Written Report #1

Baseline Investigation

Shuo Li

# Stat 435: Baseline

## Question

The problem we are going to investigate is the baseline value of the WatFactory, more specially the y300 values. In order to get a evaluate of the performance of the solution for reduce variance in WatFactory, we first need to get a accurate understanding/measurement of the baseline outputs and variance of y300.

## Plan

The plan is to collect in total 500 y300 part outputs in a 5 day period. We took measurements of 100 y300 output parts per day randomly for 5 days. We choose a 5 day period because WatFactory only works 5 days a week and 5 days is mostly likely enough data to determine if the days of the week contributes to any variance of y300. We also feel that 100 data per day spread out randomly will capture approximately equal data for all three shifts in a day, and it will enough to accurately represent y300 properties for the day. We feel that 500 y300 data gives about Relative Precision, which is enough Relative Precision to be an accurate representation for the true y300 properties (mean, variance, etc). We could have collected 1000 y300 data which will give a Relative Precision around . However, we feel that because the cost of collecting 1000 y300 data is double of what it would cost for collecting 500 y30 data, and it does not reduce Relative Precision nearly as much (only ). The cost does not justify the benefit for collection 1000 y300 data instead of 500.

## Data

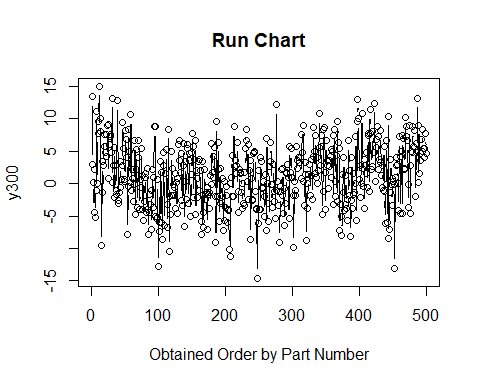
## daycount shift partnum y300   
## Min. :1 Min. :1.00 Min. : 5 Min. :-14.600   
## 1st Qu.:2 1st Qu.:1.00 1st Qu.:1793 1st Qu.: -2.200   
## Median :3 Median :2.00 Median :3555 Median : 1.400   
## Mean :3 Mean :2.02 Mean :3609 Mean : 1.265   
## 3rd Qu.:4 3rd Qu.:3.00 3rd Qu.:5416 3rd Qu.: 4.800   
## Max. :5 Max. :3.00 Max. :7198 Max. : 15.000

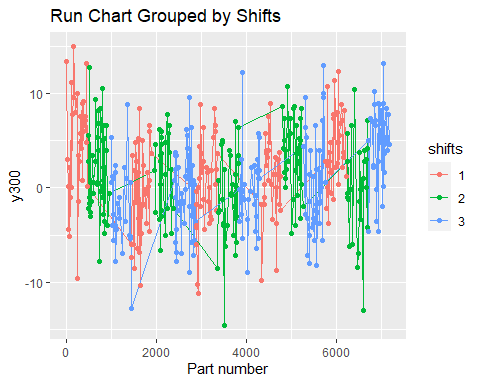
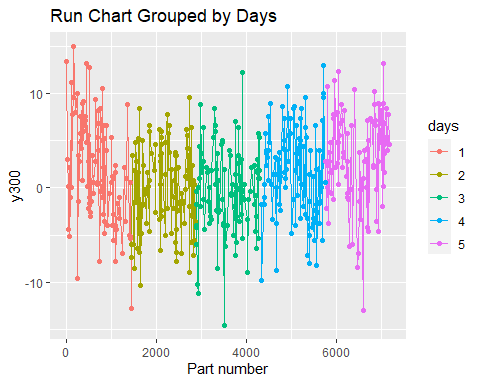
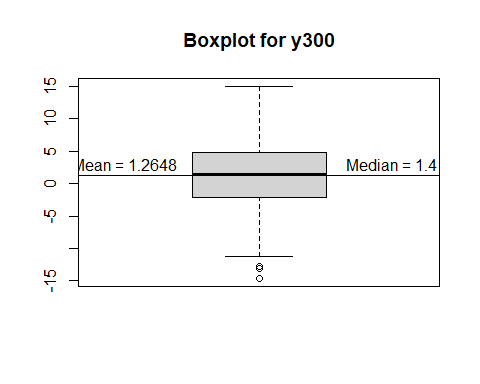
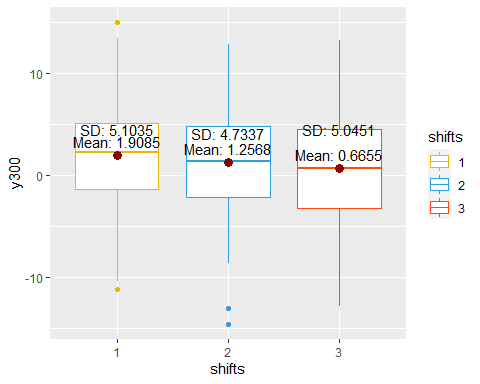
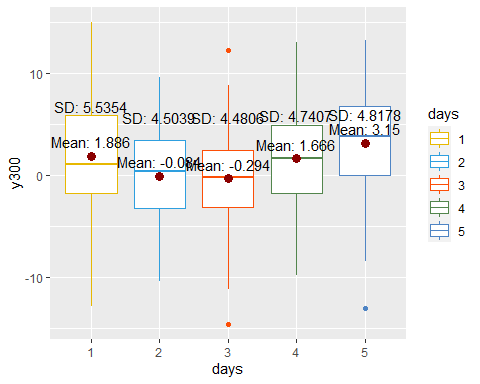
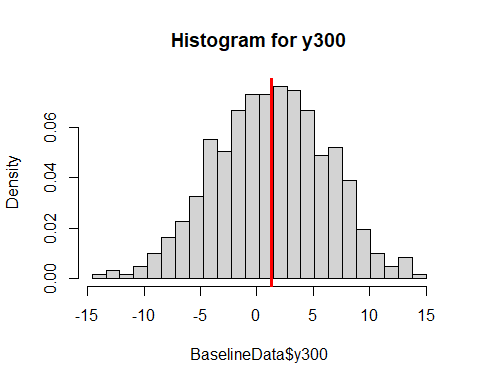
## [1] "The standard deviation of y300 is: 4.98232154883966"

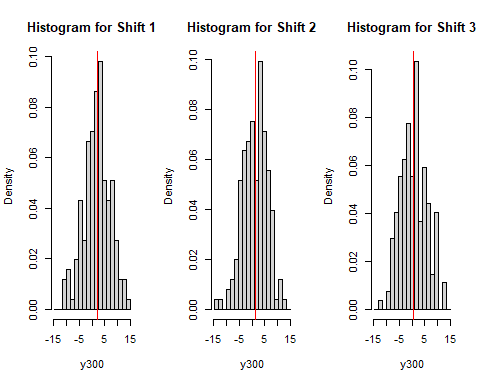
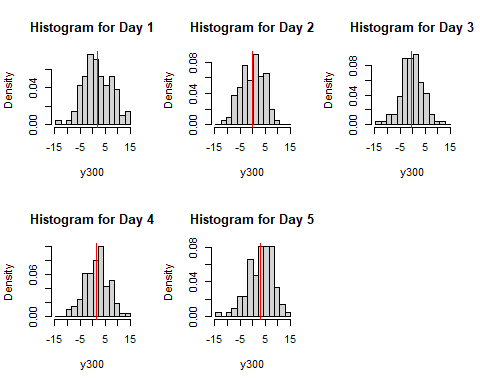
## [1] "The number of rows in this dataset is: 500"

From the R outputs, we see that we have 500 data points for y300 over 3 different shifts per day for a total of 5 days. From the output above, we see that the mean and median of y300 is similar, IQR is between -2.200 and 4.800, and the standard deviation is 4.98232154883966 The maximum output of y300 is 15.00 and the minimum output of y300 is -14.600

## Analysis

 From the Run Chart of y300 above, we can see that it does not seem to have constant mean. Rather the output of y300 goes up and down slightly as the partnumber increases.\

Next we will investigate the output of y300 conditioned on different shifts or days.  Based on the run chart by shifts, we can see that for same shift, the mean and variance varies.  From the Run chart grouped by days above, we see that there seems to be a big decrease for day 1 as the the part number increases.  From the box plot, we see that the there are a few outliers below the whiskers and no outliers above the whisker.  The plot above is the histogram of y300 grouped by shifts, this should tell us if shifts outputs different mean and/or variance for y300. We see the mean decrease as shift number increase. However, the variance does not seem to have a increase or decreasing pattern when shift number increases. The variance seems to follow quadratic shape in increase and deacrease as shift increases. We see that shift 1 have the highest mean and variance out of the three shifts, and it has one outlier above the whisker and one below the whisker. Shift 2 has the second highest mean but the lowest variance, it has two outliers and both are below the whiskers. Shift 3 has the lowest mean but second highest variance, but it does not have any outliers above or below the whiskers.  The plot above is the histogram of y300 grouped by days, this should tell us if days have outputs different mean and/or variance for y300. We see a V pattern of the mean and variance as days increases. From day 1 to day 3 the mean and variance decreases by day, and from day 3 to day 5 the mean and variance increases by day. This suggests that the mean of y300 changes depending on the day of the week it was produced. The mean change from day 1 to day 2, day 3 to day 4, and day 4 to day 5 is significantly larger in magnitude compared to the mean change from day 2 to day 3. The variance change from day 1 to day 2, and day 3 to day 4 is significantly larger in magnitude compared to the variance change from day 2 to day 3 and day 4 to day 5. There are only two days where outlier above and/or below the whiskers. These days are day 3 where there is a outlier above and below the whisker, and day 5 where there is a outlier below the whisker.  From the histogram of y300 outputs, we can see it seems to following a normal distribution with a slight left skew.

 From the Histogram by shifts plot, we see that the histogram of all three shifts seems to have normal distribution.  From the histogram by days plot, we see that all the days seems to have a normal distribution and slight left skewed.

## Conclusion

From the analysis, we see that y300 has mean of 4.98232154883966, median of 4.98232154883966 and standard deviation of 4.98232154883966 The IQR is between -2.200 and 4.800, maximum output of y300 is 15.00 and the minimum output of y300 is -14.600 We saw that the Histogram of y300, y300 grouped by shifts, and y300 grouped by days all seems to have a normal distribution with a slight left skew. From the box plot of y300 outputs, we see that the only outliers are below the whiskers. From the box plot of y300 outputs grouped by shift, we see the mean decreases as shift increase and the variance seems to follow quadratic shape as shift increases centered around the shift 2. From the box plot of y300 outputs group by days, we see that both mean and variance seems to have a quadratic shape or V shape as days increase centered around day 3. The mean increase/decrease is the biggest in magnitude for day 1 to day 2, day 3 to day 4, and day 4 to day 5. The standard deviation increase/decrease is the biggest in magnitude for day 1 to day 2, and day 3 to day 4.