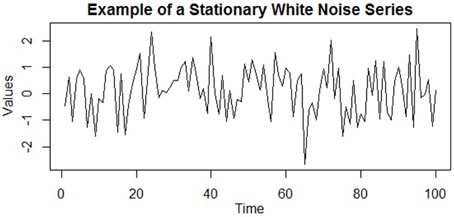
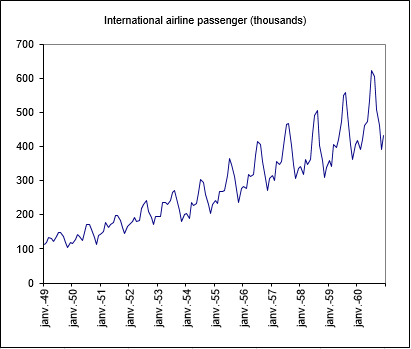
In time series, the summary statistics of observations are consistent, (e.g. mean, max, min, etc) or being stationary.

**Stationary Time Series**: data does not have any upward or downward trend or seasonal effects. Mean or variance are consistent over time



**Non-Stationary Time Series**: data show trends, seasonal effects, and other structures depend on time. Forecasting performance is dependent on the time of observation. Mean and variance change over time and a drift in the model is captured.



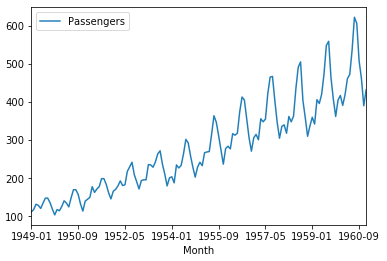
Classically, you should make your time series stationary. However, there are cases where unknown nonlinear relationships cannot be determined by classical methods. This information can be a source of information when building machine learning models. Non-stationary information can be used in feature engineering and feature selection.

**Checks for Stationarity**

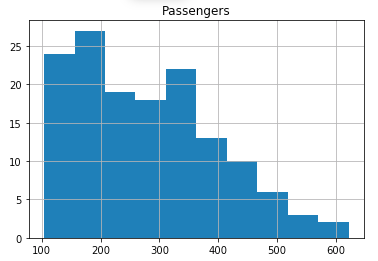
1. **Look at Plots**: plot a run sequence plot to see anything with an obvious trend or seasonal effects
2. **Summary Statistics**: partition your data into intervals and check for obvious or significant differences in summary statistics
3. **Statistical Test**: use statistical tests if the expectations of stationarity are met or violated

**Airline Passengers Dataset**

[data source](https://datamarket.com/data/set/22u3/international-airline-passengers-monthly-totals-in-thousands-jan-49-dec-60#!ds=22u3&display=line)



Look at Plots #1, run sequence plot obviously shows an upward trend as time goes by. We can also observe seasonal trends from similar shape in each year.



Look at Plots #2, if the data is stationary the summary statistics should be consistent over time. The mean should be consistent with a consistent variance indicating a Gaussian distribution. The histogram does not show Gaussian distribution and another indication of non-stationary time series data.

https://cdn-images-1.medium.com/max/800/1*jdGJN4RBbtE8kEXvsrc4bA.png

Summary Statistics, the mean and variance are very different from the first and second half of the data. Another indication of non-stationary data.

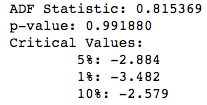
Trick, applying a log transformation to the time series will do this to our plots and summary statistics

* log run sequence plot, shows seasonal trend and non-stationary
* log histogram plot, shows sort of Gaussian distribution and might indicate stationary
* log summary statistics, show similar values and might indicate stationary
* Note, take a look at the data through different methods before deciding if your transformation converted your non-stationary data to stationary data

**Augmented Dickey-Fuller test (ADF)**

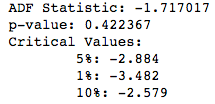
ADF tests the null hypothesis that a unit root is present in time series sample. ADF statistic is a negative number and more negative it is the stronger the rejection of the hypothesis that there is a unit root.

* **Null Hypotehsis (H0)**: If accepted, it suggests the time series has a unit root, meaning it is non-stationary. It has some time dependent structure.
* **Alternate Hypothesis (H1)**: The null hypothesis is rejected; it suggests the time series does not have a unit root, meaning it is stationary.
* **p-value > 0.05**: Accept H0, the data has a unit root and is non-stationary
* **p-value ≤ 0.05**: Reject H0. the data does not have a unit root and is stationary



More negative ADF Statistic is the more likely we reject H0. In this case the stat is positive and way above any critical values. p-value is also way above 0.05 and we cannot reject H0. The data has a unit root and is non-stationary.

What happens when we log transform our data?



ADF Stat is negative and lower, but still above our critical values. p-value is also above 0.05. The data is closer to being stationary, but still has a unit root and is non-stationary.