ECE Professional Development in Montana

A model for understanding the availability of training for Early Childhood Educators

Cassie Noble

December 12, 2018

Introduction

There are approximately 18,000 slots available in child care programs in Montana, serving an estimated 40% of the total number of children potentially in need of child care (Child Care Aware of America, 2015) (Watson, 2018). Extensive research has shown the long-term impacts of quality early childhood education and the need for high quality programs (Heckman, 2017). We also know that continued professional development is important to the success of a quality early childhood workforce (OECD). With the help of the Child Care and Development Block Grant Act, as well as a growing national movement, states across the nation are moving towards higher standards when it comes to professional development in the early care and education profession.

The Montana Early Childhood Project (ECP) is the statewide Professional Development Approval System for the early care and education workforce of Montana. The ECP also operates a statewide registry, the Practitioner Registry, to promote quality childcare and to collect workforce information about early childhood providers. Montana Child Care Licensing rule was rewritten in 2018 to require participation in the Practitioner Registry, and to increase the number of required professional development training for all licensed providers from 8 to 16 hours annually. The Practitioner Registry previously was required for individuals participating the Quality Rating Improvement System (QRIS), but was voluntary for most. Required annual training hours for the Practitioner Registry has varied over the years, but most recently required 24 hours annually.

I am interested in analyzing approved training events and hours offered per year. Since the number of required annual training hours has just recently doubled, I am interested in looking for any trends in the data, to determine if, without intervention, the number of training hours offered will be able to meet that that requirement. This has implications on training organizations and the early childhood system as a whole as to whether more investment needs to be put into creating new training events. Since online training has become more popular in recent years, I am also interested in analyzing training events by type (distance or classroom-based), to determine the effect online training has had on the number of training events offered.

In these analyses, I am primarily interested in an explanatory framework. I am interested in learning what factors have been important historically in predicting training events and training hours.

Dataset

The Practitioner Registry database contains historical data on training events submitted for approved from 1998 to present. This dataset contains the number of hours for each training event, whether the event was completed or not, and whether it was a classroom-based training or a distance training, i.e. online. Online training events have a different definition for the event date in this dataset, basically indicating that that training was available for one year prior to the end date. This is opposed to the end date for classroom-based training, which indicates the date the event occurred.

To gain an understanding of the dataset, I first aggregated the data by year and plotted the series. Counts of training events per year is shown in Figure 1, and total training hours offered per year is shown in Figure 2.

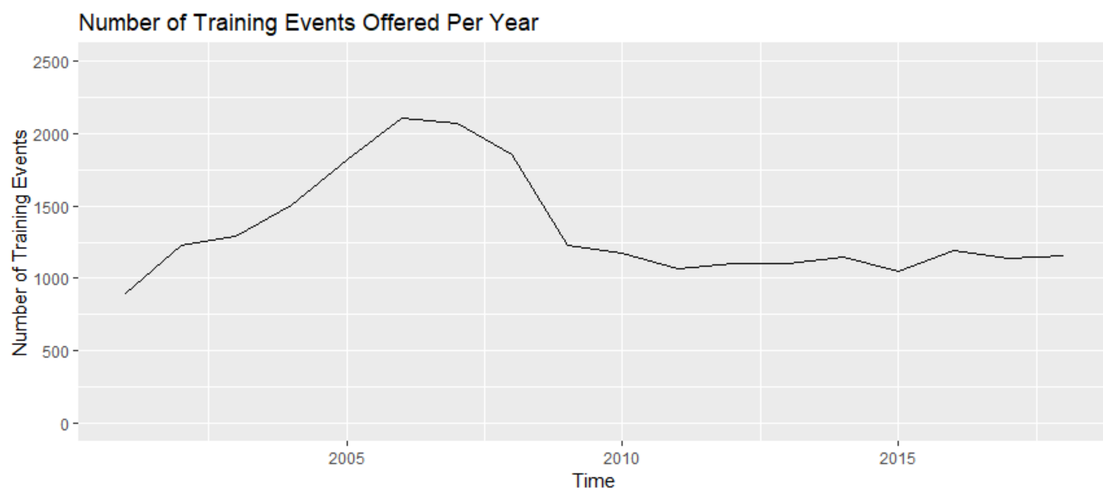

Figure 1. All approved and completed training events, aggregated by year. This dataset includes events from 2001 through 2017. Note the steep drop off from 2007 to 2010 due to the increasing availability of online training.

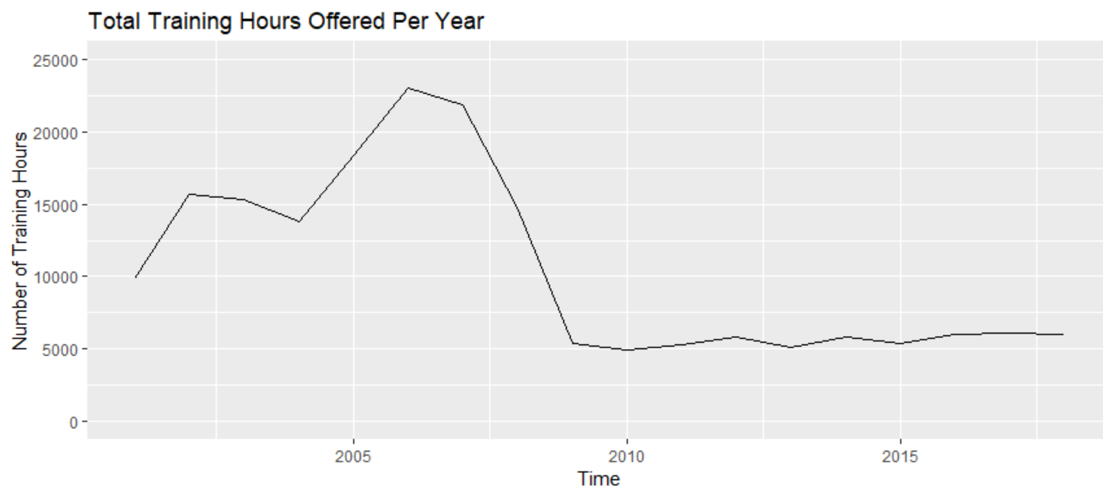

Figure 2. Total training hours for approved and completed training events, aggregated by year. This dataset includes events from 2001 through 2017. Note the steep drop of availability from 2007 to 2010 due to the increase availability of online training.

Online training started to become approved in 2005, and continued to increase in popularity until 2008. This is reflected in both Figure 1 and Figure 2, with the availability of training being offered dropping off steeply starting in 2007, and reaching a minimum in 2010. For this reason, and because I am interested in the more recent availability of training, I decided to subset the dataset to only consider training event approved from 2010 forward. Since online training has a different definition for the event date, I also decided to focus on classroom-based training. A plot of the dataset used in subsequent analyses is shown in Figure 3.

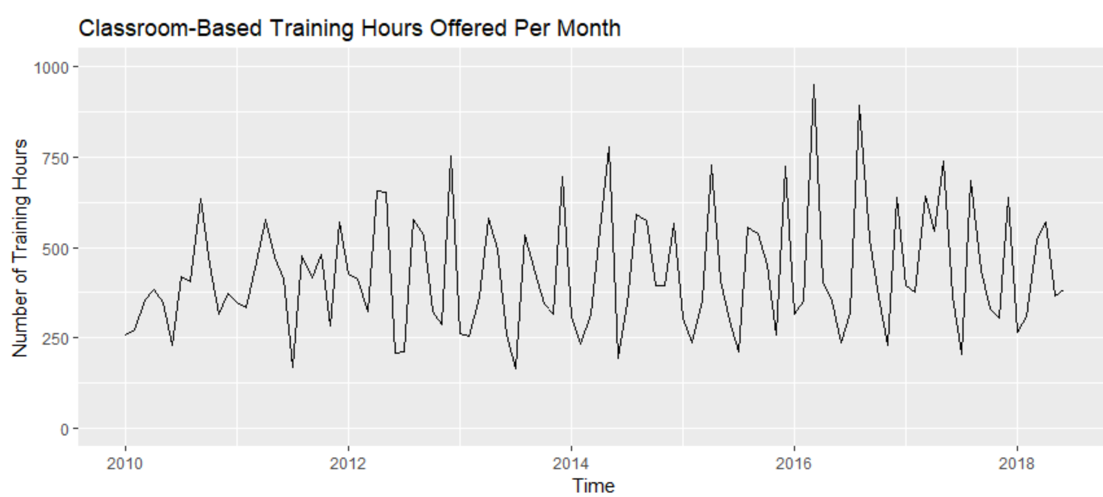

Figure 3. Classroom-Based Training Hours, aggregated by month. This series contains training hours from events approved and completed from January 2010 through June 2018. There appears to be a strong seasonal cycle.

Methods

A decomposition of the time series was constructed to get an idea of the general structure of the series, as shown in Figure 4. There appears to be an overall slight increasing trend, and a clear seasonal cycle.

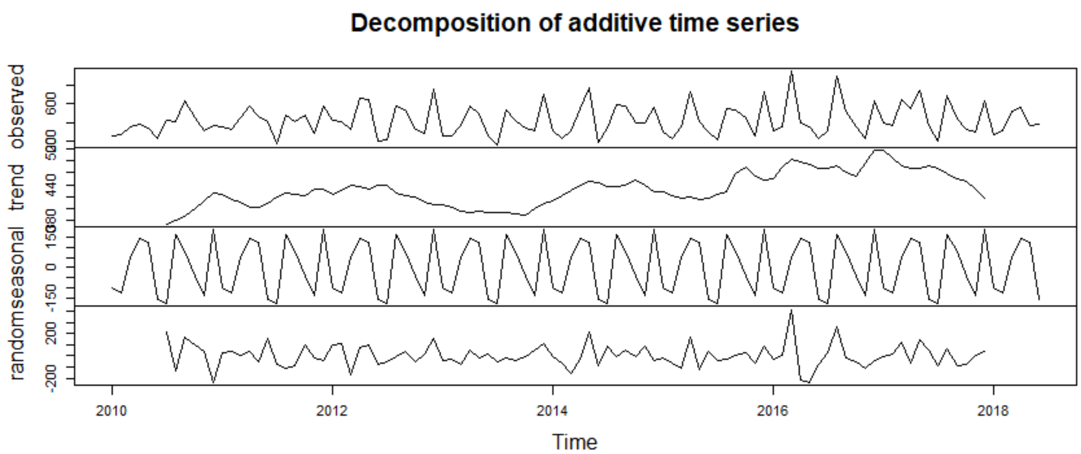

Figure 4. Decomposition of the classroom-based training hour time series. There appears to be as slight increasing trend, and a clear seasonal cycle.

Figures 5, 6, and 7 show the training data plotted as functions of theorized regressor variables. Possible regressor variable tested in this analysis included the year, and seasonal indicators for month of the year, shown in Figure 5. The month of the year could be important as some months are more popular for training. For example, July is typically a slower month as many people are on summer vacation, and August is when head starts begin their school year and have new staff training. Figure 6 shows monthly classroom training hours plotted by the number of hours offered for distance training in the previous year. Because the addition of online training had such an impact on the number of trainings offered historically, this could continue to have an effect. I am using lagged distance training (from one year previously), so that this variable could be included in future forecasting if necessary.

I also explored the possibility of including the number of annual training hours required for renewal on the Practitioner Registry as a possible regressor variable. A plot of this data is shown in Figure 7. The Practitioner Registry required 23 or 24 hours of training annually for many years, and roughly one-third of the workforce was participating in the registry by 2017. With such a higher standard than Child Care Licensing, and such a high percentage of the workforce participating, this could also have influnced the number of training hours being offered.

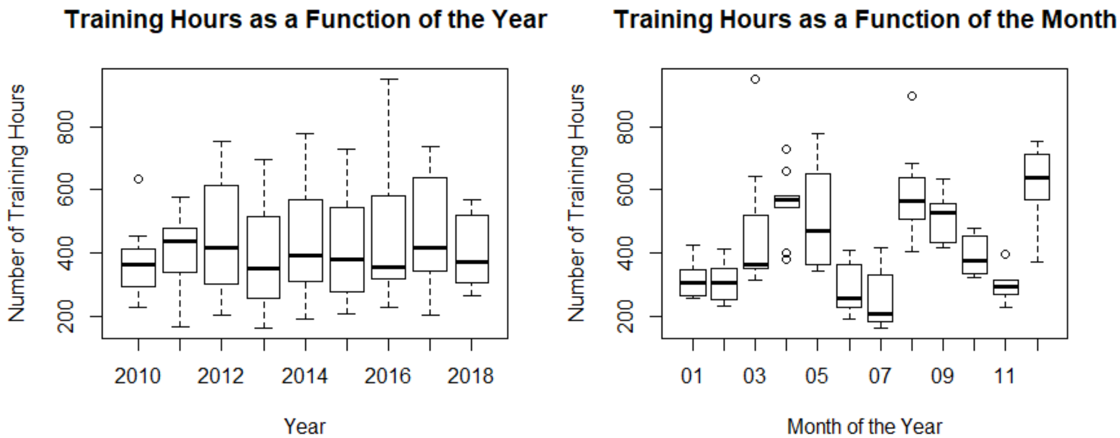

Figure 5. Total monthly classroom-based training hours plotted as a function of the year (left), and as a function of the month (right).

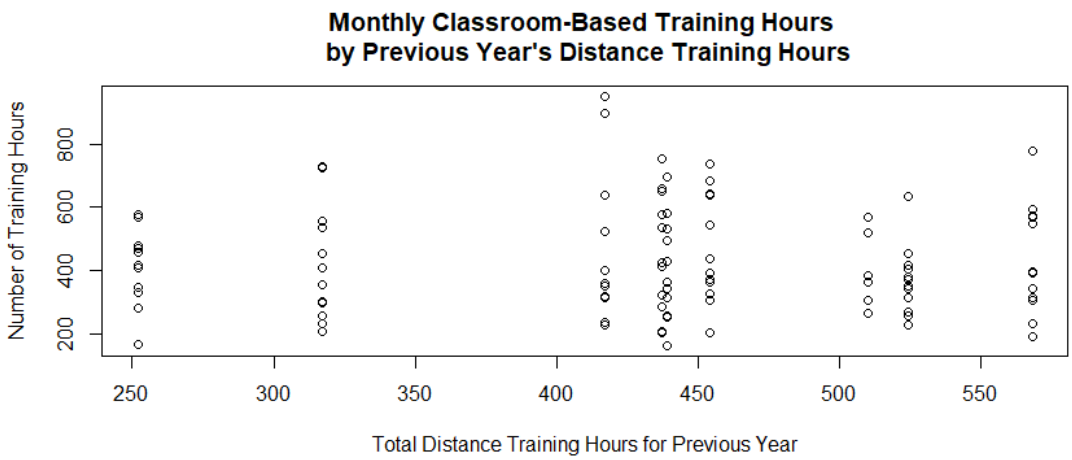

Figure 6. Total monthly classroom-based training hours plotted as a function of the number of distance training hours offered the previous year.

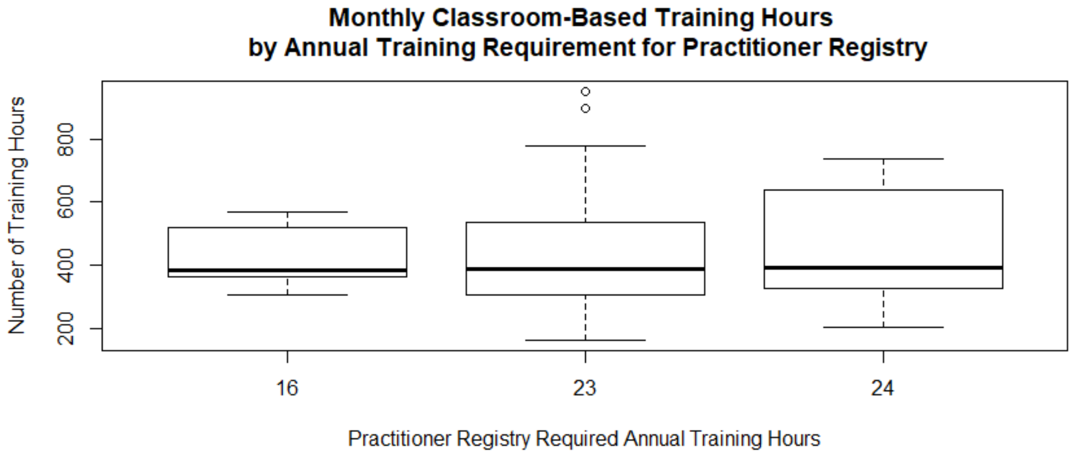

Figure 7. Total monthly classroom-based training hours plotted by the number of hours required for renewal on the Practitioner Registry. There is no clear difference between the boxplots, other than slightly different distributions.

I first fit simple regression models for each of the four variables mentioned above. Regression models for the year, lagged distance training hours, and required Practitioner Registry training hours were not significant, so they were eliminated from subsequent analyses. The model with monthly season indicators was significant and was compared to other types of models as discussed in the results section.

A regression model with monthly seasonal indicators would be as follows:

$$\hat{y}_{training} = \begin{cases} \beta_1 + z_t \\ \beta_2 + z_t \\ \vdots \\ \beta_{12} + z_t \end{cases}$$

where $\hat{y}_{training}$ is the predicted number of training hours offered, z_t is the error term, and β_i is the ith seasonal term.

I plotted the ACF and PACF to take a look at the autocorrelation of the time series. A plot of which can be seen in Figure 8. These plots show a clear correlation structure continuing throughout the plot, likely indicating a Moving Average process.

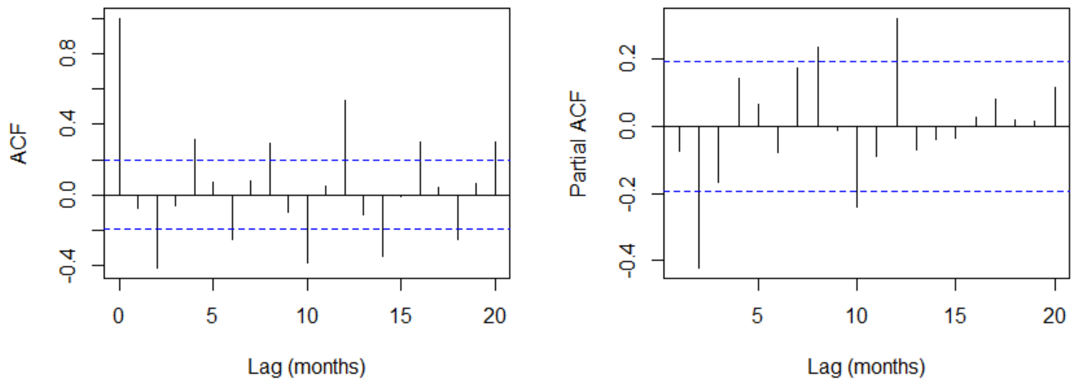

Figure 8. Autocorrelation function (left), and Partial autocorrelation function for the monthly training hour time series. There appears to be a correlation structure continuting out to a lag of 20 months.

In R, `auto.arima` was used to select the best ARIMA model, which turned out to be a Moving Average process of order 4 with a mean term. A model was also fit using the monthly seasonal regression variables along with the MA(4) with mean term. An exponential smoothing state space model was also tested using `ets()` in R.

Results

```
##
## Call:
## lm(formula = classroomHours ~ month, data = hoursMonthYr)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -246.19  -62.37  -12.11   62.66  473.39
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   320.67      38.20   8.394 6.25e-13 ***
## month02       -11.83      54.03  -0.219 0.827121
## month03       155.44      54.03   2.877 0.005009 **
## month04       234.06      54.03   4.332 3.83e-05 ***
## month05       191.50      54.03   3.545 0.000626 ***
## month06       -33.39      54.03  -0.618 0.538122
## month07       -65.10      55.69  -1.169 0.245456
## month08       269.83      55.69   4.845 5.24e-06 ***
## month09       190.15      55.69   3.414 0.000960 ***
## month10        71.27      55.69   1.280 0.203897
## month11       -21.60      55.69  -0.388 0.698970
## month12       299.52      55.69   5.379 5.89e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 114.6 on 90 degrees of freedom
## Multiple R-squared:  0.5722, Adjusted R-squared:   0.52
## F-statistic: 10.95 on 11 and 90 DF,  p-value: 1.517e-12
```

Above is the output from the regression model with monthly seasonal indicators. The overall model is significant with overall F on 11 and 90 DF $= 10.9457337, p < 0.0001$. Most months of the year are significant predictors, as is shown in the table.

```
## Series: hoursMonthYr$classroomHours
## ARIMA(0,0,4) with non-zero mean
##
## Coefficients:
##           ma1      ma2     ma3     ma4      mean
##       -0.2435  -0.3221  0.0120  0.3117  426.3213
## s.e.   0.0963   0.1041  0.1281  0.1154   10.8799
##
## sigma^2 estimated as 22262:  log likelihood=-653.03
## AIC=1318.07   AICc=1318.95   BIC=1333.82
```

The best fitting ARIMA model was a Moving Average process of order 4, with mean term. The output from that model is included above. Since that was the best ARIMA model, and the seasonal regression model was significant, an MA(4) with mean term and seasonal regression variables was also tested. Output from that model is included below.

```
##
## Call:
## arima(x = hoursMonthYr$classroomHours, order = c(0, 0, 4), xreg = hoursMon
thYr$month)
##
## Coefficients:
##           ma1      ma2     ma3     ma4  intercept  hoursMonthYr$month
##       -0.2229  -0.3189  0.0036  0.2902   415.5150              1.7365
## s.e.   0.1043   0.1045  0.1297  0.1196    21.9663              3.0757
##
## sigma^2 estimated as 21113:  log likelihood = -652.86,  aic = 1319.71
```

The final model that was tested was the exponential smoothing state space model using `ets()`. The output from that model is included below.

```
## ETS(M,N,A)
##
## Call:
##  ets(y = hoursMonYrTS)
##
##   Smoothing parameters:
##     alpha = 1e-04
##     gamma = 1e-04
##
##   Initial states:
##     l = 425.5393
```

```
##     s = 180.9247 -133.7119 -43.5824 86.8201 155.839 -166.4201
##            -147.0749 122.2369 140.0546 52.456 -134.6267 -112.9154
##
##   sigma:  0.2658
##
##      AIC     AICc      BIC
## 1440.939 1446.521 1480.314
```

```
##               df      AIC
## etsHrs        15 1440.939
## lmMonth       13 1269.961
## arimaHrsMonth  7 1319.713
## arimaHrs       6 1318.070
```

These four models were tested using AIC to find the best performing model. The results of this comparison are shown above The exponential state space model was the worst performing with the highest AIC at 1440.9393027. Both Moving Average models performed similarly, with the model that did not include the extra regression variables performing slightly better (1318.0698632 compared to 1319.712599). The best performing model was the regression model with monthly seasonal indicators with an AIC of 1269.9614761.

Model diagnostics were run for the monthly seasonal regression model. Autocorrelation and partial autocorrelation functions for the model residuals are plotted in Figure 9. There is no autocorrelation or clear structure in these plots.

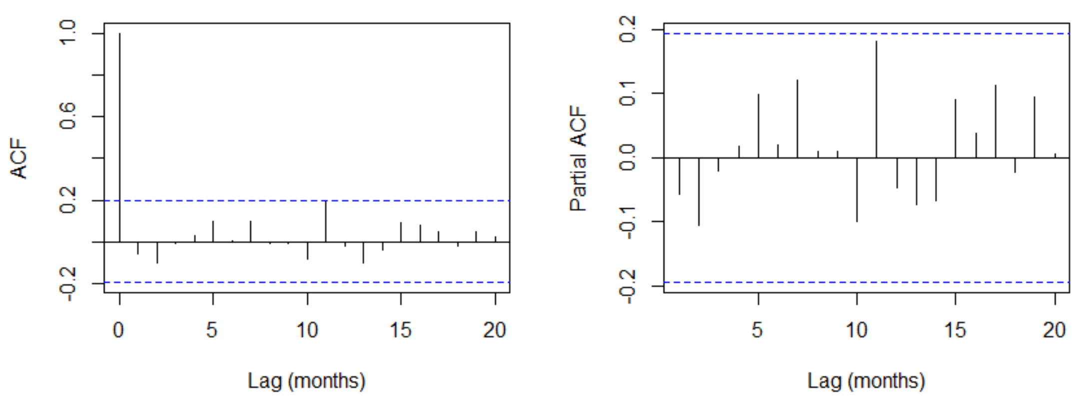

Figure 9. Autocorrelation function (left), and Partial autocorrelation function for the residuals of the regression model that includes monthly seasonal indicators to predict monthly classroom-based training hours. There is no significant autocorrelation or structure remaining in the residuals.

Conclusion

The best model in predicting monthly classroom-based training hours is a regression model including indicators for each month of the year. A plot of predictions from this model overlaid on observed data is shown in Figure 10.

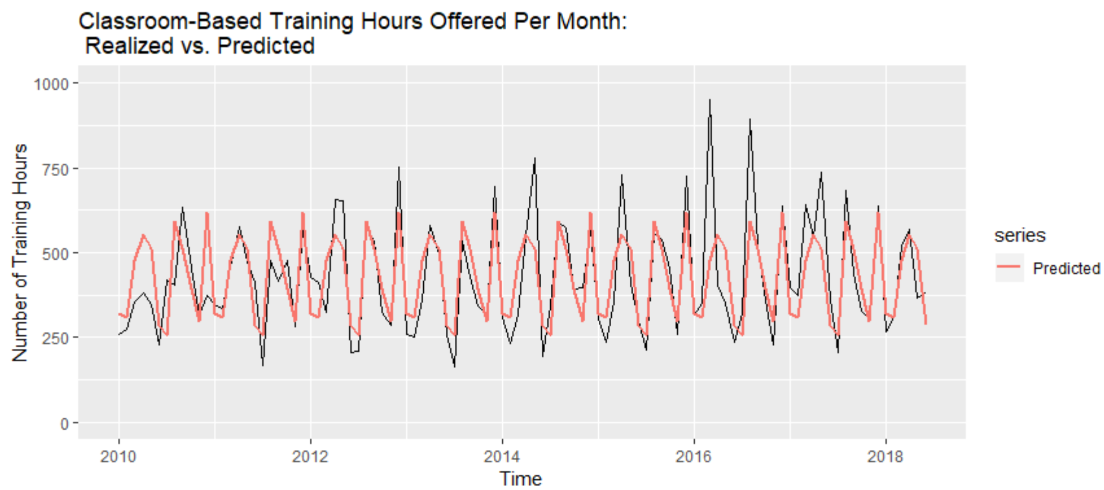

Figure 10. Plot of classroom-based training hours offered per month, with observed data in black. The red line indicates the predicted training hours offered per month based on the regression model with monthly seasonal indicators.

This model makes sense in context. Training availability varies depending on the month of the year and often can fluctuate according to traditional school year.

If I were to conduct this analysis again, I would probably attempt to fit a model based on other seasonal cycles, like the four seasons of the year, or following the school year with indicators for summer. I would also try to incorporate other regression variables, such as the total number of child care providers in the field, training attendance numbers, or total number of training hours required for Montana Quality Rating Improvement System (QRIS) participants.

The effect of online training did turn out to have a dramatic effect. Future analyses could focus more on the time period of the shift from more classroom-based training to more online training to learn more about this effect. The fact that there was only a slight trend in the data demonstrates that the amount of training being offered currently may not suffice in meeting the coming increase in demand for training.

References

- Child Care Aware of America (2015). *Child Care in the State of Montana.* Child Care Aware of America.
- Cowpertwait, P.S.P. and Metcalfe, A.V. (2009). *Introductory Time Series with R*. New York, NY: Springer Science+Business Media, LLC.
- Heckman, J.J. (2017). *There's More to Gain by Taking a Comprehensive Approach to Early Childhood Development.* Retrieved from https://heckmanequation.org/resource/research-summary-lifecycle-benefits-influential-early-childhood-program/
- Nau, R. (2018). *Statistical Forecasting: Notes on regression and time series analysis* Retrieved from https://people.duke.edu/~rnau/411home.htm
- OECD. Encouraging Quality in Early Childhood Education and Care (ECEC). OECD, Paris.
- Watson, A. (2018, September). Child Care in Montana: Supporting Montana Families and Caring for the State's Most Precious Resource. *Montana Economy at a Glance.*