

INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA

Departamento de Matemática

**Relatório para a Unidade Curricular de Introdução à Estatística Bayesiana dos
cursos de LMATE e LEIC**

População e Emprego

Docente - Paula Cristina Pires Simões

- Marta Pires N^o51726 - Daiana Lapaiescu N^o48668 - Eduardo Tavares N^o 49454
- António Pimentel N^o42146

2024-12-12

Introdução

O desemprego constitui um dos indicadores mais importantes da situação económica e social de um país, refletindo de forma direta o dinamismo do mercado de trabalho, a capacidade produtiva da economia e o bem-estar das famílias. Em Portugal, entre 2011 e 2024, este indicador apresentou variações significativas associadas à crise económica e financeira, ao período de recuperação que se seguiu e ao impacto da pandemia de COVID-19. Estas oscilações tornam o estudo do desemprego particularmente relevante e atual.

A escolha deste tema resulta precisamente da sua importância para compreender como fatores económicos e demográficos influenciam a evolução do mercado de trabalho. Variáveis como o PIB real, a estrutura etária da população ativa, a remuneração média e o peso do setor terciário desempenham papéis centrais na exploração das dinâmicas laborais e ajudam a contextualizar períodos de maior estabilidade ou vulnerabilidade económica.

Assim, este trabalho tem como objetivo analisar a evolução da taxa de desemprego em Portugal ao longo dos últimos anos e avaliar de que forma estes fatores contribuem para a sua variação. O estudo insere-se no âmbito da Unidade Curricular de Introdução à Estatística Bayesiana, aplicando os métodos aprendidos à investigação de um fenómeno real com impacto económico e social significativo.

#nao sei se por O relatório encontra-se estruturado em várias secções, incluindo o enquadramento teórico, a descrição e análise exploratória dos dados, a construção e estimativa do modelo de regressão bayesiano, a apresentação e discussão dos resultados e, por fim, as principais conclusões do estudo

Leitura de Dados

Para começar, os dados são importados:

```
# Leitura dos Dados
library(readxl)
dados <- read_excel("Base_Dados_Final.xlsx")

#Inverter a Ordem dos dados do menor para o maior
dados <- dados %>%
  arrange(desc(row_number())) %>% # inverte a ordem das linhas
  mutate(Trimestre = row_number()) # cria índice sequencial do mais antigo para o mais recente

# Visualizar primeiras linhas
head(dados)

# Verificar se existem valores omissos
na.fail(dados) #se correr é porque não há valores omissos
```

Descrição dos Dados

Variável dependente - Taxa de desemprego

A variável dependente utilizada neste estudo é a **taxa de desemprego** em Portugal, medida trimestralmente pelo INE. A escolha desta variável justifica-se por ser um dos indicadores centrais do desempenho do mercado de trabalho e um reflexo direto das condições económicas e sociais do país.

A taxa de desemprego responde rapidamente a alterações na atividade económica, à evolução dos diferentes setores produtivos, às características da população ativa e aos choques económicos, como recessões ou crises sanitárias. Por estas razões, constitui uma medida adequada para analisar de que forma os principais fatores macroeconómicos e demográficos influenciam a procura de trabalho em Portugal.

Variáveis independentes:

A seleção das variáveis independentes teve como objetivo incluir fatores que captam tanto a conjuntura económica do país como características estruturais do mercado de trabalho. Cada variável foi escolhida por representar um determinante plausível e frequentemente referido na literatura económica como influenciador do desemprego.

- **PIB real (em milhões de euros, por trimestre):** O Produto Interno Bruto real é um indicador fundamental da atividade económica. Períodos de crescimento do PIB tendem a estar associados à criação de emprego, enquanto fases de contração económica aumentam o risco de desemprego. Assim, o PIB real foi incluído para captar o efeito da conjuntura macroeconómica sobre o mercado de trabalho ao longo do tempo.
- **Remuneração média mensal bruta:** A remuneração média reflete simultaneamente o rendimento dos trabalhadores e os custos laborais enfrentados pelas empresas. Salários mais elevados podem indicar um mercado de trabalho mais dinâmico e competitivo, mas também podem representar maiores custos de contratação. A inclusão desta variável permite captar a forma como estas dinâmicas influenciam o nível de emprego e, consequentemente, o desemprego.
- **Percentagem de emprego no setor terciário:** O setor terciário representa a maior parte da estrutura económica portuguesa, abrangendo atividades como comércio, turismo e serviços. Alterações no peso relativo deste setor podem influenciar o desemprego devido à sua maior sensibilidade a flutuações económicas e choques externos. Esta variável permite captar possíveis mudanças estruturais na distribuição do emprego entre setores e o seu impacto no mercado de trabalho.
- **Proporção de jovens na população ativa:** A presença de trabalhadores mais jovens na força de trabalho é relevante para a dinâmica do desemprego, uma vez que este grupo apresenta maior rotatividade, maior taxa de entrada no mercado de trabalho e maior vulnerabilidade a ciclos económicos.

Assim, a proporção de jovens (15–24 anos) na população ativa foi incluída para medir o impacto da composição etária da força laboral sobre a evolução do desemprego.

Análise Exploratória dos Dados

Antes da construção do modelo estatístico, é essencial realizar uma análise exploratória dos dados, para compreender o comportamento das variáveis ao longo do tempo, identificar padrões e tendências, e avaliar a consistência das séries, detetando valores atípicos ou comportamentos inesperados.

Nesta secção, a análise exploratória foca-se na evolução temporal da taxa de desemprego e das variáveis explicativas selecionadas, examinando tendências gerais, picos ou quebras associados a instabilidades económicas, bem como relações simples entre variáveis, através de correlações e gráficos, fornecendo uma primeira visão das associações que serão estudadas formalmente no modelo bayesiano.

```
summary(dados %>% select(-Período, -Trimestre))
```

Séries temporais de Cada Variável

```
## Gráficos de linha ao longo do tempo (Série Temporal)
library(ggplot2)
```

Evolução Temporal das Variáveis Independentes Estas séries permitem observar como cada indicador evoluiu ao longo do período analisado e ajudam a compreender o seu possível impacto na taxa de desemprego.

Variável “PIB_real”:

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

O PIB real apresenta três fases bem definidas: - Declínio inicial (trimestres 1–8), coincidente com a crise económica e o período de austeridade. - Recuperação gradual e sustentada até 2019, com crescimento contínuo. - Queda abrupta em 2020, resultante do impacto da pandemia COVID-19, seguida de uma forte recuperação nos trimestres posteriores. Esta série mostra claramente a sensibilidade da economia portuguesa aos ciclos económicos e aos choques externos.

Variável “Proporção_15_24”: A proporção de jovens na força de trabalho apresenta uma tendência maioritariamente decrescente: - Os valores iniciais situam-se perto dos 9%. - Ao longo do período, observa-se uma redução progressiva até níveis próximos de 6%. Este padrão é consistente com fenómenos demográficos como o envelhecimento populacional, menor natalidade e migração jovem para o exterior. A ligeira recuperação final sugere um aumento recente da participação jovem no mercado de trabalho.

Variável “Emprego_Terciário”:

A proporção de trabalhadores no setor terciário mostra uma tendência claramente crescente ao longo dos trimestres. - No início do período, situava-se ligeiramente acima de 67%. - À medida que a economia portuguesa se foi tornando mais terciarizada, este valor aumentou de forma quase contínua até ultrapassar os 72%.

Este comportamento reflete a crescente importância dos serviços — como turismo, comércio e atividades especializadas — na estrutura produtiva portuguesa.

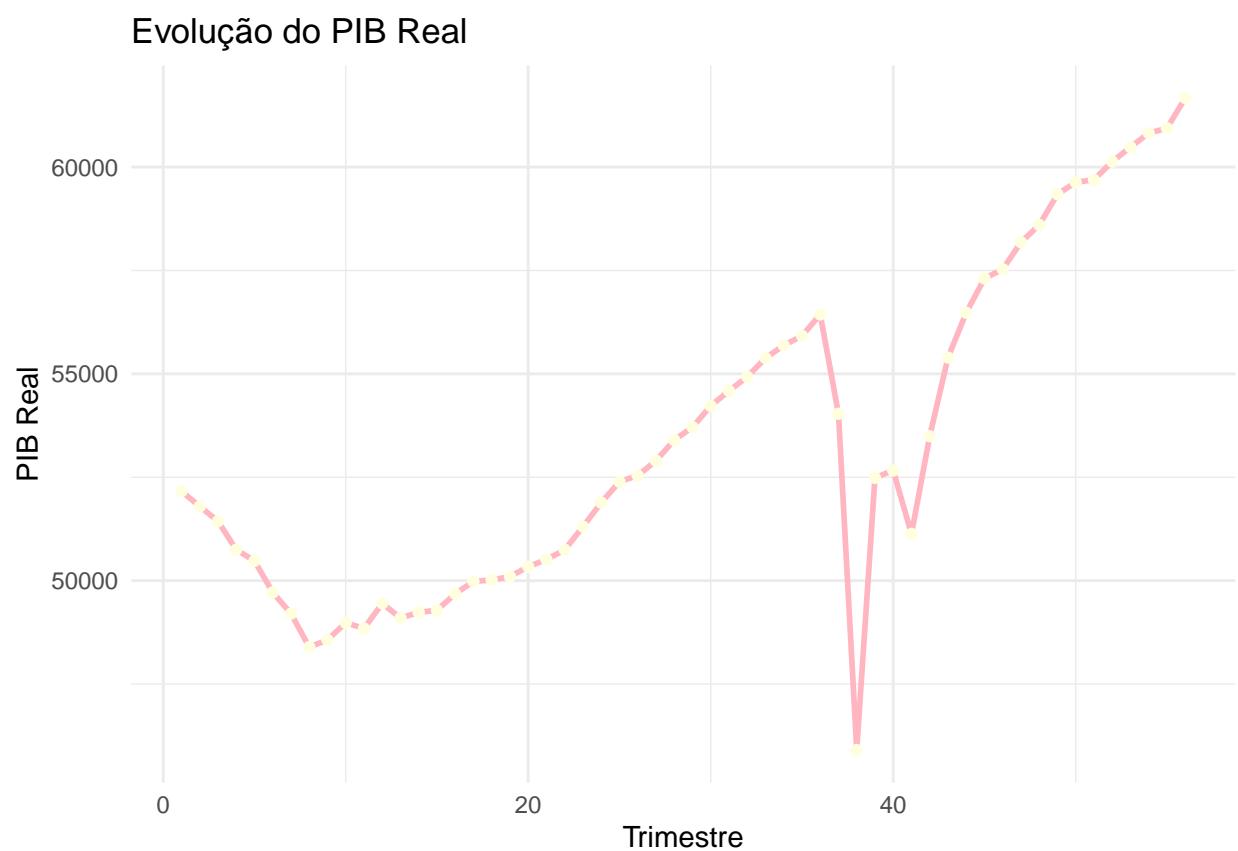


Figure 1: Série Temporal do PIB Real

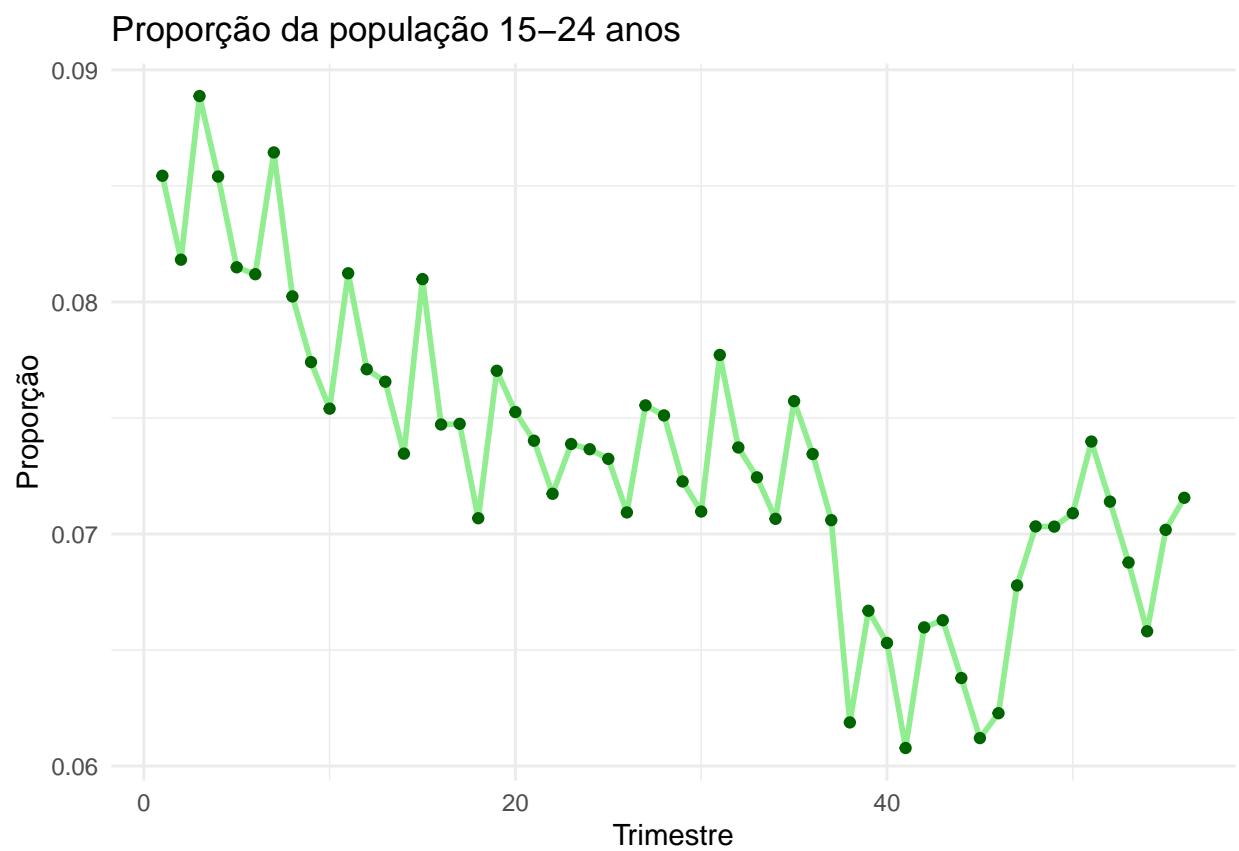


Figure 2: Série Temporal da Proporção da População entre 15-24 Anos

Emprego no setor terciário (%)

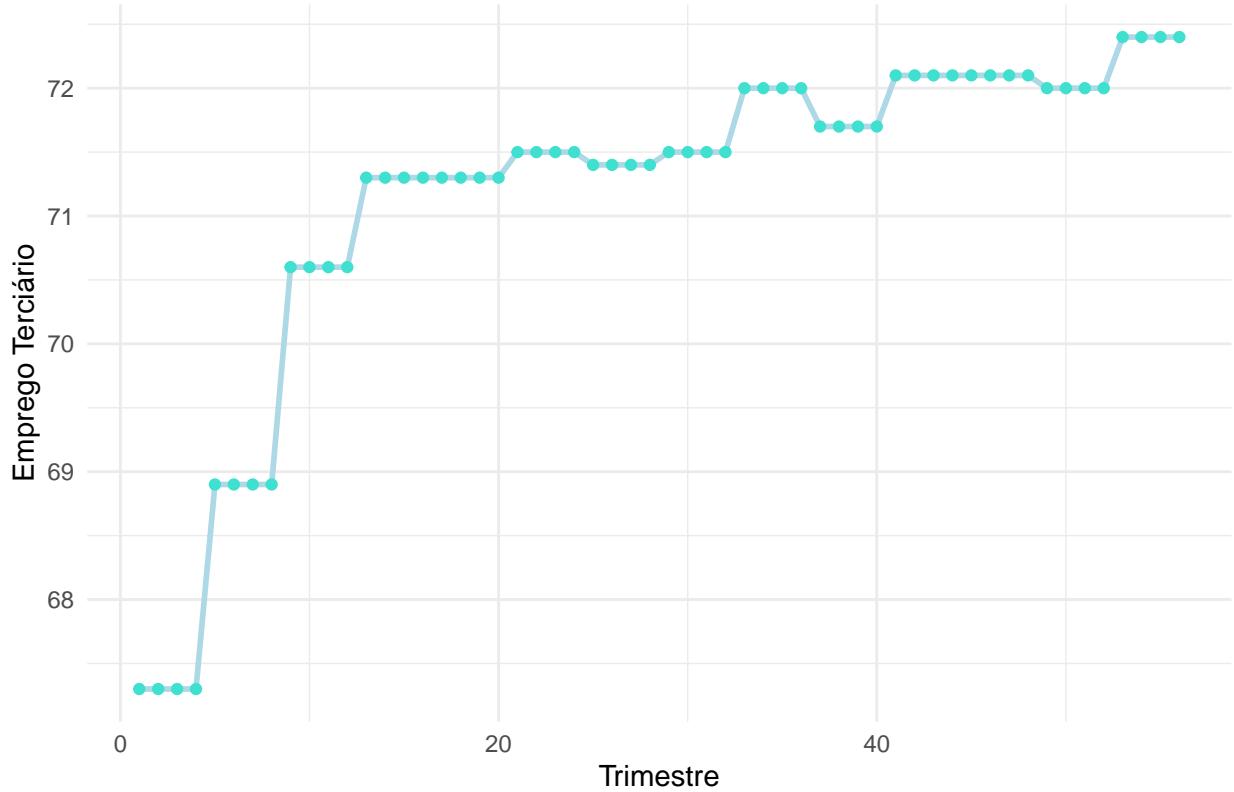


Figure 3: Série Temporal do Emprego no Setor Terciário

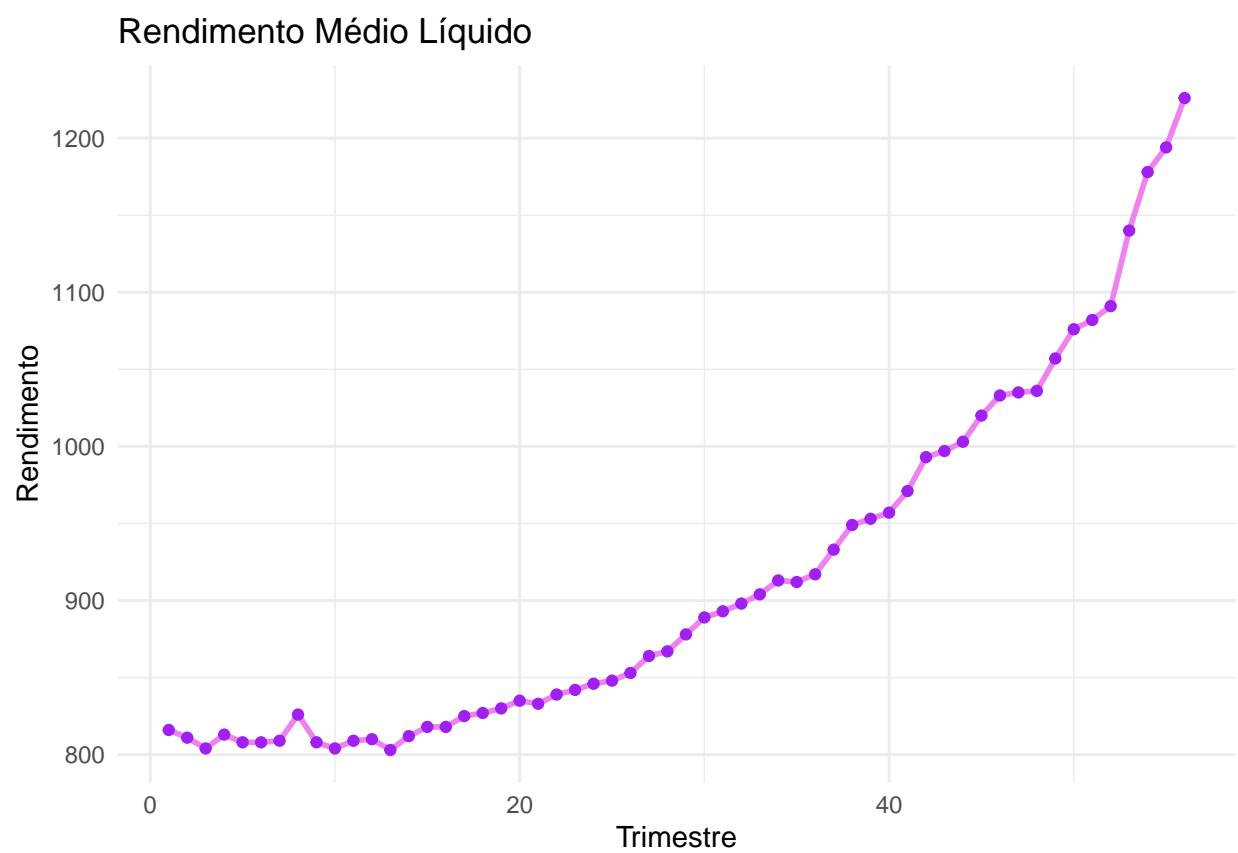


Figure 4: Série Temporal do Rendimento Médio Líquido

Variável “Rendimento_medio_liquido”: A remuneração média líquida evidencia uma tendência crescente bastante marcada: - No início do período, os valores situavam-se abaixo dos 850 €. - A partir de meados da série, o crescimento torna-se mais acentuado, ultrapassando os 1100 € nos últimos trimestres. Este aumento pode refletir fatores como melhoria salarial gradual, políticas de valorização do rendimento do trabalho e evolução positiva do mercado laboral.

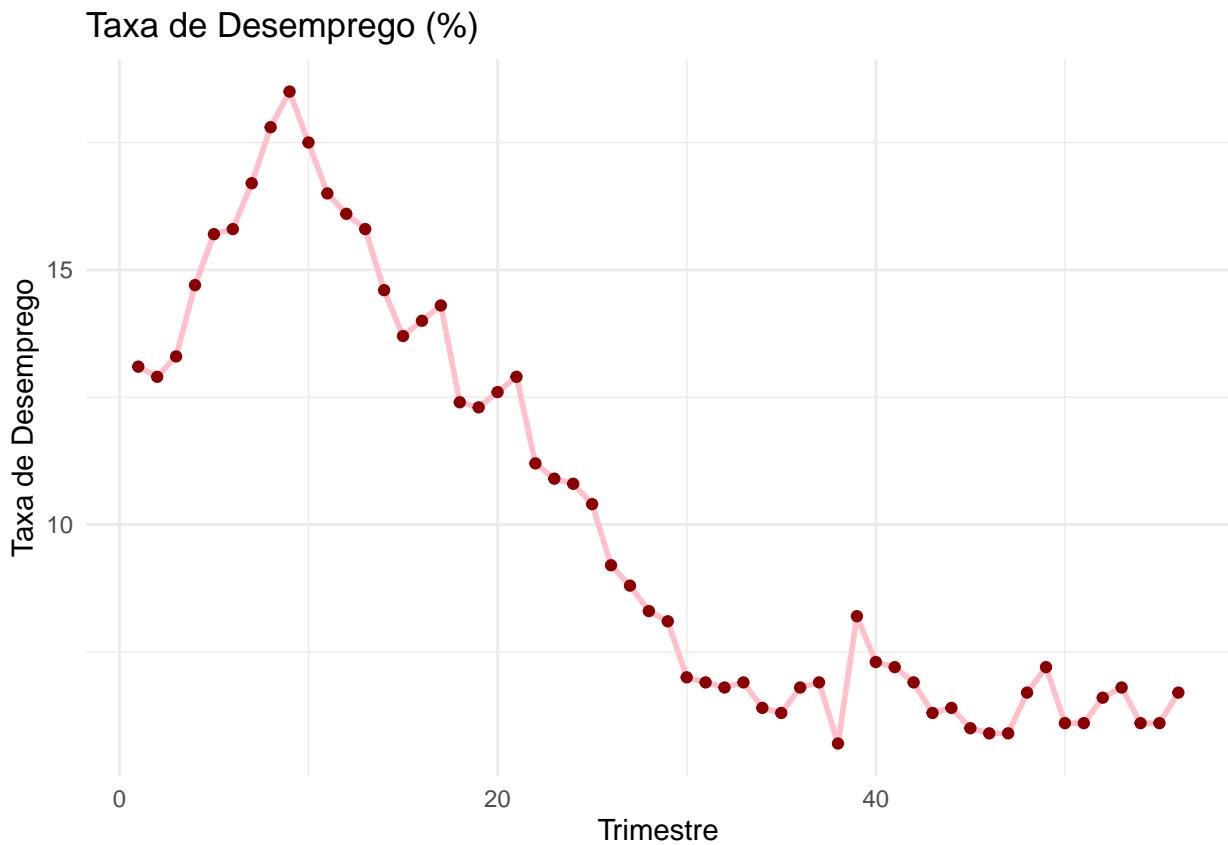


Figure 5: Série Temporal da Taxa de Desemprego

Evolução da Variável Dependente: Taxa de Desemprego O gráfico da taxa de desemprego mostra uma trajetória muito clara, composta por três momentos principais:

1. Subida significativa (2011–2013):

- A taxa de desemprego sobe rapidamente de cerca de 12% para valores superiores a 17%–18%.
- Este pico coincide com a crise económica, o período de resgate financeiro e a forte contração do mercado de trabalho.

2. Queda contínua e expressiva (2014–2019):

- Por volta do trimestre 12 (4º trimestre de 2013), observa-se uma descida quase ininterrupta.
- O desemprego recua para valores abaixo de 7%. Este período acompanha a recuperação económica, o aumento do investimento e a melhoria das condições do mercado laboral.

3. Choque da pandemia e estabilização (2020–2024)

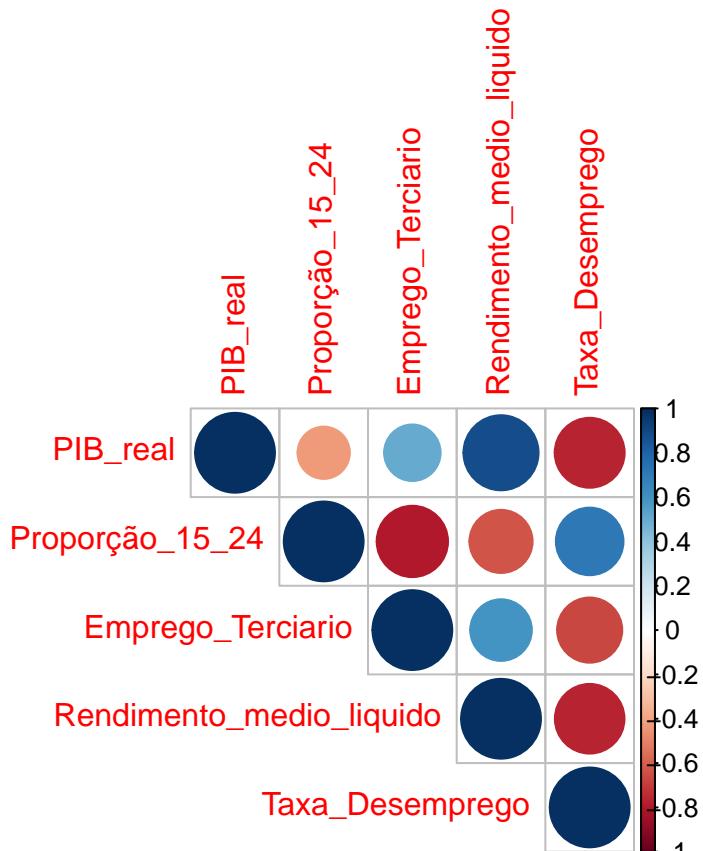
- Verifica-se um pequeno aumento em 2020, mas muito inferior ao observado em 2011–2013.
- Após essa oscilação, a taxa estabiliza entre 6% e 8%, com pequenas variações individuais por trimestre.

Este comportamento destaca a resiliência do mercado de trabalho português no pós-pandemia, apoiado por políticas de manutenção do emprego.

Correlação

A matriz de correlação permite observar a direção e a intensidade das relações lineares entre a taxa de desemprego e cada uma das variáveis explicativas. Embora a correlação não estabeleça causalidade, fornece uma primeira indicação sobre quais variáveis podem ter maior influência na evolução do desemprego e sobre a coerência das relações observadas nos gráficos temporais.

```
##                                     PIB_real Proporção_15_24 Emprego_Terciario
## PIB_real                               1.0000000 -0.4296857   0.5019000
## Proporção_15_24                          -0.4296857  1.0000000 -0.7989995
## Emprego_Terciario                      0.5019000  -0.7989995  1.0000000
## Rendimento_medio_liquido    0.8854396  -0.6263246   0.5959282
## Taxa_Desemprego                         -0.7680883   0.7173931  -0.6666016
##                                     Rendimento_medio_liquido Taxa_Desemprego
## PIB_real                               0.8854396  -0.7680883
## Proporção_15_24                        -0.6263246   0.7173931
## Emprego_Terciario                      0.5959282  -0.6666016
## Rendimento_medio_liquido    1.0000000  -0.7615427
## Taxa_Desemprego                         -0.7615427   1.0000000
## 
##  ## corrplot 0.95 loaded
```



Taxa de desemprego e PIB real — correlação negativa A correlação entre o PIB real e a taxa de desemprego é negativa, refletindo que: - Quando o PIB aumenta, o desemprego tende a diminuir. - Esta relação é consistente com a teoria económica: maior atividade económica promove criação de emprego.

Taxa de desemprego e rendimento médio — correlação negativa O rendimento médio apresenta uma correlação negativa com o desemprego, indicando que: - Em períodos de menor desemprego, o rendimento do trabalho tende a crescer. - Isto reflete a melhoria do mercado laboral e maior competitividade salarial. A relação, embora lógica, pode não ser muito forte devido à evolução gradual dos salários ao longo do período.

Taxa de desemprego e percentagem de emprego no setor terciário — correlação tendencialmente negativa A proporção de emprego no setor terciário tende a aumentar ao longo do período, enquanto o desemprego diminui, o que resulta numa correlação negativa: - Economias mais terciarizadas e orientadas para serviços tendem a absorver mais mão-de-obra. - O setor dos serviços, particularmente turismo e comércio, tem forte influência no emprego em Portugal.

Taxa de desemprego e proporção de jovens na população ativa — correlação positiva A proporção de jovens apresenta uma correlação positiva com o desemprego: - Populações mais jovens têm maior rotatividade laboral e maior taxa de entrada no mercado de trabalho. - Os jovens são historicamente mais vulneráveis a choques económicos e apresentam taxas de desemprego superiores ao total da população ativa. Esta relação confirma que mudanças demográficas podem influenciar significativamente a dinâmica do desemprego.

Especificação do Modelo Bayesiano

- **Modelo base (regressão linear múltipla):** Tendo em conta o objetivo de estudar de que forma diferentes fatores socioeconómicos influenciam a taxa de desemprego em Portugal, recorreu-se a um modelo de regressão linear múltipla em enquadramento bayesiano, que serve para diagnóstico preliminar.. Este tipo de modelo permite quantificar o efeito médio de cada variável explicativa na variável dependente enquanto incorpora, de forma explícita, a incerteza associada às estimativas dos parâmetros.

No nosso caso, especificou-se o seguinte modelo:

$$Taxa_{Desemprego}_t = \beta_0 + \beta_1(PIB_{Real}_t) + \beta_2(Proporção_{15_24}_t) + \beta_3(Emprego_{Terciário}_t) + \beta_4(Rendimento_{Médio_Liquido}_t)$$

onde, - β_0 representa o intercepto, - β_i representam os efeitos das variáveis explicativas, - $\epsilon_t \sim N(0, \sigma^2)$ representa o erro aleatório.

Uma etapa crucial consiste em determinar a distribuição adequada para modelar esta variável. Apesar de a taxa de desemprego ser uma proporção, observamos que os valores trimestrais apresentam comportamento contínuo e relativamente simétrico, sem se aproximar de 0% ou 100%. Além disso, as variações entre trimestres são moderadas, a volatilidade é baixa e não se observam caudas pesadas ou assimetrias significativas. Essas características tornam a distribuição Normal uma escolha apropriada, pois ela descreve de forma adequada variáveis contínuas, simétricas e com dispersão moderada.

Portanto, assumimos que:

\$\$

$$Y_t | \theta \sim N(\mu, \sigma^2)$$

\$\$, onde μ representa a média da taxa de desemprego e σ^2 a sua variância. Esta suposição permite-nos aplicar modelos de regressão clássicos e realizar inferência estatística de forma consistente e interpretável.

Pelo código, fazemos o modelo de regressão linear múltipla bayesiana:

```
# Modelo Múltiplo Bayesiano
library(lubridate)
modelo1 <- lm(Taxa_Desemprego ~ PIB_real + Proporção_15_24 +
               Emprego_Terciário + Rendimento_médio_liquido,
```

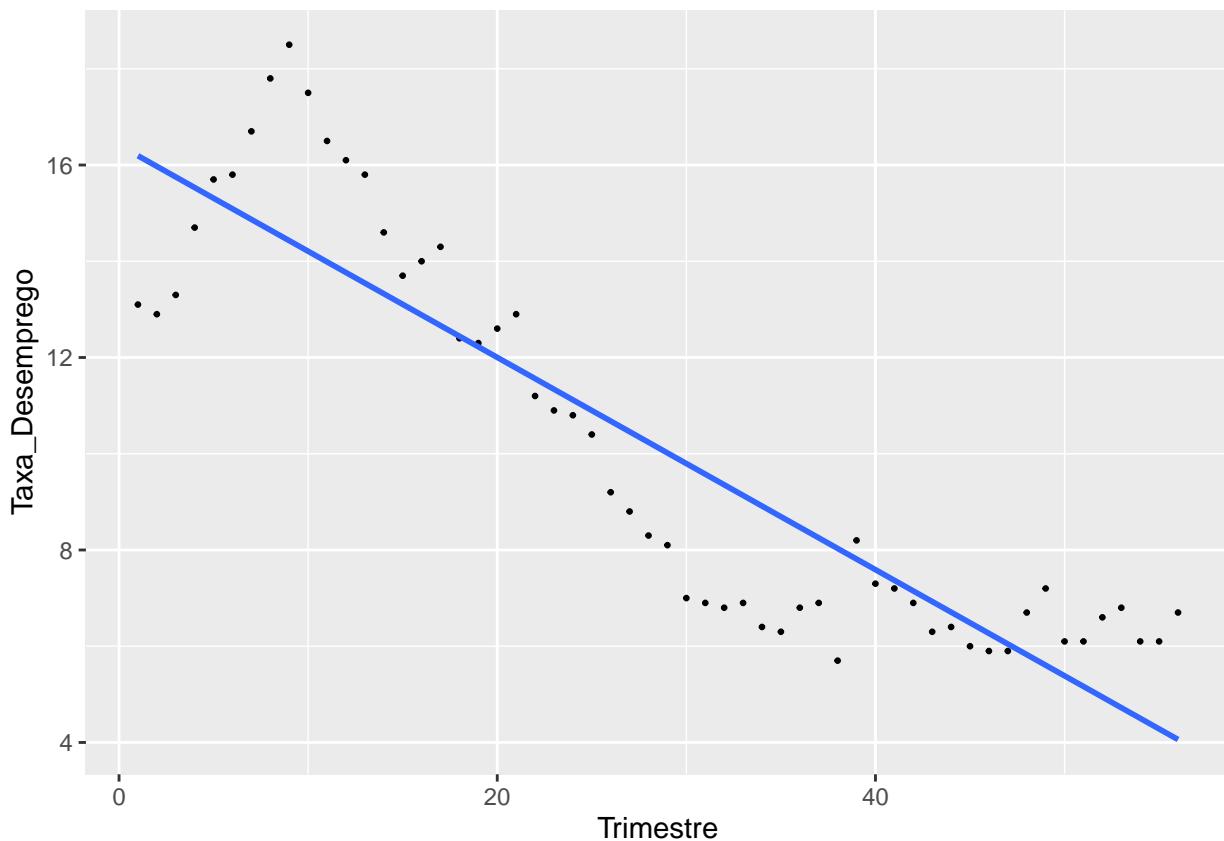
```
    data = dados)
summary(modelo1)
```

Especificação das Prioris

```
#Distribuição Posteriori
data(dados)
```

```
## Warning in data(dados): data set 'dados' not found
ggplot(dados, aes(x = Trimestre, y = Taxa_Desemprego)) +
  geom_point(size = 0.5) +
  geom_smooth(method = "lm", se = FALSE)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Precisamos especificar modelos a priori para cada uma das variáveis. Para isso, utilizaremos a estrutura padrão de modelos a priori usada pelo pacote rstanarm.

Simulação posteriori

Diagnóstico e Avaliação

Simulação via rstanarm

```
# Definir um modelo com priors fracos (nao informativos)
modelo_bayes <- stan_glm(Taxa_Desemprego ~ PIB_real + Proporção_15_24 + Emprego_Terciario + Rendimento_m...
```

```

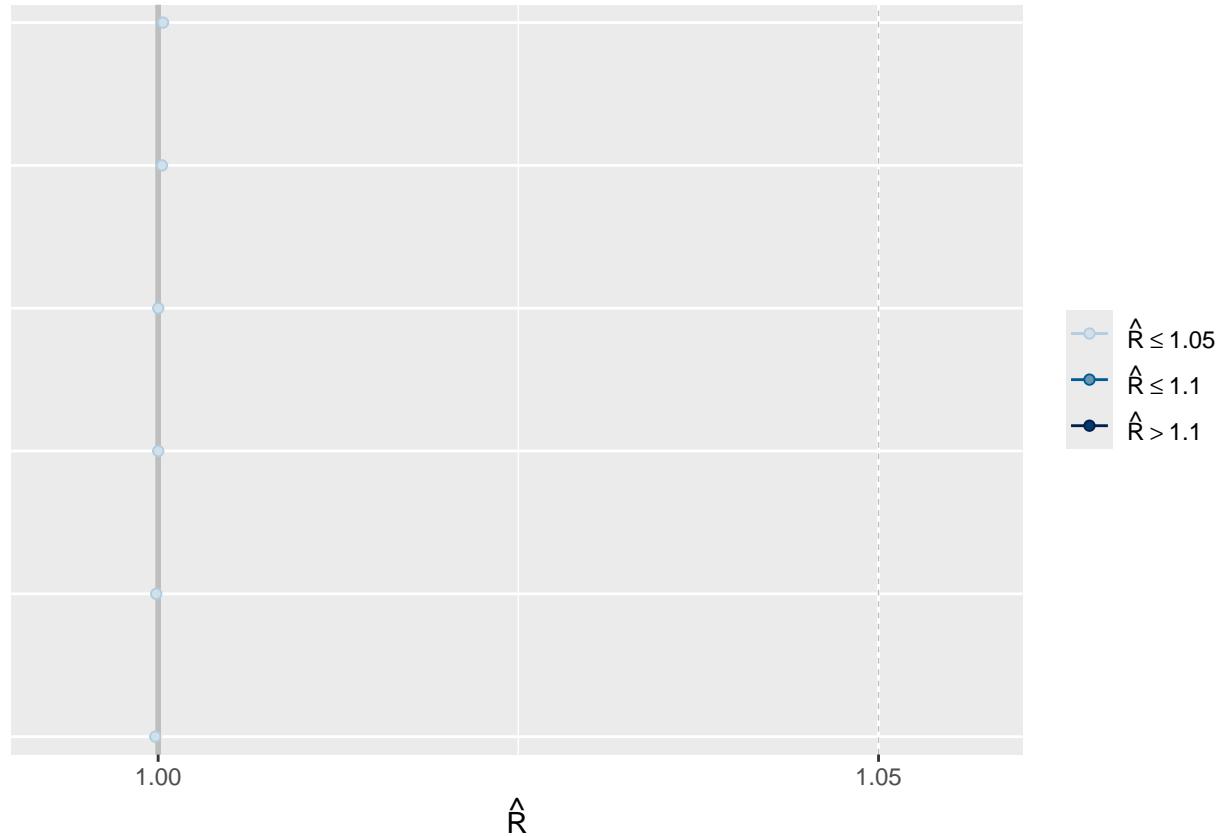
        family = gaussian(),
        prior = normal(0, 10),           # prior para os betas
        prior_intercept = normal(0, 20),  # prior para intercepto
        chains = 4,                     # numero de cadeias MCMC
        iter = 4000,                   # n° total de iteracoes
        warmup = 1000,                 # burn-in
        seed = 123
)

```

Avaliação modelo

Utilizamos a medida Rhat para comparar as 4 cadeias simuladas por stan_glm.

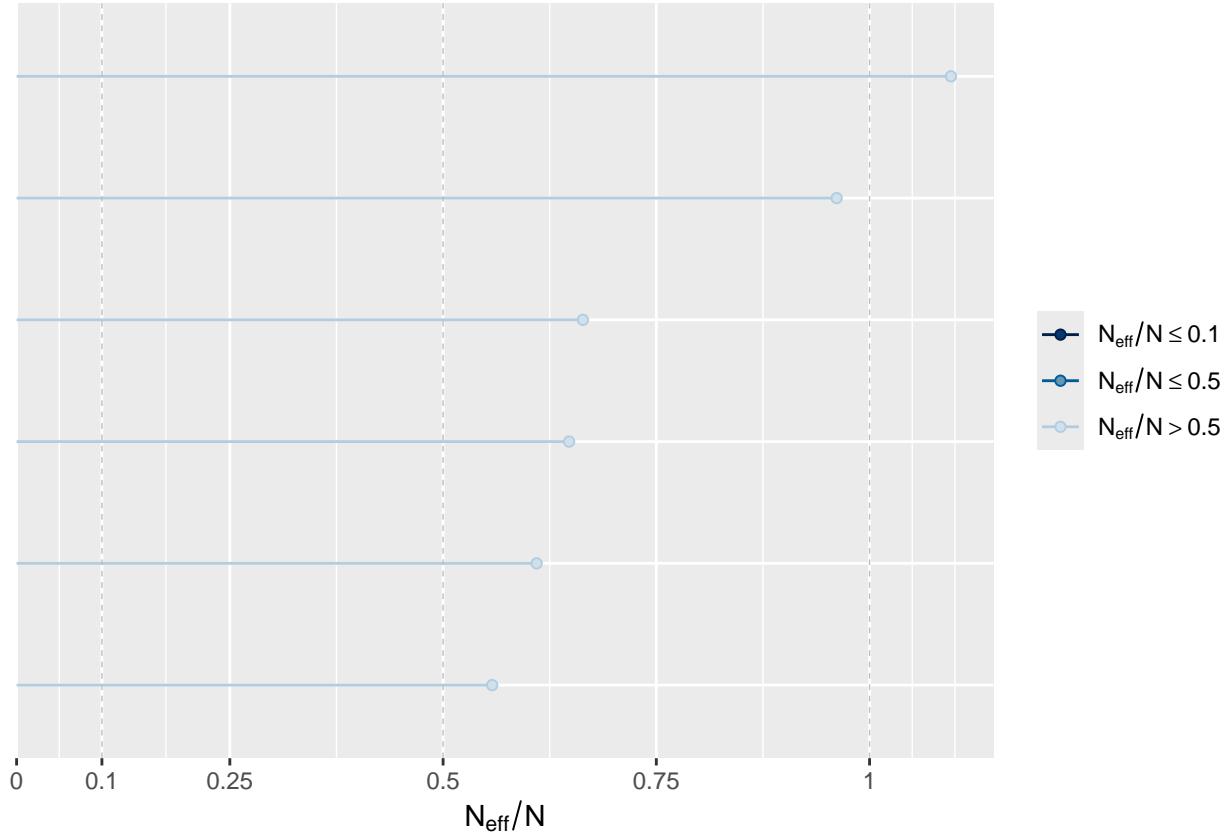
```
mcmc_rhat(bayesplot::rhat(modelo_bayes))
```



Os valores obtidos encontram-se abaixo de 1.05, indicando semelhança forte nas cadeias.

Utiliza-se também o diagnóstico neff ratio para mcmc.

```
mcmc_neff(neff_ratio(modelo_bayes))
```



Neste caso também vemos linhas azuis em todos os valores (maior que 0.5), indicando um bom rácio do tamanho de amostras efetivas.

Previsões

Com o nosso modelo avaliado, tentamos prever o desemprego com a media dos nossos posteriores. A imagem abaixo demonstra a diferença entre os valores previstos e os valores originais dos dados.

```
previsoes <- posterior_predict(modelo_bayes, dados) #cada coluna contem previsoes de um semestre

prediction_mean <- apply(previsoes, 2, mean)          # previsoes pontuais
apply(previsoes, 2, quantile, c(0.025, 0.975))       # intervalo credível

##           1        2        3        4        5        6        7        8
## 2.5% 10.13461 10.25862 10.60151 10.83205 9.629863 9.952738 10.25520 10.50863
## 97.5% 20.14896 20.35981 20.49839 20.77449 19.059757 19.360823 19.72934 20.07732
##           9       10       11       12       13       14       15
## 2.5%  8.748408 8.710731 8.723684 8.450765 7.907626 7.865059 7.679901
## 97.5% 18.228278 17.958083 18.103773 17.666197 17.161066 17.090727 17.065821
##          16       17       18       19       20       21       22
## 2.5%   7.522366 7.350739 7.276879 7.316352 7.30152 6.913853 6.791844
## 97.5% 16.912528 16.703514 16.591764 16.643993 16.49181 16.108532 16.121730
##          23       24       25       26       27       28       29
## 2.5%   6.531087 6.104034 6.121876 6.00093 5.690191 5.464703 5.143267
## 97.5% 15.732635 15.501422 15.349696 15.25594 15.041203 14.705065 14.344456
##          30       31       32       33       34       35       36
## 2.5%   4.879489 4.645382 4.425878 3.663754 3.324342 3.345844 2.988687
## 97.5% 14.147782 14.035873 13.728003 13.112397 12.678835 12.796902 12.427505
```

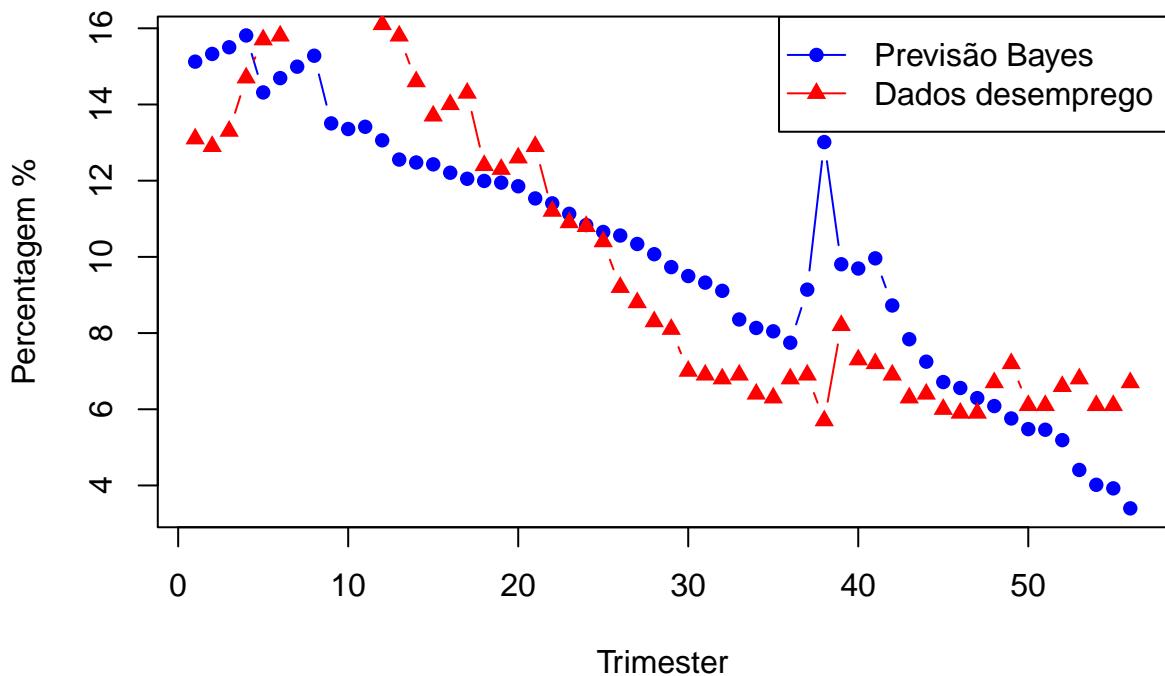
```

##          37      38      39      40      41      42      43
## 2.5%  4.502196 7.509086 5.272928 5.017328 5.11484 4.01495 3.243599
## 97.5% 13.689969 18.327462 14.394617 14.295310 14.71144 13.39231 12.497109
##          44      45      46      47      48      49      50
## 2.5%   2.554812 2.085658 1.871087 1.697038 1.490242 1.024893 0.7142266
## 97.5% 11.830464 11.241166 11.223696 10.928291 10.797774 10.485736 10.1761493
##          51      52      53      54      55      56
## 2.5%   0.6763046 0.4636601 -0.2537018 -0.8135421 -0.9383754 -1.453281
## 97.5% 10.1060640 9.8818054 9.1073456 8.7905655 8.7269548 8.265890

plot(prediction_mean, type = "b", col = "blue", pch = 16,
      xlab = "Trimester", ylab = "Percentagem %", main = "Comparação de Original vs previsão")
lines(dados$Taxa_Desemprego, type = "b", col = "red", pch = 17)
legend("topright", legend = c("Previsão Bayes", "Dados desemprego"),
       col = c("blue", "red"), lty = 1, pch = c(16, 17))

```

Comparação de Original vs previsão



Conclusão