Time Series & Forecasting,  
End Semester Practical Exam  
*(conclusions)*

Pranav Gopalkrishna, 1940223

# Dataset

The dataset used here contains daily data of female births in an unspecified region, for the entire year of 1959. This is the time series data which we will try to study and model.

# Time series components observed

Based on the time plot of the data, we observed

* No trend
* No seasonality
* No clear periodicity, either short term or long term
* Limited average deviation from a constant mean
* Finite variance

From this alone, we may conclude that the time series is stationary. But to test this more rigorously, we performed the augmented Dickey-Fuller (ADF) test, and inspected the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots.

# Testing for stationarity

The results of the ADF test lead us to conclude that, given a 0.05 significance level, the given time series is stationary. To confirm this visually, we checked the ACF and PACF plots, to observe that there was largely insignificant autocorrelation between observations separated by up to 25 lags. The occasional significant autocorrelations may be chalked down to error. Similar results were obtained from the ACF and PACF plots.

The largely insignificant autocorrelations, along with the fact that the data seems to display constant mean and finite variance, suggests moderate to strong stationarity for the time series.

# Fitting a model

Given the largely insignificant autocorrelation between observations, autoregression may be entirely absent from this time series process. Due to this, and due to its moderate to strong stationarity, we conclude that a moving average model may be more suitable for this time series. Hence, for the time series process model **ARIMA(p, d, q)**, to obtain **MA(1),** we put

* p = 0
* d = 0
* q = 1

Since the given data is a stationary time series, it is best to use stationary time series models, instead of non-stationary models such as **ARIMA(p, d, q)** where d > 0. This judgement seems to be validated by the results of fitting a non-stationary ARIMA model to our time series data, wherein the fitted values differed greatly from the actual values.

Our preferred model is **MA(1)**, even though the Akaike Information Criterion (AIC) value for the non-stationary fitted model was lower. This is because not only does it visually match the time plot of the data better, but also, we are note preforming any differencing to make the data stationary, since the time series was already concluded to be sufficiently stationary. However, the fitted non-stationary model does not consider this conclusion, and performs differencing of order 1 on the time series, which may have created the observed discrepancies between the actual values and the fitted values.

# Testing fitness of preferred model

For the residuals of the preferred **MA(1)** model, we tested for the following assumptions:

1. Residual mean is zero
2. Residuals are non-correlated
3. Residuals are normally distributed

Assumption 1 was tested using the one-sample ***t-test*** (to test for the true mean of the residuals, considering the observed residuals as a sample). Based on the results, we concluded the true mean of the residuals to be zero.

Assumption 2 was tested using the ***Portmanteau*** or ***Box-Pierce*** test. Based on the results, we concluded the residuals to be uncorrelated.

Assumption 3 was tested using the ***Shapiro-Wilk*** test. Based on the results, we concluded the residuals to not be normally distributed.

Due to non-normal distribution of the residuals, we may assume that the residuals may be caused by more than random error, hence indicating that our model may not be the fittest possible model for our data.

# Speculations

The conclusion that the residuals were not drawn from a normal population could be due to the following non-exhaustive reasons:

* Random chance i.e. the current set of residuals may be non-representative of the residuals from this model in general
* Overlooked factors that may be significant to the observations, and may have caused a non-normal distribution of the residuals

To correct for the first reason, we could increase our data sample and then create and test an MA(1) model. Alternatively, it could be that MA(1) is indeed not the most suitable stationary model for our time series.