PS6_answers

vkvats

4/15/2020

Introduction

In class we observed that there was a doubt about the abnormalty observed at site Morris for the year 1931 and 1932. It was hypothesized by Cleveland (1993, pp. 5, 338) that "Either an extraordinary natural event, such as disease or a local weather anomaly, produced a strange coincidence, or the years for Morris were inadvertently reversed" and "on the basis of the evidence, the mistake hypothesis would appear to be the more likely." this hypothesis was formulated after a through analysis of the data which showed, the yields at Morris were higher in 1932 than in 1931 but for at all the other sites the opposite was true. We explore the same data set which has ten years of data from 1927 to 1936, which is equally spread to either side of year 1931 and 1932.

To explore the data first, we observe the pattern of total yeild of barley for all species combined for each year, this is shown in figure 1 below:

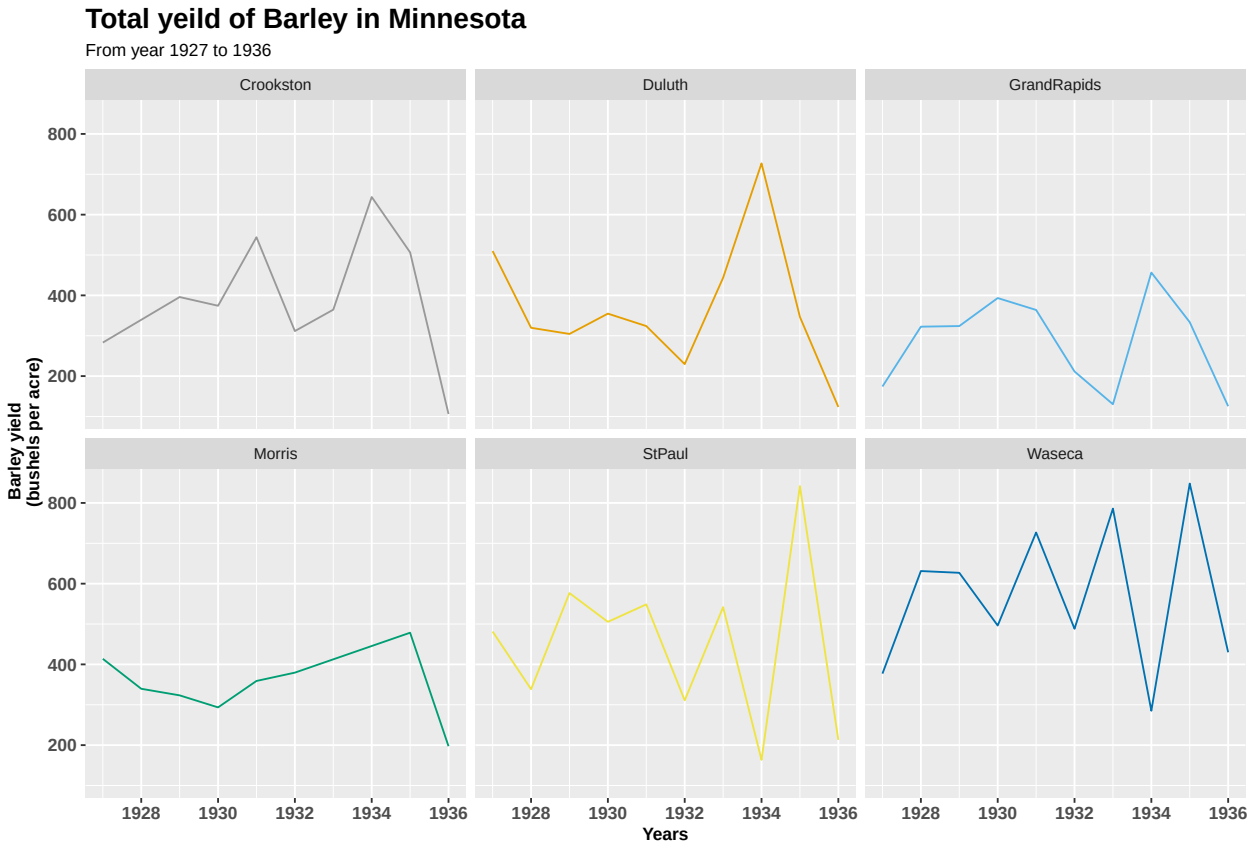

Figure 1

The plot above, in general, doesn't shows any pattern as such. For sites like Waseca, StPaul and Crookston, we can see random increase and decrease in the for consecutive years. looking at this plot we can say that it is not at all common for the yields to move in the same direction in successive years, to quote some examples, at StPaul site, the total yield has increase by four fold from year 1934 to 1935, Waseca has also observed its highest increase in total production in the same year (thought it observed its sharpest decrease in the previous year of 1933 to 1934 production), but sites like Crookstone, GrandRapids and Duluth has observed decrease in barley production in those years. In last two previous years, from 1932 to 1934, Duluth has observed sharp increase, which contrasts with StPaul, GrandRapids and Waseca. In similar way we can find many such examples which support randomness in production yield. Wright, 2013 also states that at some sites, federal funding after drought might have been the reason is sharp jumps and when the federal aid was removed, the production fell down.

Before we start investigating Morris site, it is important to note that this data doesn't have the yield data for site Morris in the years 1933 and 1934, also, from the expanded data set, we observe that each genotype of barley which was grown at different sites were not same throughout the years. These two effects makes is a little complicated to observe the pattern for individual genotype yeild at each site. The fact that not the same genotypes of barley was grown at same site every year, brings some randomness in the data, it might be possible that, some genotypes are more suited for a particular site and not that much suited for some other site based on different soil types and environmental condition (Wright (2013)). Assuming that this randomness is there, we leave the effects of this phenomena to randomness.

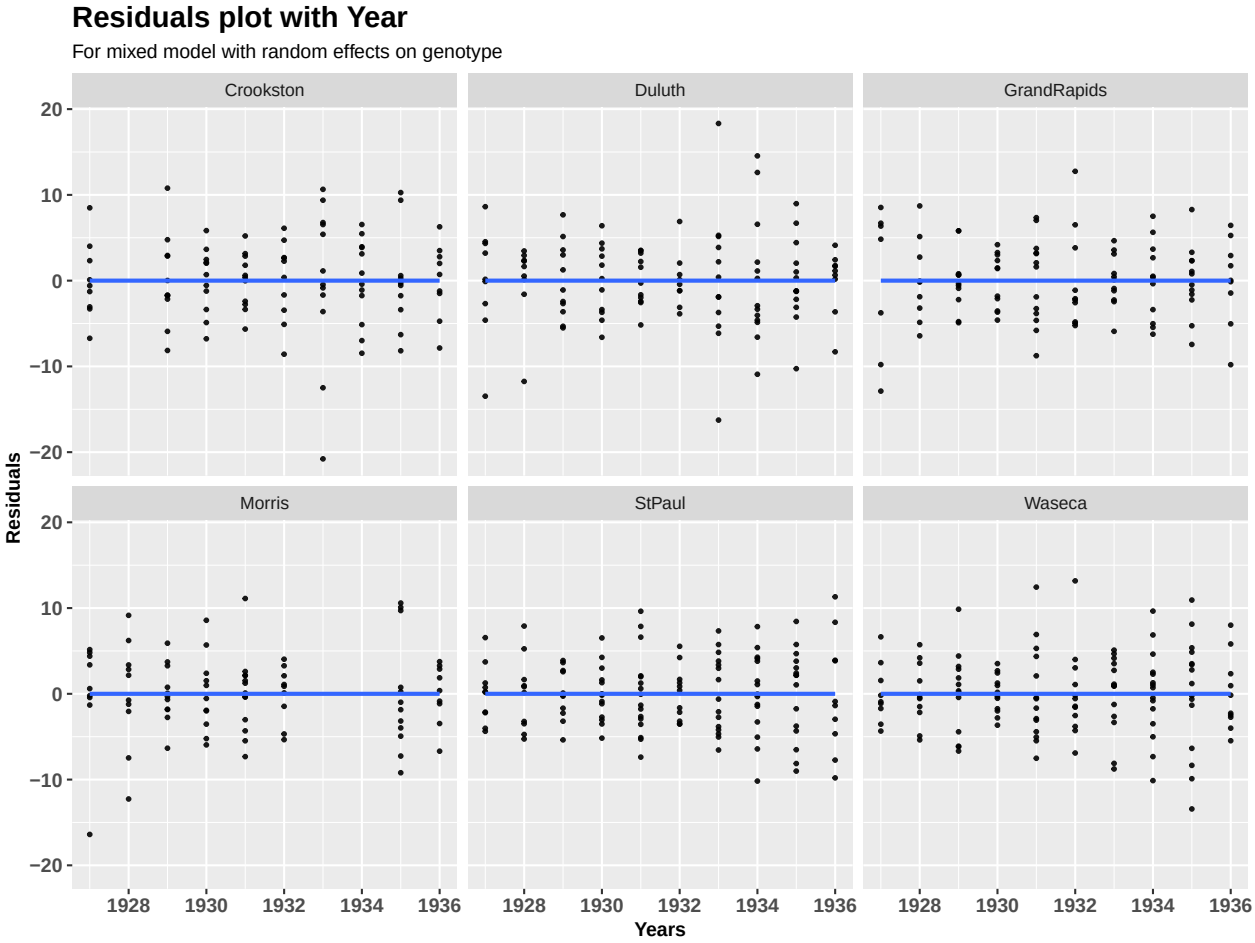

Figure 2

Model selection

Keeping in mind the randomness that we observed in site selection for genotype and other external facotors, I decided to fit a random mixed model and applied randomness on genotype selected for that site. I also found that interaction between site and year significantly improves the model output, so I have included this interaction in my model.

This model explains 87.12% of the variance of the data. This is evident from the residual plot shown in figure 2 above. The loess smoother of degree 1, perfectly fits the data, which might seems suspicious at first as attributing this explanation only to randomness can be doubted. But this is what my model is telling, and based on this, it becomes quite difficult to deny the effects of randomness in the data. Keeping that it mind, it can been seen on the residual plot that even for Morris, we don't see any deviation or kind on smoother, perhaps randomness is what is the real reason and not any advertant or inadvertant swapping of data in year 1931 and 1932 at Morris site. Has it been the case, it should have been visible in our model. The randomness explainss most of the data which can be explained as natural variation.

Referenses

[1]: Kevin Wright (2013) Revisiting Immer's Barley Data, The American Statistician, 67:3, 129-133, DOI: 10.1080/00031305.2013.801783