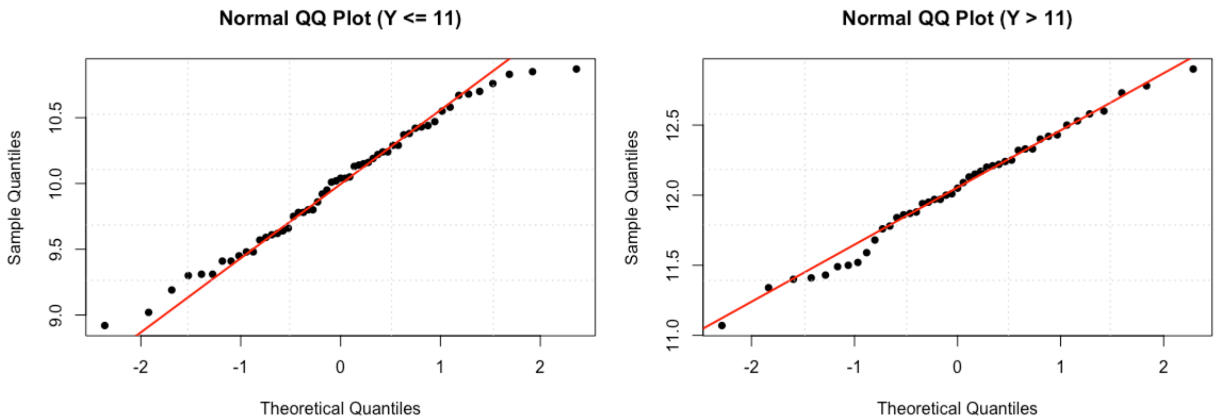

Figure 2.0. Normal QQ Plot Without Super Discounts (Left) & With Super Discounts (Right)

The values for both cases are mostly close to their respective theoretical normal quantities. We further confirm normality by performing the Shapiro-Wilk Normality Test (using the `shapiro.test()` command in R), whose null hypothesis states that values come from a normal population. Both cases resulted in p-values of larger than 0.05, hence we accept the null hypothesis and conclude both cases are normally distributed.

Next, we consider the possibility that super discounts actually decrease earnings. If we assume cases where $y \leq 11$ as earnings with super discounts, we will have super discount days' proportion at 0.55, which is significantly different from our given information. Alternatively, if we assume the lowest 40% of all earnings (approximately when $y \leq 10.4$) as super discounts earnings, then the earnings without super discounts will fail the Shapiro-Wilk Normality Test and this again clashes with the given information that both cases are normally distributed. Hence, from now onwards, we will assume that super discounts in general positively affect earnings. Below are some R outputs for the summary statistics.

Earnings assumed without super discounts ($y \leq 11$):

```
 Min.       1st Qu.    Median     Mean       3rd Qu.    Max.
 8.920      9.615      10.040     9.997      10.375     10.870
```

Earnings assumed with super discounts ($y > 11$):

```
  Min.      1st Qu.    Median     Mean       3rd Qu.    Max.
  11.07     11.78      12.05      12.04      12.33      12.90
```

Both set of samples have standard deviations of 0.495 and 0.421, respectively. We have each samples' population mean $\mu_i = \alpha + \gamma d_i$; $i = 1,2,\ldots,n$. Intuitively, α will represent the earnings' mean without super discounts, while γ represents the mean extra earnings with super discounts (i.e., when $d_i = 1$). Hence, from our data, we can estimate α to be around 10, and γ around 2.

PART 2: Likelihood Function

We have $y_i \sim N(\alpha + \gamma d_i\,, \psi)$; $i = 1,2,\ldots,n$ where ψ denotes the precision (i.e., $\frac{1}{\sigma^2}$). To derive the likelihood of our observed data, we first consider each observation to be a random draw from the following mixture distribution:

$$g(y_i; p, \alpha, \gamma, \psi) = p f_I(y_i) + (1-p) f_O(y_i)$$

where $f_I(y_i) = \sqrt{\frac{\psi}{2\pi}}\, e^{-\frac{1}{2}\psi(y_i-\alpha-\gamma)^2}$ i.e., the density for earnings on super discount days, and $f_O(y_i) = \sqrt{\frac{\psi}{2\pi}}\, e^{-\frac{1}{2}\psi(y_i-\alpha)^2}$ i.e., the density for earnings on days without super discounts. Thus, in the absence of $\underline{d}$, the likelihood of our observed data is:

$$L = \prod_{i=1}^{n} g(y_i; p, \alpha, \gamma, \psi)$$

In the presence of $\underline{d}$, our likelihood becomes:

$$\begin{aligned} L &= \prod_{i=1}^{n} \big(p f_I(y_i)\big)^{I(d_i=1)} \big((1-p) f_O(y_i)\big)^{I(d_i=0)} \\ &= p^{n_I(d)} (1-p)^{n_O(d)} \prod_{i=1}^{n_I(d)} f_I(y_i) \prod_{i=1}^{n_O(d)} f_O(y_i) \end{aligned}$$

where $n_I(d)$ and $n_O(d)$ denote the number of observations with and without super discounts, respectively.

Substituting the formula for $f_I(y_i)$ and $f_O(y_i)$ into the equation and simplifying the equation, the complete likelihood (in the presence of $\underline{d}$) becomes:

$$L\left(\underline{d}, \underline{y}; p, \alpha, \gamma, \psi\right) \;\propto\; p^{n_I(d)} \cdot (1-p)^{n_O(d)} \cdot \psi^{\frac{n}{2}} \cdot e^{-\frac{\psi}{2}\sum_{i=1}^{n_I(d)}(y_i-\alpha-\gamma)^2} \cdot e^{-\frac{\psi}{2}\sum_{i=1}^{n_O(d)}(y_i-\alpha)^2}$$

where $n = n_I(d) + n_O(d)$.

PART 3: Prior Distributions

We now choose our priors. First, for α and γ, we use a N(0, 100) distribution as our prior. We choose this prior specification because before seeing the data, we have very little knowledge on the location of the regression parameters. So, we choose a prior with a large variance. The choice of the mean will not have much importance here. Since it is a vague prior, it will have little impact on the posterior.

Next, for p, we choose a Beta distribution with a mean of 0.4 and a variance of 0.12 as our prior. The choice of the mean was to comply with the given information that super discount occurs on around 40% of the days. Furthermore, the conjugacy property of a Beta prior on a binomial likelihood will also make the inference on p more convenient. To get the exact distribution, we solve the following equations.

$$\frac{\alpha_1}{\alpha_1+\alpha_2} = 0.4 \quad and \qquad \frac{\alpha_1\alpha_2}{(\alpha_1+\alpha_2)^2(\alpha_1+\alpha_2+1)} = \frac{0.4\alpha_2}{(\alpha_1+\alpha_2)(\alpha_1+\alpha_2+1)} = 0.12$$

The 1st equation tells us that $\alpha_1 = \frac{2}{3}\alpha_2$. Substituting this to the 2nd equation allows us to find the value of α_2, which we can use to find α_1. Doing that, we find the exact prior distribution for p to be Beta(0.4, 0.6).

Finally, for ψ, we will use a $\Gamma(0.01, 0.01)$ prior. This prior will have a mean of 1 and a variance of 100. Similar to the case of α and γ, this is a vague prior.

PART 4: Posterior Function

We first assume that $\underline{d}$ was observed (we will use data augmentation to explore it later). Thus, the joint posterior can now be written as follows:

$$\pi(p, \alpha, \gamma, \psi \mid \underline{d}, \underline{y}) \propto \pi(p)\pi(\alpha)\pi(\gamma)\pi(\psi)p^{n_I(d)}(1-p)^{n_O(d)}\psi^{\frac{n}{2}}e^{-\frac{\psi}{2}\sum_{i=1}^{n_I(d)}(y_i-\alpha-\gamma)^2}e^{-\frac{\psi}{2}\sum_{i=1}^{n_O(d)}(y_i-\alpha)^2}$$

where: $\pi(p) \propto p^{-0.6}(1-p)^{-0.4}$

$\pi(\alpha) \propto e^{-\frac{\alpha^2}{200}}$

$\pi(\gamma) \propto e^{-\frac{\gamma^2}{200}}$

$\pi(\psi) \propto \psi^{-0.99}e^{-0.01\psi}$

Hence, we have:

$$\pi(p,\alpha,\gamma,\psi \mid \underline{d},\underline{y}) \propto p^{n_I(d)-0.6}(1-p)^{n_O(d)-0.4}\,\psi^{\frac{n}{2}-0.99}e^{-\frac{\psi}{2}\sum_{i=1}^{n_I(d)}(y_i-\alpha-\gamma)^2}e^{-\frac{\psi}{2}\sum_{i=1}^{n_O(d)}(y_i-\alpha)^2}e^{-0.01\psi}e^{-\frac{\alpha^2}{200}}\,e^{-\frac{\gamma^2}{200}}$$

We can now find the conditional posterior for each parameter by extracting only the factors that depend on each parameter.

$$\pi(p\mid \alpha,\gamma,\psi\ \underline{d},\underline{y}) \propto\ p^{n_I(d)-0.6}(1-p)^{n_O(d)-0.4}$$

$$\pi(\alpha\mid p,\gamma,\psi\ \underline{d},\underline{y}) \propto e^{-\frac{\psi}{2}\sum_{i=1}^{n_I(d)}(y_i-\alpha-\gamma)^2-\frac{\psi}{2}\sum_{i=1}^{n_O(d)}(y_i-\alpha)^2-\frac{\alpha^2}{200}} \qquad \text{(i)}$$

$$\pi(\gamma\mid p,\alpha,\psi,\underline{d},\underline{y}) \propto e^{-\frac{\psi}{2}\sum_{i=1}^{n_I(d)}(y_i-\alpha-\gamma)^2-\frac{\gamma^2}{200}} \qquad \text{(ii)}$$

$$\pi(\psi\mid p,\alpha,\gamma,\underline{d},\underline{y}) \propto\ \psi^{\frac{n}{2}-0.99}e^{-\psi[\frac{1}{2}\sum_{i=1}^{n_I(d)}(y_i-\alpha-\gamma)^2+\frac{1}{2}\sum_{i=1}^{n_O(d)}(y_i-\alpha)^2+0.01]}$$

Since we actually don't have the values for $\underline{d}$, we treat it as an additional parameter to explore and its conditional posterior is as follows.

$$\pi(\underline{d}\mid p,\alpha,\gamma,\psi,\underline{y}) \propto \prod_{i=1}^{n}\left(pf_I(y_i)\right)^{I(d_i=1)}\left((1-p)f_O(y_i)\right)^{I(d_i=0)} \propto \prod_{i=1}^{n}\pi(d_i\mid p,\alpha,\gamma,\psi,y_i)$$

where $\pi(d_i\mid p,\alpha,\gamma,\psi,y_i)=\frac{pf_I(y_i)}{pf_I(y_i)+(1-p)f_O(y_i)}$ when $d_i=1$ and $\pi(d_i\mid p,\alpha,\gamma,\psi,y_i)=\frac{(1-p)f_O(y_i)}{pf_I(y_i)+(1-p)f_O(y_i)}$ when $d_i=0$, with $f_I(y_i)$ and $f_O(y_i)$ defined earlier in question 2. Next, we observe that the conditional posteriors of p and ψ follow the following distributions:

$$(p\mid \alpha,\gamma,\psi\ \underline{d},\underline{y})\sim Beta(n_I(d)+0.4\,,n_O(d)+0.6) \qquad \text{(iii)}$$

$$(\psi\mid p,\alpha,\gamma,\underline{d},\underline{y})\sim\Gamma(\tfrac{n}{2}+0.01\,,\tfrac{1}{2}\textstyle\sum_{i=1}^{n_I(d)}(y_i-\alpha-\gamma)^2+\tfrac{1}{2}\sum_{i=1}^{n_O(d)}(y_i-\alpha)^2+0.01) \qquad \text{(iv)}$$

The conditional posterior of α and γ don't seem to follow any standard distribution. So, to sample from them, we can use the Metropolis-Hastings algorithm. The full algorithm works as follows:

1) Set initial values for $p,\alpha,\gamma,\psi\ \underline{d}$ (will be discussed in question 5) as well as for Δ_α and Δ_γ.
2) Draw a sample for p and ψ from the distributions in (iii) and (iv), respectively. These are Gibbs steps.
3) For each j[th] iteration, propose a new value for α from a U($\alpha_{j-1}-\Delta_\alpha\,,\alpha_{j-1}+\Delta_\alpha$). Denote this value by α'. Compute the acceptance probability $P^{Acc}=\min\left(\frac{\pi(\alpha'\mid p,\gamma,\psi\ \underline{d},\underline{y})}{\pi(\alpha_{j-1}\mid p,\gamma,\psi\ \underline{d},\underline{y})}\,,1\right)$. The formula for $\pi(\alpha\mid p,\gamma,\psi\ \underline{d},\underline{y})$ can be found in (i). When calculating, use the updated values for ψ.
4) Generate a random number from U(0,1). If this number is less than or equal to the P^{Acc}, set $\alpha_j=\alpha'$. Otherwise set $\alpha_j=\alpha_{j-1}$.
5) Repeat steps 3 and 4 for sampling γ_j. Use updated values of ψ and α in the calculations. The conditional posterior can be found in (ii). Steps 3-5 are M-H steps.
6) For each observation, calculate $p_i=\frac{pf_I(y_i)}{pf_I(y_i)+(1-p)f_O(y_i)}$, $i=1,2,\dots,100$ using the updated values of each parameter.

7) Generate $U_1, U_2, \dots, U_{100}$ $where\ U_i \sim U(0\,,1)\ i.i.d.$ For each observation, if $U_i \le p_i$, set $d_i = 1$. Otherwise set $d_i = 0$.
8) Repeat steps 2-7 for a large number of iterations (in our case, 20 000). Analyze chain and distribution for each parameter, then apply thinning and parameters (Δ_α, Δ_γ, number of iterations, burn-in, etc.) tuning where necessary.

PART 5: Running the MH Algorithm

We now implement the algorithm discussed in question 4. First, we choose the starting values for the parameters. For p, we set it to 0.4 to comply with the store's beliefs. For α and γ, we choose 10 and 2 (as estimated in the EDA process in question 1). Finally, for ψ, we choose the value $\frac{1}{0.45^2} = 4.938$. We obtain this value by first computing the confidence interval for population variance of a Normal distribution, for both the samples where $y \le 11$ and $y > 11$. The 95% confidence interval is given by this formula:

$$95\%\ CI\ for\ \sigma^2 = \left(\frac{(n-1)s^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)s^2}{\chi^2_{n-1,\frac{\alpha}{2}}} \right)$$

where n denotes the sample size, s^2 the sample variance, and χ^2_{n-1} denotes a chi-square distribution with (n-1) degrees of freedom. Doing so and taking the square root, we obtain the 95% interval for σ as (0.417, 0.61) for the samples where $y \le 11$ and (0.35, 0.532) for the samples where $y > 11$. We then pick any value that is present in both intervals, in our case, 0.45, and calculate $\psi_1 = \frac{1}{0.45^2}$. For $\underline{d}$, we initially set $d_i = 1$ if $y_i > 11$ and $d_i = 0$ otherwise.

Running the algorithm for 20000 iterations with Δ_a and Δ_γ at 0.25, we obtain the following trace plots.

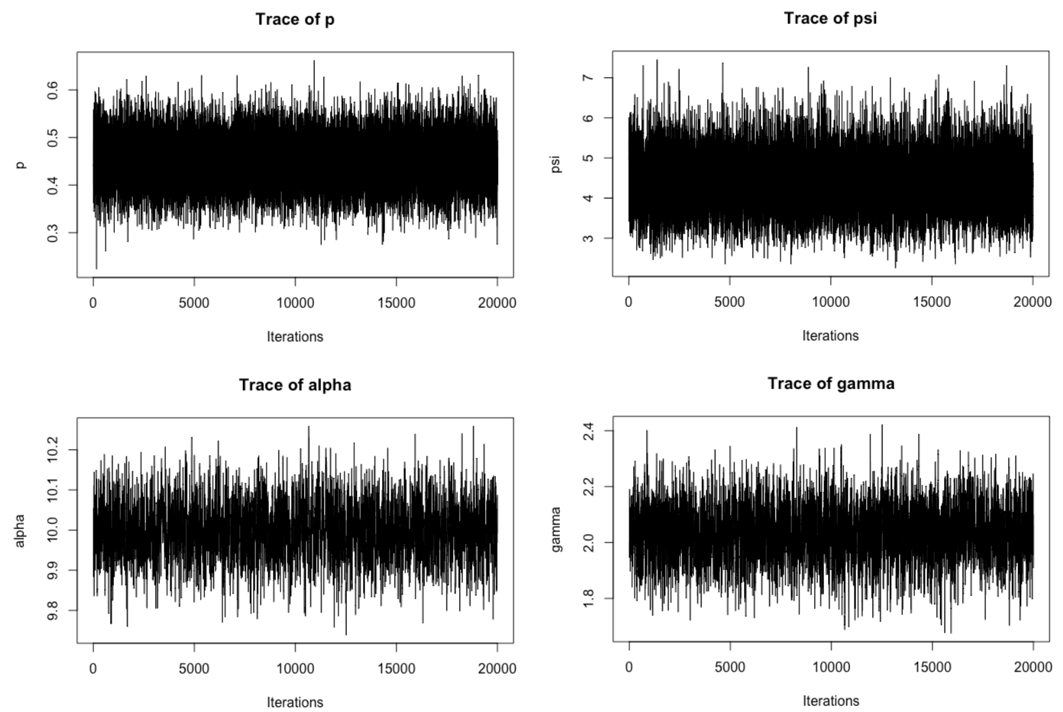

Figure 3.0. Trace for $p, \psi, \alpha, and\ \gamma$

We observe that the chains have stabilized not far from their respective initial values. While it does seem that our chains have converged, we would like to confirm this further. So, we run the algorithm for a second time, but with the initial values of p, ψ, α, γ all set to 0.1. If our algorithm works correctly, then we would see the chains converge around the same values as above. The results from this second trial are shown below.

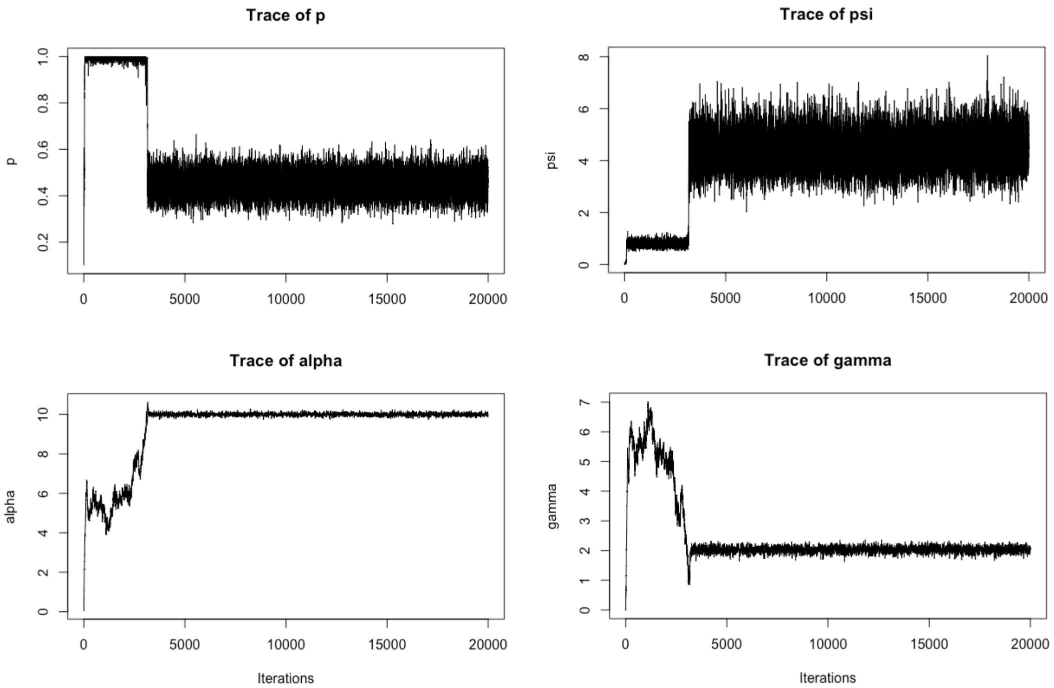

Figure 4.0. Trace for $p, \psi, \alpha, and\ \gamma$ with starting values all set to 0.1

All chains converged to similar values as in the first trial. Hence, we conclude that our algorithm indeed works. We further set the burn-in period to 5000 and apply thinning with m = 20. This is to ensure that our samples have converged and are independent. With the generated samples, we observe each parameter's histogram below.

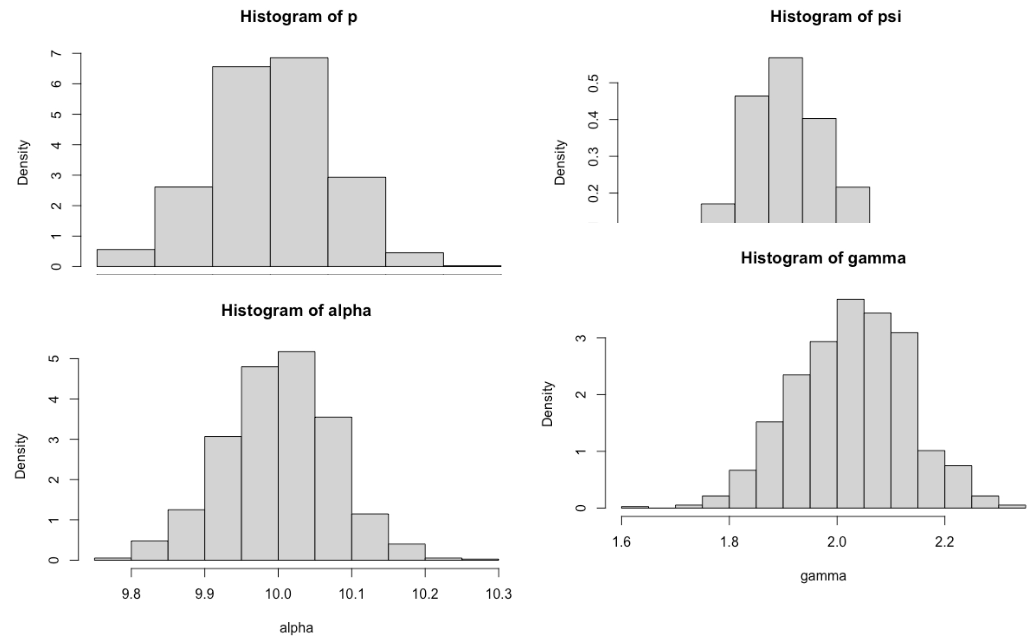

Figure 5.0. Histogram of $p, \psi, \alpha, and\ \gamma$

Now we plot the autocorrelation using R's `ggAcf()` command from the `forecast` library.

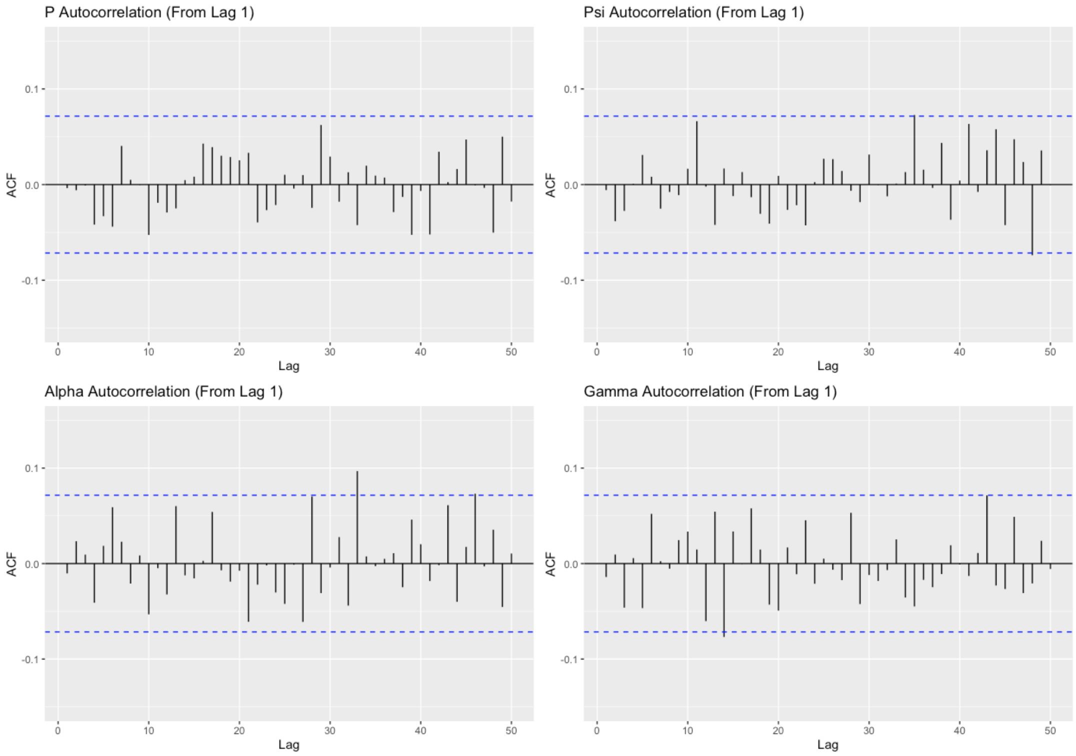

Figure 6.0. Autocorrelation of $p, \psi, \alpha, and\ \gamma$

There is no significant correlation. Hence, the samples can be said to be independent. Now, we provide the estimates for the mean, standard deviation and 95% CI of each parameter.

- p : mean = 0.4514, standard deviation = 0.0516, CI = (0.3497, 0.5488)
- ψ: mean = 4.3426, standard deviation = 0.7293, CI = (3.085, 5.8685)
- α: mean = 10.0012, standard deviation = 0.0749, CI = (9.8483, 10.1457)
- γ: mean = 2.02732, standard deviation = 0.1057, CI = (1.818, 2.2322)

With the obtained $\underline{d}$, we plot the histograms of the earnings with & without super discounts.

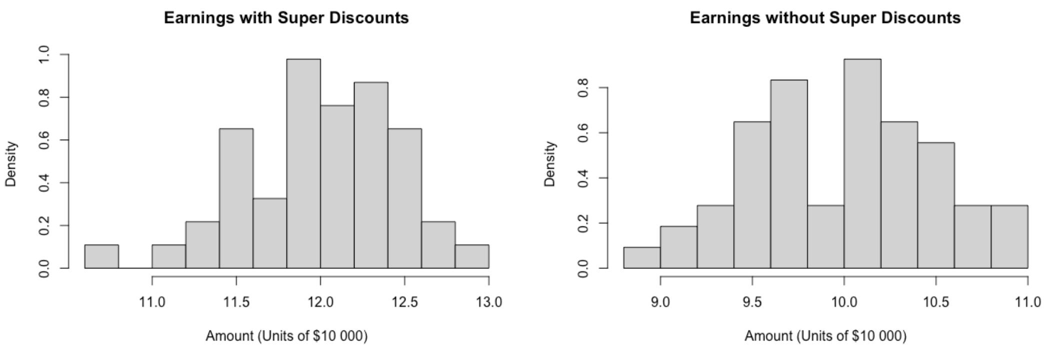

Figure 7.0. Earnings with (left) and without (right) super discounts

PART 6: Conclusion

We conclude that super discounts occur on around 45% of all days, and on average, brings in an extra $20 263. Without super discounts, daily sales have a mean of $99 844. With super discounts, they have a mean of $120 107. We estimate the standard deviation of the earnings in both cases to be $4 799.