Econometria e Séries Temporais
- Aula 6 -

Prof. Mestre. Omar Barroso

Instituto Brasileiro de Educação, Pesquisa e Desenvolvimento

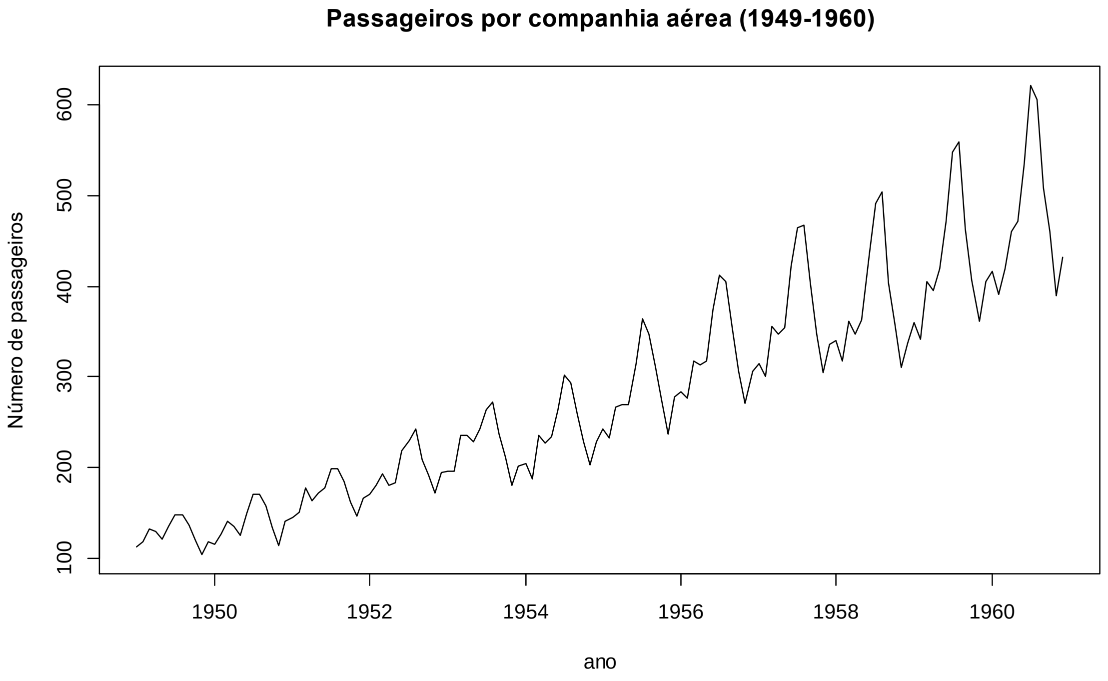
SARIMA

- S: Seasonal
- AR: AutoRegressive
- MA: Movivng Average
- Série Sazonal, Autoregressiva com média móvel.

SARIMA

- Em econometria, o modelo **SARIMA** é usado para modelar e prever dados de séries temporais que exibem padrões de tendência e sazonais, como PIB, taxas de inflação ou dados de vendas que se repetem periodicamente (por exemplo, trimestralmente, mensalmente).

Sazonalidade de passageiros de companhias aéreas



Metodologia

- **Estacionariedade:** SARIMA, como ARIMA, requer que a série temporal seja estacionária, o que significa que suas propriedades estatísticas (média, variância e autocovariância) não mudam ao longo do tempo.
- Para dados não estacionários, SARIMA aplica diferenciação tanto no nível regular (d) quanto no nível sazonal (D) para remover tendências e padrões sazonais.

Metodologia

- **Autoregressão (AR):** O componente AR modela a relação entre um valor e seus valores passados. Por exemplo, se o PIB depende de seus valores de trimestres anteriores, um termo autorregressivo é introduzido.

Metodologia

- **Média Móvel (MA):** O componente MA modela a relação entre um valor e erros de previsão passados (choques). Por exemplo, se erros de previsão passados de períodos anteriores afetam o valor atual do PIB, termos de média móvel são incorporados.

Metodologia

- **Sazonalidade:** Os componentes sazonais (P, D, Q) são como os componentes não sazonais (p, d, q), mas modelam especificamente padrões repetidos em intervalos regulares, como ciclos anuais ou trimestrais em dados.
- Por exemplo, o PIB trimestral pode apresentar variação sazonal devido a atividades econômicas cíclicas, como vendas de fim de ano ou desacelerações na produção.

Componentes

- $SARIMA\ (p, d, q)(P, D, Q)s$
- P: Ordem dos termos autoregressivos (AR).
- d: Número de diferenças aplicadas para tornar uma série estacionária (integrar).
- q: Ordem de termos da média móvel (MA).
- P: Ordem sazonal dos termos Autoregressivos.
- D: Número de diferenças sazonais.
- Q: Ordem das médias móveis sazonais.
- s: O tamanho do ciclo sazonal (e.g., 12 para anual e 4 para trimestral).

Construção do modelo

- **Diferenciação sazonal:** para contabilizar a sazonalidade, o SARIMA aplica a diferenciação sazonal (D), que subtrai o valor do mesmo período na temporada anterior. Isso remove quaisquer efeitos sazonais regulares. $y'_t = y_t - y_{t-1}$

- **Combinando termos regulares e sazonais:** o SARIMA combina termos ARIMA regulares (p, d, q) com termos ARIMA sazonais (P, D, Q) para capturar correlações sazonais e de curto prazo nos dados.

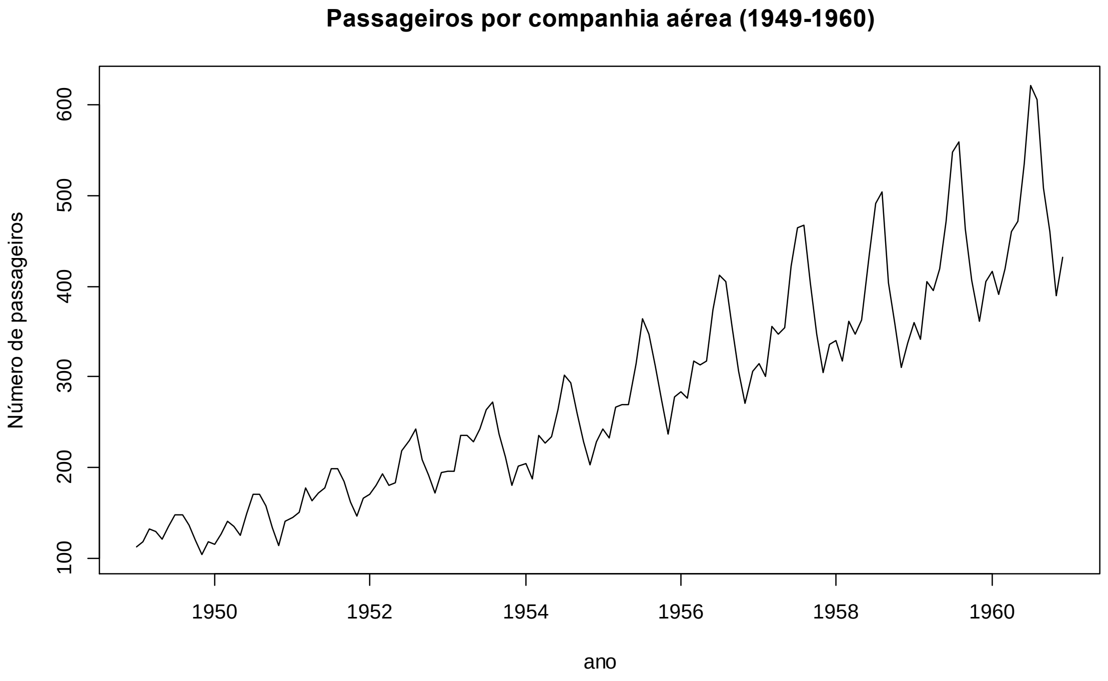
Suposições Fundamentais

- **Linearidade:** os modelos SARIMA assumem relações lineares nos dados.
- **Estacionariedade:** a série temporal deve se tornar estacionária após a diferenciação apropriada.
- **Sazonalidade:** o modelo presume que o componente sazonal é periódico e pode ser capturado por padrões repetidos.

Exemplo Teórico

- Vamos supor que queremos aplicar um SARIMA no PIB de uma nação.
- Para dados mensais do PIB com um **ciclo anual**, um modelo **SARIMA(1,1,1)(1,1,1)[12]** indica:
- Um componente AR(1) e MA(1) para efeitos de curto prazo (não sazonais).
- Um componente sazonal AR(1) e MA(1) para contabilizar a influência de valores e erros do mesmo mês do ano anterior.
- Diferenciando tanto no nível regular (d=1) quanto no nível sazonal (D=1) para atingir a estacionariedade.

Vamos testar o SARIMA sob os dados dos passageiros aéreos



Utilizando o ADF (Relembrando a Aula 3)

- *(Mesmo com series sazonais temos que testar a estacionáriedade).*
- H_0: A série temporal é não estacionária. Em outras palavras, ela tem alguma estrutura dependente do tempo e não tem variância constante ao longo do tempo. Ela apresenta RU. (P-valor > α)
- H_a: A série temporal é estacionaria. (P-valor < α)
- Caso H_0 seja rejeitada e Podemos dizer que a série apresenta um padrão estacionário.
- Em outras palavras, caso H_a falhe em ser rejeitada teremos um padrão estacionário.

Resultados ADF

- Aparentemente o P-valor é mais baixo do que o nível de alpha.
- Conforme os resultados, podemos rejeitar H0?
- Desta maneira, o teste de ADF sugere que a série é estacionária?

- Augmented Dickey-Fuller Test

- data: AirPassengers
- Dickey-Fuller = -7.3186, Lag order = 5, **p-value = 0.01**
- alternative hypothesis: stationary

- Warning message:
- In adf.test(AirPassengers) : p-value smaller than printed p-value

Resultados ADF

- Em outras palavras, não podemos confiar totalmente nos resultados do teste de ADF!
- Isso ocorre porque apesar do P-valor ser menor do que alpha a série apresenta um padrão sazonal.
- De forma simplória: verificando pelo “olhonômetro” mesmo.
- Desta maneira, teremos que diferenciar!

- Augmented Dickey-Fuller Test

- data: AirPassengers
- Dickey-Fuller = -7.3186, Lag order = 5, **p-value = 0.01**
- alternative hypothesis: stationary

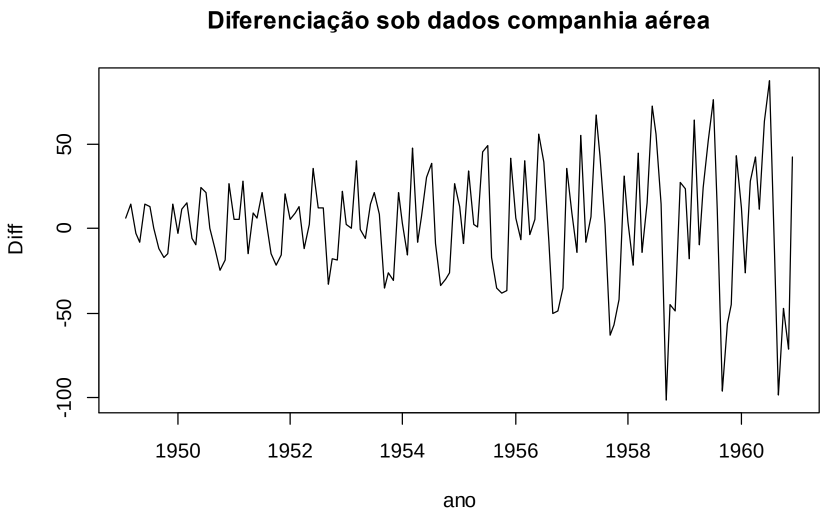
- Warning message:
- In adf.test(AirPassengers) : p-value smaller than printed p-value

Teremos que diferenciar

- Para remover as tendências aparentes na série temporal.
- Faremos isso para um melhor ajuste dos dados para nossa modelagem econométrica.

Após a diferenciação...

- Obtemos esse padrão.
- Podemos supor que agora seguimos um padrão estacionário?



ADF novamente

- Vamos utilizar o ADF na série diferenciada e verificar a estacionariedade.
- Agora podemos sugerir que a série é estacionária!
- A estatística Dickey-Fuller é -7,0177, que é um grande valor negativo. Isso indica que o teste **rejeita fortemente a hipótese nula** de uma raiz unitária (não estacionariedade).
- O valor p é relatado como 0,01, que é menor do que o nível de significância comum de 0,05. Isso significa que você pode rejeitar a hipótese nula de não estacionariedade, o que implica que a série diferenciada é estacionária.

- Augmented Dickey-Fuller Test

- data: diff_AirPassengers
- Dickey-Fuller = -7.0177, Lag order = 5, p-value = 0.01
- alternative hypothesis: stationary

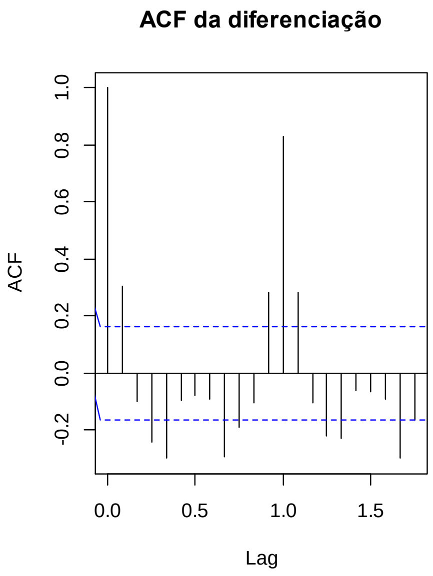
- Warning message:
- In adf.test(diff_AirPassengers) : p-value smaller than printed p-value

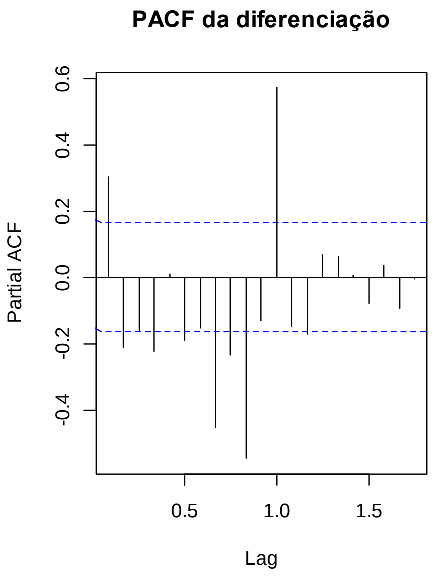
Agora Podemos aplicar o SARIMA...

- Vamos determinar os parâmetros do modelo SARIMA (p, d, q) e sazonais (P, D, Q)
- Após utilizaremos os gráficos da função de autocorrelação (ACF) e da função de autocorrelação parcial (PACF) para determinar os termos AR e MA.

ACF e PACF

- Os gráficos da Função de Autocorrelação (ACF) e da Função Parcial de Autocorrelação (PACF), demonstram uma forte correlação na defasagem 1.
- Nesse caso, podemos escolher a defasagem 1 para modelar.





SARIMA no R

- Utilizaremos a função auto.arima() do pacote forecast, que selecionará automaticamente o melhor modelo SARIMA com base nos critérios AIC/BIC.
- Akaike Information Criterion.
- Bayesian Information Criterion.

SARIMA no R

- ARIMA(2,1,1)(0,1,0)[12]:
- (2,1,1): Estes são os parâmetros **ARIMA não sazonais:**
- p=2: Dois termos autorregressivos (AR).
- d=1: A série foi diferenciada uma vez para remover a tendência (conforme você aplicou a diferenciação regular).
- q=1: Um termo de média móvel (MA).
- (0,1,0)[12]: Estes são os parâmetros **ARIMA sazonais:**
- P=0: Nenhum termo autorregressivo sazonal.
- D=1: A série foi diferenciada sazonalmente uma vez (para remover a sazonalidade com uma periodicidade de 12 meses).
- Q=0: Nenhum termo de média móvel sazonal.
- [12]: Indica um padrão sazonal anual com 12 meses como período de sazonalidade.

- Series: AirPassengers
- ARIMA(2,1,1)(0,1,0)[12]

- Coefficients:
- ar1 ar2 ma1
- 0.5960 0.2143 -0.9819
- s.e. 0.0888 0.0880 0.0292

SARIMA no R (Coeficientes)

- ar1 = 0,5960: Este é o coeficiente para o primeiro termo autorregressivo. Isso significa que 59,6% do valor de um ponto da série temporal depende do seu valor anterior (lag 1).
- ar2 = 0,2143: Este é o coeficiente para o segundo termo autorregressivo, indicando que 21,4% do valor atual é influenciado pelo valor de dois períodos atrás (lag 2).
- ma1 = -0,9819: Este é o coeficiente para o primeiro termo da média móvel, que representa o impacto do erro de previsão anterior. Um valor próximo a -1 sugere um forte efeito corretivo com base no erro anterior.
- Os erros padrão (s.e.) são relativamente pequenos, indicando que esses coeficientes são estimados com um bom grau de precisão.

- Series: AirPassengers
- ARIMA(2,1,1)(0,1,0)[12]

- Coefficients:
-         ar1   ar2   ma1
-       0.5960 0.2143 -0.9819
- s.e. 0.0888 0.0880   0.0292

- sigma^2 estimated as 132.3: log likelihood=-504.92
- AIC=1017.85   AICc=1018.17   BIC=1029.35

SARIMA no R (ajustes do modelo)

- sigma^2 estimado como 132,3: Esta é a variância estimada dos resíduos (erros). **Um valor menor de sigma^2 indica um melhor ajuste do modelo.**
- Log-verossimilhança = -504,92: Isto indica o quão bem o modelo se ajusta aos dados. Valores maiores (mais próximos de zero) são melhores, mas este valor é geralmente interpretado no contexto dos critérios de seleção do modelo.

sigma^2 estimated as 132.3: log likelihood=-504.92
AIC=1017.85   AICc=1018.17   BIC=1029.35

```
Training set error measures:
                 ME     RMSE     MAE      MPE     MAPE     MASE        ACF1
Training set 1.3423 10.84619 7.86754 0.420698 2.800458 0.245628 -0.00124847
>
```

SARIMA no R (ajustes do modelo)

- AIC = 1017,85 e BIC = 1029,35:
- AIC (Critério de Informação de Akaike) e BIC (Critério de Informação Bayesiano) são usados para comparação de modelos. **Valores menores indicam um melhor ajuste**, mas o BIC penaliza modelos mais complexos (mais parâmetros) mais pesadamente do que o AIC. Na prática, estes valores são comparados com aqueles de outros modelos para decidir qual modelo é mais adequado aos dados.
- AICc = 1018,17: Este é o AIC corrigido para tamanhos de amostra pequenos. Neste caso, é muito próximo do AIC.

sigma^2 estimated as 132.3: log likelihood=-504.92

AIC=1017.85   AICc=1018.17   BIC=1029.35

```
Training set error measures:
                   ME     RMSE     MAE      MPE     MAPE     MASE        ACF1
Training set 1.3423 10.84619 7.86754 0.420698 2.800458 0.245628 -0.00124847
>
```

SARIMA no R (ajustes do modelo)

- ME (Erro Médio): 1,3423. Este é o erro médio de previsão, e um **valor próximo a 0 é desejável**.
- RMSE (Erro Quadrático Médio): 10,84619. Esta é uma medida comumente usada de precisão de previsão, com **valores mais baixos indicando melhor desempenho**.
- MAE (Erro Médio Absoluto): 7,86754. Isto representa a **magnitude** média dos erros nas previsões, sem considerar sua direção (positiva ou negativa).
- MPE (Erro Médio Percentual): 0,420698. Este é o erro percentual médio nas previsões.

sigma^2 estimated as 132.3: log likelihood=-504.92
AIC=1017.85   AICc=1018.17   BIC=1029.35

```
Training set error measures:
                 ME     RMSE     MAE      MPE     MAPE     MASE        ACF1
Training set 1.3423 10.84619 7.86754 0.420698 2.800458 0.245628 -0.00124847
>
```

SARIMA no R (ajustes do modelo)

- MAPE (Erro Médio Percentual Absoluto): 2,800458. Isto mede a precisão como uma porcentagem, com **valores mais baixos indicando melhor desempenho**.
- MASE (Erro Médio Absoluto Escalado): 0,245628. Esta é uma métrica de erro relativo que compara o MAE do seu modelo com um modelo ingênuo. **Valores menores que 1 indicam que seu modelo é melhor que um modelo ingênuo.**
- ACF1 (First Lag of Residuals ACF): -0,00124847. Esta é a primeira autocorrelação dos resíduos. **Um valor próximo a 0 indica que não há autocorrelação significativa restante nos resíduos**, o que significa que o modelo capturou a maioria dos padrões nos dados.

sigma^2 estimated as 132.3: log likelihood=-504.92

AIC=1017.85 AICc=1018.17 BIC=1029.35

```
Training set error measures:
                 ME     RMSE     MAE      MPE     MAPE     MASE        ACF1
Training set 1.3423 10.84619 7.86754 0.420698 2.800458 0.245628 -0.00124847
>
```

Interpretação

- O modelo inclui dois termos autorregressivos e um termo de média móvel, o que indica que o valor atual da série temporal é explicado por seus dois valores passados e pelo erro de previsão passado.
- A diferenciação sazonal (D=1) foi responsável pela sazonalidade com sucesso, pois não há termos AR ou MA sazonais adicionais necessários (P=0 e Q=0).
- Os erros padrão relativamente pequenos para os coeficientes sugerem que os parâmetros são bem estimados.

Interpretação

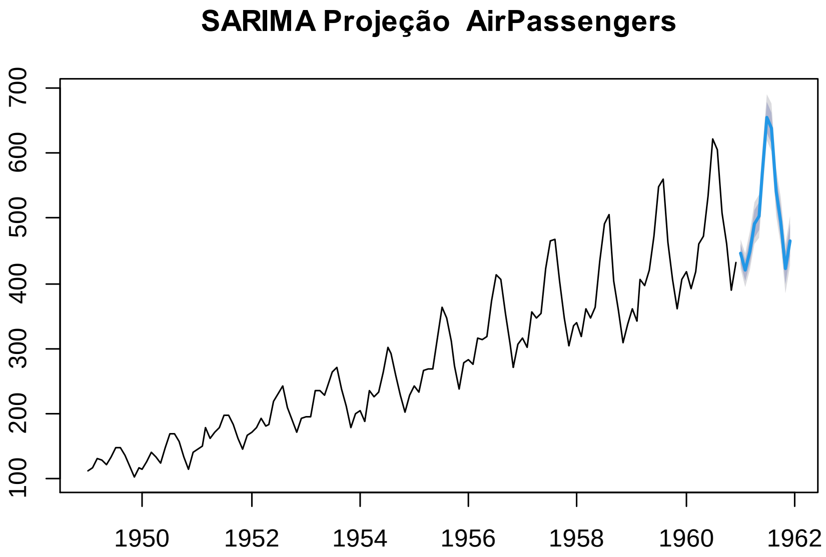
- O diagnóstico residual (ACF1) mostra que os resíduos não são autocorrelacionados, sugerindo que o modelo se ajusta bem aos dados.
- Medidas de erro como RMSE, MAE e MAPE fornecem uma maneira de avaliar a precisão do modelo. O MAPE de 2,8% indica que, em média, as previsões do modelo desviam dos valores reais em cerca de 2,8%, o que é muito bom.

Interpretação

- Este modelo SARIMA se ajusta razoavelmente bem aos dados do AirPassengers, capturando tanto a tendência (por meio de diferenciação regular) quanto a sazonalidade (por meio de diferenciação sazonal). A precisão do modelo (conforme visto por meio do MAPE e do RMSE) sugere que ele pode ser usado para previsões, e a falta de autocorrelação em resíduos implica que o modelo capturou adequadamente a estrutura subjacente da série temporal.
- Você pode usar este modelo para prever valores futuros de passageiros de companhias aéreas ou compará-lo com modelos alternativos (como diferentes configurações ARIMA/SARIMA) para ver se um ajuste melhor pode ser obtido.

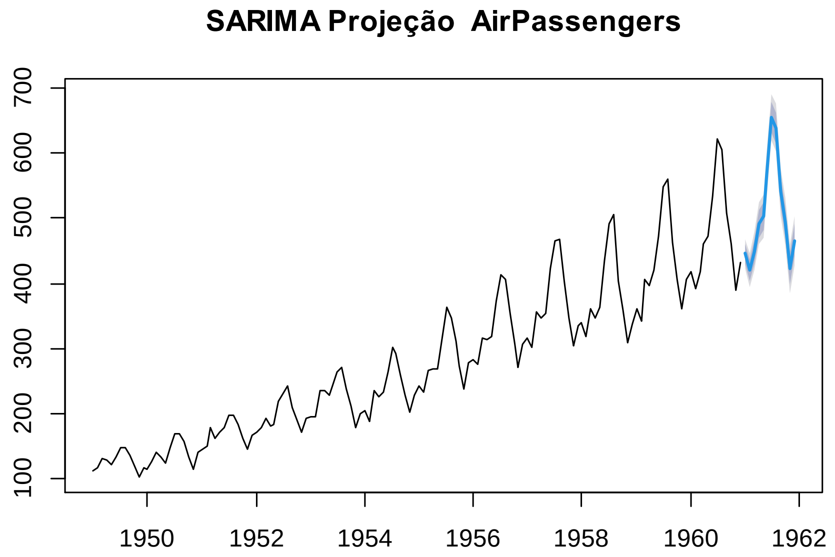
Projeções

- A função forecast() em R é usada para gerar previsões futuras com base em um modelo de série temporal ajustado (como SARIMA). A função funciona alavancando os parâmetros do modelo (como AR, MA, componentes sazonais) e dados passados para projetar valores futuros da série temporal.



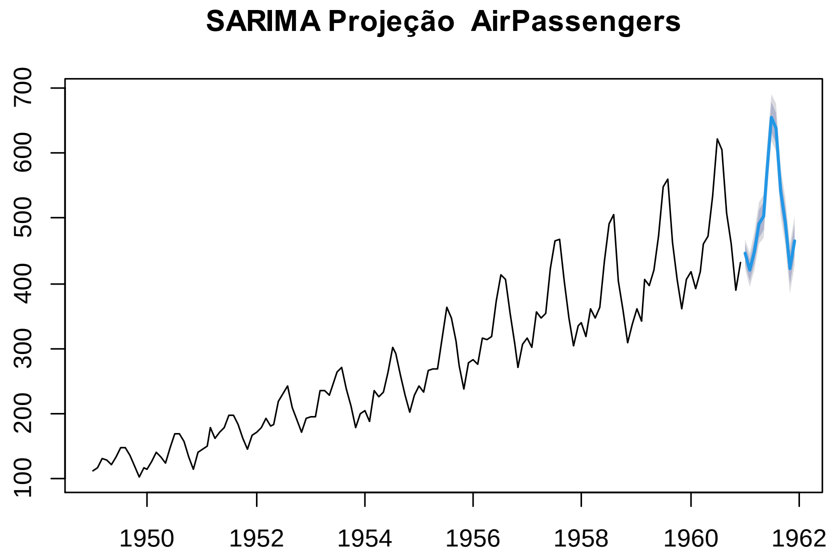
Projeções

- O valor médio representa os valores previstos da série temporal para períodos futuros (por exemplo, os próximos 12 meses para o conjunto de dados AirPassengers). As previsões são feitas iterando as equações do modelo para frente, usando os valores observados no passado e os coeficientes ajustados.
- Para SARIMA, isso envolve prever os componentes sazonais e não sazonais.

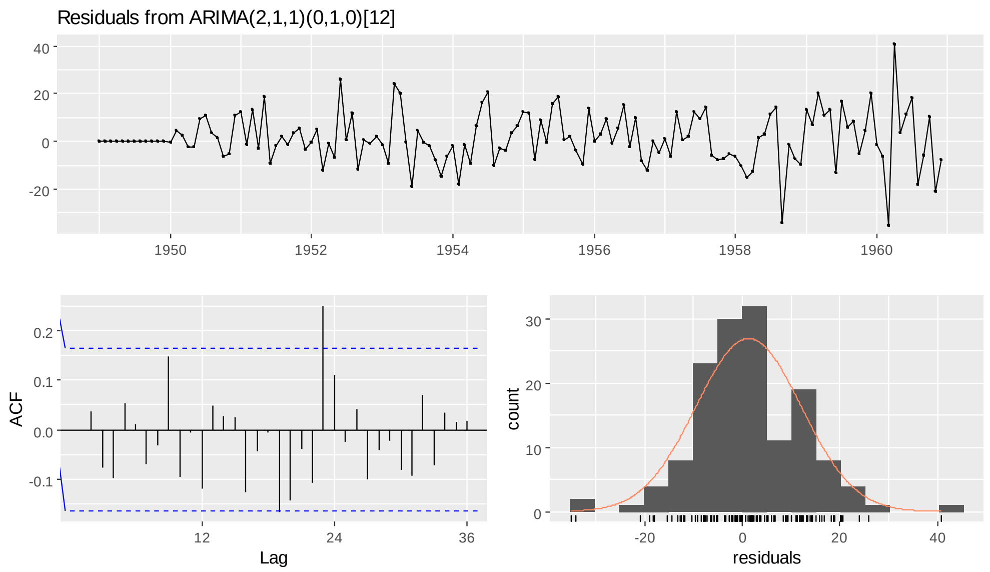


Projeções

- Os valores inferior e superior representam os intervalos de confiança de 80% e 95% para cada valor previsto. Esses intervalos refletem a incerteza nas previsões.
- Por exemplo, é possível que previsão média para um período futuro de 1962, o intervalo de confiança de 80% pode ser [611, 647], o que significa que há 80% de chance de que o valor verdadeiro caia dentro desse intervalo.



Diagnostico do Modelo



- Agora vamos verificar se os resíduos se ajustam corretamente ao modelo.
- p-value = 0,8998: Isso é muito maior do que os níveis de significância típicos (por exemplo, 0,05 ou 0,01). Portanto, falhamos em rejeitar a hipótese nula.
- Isso sugere que não há evidência significativa de autocorrelação nos resíduos, o que significa que os resíduos provavelmente não são correlacionados e seguem um processo de ruído branco.

Box.test(residuals(sarima_model), lag = 2, type = "Ljung-Box")

Box-Ljung test

data: residuals(sarima_model)
X-squared = 0.21113, df = 2, p-value = 0.8998

Questões para revisão

- Como é construído o modelo SARIMA?
- Quais são as suposições fundamentais do modelo SARIMA?
- Devemos diferenciar uma série sazonal? Porquê ?
- No caso da nossa análise de SARIMA, quais estratégias ou políticas comerciais ou de investimento se você fosse o CEO de uma empresa dessa indústria?

Referências

- - Box, G.E.P. and Jenkins, G.M., 1970. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- Box, G.E.P. and Pierce, D.A., 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332), pp.1509-1526.
- Dickey, D.A. and Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), pp.427-431.

- - Dickey, D.A. and Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), pp.427-431.

- - Ljung, G.M. and Box, G.E.P., 1978. On a measure of lack of fit in time series models. *Biometrika*, 65(2), pp.297-303.