**DSO 522: HW 2 template**

**Due 11:59 PM September 17, 2021**

**Please list your team members (Official Last name, Official First name):**

**1. Kate Donenfeld**

**2. Alex Rocouso**

**3.**

**Instructions:** Provide answers to HW questions. Do not include code in this file. Below in red **Answer:** helps you understand which questions you need to answer in this file. You will submit R code file separately. Use the link on blackboard **HW #2 word file submission link** to submit this file with your answers.

**Case 1**

**Questions and your answers:**

1. – See code.
2. – See code.
3. Create a graph of your KPI defined in the previous question and describe the pattern(s) you see in data.

**Answer:**

The points on the graph vary up and down (higher and lower KPI), which seems like noise. There is a slight downward trend with average KPI decrease and the year increases.

A picture containing chart

Description automatically generated

1. Meteorologist who study weather patterns, hired your consulting firm to help them quantitatively assess the hypotheses of global warming. It had been known since 1960 (judgmental call) that humans were increasing the amount of heat-trapping greenhouse gases in the atmosphere that led to the questions: Does this human activity warm the climate noticeably? Judgmentally1960 has been selected to be a year of a long term temperature change(You may ask your parents or grandparents why 1960). Does the data provide statistically significant evidence of global warming? Answer this question by assessing the significance in the abrupt change in KPI.

**Answer:**

Having only one trend line fits the observations, and the residuals all look like noise. All observations are correlated with the previous observations so there doesn’t seem to be abrupt change. When I graph auto-covariance of the data, only one point has a correlation coefficient that exceeds the blue dashed line which indicates that it may be significant. However, that points exceeds the blue line by such a small amount that I would argue that there is no significant abrupt change in KPI.

Chart, box and whisker chart

Description automatically generated

Looking at Pacf, we see the same singular point has significance.

Chart, box and whisker chart

Description automatically generated

Finally, looking at a graph of residuals, they all look like white noise, which would indicate there is no significance in ACF and that there is no significance in any abrupt change.

A picture containing text, antenna

Description automatically generated

1. Answer the previous question by assessing the significance of the gradual change in KPI.

**Answer:**

Running a linear regression gives a slope of -0.06682. This is a small number in magnitude so the human activity probably does not warm the climate noticeably. However, this warming (decrease in time before the ice breaks) is there because the slope is negative and statistically significant. It has a p-vale of .00825, which means we can conclude that the slope is less than 0 with 99.17% certainty. Therefore, the data provides statistically significant evidence of global warning.

Text, letter

Description automatically generated

Additionally, in the graph, you can see that the red trendline is sloped downward, indicating a reduction in the time before the ice breaks, which indicates significant gradual global warming.

Chart

Description automatically generated

1. Report and creatively interpret the estimates of both abrupt and gradual change in climate. Which one do you think it is more appropriate and meaningful?

**Answer:**

The estimate of climate change that we get from our data is -0.06682 days decrease in time before the ice breaks per year. This shows gradual ice melting. I think the estimate of gradual change is more appropriate and meaningful because climate does not change all at once, it happens over time. The ice melts faster and faster each year, but the ice isn’t abrupted all melted.

1. You recall from intro to business statistics course that statistical significance ≠ practical importance. Suggest and implement a way(s) to quantify the evidence of global warming from a practical perspective. If you use terms such as “result is statistically significant or in/nonsignificant”, many people won’t have a clue what they mean and will ask you to put it Layman terms.

**Answer:**

We can conclude with 99.17% certainty, that the ice is melting faster each year. In particular, we estimate that time to melt decreases by .067 days each year.

1. Compare your approach with the approach we used to assess the effectiveness of the batmobile program. Have you observed any differences? Briefly comment.

**Answer:**

In Batmobile program, we used a test with a null and alternative hypothesis.

We used:

t.test(x= data$KPI[1:29],y=data$KPI[30:52],alternative="greater")

test\_statistic=(mean(data$KPI[1:29]) - mean(data$KPI[30:52]))/

sqrt(var(data$KPI[1:29])/29 + var(data$KPI[30:52])/23)

test\_statistic

The null hypothesis was the BM program was ineffective and the alternative hypothesis was the BM program is effective. So we are testing to see if the KPI called mu is greater than is was before: mu\_before - mu\_after >0.

This is different than what we did for nenana1. For nenana1, we created a linear model and observed the coefficient, which was the slope. The null hypothesis is inherently that the slope is 0. We then looked at our P-value and test statistic and concluded that the p-value was so small that the slope must be significant.

These two different tests have the same results in indicating the statistical significant of a measurement.

1. Read the original article published by two Stanford professors:   
   [**https://science.sciencemag.org/content/sci/294/5543/811.full.pdf**](https://science.sciencemag.org/content/sci/294/5543/811.full.pdf)

And also a Critique of the article and the comments below the critique:   
[**https://www.john-daly.com/nenana.htm**](https://www.john-daly.com/nenana.htm)

**Write a commentary on the above articles. (One paragraph is enough. But don’t limit yourself!)**

**Answer:**

**Commentary on Article 1: “Climate Change in Nontraditional Data Sets” by Raphael Sagarin and Florenza Micheli**

I found this article to be a really good supplement to the data and understanding I got from working with the Nenana1 data in the questions above. I found it interesting that at the start of this competition, I am sure the residents did not realize how significant the data would be and the impact it could make on research. In playing this game, they created a consistent definition for ice break for measurement purposes and gathered the data that we are now using an analyzing. The article even said, “Records of ice melt are more accurate long-term indices of air temperature than air temperature records themselves.” I thought it was cool how this captured both thermal and dynamic effects and both were created by climate change. Thermal effects are direct melting which comes from warming temperatures. Dynamic effects are the mechanical forces from upstream drift, which comes from runoff from snow melting in the mountains. This article also noted the impact of ice melt, which not only includes changes for the ecosystem like availability of resources and habitats, but also the start of water commerce for the season. I thought it was interesting that they found periods of advancement and delay highlighted by polynomial regression with the data. I mostly saw noise in the fluctuations; however, after reading this, I see in the graph how the data moves up and down for cooling and warming. This effect is very slight because only one point was significant in my ACF graph, but I am sure more sophisticated data visualization would reveal the pattern better. Lastly, I was surprised that the article said trends for precipitation and snowfall were insignificant because I would expect those to be a function of climate change as well.

**Commentary on “The Nenana Ice Classic: Betting on Warming” by John L. Daly**

This article critique research on Nenana similar to what we worked on. The data indeed found a slight trend toward earlier breakup dates, just as we found above, but the change is very slight. The authors attribute recent earlier breakups to climatic warming, but data from Alaska Climate Research Center show little warming significant increases in snowfall. I found this interesting because the authors of the research and the other article argue that there is change in temperature and no change in rain or snowfall. The article notes that correlations between river breakup dates and climatic warming only use part of the available data and exclude the peak warm year in 1940. It also excludes the coldest year in 2001. This must have affected the data and analysis. Daly makes a harsh point that the authors might not have gotten published had they not altered the data to get the results they wanted. We think this shows some politics in research and the problem of the recognition that goes along with publishing. It should be more about advances in science and human knowledge rather than fame and recognition.

1. This summer I received an email from NASA National Snow and Ice Data Center, that the data set was update by including more data. The data set that contains the most recent information is in **Nenana2.txt**. Does the data set that includes more recent info provide statistically significant evidence of global warming? Answer this question by assessing the significance of both abrupt and gradual change in KPI. Does more recent that change your conclusions compared to the previous conclusions that you obtained using data in **Nenana1.txt**?

**Answer:**

Initially looking at the plot of KPI vs. Year, we can see a stronger downward trend

Chart

Description automatically generated

When getting summary of the lags, neither the first nor second lag is statistically significant, so there is no abrupt change.

Table

Description automatically generated

Running the autocorrelation function, there is only one point that goes beyond the blue dashed line at a 95% confidence interval, so abrupt change is not statistically significant.

Chart, box and whisker chart

Description automatically generated

Partial autocorrelation has 2 points going beyond the blue dashed line.

Chart, box and whisker chart

Description automatically generated

Finally, a plot of the residuals looks like noise so there is not abrupt change.

A picture containing text, antenna

Description automatically generated

When running a linear model for nenana2, we get a slope of -0.0865. This is really close to the slope we got for nenana1. The p value for this value is very small and t=-4.451, which is large in magnitude so the small negative slope of -0.086 is statistically significant, shown as the red line on the plot. This means gradual change is statistically significant.

Text

Description automatically generated

Chart

Description automatically generated

**Case 2**

**Questions and your answers:**

1. -
2. Graph the data and describe the patterns you see in the data set?

**Answer:**

In the first graph we can appreciate the 12 data points we have for each month of each year. We can clearly see that after 1960, when the first intervention took place, no measure is again above level 6 of ozone (unlike before 1960, which was quite common). We can also see how the lowest measures get even lower as time goes by.

In the second graph, we can get a better sense of the medians, quartiles, and outliers by using boxplots for each year. We can clearly see how after 1960 the boxplots are placed on lower ozone levels and keep decreasing especially after 1966.

1. Does the data provide statistically significant evidence that the opening of the Golden State Freeway and the implementation of Rule 63 in 1960 reduced the pollution statistically significantly? Evaluate both the abrupt and the gradual changes? Report and interpret the estimates of both the abrupt and the gradual changes in ozone. Which one do you think it is more appropriate and meaningful?

**Answer:**

*Ha: Intervention was effective and reduced pollution*

*Ho: Intervention was not effective and did not reduce pollution*

*Tolerance level alpha = 0.05 = 5%*

Carried out 2 sample t-test to assess whether the intervention was effective, found that:

**p-value = 0.00000002391 < alpha = 0.05**

so the data provides statistically significant evidence against Ho, hence the intervention of 1960 was effective.

After implementing the abrupt change using regression model, we can say the highway intervention of 1960 is statically significant since in R function linear model output p-values are two sided, so:

**p-value = 2.4e-8/2 = 0.0000000024**

Now let’s see if gradual effect is significant:

*Ho: Intervention was NOT effective (slope of Ramp is NOT negative)*

*Ha: Intervention is effective (slope of Ramp is negative)*

After implementing the model, we can see the slope of the ramp is negative (-0.019370) with statically significant confident since:

**p-value 0.000156/2 = 0.000078**

However, despite the gradual change has mathematical significance, we should be careful and note that the abrupt change (caused probably by the highway intervention) may be causing results to yield a gradual significancy without having it (meaning the regulation intervention was not effective).

To prove this, I implemented a model explaining both abrupt and gradual changes (M3). When we visualize this new model, we can see a much better explanation of what is going on: a big abrupt change caused by the highway (because it was effective), followed by a slightly positive slope (because the regulation was probably not effective).

Hence, I think this linear model combining both abrupt and gradual change is more appropriate and meaningful to explain the situation.

1. Does the data provide statistically significant evidence that the regulations implemented in 1966 requiring engine changes in new cars reduced the pollution statistically significantly? Evaluate both the abrupt and the gradual changes? Report and interpret the estimates of both the abrupt and the gradual changes in ozone. Which one do you think it is more appropriate and meaningful?

**Answer:**

*Ha: Intervention was effective and reduced pollution*

*Ho: Intervention was not effective and did not reduce pollution*

*Tolerance level alpha = 0.05 = 5%*

Carried out 2 sample t-test to assess whether the intervention was effective, found that:

**p-value = 3.695e-08 = 0.00000003695 < alpha = 0.05**

so the data provides statistically significant evidence against Ho, hence the intervention is effective.

After implementing the abrupt change using regression model, we can say it is statically significant since in R function linear model output p-values are two sided, so

**p-value 2.38e-07/2 = 0.000000119**

However, after implementing an abrupt change comparing periods 1960-1966 versus after 1966, we see the statical significance is much smaller. This may suggest the abrupt change is not as relevant.

Now let’s see if gradual effect is significant:

*Ho: Intervention was NOT effective (slope of Ramp is NOT negative)*

*Ha: Intervention is effective (slope of Ramp is negative)*

After implementing the model, we can see the slope of the ramp is negative (**-0.020524**) with statically significant confident since **p-value 1.6e-08/2 = 0.000000008**

In this case, I think again a model that records both abrupt and gradual change is more appropriate and insightful, as implemented in M3. Despite implementation of a new regulation is mostly gradual in nature (as more people become aware of it, or in this case as new cars are manufactured under the new regulation), it might also incorporate a slight component of abrupt change (as we can argue it has some mathematical significance), although we could argue it is irrelevant from a qualitative perspective.

1. Can you capture both interventions, the opening of the Golden State Freeway and the implementation of Rule 63 in 1960, and the regulations for new car engines implemented in 1966 in one model? Can you model both interventions and both abrupt and gradual changes? What insights do you gain from such model? Summarize the results.

**Answer:**

First, we implemented a model to model for both abrupt changes. As explained in the previous exercises, the abrupt change in 1960 is highly relevant, while the abrupt change in 1966 is very light.

Then, when we implemented a model for both gradual changes, we incur in the same flaw that we incurred when making a gradual-only model for the first intervention (the strong abrupt change is messing up the gradual model and making it inaccurate to reality).

Finally, when we modeled both abrupt and gradual changes for both interventions, we see a model that is quite accurate to reality: an abrupt change in 1960 with a light upward-sloped gradual trend, then a gradual change since 1966.