Time Series HW 2 - Due Friday 9/9 at 6 pm

Andrea Mack

## HW 2

This will not be written in report style - just provide numbered answers to each question. Include your edited R code either in-line (using something like R-markdown - the newer option to compile directly to a Word document is quite nice for controlling formatting after compiling) or as an appendix to your homework.

**You can not work with anyone that you worked with on the first homework.** You can discuss it with old partners but should try work as much as possible with new collaborators.

You will get a 5% bonus on this homework if you do this homework in a group of 2 or 3 but not if it is any larger. If you discuss the assignment with others but turn in separate assignments, you need to document any discussions you had and how it impacted your answers - treat this like citing your sources.

We'll continue working with temperatures in degrees F in the column labeled MMXT (<https://dl.dropboxusercontent.com/u/77307195/rawbozemandata.csv>) for Bozeman's MSU weather station.

Either use my code or your code to generate a Year variable that includes the fractional information on the month. For example, January would be 1900+0/12 and December would be 1900+11/12 in the Year variable.

1. Re-estimate and then use the output to write out the estimated model that incorporates the linear time trend and the month component from HW 1. Use indicator/dummy variable notation and define all model aspects.

rawd <- NULL  
rawd <- data.frame(rawbozemandata)  
rawd$year <- as.numeric(substring(rawd$DATE, 1,4))  
rawd$month <- as.numeric(substring(rawd$DATE, 5,6))  
rawd$month.dec <- rawd$month/12  
rawd$date.new <- rawd$year + rawd$month.dec   
rawd$month <- month.abb[((rawd$month))]  
rawd$temp <- rawd$MMXT  
  
  
mod1 <- lm(temp ~ month + year, data = rawd)  
summary(mod1)

1. What is the model for an observation in a January? July? Simplify the full model for each specific month to just provide a function of Year.
2. Has the mean temperature changed differently in different months? Load the car package and make a plot using something like scatterplot(MMXT~Year|Month,data=rawd1,legend.plot=T,smoother=F). Discuss the results in the plot.

It appears that December and November average temperatures have not changed much from 1900 to 2015. Average temperature appears to have increased at similar rates in other months over the last 115 years. It appears that average monthly temperature increased about 5 $\expression{degrees F}$ over the 115 years, which is quite similar to the average rate of increase we discussed in the Oregon State temperature data in class.

require(car)  
  
scatterplot(temp~date.new|month,data=rawd,legend.plot=T,smoother=F)

1. Explore the same research question as in #3 by fitting a model with a Year by Month interaction (include main effects too). Generate and report an F-test for the interaction in the model using either anova or Anova from the car package. Write out a one-sentence conclusion that summarizes the results of the test including information on the distribution of the test statistic under the null hypothesis in that sentence. : for all i, = 0 : at least one 0

compared to the F distribution with 11 and 1350 df lead to a pvalue of 0.016.

There is some evidence that the effect of year on true average temeperatures changes by month.

mod2 <- lm(temp~year\*month, data = rawd)  
Anova(mod2, type = "II")

1. Return to the additive model (Year+Month) and let's consider the potential evidence for specific differences in the pairs of months. We can test the 12 choose 2 comparisons, maintaining overall or "family-wise" error rates of, say, 5% by using Tukey's Honest Significant Difference. This is a little more complicated to use than in a simpler One-Way ANOVA model because we need to adjust for the linear trend. But the method is easily extended to handle more complicated models and those with multiple predictors using the same code you can use in the One-Way ANOVA case (See Greenwood and Banner (2016) Section 2.5 and 2.6 (<https://scholarworks.montana.edu/xmlui/handle/1/2999>) for more on this method in the simpler scenario). Your code will be something like the following, after you install the multcomp package. Note that you may need to modify the model name (model1 below) and the variable to perform Tukey's HSD on (Monthf below).

require(multcomp)  
  
Tukey\_results<-glht(mod1,linfct=mcp(month="Tukey"))  
  
summary(Tukey\_results)  
cld(Tukey\_results)  
  
confint(Tukey\_results)  
  
require(stats)  
TukeyHSD(aov(temp~year+month, data=rawd), which = "month")

You can also plot the results but with so many comparisons, the plot is not very useful. Based on these results, what can you say about evidence for differences in the months after controlling for the linear year trend? A fairly simple pattern should arise in the results. I find it useful to switch to saying differences were or were not "detected" when dealing with a large suite of Tukey's results.

%This is looking for evidence in differences in average month temperatures in a given year, which %does not seem to be the most interesting thing.

Overall, there is strong evidence that August and July were the warmest months on average and after accounting for yearly temperature changes. With no evidence of a difference in temperatures between the two months on average and after accounting for yearly temperature changes. There was strong evidence that June was the next warmest, followed by September, then May, then October, then April, then March, and then November on average and after accounting for yearly temperature changes.

There is strong evidence that December and Feburary were the coolest months on average and after accounting for yearly temperature changes, with no evidence of a difference in temperatures between the two months on average and after accounting for yearly temperature changes. There was strong evidence that January was cooler than all months aside from December and Feburary, which there was strong evidence that January was warmer than these two months on average and after accounting for yearly changes in temperature.

1. One other issue that wasn't obvious in either the initial plots or in the model diagnostics is whether the linear trend really is a good description of the changes over time. There are a couple of ways to add polynomial terms to linear models. The simplest is just to create a squared version of the variable and include it as an additional variable. We'll see that this isn't an optimal choice in the next homework, but let's start this way for now. Once you are considering incorporating polynomials, we need to consider some sort of model refinement to decide on the polynomial order (linear, quadratic, cubic, quartic, etc.) to use. One technique is to sequentially add higher order versions of the variable to the lower order (linear, quadratic, etc.) versions of it and stop when the highest order term has a "large p-value" and drop back to the next lower order model that has the highest order term having a small p-value (checking diagnostics to make sure nothing really high order was missed). Employ this approach, starting with adding a quadratic Year variable to the Year+Month model. Report the test for the quadratic component (with distribution, test statistic, and p-value) and interpret this result, remembering that t-tests are conditional on other stuff in the model.

mod3 <- lm(temp~month + poly(year,2), data = rawd)  
summary(mod3)  
anova(mod3)

The t-statistic was -0.518, and based on the t-distribution with 1360 df, this lead to a pvalue of 0.604. There is no evidence of a quadratic relationship between year and average temperature after accounting for monthly variation and the linear relationship between year and average temperature.

1. For the model with the linear and quadratic Year components and an additive Month, produce the 2x2 diagnostic plots from plot(model) as in HW 1. This time, discuss the top left panel (Residuals vs Fitted). Discuss what you can generally assess in this plot and then discuss the specific results for this model.

The monthly discretness is apparent in the clumps in the residuals vs. fitted plot. It does appear that there is less spread in the right of the plot than the left. However, I do not know the monthly order that is plotted. If it is how the linear model was fit, the months are ordered alphabetically, so not having the usual seasonal ordering we typically see in the calendar year, so it is hard to jump to conclusions because saying there is decreasing variation wouldn't be true if I had ordered the months differently than alphabetically. If we were to account for this in the model, it would be more appropriate to fit a model that allowed for different variances within each month, since a transformation wouldn't help with the discrete predictor. Also note the scale of the residuals. The plot shows a -20 residual. I am not concerned because this is temperature data that was not standardized, so the standard empirical rule +/-3 does not help here.

par(mfrow=c(2,2))  
plot(mod3)

1. Run the following code so I can see what version of R you are using:

### Documenting R version

getRversion()

## [1] '3.3.1'