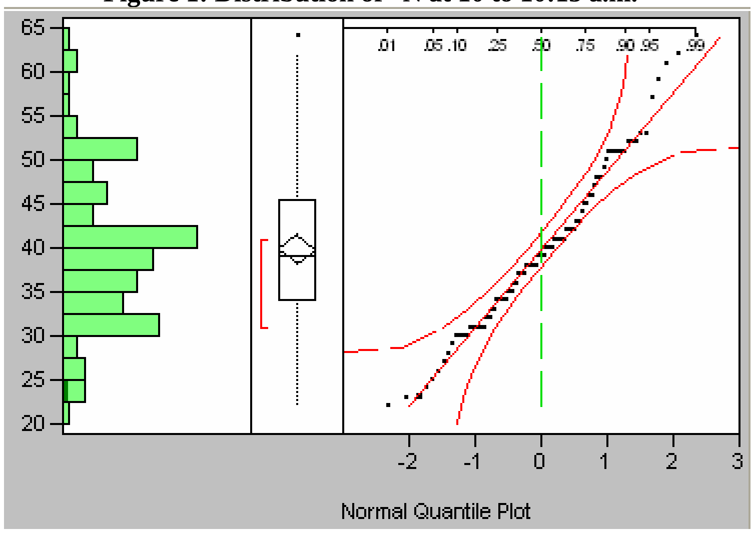
Solution of Project 3 – Call Center Project

Part 1: Using ANOVA to predict incoming calls during 10-10:15 a.m.

1. Since we can not use any facts about the hypothesized day other than it is Wednesday, so we will use ANOVA to do the prediction. From the figure 1,

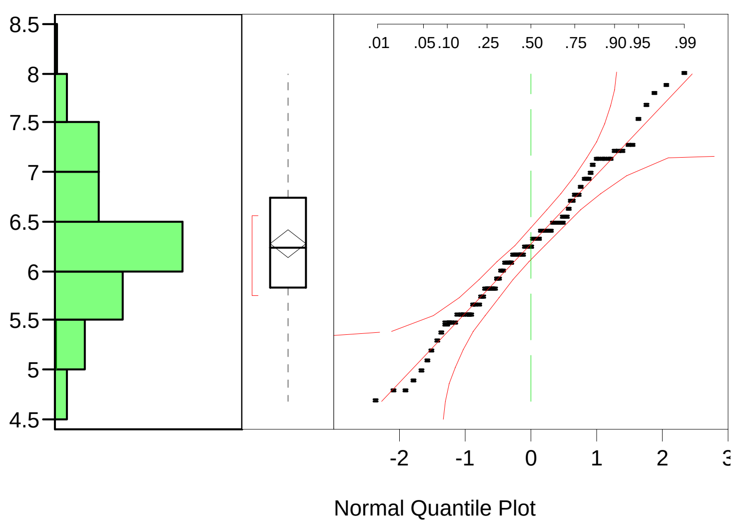
Figure 1: Distribution of "N at 10 to 10:15 a.m."



there is no significant outliners and "N at 10 to 10:15a.m." seems normal but a little bit skewed to the right.

Let's try the square-root transformation of N at 10 to 10:15a.m.

Distribution of sqrt(N at 10 to 10:15a.m.)

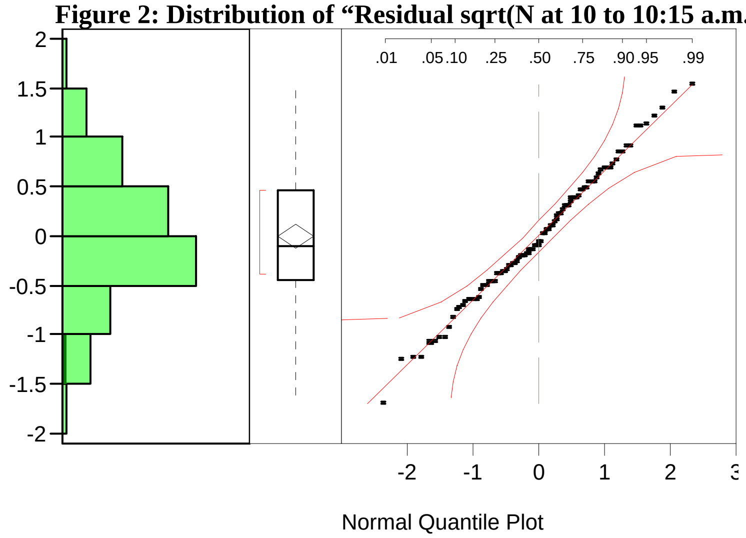


The transformed data looks more like normal. So we will use the transformed data as the response variable.

2. Two assumptions are needed for ANOVA –
   a. Residuals follow normal distribution.
   b. Equal variance across groups.

   From figure 2, assumption a is valid. From ANOVA result, equal variance across groups is also valid.

Figure 2: Distribution of “Residual sqrt(N at 10 to 10:15 a.m.)”



Oneway Anova

Summary of Fit

Rsquare	0.142107
Adj Rsquare	0.106361
Root Mean Square Error	0.663622
Mean of Response	6.27246
Observations (or Sum Wgts)	101

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Day of week	4	7.003167	1.75079	3.9755	0.0050
Error	96	42.277799	0.44039		
C. Total	100	49.280966			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
1	19	6.57811	0.15225	6.2759	6.8803
2	20	6.29293	0.14839	5.9984	6.5875
3	22	6.53643	0.14148	6.2556	6.8173
4	20	5.92939	0.14839	5.6348	6.2239
5	20	6.01431	0.14839	5.7198	6.3089

Std Error uses a pooled estimate of error variance

3. From above ANOVA analysis, we reject that there is no difference of mean of number of incoming calls among different day of the week.

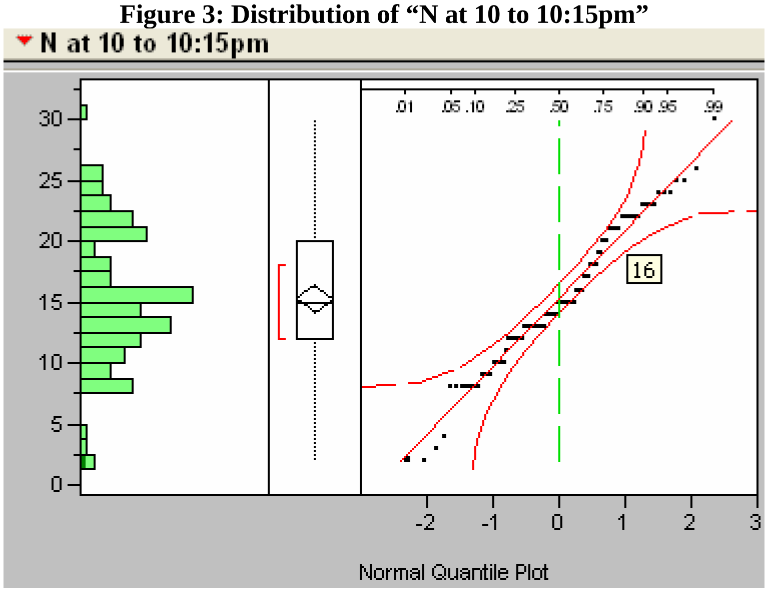
$\hat{y}$ = 6.29 since $\sqrt{s^2 + s_{\hat{y}}^{\,2}}$ =0.68 (In JMP IN you can get this standard error by save column -> Std Error of Individual), so 90% safety value:

$(\hat{y} + t_{.10,96}\sqrt{s^2 + s_{\hat{y}}^{\,2}})^2 = (6.29 + 1.29*.68)^2 = (7.17)^2 = 51.41$

Part 2: Using Multiple Regression to predict incoming calls during 10-10:15 p.m.

1. We first check the distribution of “N at 10-10:15pm”. From figure 3, there is no significant outlier and it is approximately normal. By checking other variables, there are no significant outliers either.

Figure 3: Distribution of “N at 10 to 10:15pm”



Summary 2: Multiple Regression

Response N at 10 to 10:15pm

Summary of Fit

RSquare	0.273444
RSquare Adj	0.227068
Root Mean Square Error	4.908461
Mean of Response	15.29703
Observations (or Sum Wgts)	101

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	852.3485	142.058	5.8962
Error	94	2264.7407	24.093	Prob > F
C. Total	100	3117.0891		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	6.652778	2.917186	2.28	0.0248
Day of week[1]	-3.333987	1.024715	-3.25	0.0016
Day of week[2]	2.6643136	0.985426	2.70	0.0081
Day of week[3]	1.8572142	0.948538	1.96	0.0532
Day of week[4]	0.1971802	0.992266	0.20	0.8429
N before 12	-0.038755	0.014534	-2.67	0.0090
N before 6	0.0243232	0.006534	3.72	0.0003

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Day of week	4	4	463.50065	4.8095	0.0014
N before 12	1	1	171.31469	7.1106	0.0090
N before 6	1	1	333.87725	13.8579	0.0003

Residual by Predicted Plot

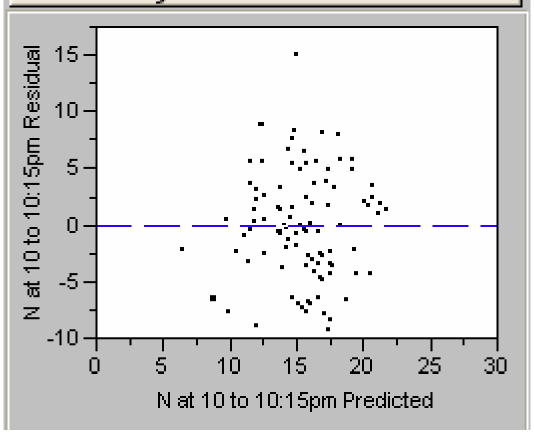
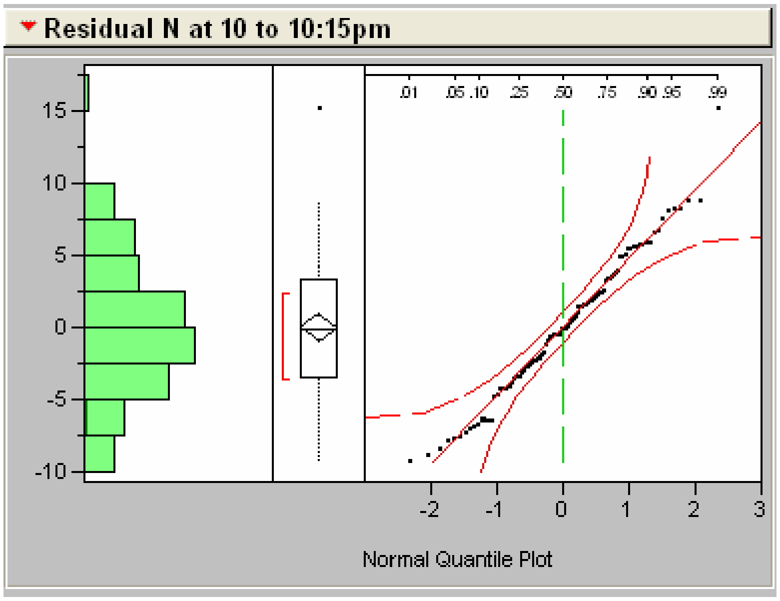


Figure 4: Residual Plot



2. By adding significant variables one by one and remove non-significant ones, we get out final model. (Summary 2 of Multiple Regression) From residuals normal quantile plot, residuals are approximately normal. From residual vs predicted, constant variance of residuals can't be rejected.

$\hat{y}$ = 13.54,

and 90% safety value is: 13.54+6.60=20.14