Case 5 - Treatment Facility:
Two Sample Means and Time Series

Marlene Smith, University of Colorado Denver
Business School

Treatment Facility[1]:
Two Sample Means and Time Series

Background

The New Life Residential Treatment Facility is a nonprofit foundation that treats teenagers who have shown signs of mental illness. It provides housing and supervision of teenagers who are making the transition from psychiatric hospitals back into the community. Because many of the teenagers were severely abused as children and have been involved with the juvenile justice system, behavioral problems are common at New Life. Employee pay is low and staff turnover is high.

A reengineering program was instituted at New Life with the goals of lowering behavioral problems of the kids and decreasing employee turnover rates. As a part of this effort, the following changes were made:

- Employee shifts were shortened from 10 hours to 8 hours each day.
- Employees were motivated to become more involved in patient treatments. This included encouraging staff to run various therapeutic treatment sessions and allowing staff to have more say in program changes.
- The activities budget was increased.
- A facility-wide performance evaluation system was put into place that rewarded staff participation and innovation.
- Management and staff instituted a program designed to raise expectations about appropriate behavior from the kids. This included strict compliance with reporting of behavioral violations, insistence on participation in therapeutic sessions, and a lowered tolerance for even moderate behavioral infractions.

To determine the effectiveness of the reengineering effort, a data set comprised of pre- and post-reengineering periods was compiled. The information contains two measures of behavioral problems. A *critical incident* occurs when a resident goes AWOL (leaves the premises without permission), destroys property (e.g., punching a hole in a wall or throwing furniture through windows), is caught in possession of street drugs, or engages in assault against other residents or staff members. A teenager is *temporarily removed from the facility* when s/he is sent to jail or back to a psychiatric hospital.

The Task

Determine what effect, if any, the reengineering effort had on the incidence behavioral problems and staff turnover.

The Data Treatment Facility.jmp

The data set contains 20 months of data; the first 13 months were prior to reengineering. The variables in the data include:

Reengineer	Whether the month was before (Prior) or after (Post) reengineering
Employee Turnover	The percentage of employees who quit in a given month, out of the total number of employees

[1]Tiffany K. Espinosa provided the data and scenario. The facility name has been changed.

TRFF(%)	The percentage of residents who were temporarily removed from the facility, out of the total number of residents
CI (%)	The percentage of critical incident reports written that month, out of the total number of residents

Analysis

We begin by analyzing **Employee Turnover**. Surprisingly, Exhibit 1 shows that employee turnover *increased* on average from 11.7% before the re-engineering to 18.7% afterwards. The typical fluctuation around the mean in turnover has also increased from 7% to 10.6%.

Exhibit 1 Summary Statistics for **Employee Turnover**

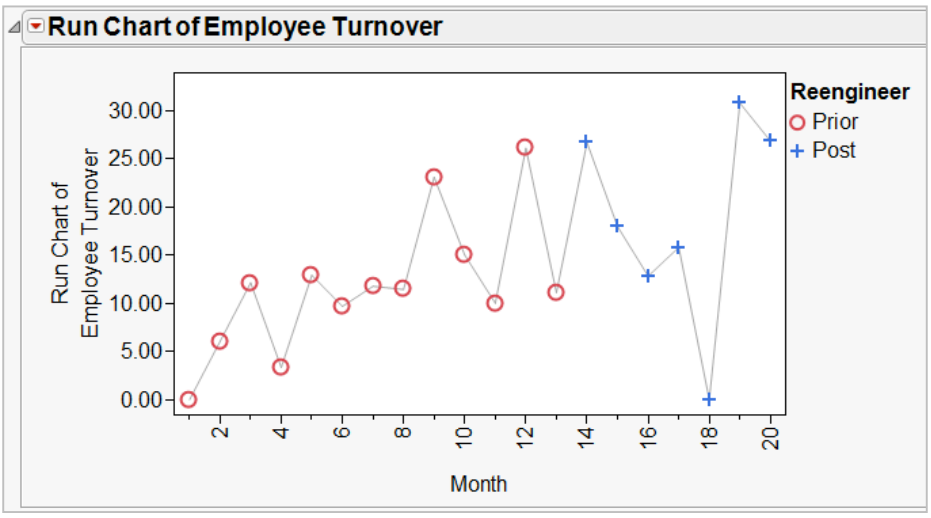
	Employee Turnover	
Reengineer	**Mean**	**Std Dev**
Prior	11.74	7.04
Post	18.69	10.56

*(Analyze > Tabulate; drag **Reengineer** into drop zone for rows, and **Employee Turnover** into the drop zone for columns as an analysis column. Then, drag Mean and Std Dev from the middle panel to the middle of the table. Note: To have Prior appear first, use Column Info >Column Properties, Value Ordering.*

Note that in JMP versions 10 and earlier Tabulate is under the Tables menu.)

The time series plot in Exhibit 2 shows the pre- and post-reengineering results. Turnover got steadily worse during the thirteen months prior to reengineering. After reengineering, the results became more erratic (i.e., the standard deviation increased).

Exhibit 2 Time Series of **Employee Turnover**

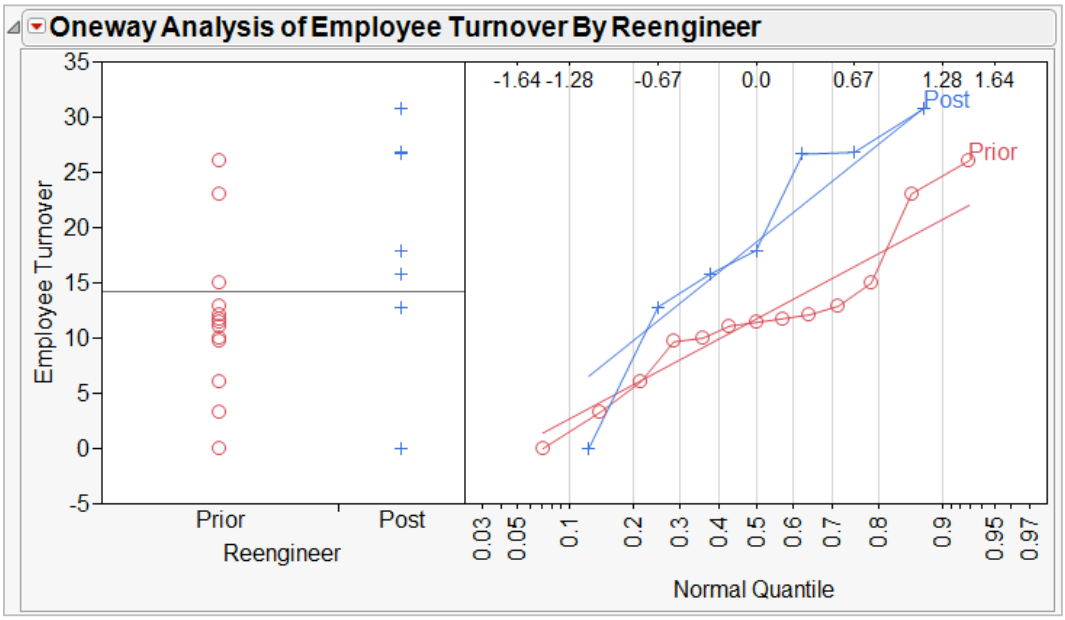


*(Analyze > Quality and Process > Control Chart > Run Chart; Use **Employee Turnover** as Process, and **Month** as Sample Label. To color points or change markers, right-mouse-click in the chart, select Row Legend, and select **Reengineer**. Under Markers, select an option.)*

Although the averages are *numerically* different (11.7 versus 18.7), a Two-Sample t-Test will be used to determine whether the averages are *statistically* different. First, we explore the normality assumption. Exhibit 3 shows the normal quantile plots for each group. Because the observations follow more-or-less a

straight line, with no obvious non-linear pattern, the normality assumption for the Two-Sample t-Test is met. (Note that the Two-Sample t-Test is fairly robust to departures from normality.)

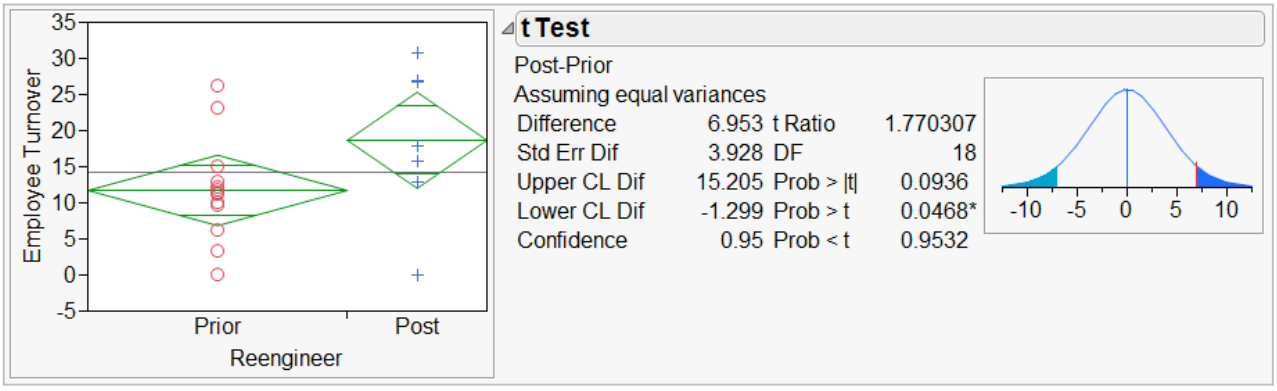
Exhibit 3 Normal Quantile Plots **of Employee Turnover** by **Reengineer**



*(Analyze > Fit Y by X; select **Employee Turnover** as Y, Response and **Reengineer** as X, Factor. Then, select the first Normal Quantile Plot option.)*

From Exhibit 4, we see that the difference in the average turnover rate before and after reengineering is 6.953%. However, the means diamonds (95% confidence intervals Prior and Post means) overlap, the p-value is above 0.05 (Prob > |t| = 0.0936), and the 95% confidence interval for the difference in the means (-1.299, 15.205) contains zero. Thus, we conclude that employee turnover rates before and after reengineering are not significantly different at a significance level of 0.05.

Exhibit 4 Two-Sample t-Test for **Employee Turnover** by **Reengineer**



(From the previous output, select Means/ANOVA/Pooled t from the red triangle.)

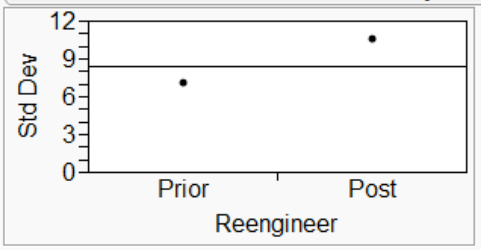
Note that in Exhibit 4 we are given the following reminder: “Assuming equal variances.” The Two-Sample t-Test presumes that the variances (or standard deviations) are equal. Before putting faith in the test

results, the equal variances assumption should be assessed. Exhibit 1 shows that the calculated standard deviations are *not* equal—they are 7% for the pre-reengineering data and 10.6% for post-reengineering. Even so, just because the standard deviations are *numerically* different doesn't mean that they are *statistically* different; we will use a statistical test to determine whether the variances (standard deviations) are equal. Exhibit 5 shows the equal variance test results for **Employee Turnover**.

While the graph of the standard deviations indicates that the standard deviations are different, the p-values for the five statistical tests indicate that the differences are not *statistically* significant (they are all above 0.05). So, we can trust the results provided by the Two-Sample t-Test in Exhibit 4.

Exhibit 5 Unequal Variances Test for **Employee Turnover** by **Reengineer**

Tests that the Variances are Equal



Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
Prior	13	7.04165	4.698284	4.699085
Post	7	10.55551	8.056518	8.259200

Test	F Ratio	DFNum	DFDen	p-Value
O'Brien[.5]	1.4249	1	18	0.2481
Brown-Forsythe	2.0402	1	18	0.1703
Levene	1.7687	1	18	0.2002
Bartlett	1.3141	1	.	0.2517
F Test 2-sided	2.2470	6	12	0.2193

Welch's Test

Welch Anova testing Means Equal, allowing Std Devs Not Equal

F Ratio	DFNum	DFDen	Prob > F
2.4504	1	8.9628	0.1521

t Test

1.5654

(From the previous output, select Unequal Variances under the top red triangle.)

Let's now turn our attentions to the variable **TRFF(%),** Temporarily Removed from Facility. In Exhibit 6 note that the average rate of removal after reengineering dropped from 20.54% to 9.23%. The standard deviation also dropped, from 10.5% to 2.6%.

Exhibit 6 Summary Statistics for **TRFF(%)**

	TRFF(%)	
Reengineer	**Mean**	**Std Dev**
Prior	20.54	10.45
Post	9.23	2.63

A time series plot makes these differences more evident.

Exhibit 7 Time Series of **TRFF(%)**

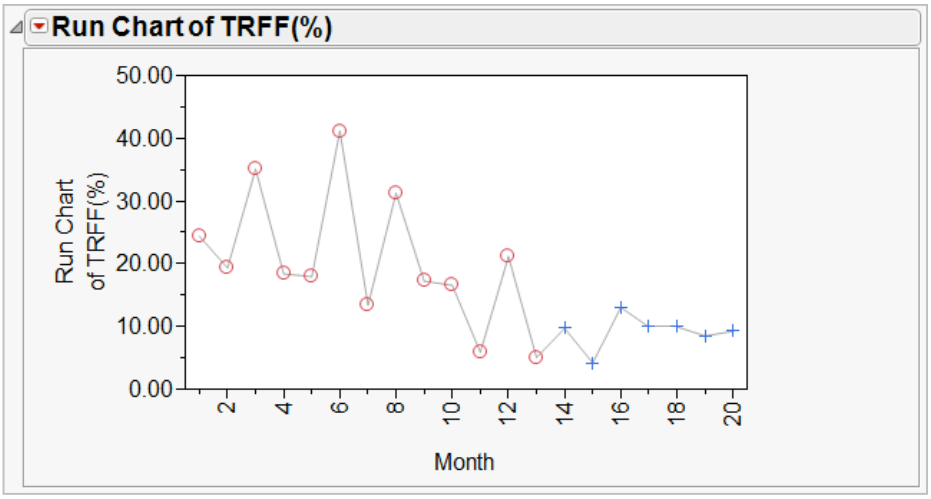
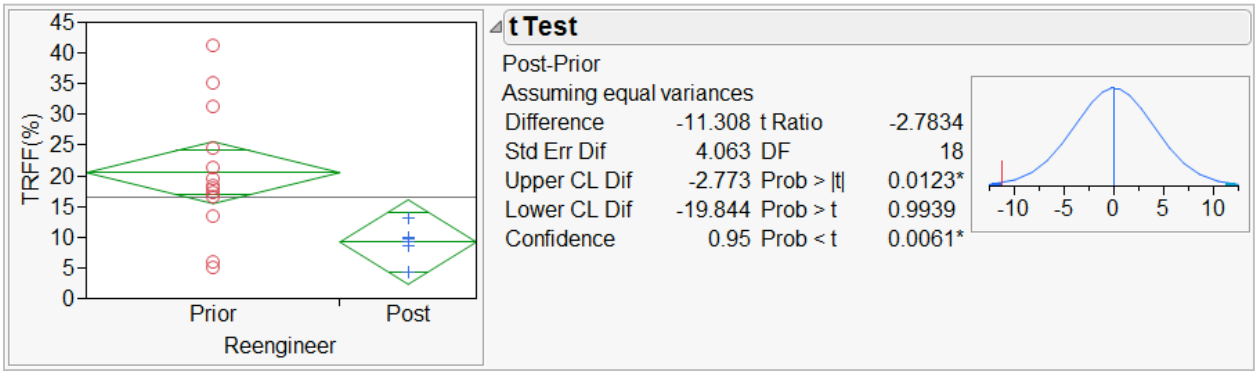


Exhibit 8 shows the result of the Two-Sample t-Test for **TRFF(%).** The results are unequivocal (Prob > |t| = 0.0123)—there has been a *statistically significant* drop in the rate of removal from the facility at the 0.05 significance level.

Exhibit 8 Two-Sample t-Test for **TRFF(%)** by **Reengineer**



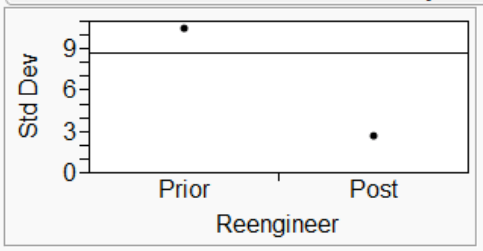
In Exhibit 6 we saw that the standard deviations appear to be different. This suggests examining the assumption of equal variances.

The tests for equal variances (Exhibit 9) are mixed. Two of the five tests (O'Brien and Brown-Forsythe) find no difference in the variances, since their p-values exceed 0.05. The others indicated that there *is* a statistically significant difference in the variances. This ambiguity suggests examination of the Welch Test, which is appropriate if the variances are not equal.

Because the p-value for the Welch Test is less than the 0.05 significance level (Prob > F = 0.0023), this confirms our conclusion that the means are statistically different from each other. Both the Two-Sample t-Test and the Welch Test indicate that there is a *statistically lower* rate of removal from the facility after implementation of the reengineering program.

Exhibit 9 Unequal Variances Test for **TRFF(%)** by **Reengineer**

Tests that the Variances are Equal



Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
Prior	13	10.44922	7.748735	7.435998
Post	7	2.63349	1.643361	1.601987

Test	F Ratio	DFNum	DFDen	p-Value
O'Brien[.5]	3.1173	1	18	0.0944
Brown-Forsythe	4.1224	1	18	0.0574
Levene	5.5267	1	18	0.0303*
Bartlett	9.2063	1	.	0.0024*
F Test 2-sided	15.7436	12	6	0.0029*

Welch's Test

Welch Anova testing Means Equal, allowing Std Devs Not Equal

F Ratio	DFNum	DFDen	Prob > F
13.6189	1	14.592	0.0023*

t Test

3.6904

Summary

Statistical Insights

This case uses the Two-Sample t-Test to examine differences between means and shows how to assess the normality and equal-variances assumptions. To test for normality, examine the normal quantile plot of each group. To test the equal variances assumption, examine the five equal-variance tests displayed by JMP. The Welch Test should be used instead of the Two-Sample t-Test if we conclude that the variances for the groups are not equal.

The number of observations in both the Prior and Post periods is small. Keep in mind that the power of statistical tests is based largely on the sample size. Larger data sets, in general, have more power; that is, a better ability to detect differences.

For non-normal data, or with small data sets, nonparametric tests can also be used. Nonparametric tests tend to be more powerful when data are highly skewed.

The time series plot is an effective tool for plotting time-ordered data. Critical metrics, such as employee turnover, removals and critical incidents should be tracked over time and monitored using these plots. This would allow management to explore and anticipate seasonal or cyclical changes.

Managerial Implications

The reengineering effort had a statistically significant influence on the temporary removal rate from the facility, which dropped by an average of 11%. Employee turnover, on the other hand, did not change (statistically) as a result of reengineering. This was a surprising result, which should be investigated further. We'll see whether the rate of critical incidents improved in an exercise.

JMP Features and Hints

We used Tabulate to produce summary statistics, and Run Chart to create time series plots. In Fit Y by X we generated normal quantile plots to assess the normality of the groups for the variables of interest. Note that normal quantile plots can also be generated from Distribution. For this data, we would put **Reengineer** in the "By" field to produce separate normal quantile plots for each group.

In Fit Y by X, we ran a Two-Sample t-Test to test for differences between two means (Means/ANOVA/Pooled t), and used UnEqual Variances to test the assumption of equal variances.

Exercises

Use the tools illustrated in this case to determine if the reengineering effort changed the critical incidence rate.

1. Is there evidence that the critical incidence rate improved?
2. Are the results statistically significant? What is the confidence interval for the difference? What is the p-value for the Two-Sample t-Test?

SAS Institute Inc. World Headquarters +1 919 677 8000

JMP is a software solution from SAS. To learn more about SAS, visit **www.sas.com**
For JMP sales in the US and Canada, call 877 594 6567 or go to **www.jmp.com**