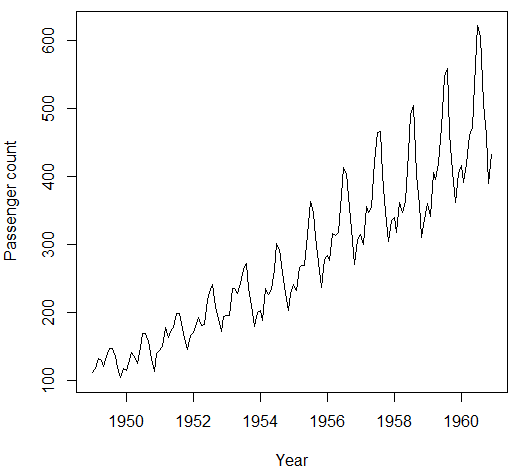
Student Deniss Belovs 4801BD

***REPORT***

***Laboratory Work N7***

***Name of the data set: Air Passengers***

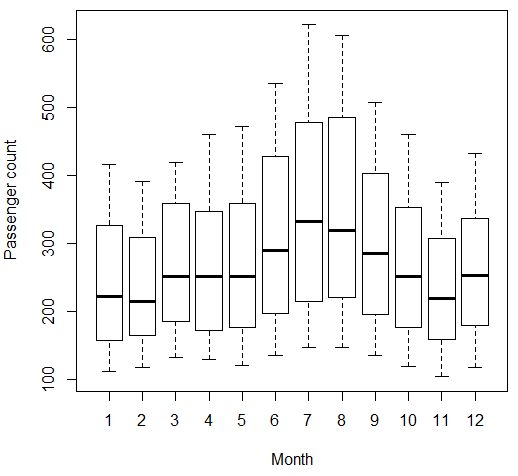
1. ***Line plot of the data***



***Describe plot:***

Рассматривая данную диаграмму можно отметить наличие тенденции возрастания у имеющихся данных (выраженный возрастающий тренд). Значения кратковременно и периодически изменяются с фиксированной частотой, что указывает на сезонность временного ряда. Также, стоит отметить увеличение амлитуды изменения данных с каждым последующим годом (возрастает дисперсия y во времени), из чего можно сделать вывод, что имеющийся временной ряд является **нестационарным,** и следует использовать мультипликативную модель.

1. ***Boxplot of the data***

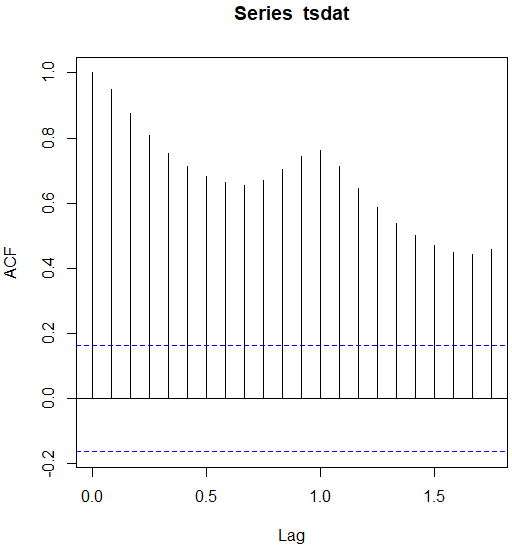


***Make a conclusion about boxplot:***

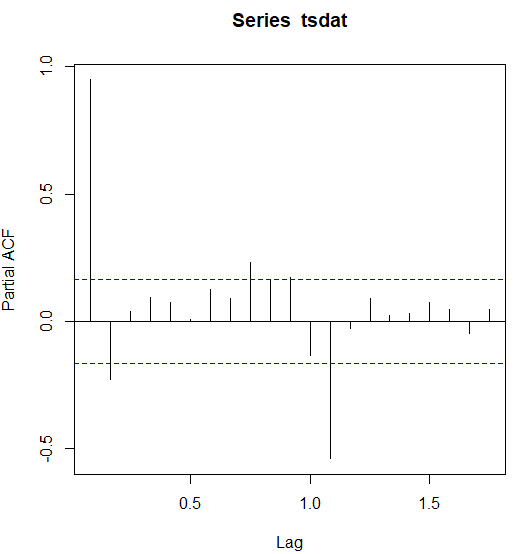
На графике видно отсутствие каких-либо выбросов. Наибольшее значение медианы, и максимальное количество пассажиров приходится на 7-ой месяц (июль). Минимальное количество пассажиров зафиксировано на 11-ом месяце (ноябрь). Верхний квартиль межквартильного размаха 8-ого месяца (август) имеет наибольшее значение количества пассажиров среди всех остальных межквартильных размахов. Высота всех ящиков (межквартильных размахов) отличается.

1. ***Autocorrelation and partial autocorrelation***

Autocorrelation



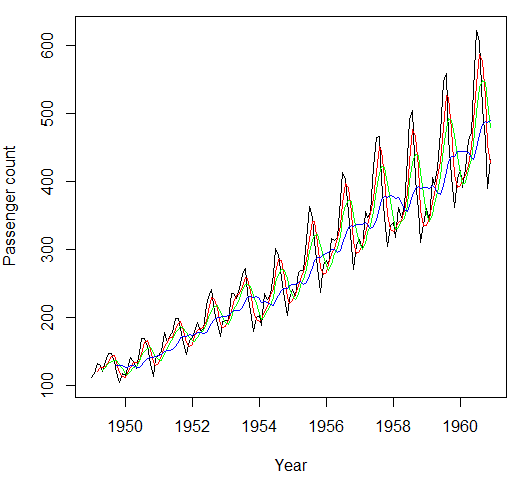
Partial autocorrelation



***Make a conclusion:***

Значения автокорреляции и частной автокорреляции уменьшаются с увеличением значения лагов. Можно обнаружить наличие сезонности. Сильная, положительная автокорреляция наблюдается на протяжении всего временного ряда, в то время как сильная частная автокорреляция наблюдается только при значениях лага равным *0.1667, 0.7500, 1.0833*.

1. ***Simple moving average***

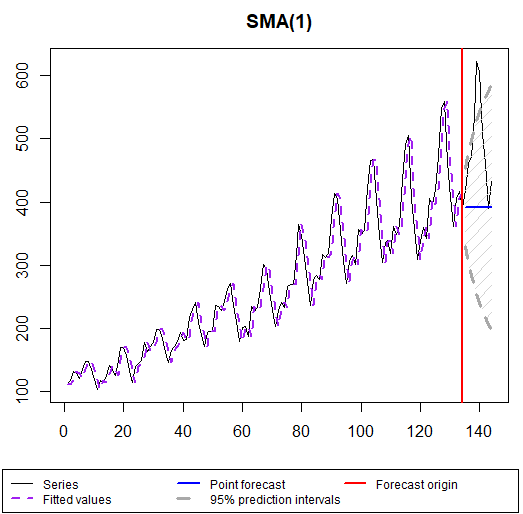


***Make a conclusion:***

С увеличением количества лагов (5 – зелёная линия, 10 – синяя линяя) наблюдается более выраженное сглаживание, уменьшается количество скачков (дисперсия), линия среднего - выпрямляется.

1. ***Simple moving average (built-in) forecast (for last 10 observations)***

***Plot:***



***Forecast:***

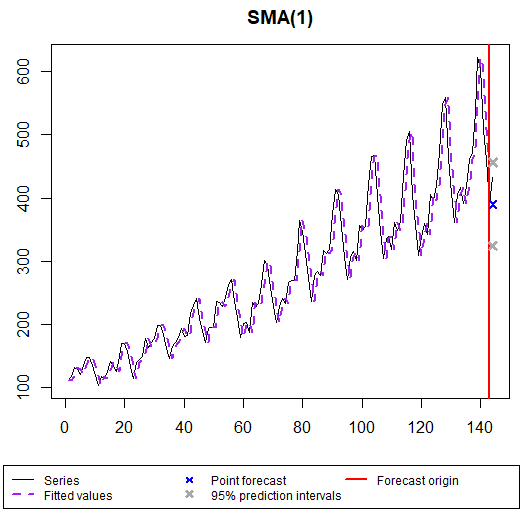
|  |  |  |  |
| --- | --- | --- | --- |
| ***MSE*** | ***MAE*** | ***MAPE*** | ***RMSE*** |
| ***15283.8*** | ***99.8*** | ***18.65237%*** | ***123.6277*** |

***Make a conclusion:***

Значение метрики *MSE (Mean Squared Error)* достаточно большое, из чего можно сделать вывод о том, что имеющаяся модель имеет малую точность прогнозирования. Данная модель недостаточно хорошая, так как не все реальные значения попали в доверительный интервал, из чего следует, что не все данные будет возможно правильно спрогнозировать.

1. ***Simple moving average (built-in) forecast (for the last observation)***

***Plot:***



***Forecast:***

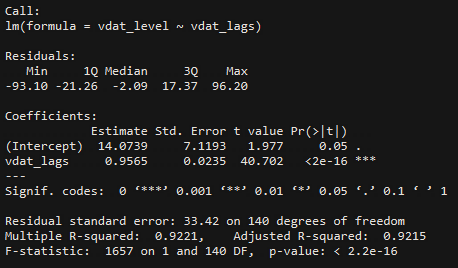
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***MSE*** | ***MAE*** | ***MAPE*** | ***RMSE*** | ***Actual value*** | ***Forecast*** | ***Confidence interval*** |
| ***1764*** | ***42*** | ***9.722222%*** | ***42*** | ***432*** | ***390*** | ***(323.2511; 456.7489)*** |

***Make a conclusion:***

Данная модель лучше предыдущей, так как все значения её ошибок меньше, а реальное значение попадает в доверительный интервал. Следовательно, данная модель будет лучше прогнозировать последующие значения.

1. ***Autoregression model***

***Model summary:***



***Forecast:***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***MSE*** | ***MAE*** | ***MAPE*** | ***RMSE*** | ***Actual value*** | ***Forecast*** | ***Confidence interval*** |
| ***22.30816*** | ***4.723151*** | ***1.093322%*** | ***4.723151*** | ***432*** | ***427.2768*** | ***(418.2418; 436.3119)*** |

***Make a conclusion:***

Значения ошибок в данной модели самые маленькие, по сравнению с предыдущими моделями. Реальное значение попадает в доверительный интервал, а спрогнозированное значение наиболее близко к реальному из всех рассмотренных моделей. Данная модель описывает *~92%* сумм квадратов, что свидетельствует о достаточно высоком качестве модели. Значение *p-value* переменной *vdat\_lags* меньше уровня значимости, из чего следует что коэффициенты данной модели не равны нулю. Эта модель лучше всего прогнозирует будущие значения.

1. ***Comparison of models obtained in the previous stages.***

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***Model*** | ***MSE*** | ***MAE*** | ***MAPE*** | ***RMSE*** | ***Actual value*** | ***Forecast*** | ***Confidence interval*** |
| SMA(1) | ***1764*** | ***42*** | ***9.722222%*** | ***42*** | ***432*** | ***390*** | ***[323.2511; 456.7489]*** |
| AR (1) | ***22.30816*** | ***4.723151*** | ***1.093322%*** | ***4.723151*** | ***432*** | ***427.2768*** | ***[418.2418; 436.3119]*** |

***Make a conclusion:***

Сравнивая имеющиеся модели я пришёл к выводу, что **авторегрессионная модель** показала себя лучше, чем **модель скользящего среднего**, так как:

* значения её ошибок наименьшие,
* прознозируемое значение наиболее близко к реальному,
* реальное значение попало в доверительный интервал, в то время как не все реальные значения в первой модели скользящего среднего попали в доверительный интервал.