

---

# Análise de Séries Temporais 2016/1

## Relatório 1: Modelagem ARIMA

---

*Alunos:*

Ivani IVANOVA e  
Jonathas FERREIRA

*Professora:*

Alexandra M. SCHMIDT

Universidade Federal do Rio de Janeiro

Julho de 2016

# 1 Introdução

A variação temperatura da superfície terrestre ao longo do tempo tem-se tornado um tema de importância central em estudos meteorológicos e estatísticos, uma vez que verificamos alterações substanciais no último século. Seja para prever o comportamento futuro da temperatura ou identificar fatores correlacionados, este é um assunto de destaque.

Neste trabalho, buscamos analisar a série temporal de anomalias globais da temperatura do ar superficial entre os anos de 1880 e 1985 que foi obtida por J. Hansen e S. Lebedeff (vide Bibliografia [2]). Usaremos ferramentas de especificação de parâmetros e estimação de coeficientes para ajustar um modelo  $ARIMA(p,d,q)$  satisfatório. Por fim, faremos a previsão da série para anos seguintes ao seu término.

## 2 Considerações Meteorológicas

J. Hansen e S. Lebedeff construíram a série temporal de anomalias globais da temperatura do ar superficial que investigaremos a seguir. Essa série é referente a medições de temperatura do ar superficial, isto é, da temperatura ambiente mostrada por um termômetro exposto ao ar, mas protegido da radiação solar direta. Tais temperaturas são medidas em estações meteorológicas localizadas ao redor do globo e que até meados do século passado, estavam concentradas em áreas terrestres do hemisfério norte. Portanto, essa análise de temperaturas globais teve um grande problema que é a cobertura incompleta das estações existentes, principalmente para dados antigos. Em 1900, ainda existiam grandes áreas da África e da América do Sul que não eram alcançadas pelas estações e medições substanciais na Antártida começaram por volta de 1950. Até o trabalho dos autores, grandes áreas oceânicas ainda não tinham estações fixas.

O termo "anomalia" que empregamos é devido ao fato que as temperaturas são comparadas a um valor padrão ou média a longo prazo, porém o eixo  $y = 0$  pode ser uma escolha arbitrária também, como é o caso dos dados que analisaremos. O que, de fato, concluímos ao examinar a anomalia de temperatura é se a temperatura aumenta ou diminui num ano em relação a outro.

O método usado para medir a temperatura consiste em dividir a superfície terrestre em 80 regiões de igual área (na projeção cilíndrica de Miller essas regiões são retângulos), que ainda são subdivididas. Após isso, em cada região, usa-se um método iterativo para fazer a média das temperaturas registradas por cada estação que busca utilizar toda a informação disponível, visto que cada estação teve seu próprio período de funcionamento (que, geralmente, não corresponde a 1880-1985). Depois disso, faz-se uma média ponderada das médias nas regiões, com pesos proporcionais às regiões, exceto as que não contêm estações meteorológicas, as quais não são usadas.

### 3 O Problema

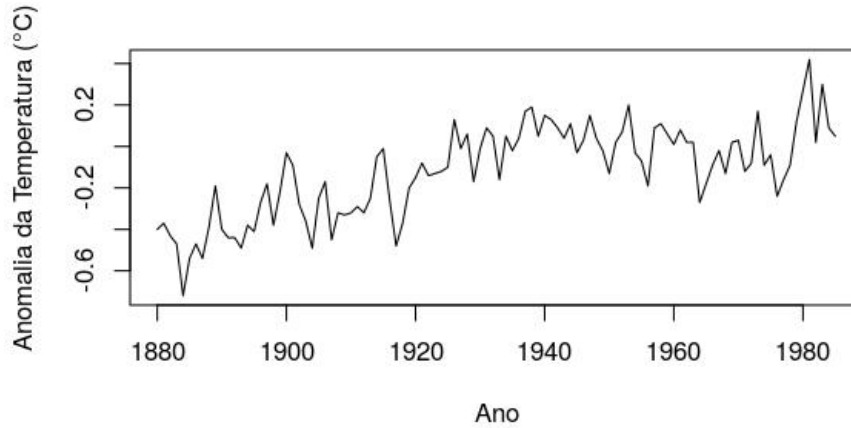
Queremos modelar a série antes mencionada, que apresentamos no Gráfico 1, com um modelo  $ARIMA(p,d,q)$ . Ou seja, procuramos um modelo que explique o comportamento observado da série. Seja  $\{Y_t\}_{t \in \{1880, \dots, 1985\}}$  a série de anomalias globais da temperatura (medidas em  $^{\circ}C$ ), o modelo que atribuímos a  $Y_t$  é dado por

$$\nabla^d = \phi_1 \nabla^d Y_{t-1} + \dots + \phi_p \nabla^d Y_{t-p} + e_t - \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} \quad (1)$$

onde  $\nabla X_t = X_t - X_{t-1}$  é o operador diferença e  $e_t$  é uma série de ruído branco, ou seja, uma sequência independente e identicamente distribuída de média 0 e variância  $\sigma_e^2$ . A seguir, buscaremos coeficientes  $p$ ,  $d$  e  $q$  possíveis de terem gerado uma série como  $Y_t$  e estimaremos os coeficientes  $\phi_i$ ,  $\theta_j$  e  $\sigma_e^2$ .

---

**Gráfico 1:** Série temporal  $Y_t$  da anomalia da temperatura global

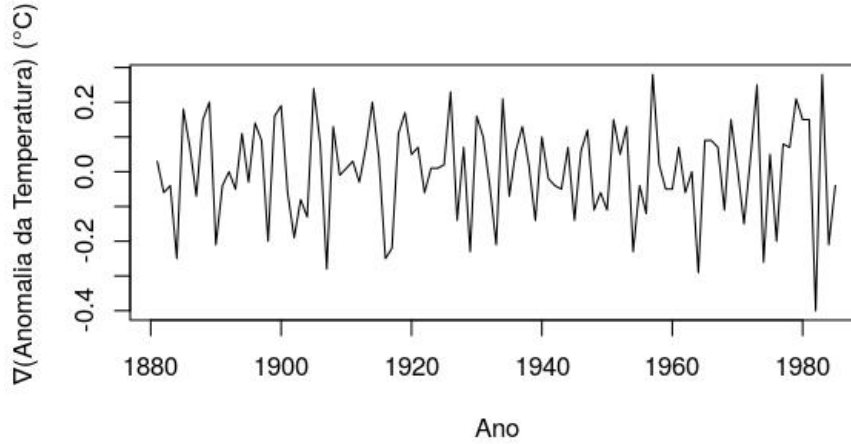


## 4 Modelagem

Como o Gráfico 1 evidencia, a série temporal  $Y_t$  das anomalias na temperatura apresenta um comportamento não estacionário, observamos uma tendência ascendente linear, que devemos remover. Para isso, vamos verificar se a série da primeira diferença de  $Y_t$ ,  $\nabla Y_t = Y_t - Y_{t-1}$  possui propriedades melhores. Um plot da série  $\nabla Y_t$  é mostrada na Gráfico 2.

---

**Gráfico 2:** Série temporal  $\nabla Y_t$  da primeira diferença da anomalia da temperatura global

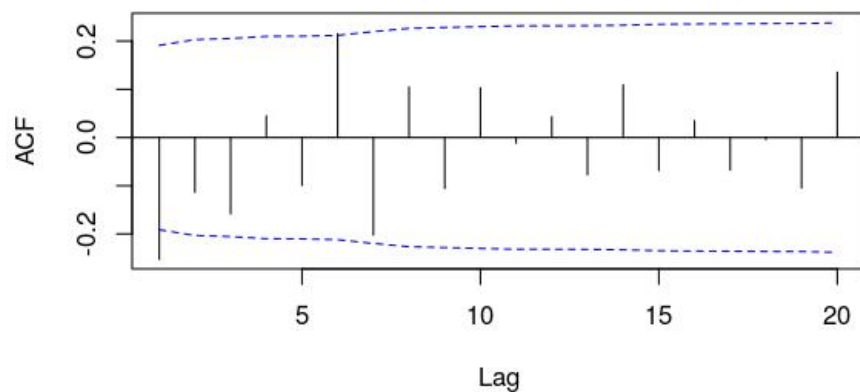


Esta série apresenta, aparentemente, um comportamento suficientemente estacionário e nos basearemos nela para a modelagem a seguir. Como queremos ajustar um modelo  $ARIMA(p,d,q)$  para  $Y_t$  e, portanto, um modelo  $ARIMA(p,d-1,q)$  para  $\nabla Y_t$ , vamos analisar a função de autocorrelação amostral e a função de autocorrelação parcial amostral de  $\nabla Y_t$ , cujas informações encontram-se resumidas nos Gráficos 3 e 4.

Notamos que a função de autocorrelação amostral apresenta apenas um coeficiente significativo, o coeficiente de lag igual a 1, e este muito próximo dos limites de significância. Não há um decaimento linear, com vários coeficientes significativos, o que confirma que uma diferença foi suficiente para tratar da não estacionariedade de  $Y_t$ . Por outro lado, a função de autocorrelação parcial amostral tem coeficientes significativos nos lags 1,3,5 e 7.

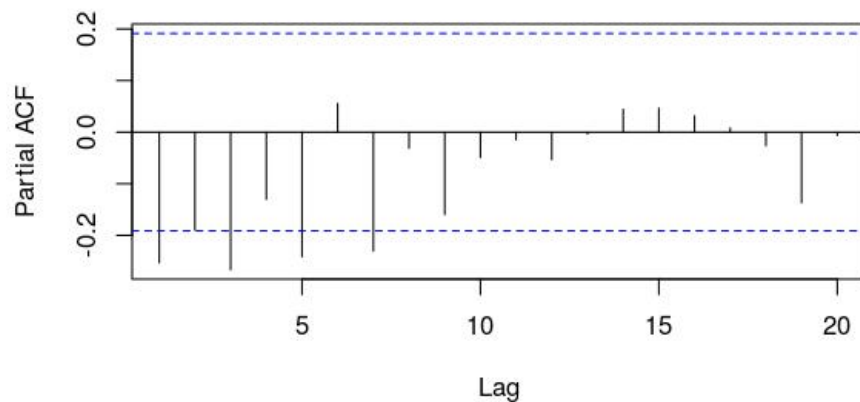
---

**Gráfico 3:** Função de autocorrelação amostral de  $\nabla Y_t$



---

**Gráfico 4:** Função de autocorrelação parcial amostral de  $\nabla Y_t$



Relembremos que um modelo média móvel puro  $MA(q)$  possui autocorrelação nula após o lag  $q$  e autocorrelação parcial que decai na medida que o lag cresce. Dado isso, nossa análise inicial sugere um modelo  $ARIMA(0,1,1)$  para  $Y_t$ . Vamos, entretanto, investigar se a função de autocorrelação estendida amostral, apresentada na Tabela 1, nos fornece alguma outra informação.

---

**Tabela 1:** Função de autocorrelação estendida amostral de  $\nabla Y_t$

```
> eacf(diff(tempts))
AR/MA
0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x o o o o x x o o o o o o
1 x o o o o o o o o o o o o
2 x x o o o o o o o o o o o
3 x x o o o o o o o o o o o
4 x x o o o o o o o o o o o
5 x x x x x o o o o o o o o
6 x x x o x o o o o o o o o
7 o x o o x o o o o o o o o
```

Sabemos que a função de autocorrelação teórica para um modelo  $\text{ARMA}(p,q)$ , apresenta um padrão triangular de zeros com vértice na posição  $(p,q)$ . Portanto, para  $\nabla Y_t$  a função de autocorrelação estendida amostral sugere um modelo misto com  $p = 1$  e  $q = 1$  ou, vagamente,  $p = 0$  e  $q = 1$ . Vamos examinar estes dois possíveis modelos a seguir.

#### 4.1 Modelo 1: ARIMA(0,1,1)

Por parcimônia, começamos pelo modelo mais simples que poderia ser ajustado a nossa série:

$$\nabla Y_t = e_t - \theta_1 e_{t-1} \quad (2)$$

Da Tabela 2, vemos que estimamos, por máxima verossimilhança (método que usaremos até o final do relatório),  $\theta_1$  igual a -0.614 com desvio padrão 0.1348, portanto, é significativo. Por outro lado, o Gráfico 5 nos mostra três ferramentas de diagnóstico que apontam para problemas do modelo. O gráfico dos resíduos padronizados não mostra uma correlação evidente, entretanto, a autocorrelação amostral indica que o coeficiente de lag 3 é significativo, mesmo que pouco. Isto é reforçado pelos valores-p do teste de Ljung-Box, que rejeitam a hipótese nula de que os termos de erro são descorrelacionados para vários valores. Assim sendo, concluímos que este modelo não explica bem as variações da série.

---

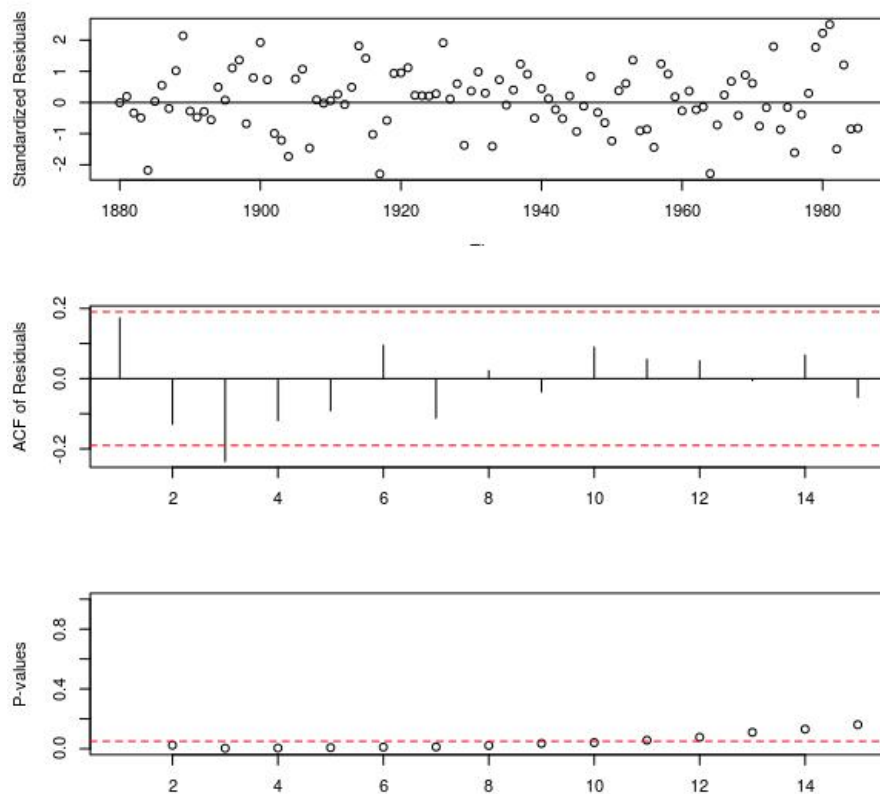
**Tabela 2:** Resumo do ajuste do modelo ARIMA(0,1,1) para  $Y_t$

Coefficients :  
           mal  
           -0.6140  
 s.e.      0.1348

$\sigma^2$  estimated as 0.01743: log likelihood = 63.39,  
 aic = -122.78

---

**Gráfico 5:** Resíduos padronizados, função de autocorrelação amostral e valores-p do teste de Ljung-Box referentes ao Modelo 1



## 4.2 Modelo 2: ARIMA(1,1,1)

A seguir, ajustamos um modelo ARIMA(1,1,1) para  $Y_t$ , ou seja,

$$\nabla Y_t = \phi_1 \nabla Y_{t-1} + e_t - \theta_1 e_{t-1} \quad (3)$$

O resumo do ajuste, encontrado na Tabela 3, mostra que temos  $\hat{\phi}_1 = 0.3737$  e  $\hat{\theta}_1 = -0.8348$  sendo ambos os coeficientes significativos. No Gráfico 6, não observamos correlação entre os resíduos do modelo, o que é confirmado pelos valores da autocorrelação amostral. Além disso, pelos valores-p obtidos no teste de Ljung-Box, não podemos rejeitar a hipótese nula de os erros serem descorrelacionados. Sob esses fatores, o modelo parece razoável, portanto vamos investigar melhor o comportamento dos resíduos.

Aplicando o teste Runs aos resíduos do modelo, obtemos um valor-p de 0.849, logo não podemos rejeitar independência dos resíduos. Nos Gráficos 7 e 8, observamos que o Q-Q plot é aproximadamente uma linha reta e o histograma se assemelha a uma gaussiana, o que apoia a hipótese de normalidade dos resíduos. Além disso, o teste de Shapiro-Wilk de normalidade nos retorna uma estatística  $W = 0.9921$  e um valor-p de 0.798, portanto não rejeitamos a hipótese de que os erros são normalmente distribuídos. Desse modo, verificamos que este modelo é bastante satisfatório.

---

**Tabela 3:** Resumo do ajuste do modelo ARIMA(1,1,1) para  $Y_t$

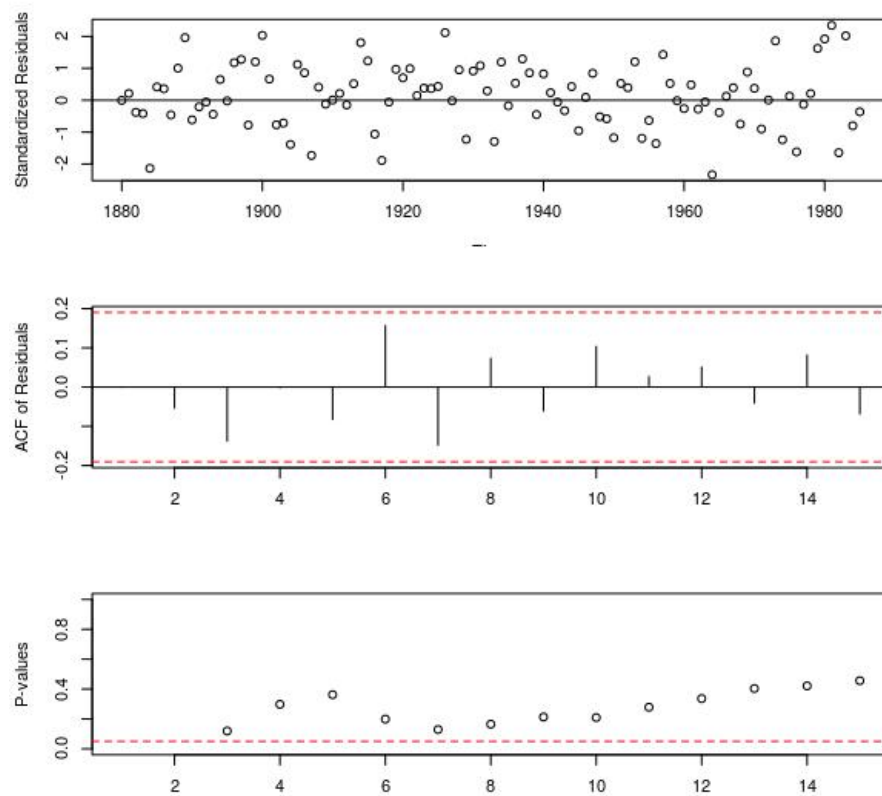
Coefficients :

	ar1	ma1
	0.3737	-0.8348
s.e.	0.1207	0.0643

sigma^2 estimated as 0.01604: log likelihood = 67.68,  
aic = -129.36

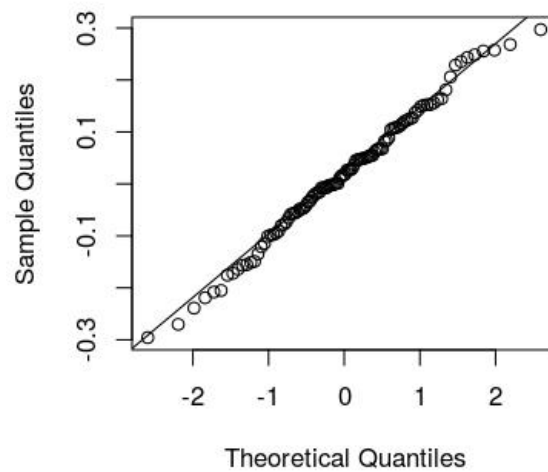


**Gráfico 6:** Resíduos padronizados, função de autocorrelação amostral e valores-p do teste de Ljung-Box referentes ao Modelo 2



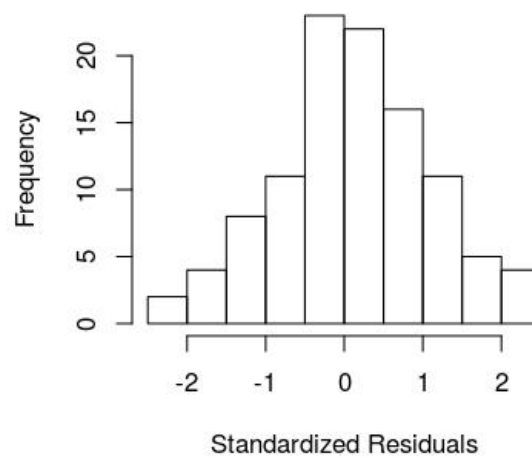
---

**Gráfico 7:** Q-Q Plot dos resíduos padronizados do Modelo 2



---

**Gráfico 8:** Histograma dos resíduos padronizados do Modelo 2



Dado isso, constatamos que este modelo é bastante adequado. Nesse caso, o modelo estimado é

$$\nabla Y_t = 0.3737\nabla Y_{t-1} + e_t + 0.8348e_{t-1} \quad (4)$$

ou, equivalentemente,

$$Y_t = 1.3737Y_{t-1} - 0.3737Y_{t-2} + e_t + 0.8348e_{t-1} \quad (5)$$

com variância do ruído estimada  $\hat{\sigma}_e = 0.01604$ .

### 4.3 Modelos com mais coeficientes

A seguir, nas Tabelas 4 e 5, mostramos o ajuste de modelos ARIMA(2,1,1) e ARIMA(1,1,2), respectivamente, para  $Y_t$ . Verificamos que os coeficientes adicionais, se comparados ao modelo 2 não são significativos e as mudanças nas estimativas dos coeficientes presentes no modelo ARIMA(1,1,1) também não o são. A verossimilhança praticamente não se altera e o AIC é ligeiramente maior que no modelo 2. Como buscamos um modelo parcimonioso, optamos pelo ajuste do modelo 2.

---

**Tabela 4:** Resumo do ajuste do modelo ARIMA(2,1,1) para  $Y_t$

```
Coefficients :
      ar1      ar2      ma1
      0.3692  -0.0849  -0.8063
s.e.    0.1232   0.1088   0.0803
```

```
sigma^2 estimated as 0.01594:  log likelihood = 67.98,
aic = -127.96
```

---

**Tabela 5:** Resumo do ajuste do modelo ARIMA(1,1,2) para  $Y_t$

```
Coefficients :
      ar1      ma1      ma2
      0.2460  -0.6912  -0.1076
s.e.    0.2577   0.2511   0.1775
```

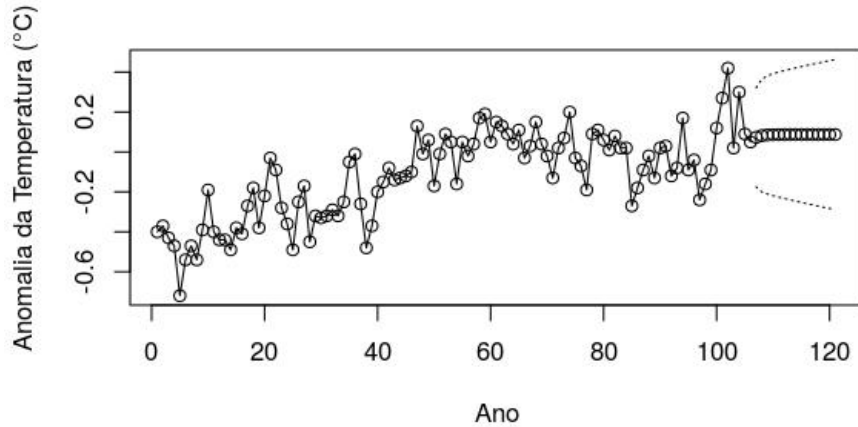
```
sigma^2 estimated as 0.01598:  log likelihood = 67.85,
aic = -127.71
```

## 5 Previsão

No Gráfico 9, mostramos a previsão de 15 passos a frente da série de anomalia da temperatura global com limites de previsão de 95% de confiança. Apesar de rapidamente as estimativas convergirem para um valor constante de aproximadamente 0.0873 (isso se deve ao fato de previsões de modelos  $ARMA(p,q)$  estacionários convergirem para a média em poucos passos), os limites de confiança nos dão uma estimativa de onde esperamos que os próximos valores sejam encontrados.

---

**Gráfico 9:** Previsão da série  $Y_t$



## 6 Conclusão

Neste trabalho, nos propomos usar técnicas de análise de séries temporais para modelar a mudança na temperatura global com dados de 1880 a 1985 e fazer uma previsão a curto prazo da série. Usando ferramentas de investigação disponibilizadas no R, ajustamos um modelo  $ARIMA(1,1,1)$  para a anomalia da temperatura global. Apesar de as previsões desse modelo serem relativamente vagas, obtemos um intervalo de confiança para dados futuros que pode ser útil. Um modelo diferente ou mais geral pode ser uma solução para contornar esses defeitos.

## 7 Bibliografia

[1] - Cryer, J. D. & Chan, K. S. - *Time Series Analysis With Applications in R*, Springer, 2nd edition, 2010.

[2] - Hansen, J. & Lebedeff, S. - *Global Trends of Measured Surface Air Temperature*, Journal of Geophysical Research, Vol. 92, NO. D11, pages 13,345-13,372, 1987.

[3] - <https://datamarket.com/data/list/?q=provider:tsdl>