

BZAN 535 - Homework 5

Kevin Gardner

9/30/2016

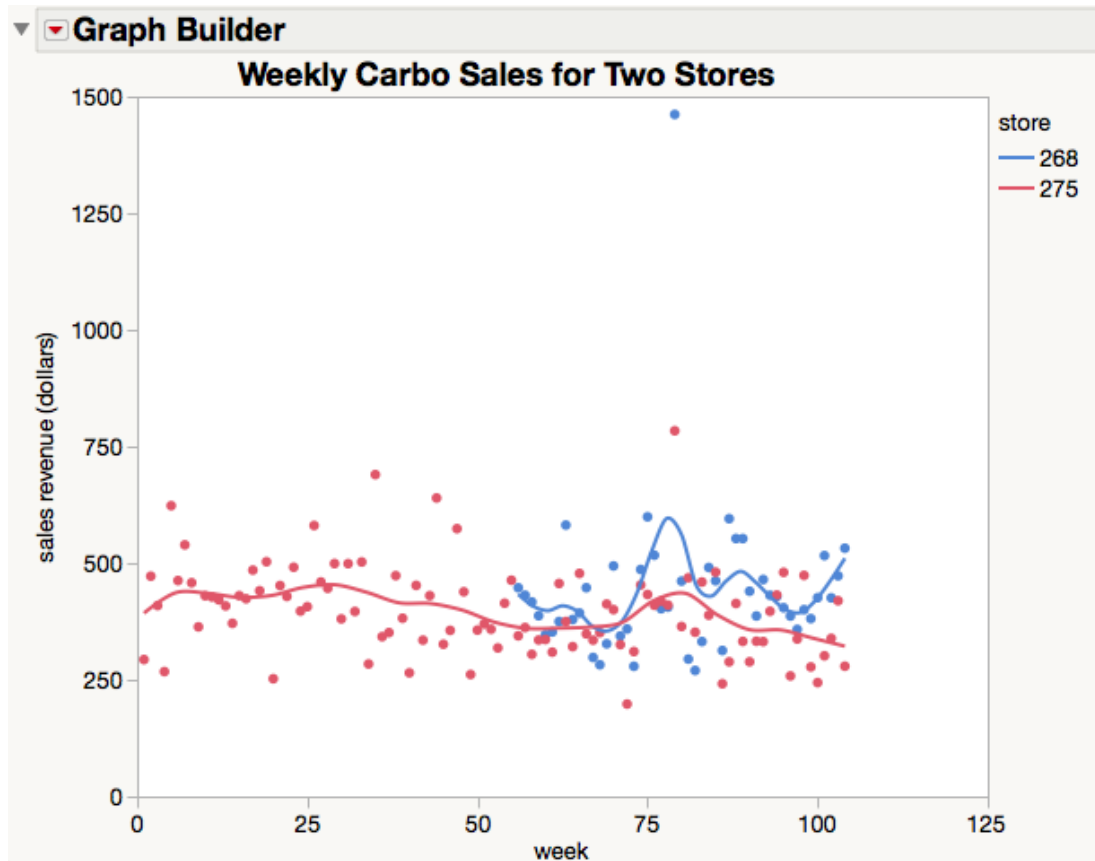
Question 1

The carbo data lists two Kroger stores in zipcode 37075. This is Hendersonville, TN, just north of Nashville. The stores are #268 and #275.

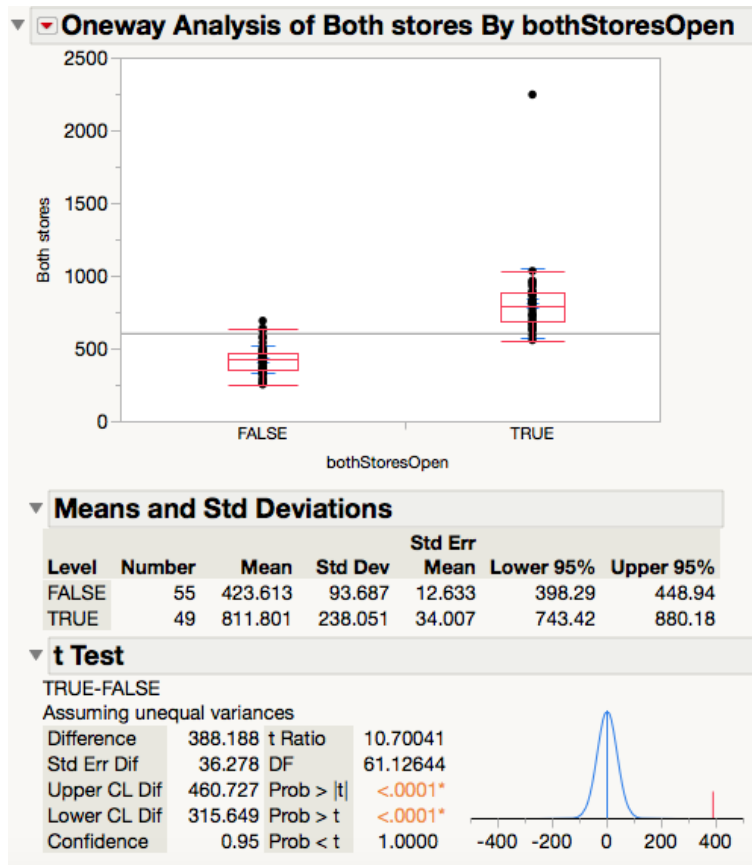
- a) Write a query to capture the weekly carbo sales revenue for each of these stores. Show your query but not the resulting data. Export the data to JMP.

```
select store
, week
, sum(dollar_sales*units) as sales_revenue
from carbo_transactions
where store in (268,275)
group by store, week;
```

- b) Use Graph Builder to plot the weekly carbo sales revenue for the stores. The plot should show the data and a smoother for each store.



- c) What is the % increase in sales volume for Kroger (in Hendersonville) due to opening the second store? Give a single answer (a “point estimate”) and explain how you computed this number. Also explain any assumptions upon which your answer depends.



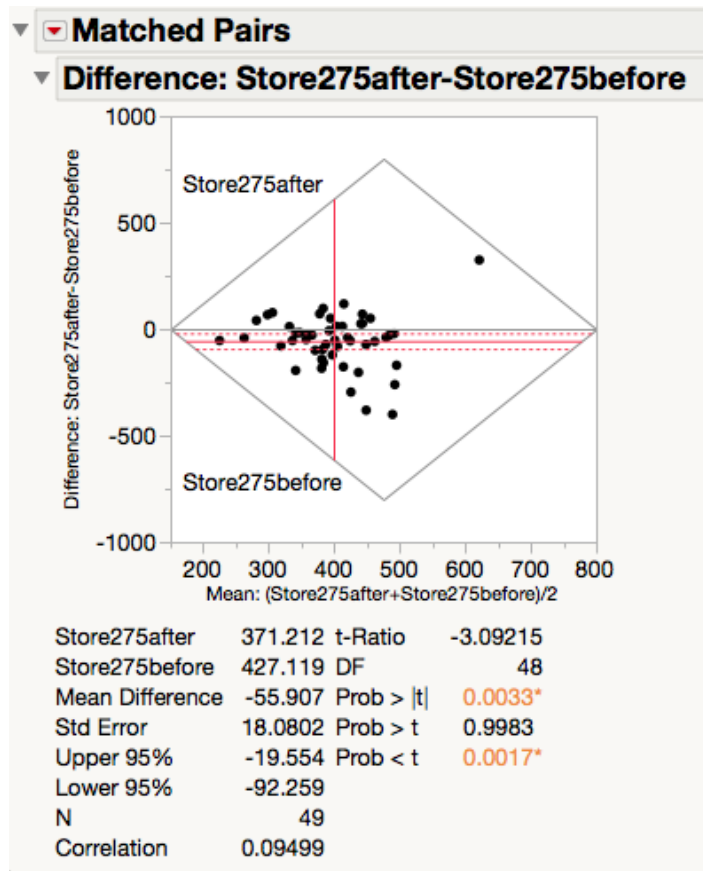
Here we will treat the average (combined) sales before and after opening the second store as independent samples and perform a t-test. Average weekly sales before and after opening the second store were \$423.61 and \$811.80 respectively. The two means represent average weekly sales for weeks 1–56 and average (combined) weekly sales for weeks 57–104. For this comparison we assume that the distribution is normal.

$$(811.80 - 423.61) / 423.61$$

```
## [1] 0.9163854
```

- d) Did store #275 appear to suffer any loss in sales that coincided with the presence of this new store? (This might occur due to customers switching from #275 to #268.) Give a 95% confidence interval for the change in mean weekly sales. Explain in detail how you approached this problem. Finally, interpret the interval in the language of the problem.

Here we use a matched pairs t-test to determine a 95% confidence interval for the difference in means. Observations are matched based on week of the year. For example week 1 is paired with 53, 2 with 54, 3 with 55, etc. Store #268 opened the fourth week of the year (week 56), so weeks 1–3 are excluded and the first pair consists of week 4 and 56. From the results we conclude with 95% confidence that the mean difference in sales is between -92.259 and -19.554. The interval does not contain 0, so our analysis suggests that store #275 did suffer a loss in sales.



- e) Was it necessary to adjust the standard error(s) in d) for autocorrelation? Justify your answer here.

Here we did not adjust for autocorrelation.

Question 2

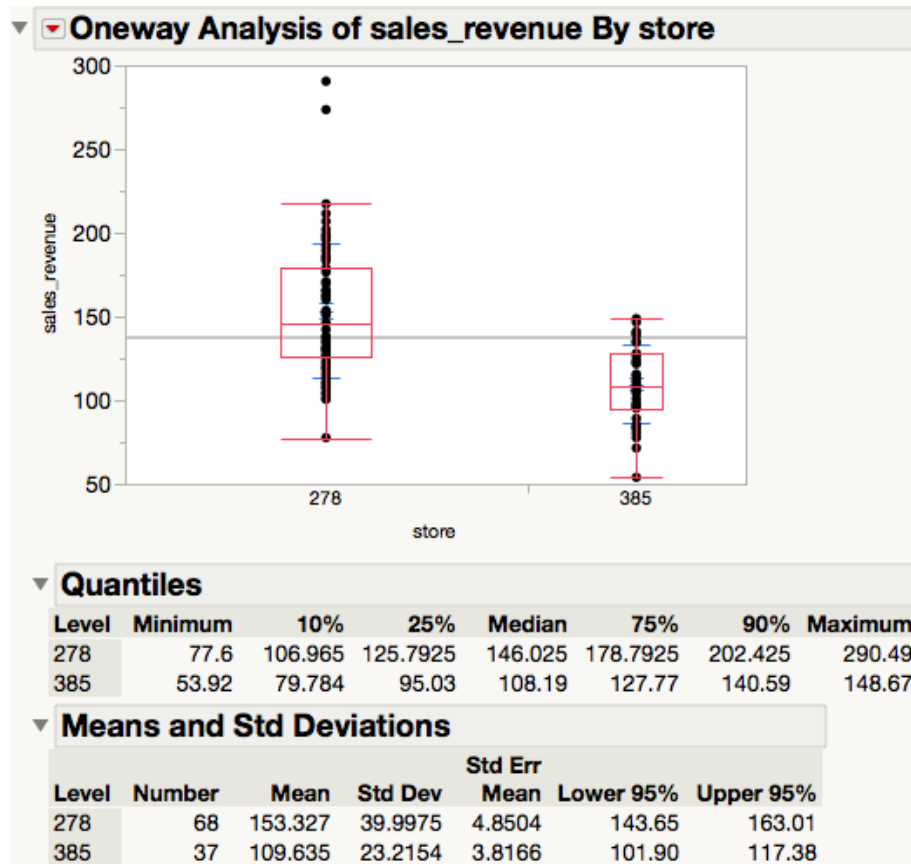
Compare stores 229 and 321 in terms of carbo sales revenue. In particular, can we say from the available data that one store has higher sales on average than the other on an ongoing basis?

- Answer this question either with a 95% confidence interval for the difference in mean weekly sales or with a two-sided hypothesis test. (Was it necessary to adjust the standard error for autocorrelation? Justify your answer.)
- Also, comment about what assumptions you are making regarding this historical data when you make your assertion about any difference “on an ongoing basis.”

Question 3

Store #385 closed and a new store (#278) opened during the period of our data. Both stores are in Shelbyville, TN 37160.

- a) Plot weekly sales for these stores. Discuss what you see and how this makes answering 3b) a challenge. Also, what should we do with week 37 data? Why?
- b) Estimate the \$ increase and the % increase in weekly sales based on this store closure/opening. Explain what data you used to answer this question and justify this choice.



- c) Construct a 95% confidence interval for the mean \$ increase in weekly sales. Show how you calculated the interval. Be sure to take into account any autocorrelation if it is statistically significant.

▼ **Response sales_revenue store=278**

▼ **Whole Model**

▶ **Actual by Predicted Plot**

▶ **Summary of Fit**

▶ **Analysis of Variance**

▶ **Parameter Estimates**

▶ **Residual by Predicted Plot**

▼ **Durbin-Watson**

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
1.4299542	68	0.2648	0.0080*

▼ **Response sales_revenue store=385**

▼ **Whole Model**

▶ **Actual by Predicted Plot**

▶ **Summary of Fit**

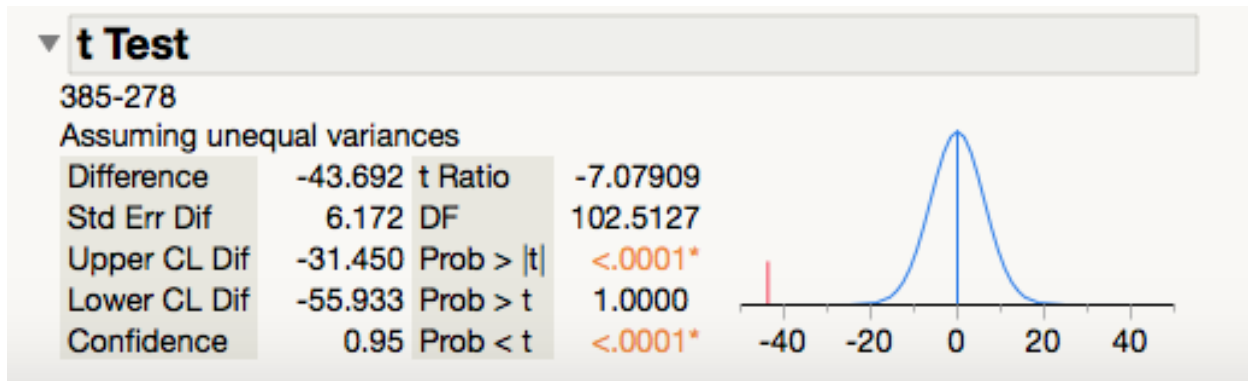
▶ **Analysis of Variance**

▶ **Parameter Estimates**

▶ **Residual by Predicted Plot**

▼ **Durbin-Watson**

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
1.7728058	37	0.0302	0.2424



d) Comment about what might make the actual average weekly gain not coincide with our interval estimate.

For the adjusted standard error we get $SE = 6.99$. So the upper and lower endpoints for the adjusted CI are given by

```
# Autocorrelation = mean + 1.96*SE or mean - 1.96*SE
c(4.8504 - 1.96*6.99, 4.8504 + 1.96*6.99)
```

```
## [1] -8.8500 18.5508
```

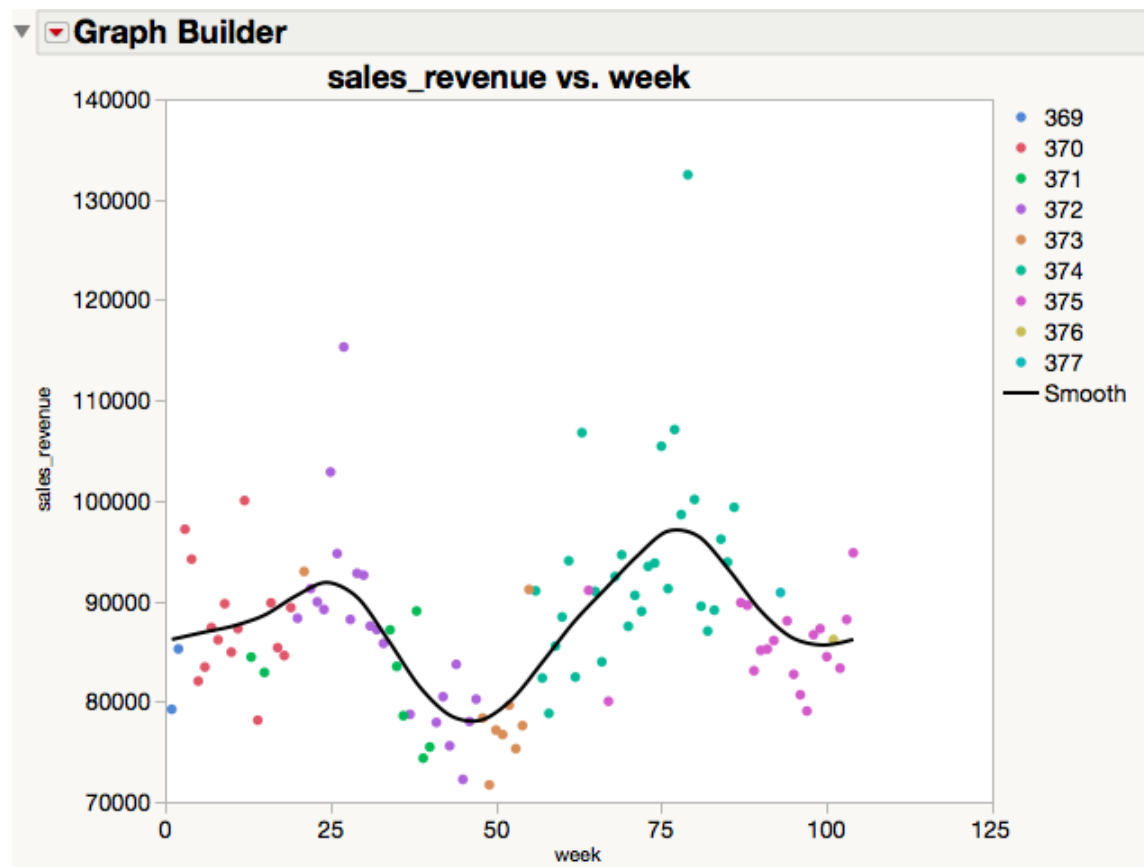
Question 4

This question involves computing an estimate for the expected weekly sales for Kroger in these four categories (in the available Kroger data).

- a) Show a query that computes the weekly total carbo sales and the number of Kroger stores with sales each week.

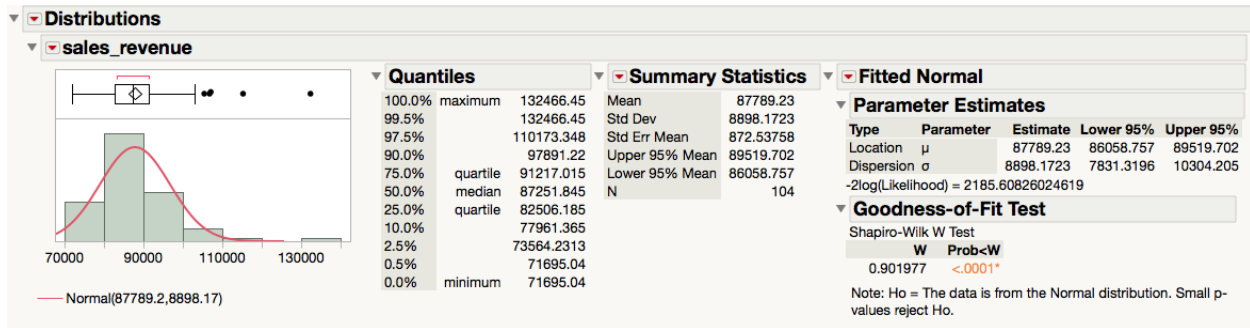
```
SELECT count(distinct store) as num_stores
, week
, sum(dollar_sales) as sales_revenue
FROM carbo_transactions
group by week;
```

- b) Plot the data in an informative way and discuss what this reveals?

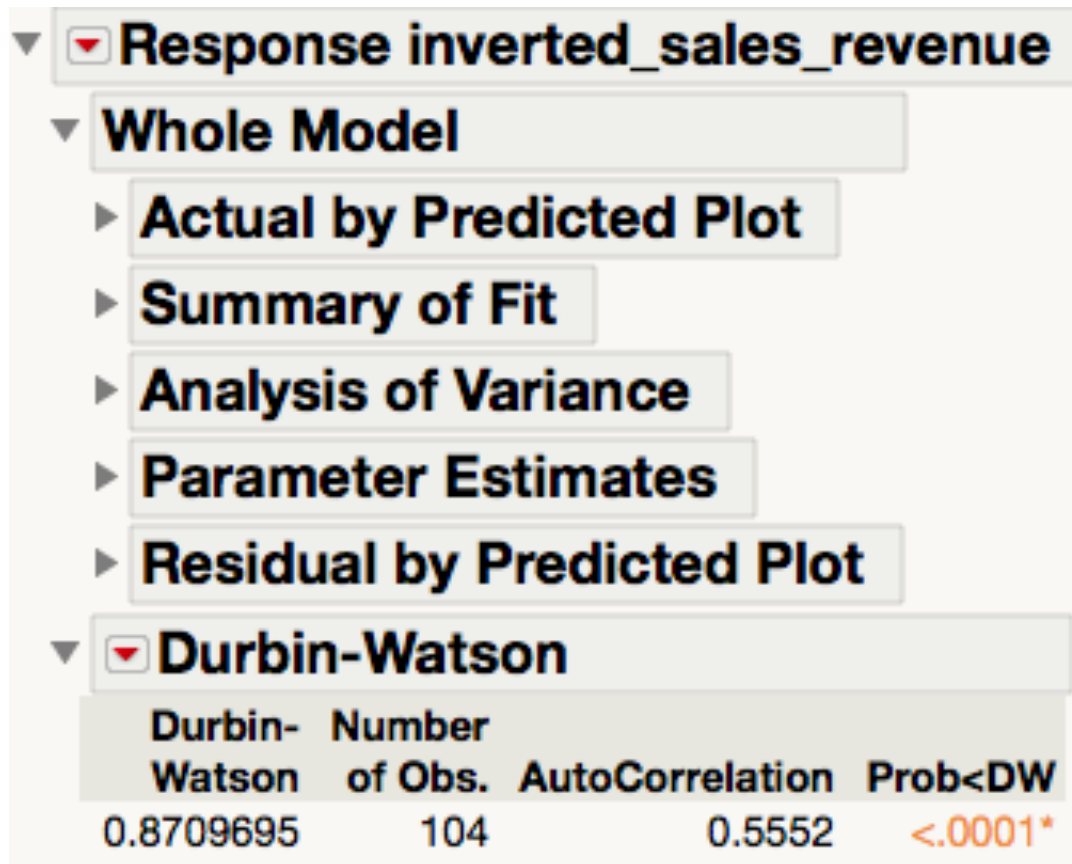
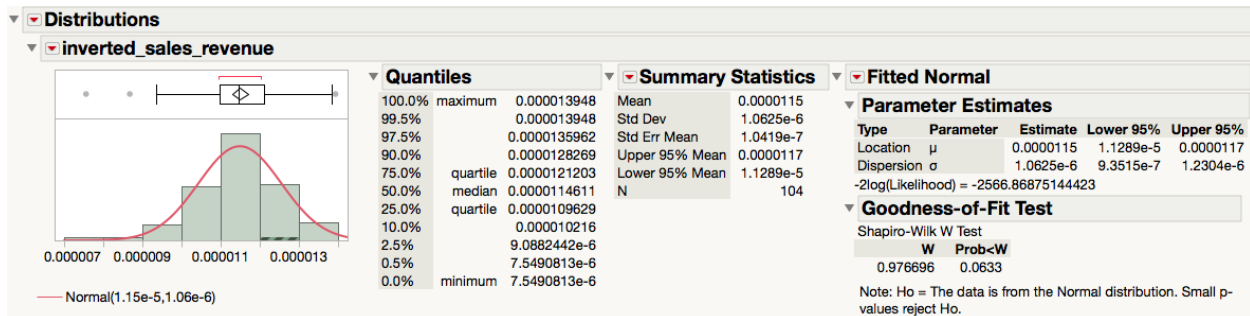


- c) JMP provides a 95% confidence interval of [\$86,059, \$89,520] for the expected Kroger weekly sales in these categories. Why is this interval not valid? That is, what assumption(s) is (are) violated?

From the Goodness-of-Fit Test we reject the null hypothesis that the distribution is normal, thus the CI is not valid.



d) Provide an adjusted confidence interval. No need to interpret the interval, just show the computations.



Calculations: Variance=1.129e-12 N=104 Autocorrelation=0.5552 Using the formula we obtain

Formula for the adjusted standard error

$$SE = \sqrt{\frac{1+r}{1-r} * s_1^2}$$

Lower and upper bounds for the adjusted CI

$$[mean - 1.96 * SE, mean + 1.96 * SE]$$

For the adjusted standard error we have

$$SE = \sqrt{\frac{1+r}{1-r} * s_1^2} = 3.8\text{e-}7.$$

For the 95% confidence interval we have

```
# [mean + 1.96*SE or mean - 1.96*SE]  
c(0.0000115 - 1.96*3.8e-7, 0.0000115 + 1.96*3.8e-7)
```

```
## [1] 1.07552e-05 1.22448e-05
```