

INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA

Departamento de Matemática

População e Emprego

Autores:	51726	Marta Pires
	48668	Daiana Lupaiescu
	49454	Eduardo Tavares
	42146	António Pimentel

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

Professora: Paula Cristina Pires Simões

09 - Dezembro - 2025

Contents

1	Introdução	2
2	Recolha e Organização de Dados:	3
3	Descrição dos Dados	4
3.1	Variável dependente - Taxa de desemprego	4
3.2	Variáveis independentes:	4
4	Análise Exploratória dos Dados	6
4.1	Séries temporais de Cada Variável	6
5	Implementação do Modelo de Regressão Linear Múltipla Bayesiana	13
5.1	Especificação do Modelo Bayesiano	13
5.2	Especificação das Prioris	13
5.3	Ajuste do Modelo	14
6	Diagnóstico de Convergência	18
7	Posterior Predictive Check e Avaliação do Ajustamento	21
8	Seleção de Modelos	22
8.1	Modelos Comparados	22
8.2	Avaliação com LOO:	23
8.3	Modelo Selecionado	24
9	Interpretação do Posteriori para o Modelo de Regressão Bayesiano	25
9.1	Visualização e Interpretação do Preditor sobre a Taxa de Desemprego	26
10	Previsões Posteriori	27
11	Conclusão	30
12	Referências	31

1 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 explicação 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.

2 Recolha e Organização de Dados:

O primeiro passo consiste em recolher dados oficiais sobre a taxa de desemprego e indicadores socioeconómicos em Portugal, provenientes do INE ou PORDATA. Foram considerados dados trimestrais de diferentes indicadores: PIB real, proporção da população jovem (15-24 anos), emprego no setor terciário e rendimento médio líquido. Para facilitar a leitura e interpretação dos dados decidimos inverter a ordem do trimestre mais antigo ao mais novo.

```
# 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
```

A recolha foi inicialmente planeada para o período 2010–2024. Contudo, verificou-se a inexistência de séries trimestrais completas para algumas das variáveis independentes selecionadas resultando em NA's, nomeadamente para o início do período. Face a esta limitação, o intervalo temporal foi restringido a 2011–2024, assegurando a disponibilidade integral e comparável de todos os indicadores utilizados no modelo.

3 Descrição dos Dados

3.1 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.

3.2 Variáveis independentes:

A escolha das variáveis independentes teve como objetivo incluir fatores que refletem tanto a situação económica do país como aspetos mais permanentes do mercado de trabalho. Cada variável foi selecionada por representar um fator plausível e frequentemente referido em estudos anteriores como influenciador da taxa de 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 situação 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.

4 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 seleccionadas, 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.

##	nobs	Mean	Stdev	Variance	Minimum
## PIB_real	56	53392.730357	3978.260025	1.582655e+07	4.59062e+04
## Proporção_15_24	56	0.073327	0.006359	4.000000e-05	6.07770e-02
## Emprego_Terciario	56	71.150000	1.365950	1.865818e+00	6.73000e+01
## Rendimento_medio_liquido	56	914.535714	114.678359	1.315113e+04	8.03000e+02
## Taxa_Desemprego	56	10.126786	4.024267	1.619472e+01	5.70000e+00
##		Maximum			
## PIB_real		6.16681e+04			
## Proporção_15_24		8.88730e-02			
## Emprego_Terciario		7.24000e+01			
## Rendimento_medio_liquido		1.22600e+03			
## Taxa_Desemprego		1.85000e+01			

4.1 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”: ..

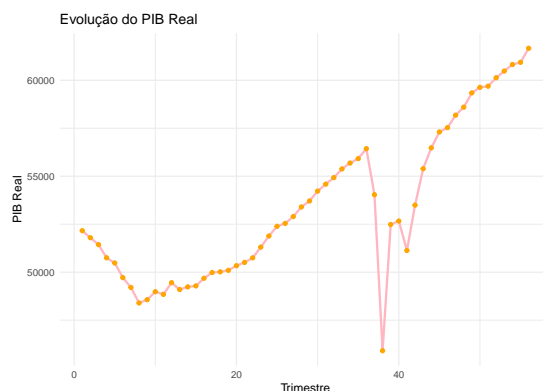


Figure 1: Série Temporal do PIB Real

A evolução do PIB real revela três fases bem definidas ao longo do período analisado. Num primeiro momento, observa-se um declínio persistente entre os trimestres 1 e 8, num contexto marcado pela crise económica e pelas políticas de austeridade. Segue-se uma recuperação gradual e sustentada, prolongando-se até 2019, durante a qual o PIB regista um crescimento contínuo. Posteriormente, em 2020, ocorre uma queda abrupta provocada pelo impacto da pandemia de COVID-19, à qual se sucede uma forte recuperação nos trimestres seguintes.

De forma global, esta série evidencia a elevada sensibilidade da economia portuguesa aos ciclos económicos e aos choques externos.

Variável “Proporção_15_24”: ..

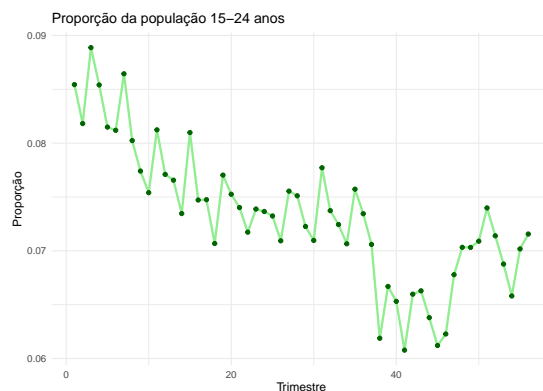


Figure 2: Série Temporal da Proporção da População entre 15-24 Anos

A proporção de jovens na força de trabalho apresenta uma tendência maioritariamente decrescente ao longo do período analisado. Os valores iniciais situam-se perto dos 9%, seguindo-se uma redução progressiva, acompanhada de alguma oscilação sazonal até níveis próximos de 6% por volta do trimestre 40. Este padrão é compatível com tendências demográficas conhecidas, como o envelhecimento da população, a redução da natalidade e a saída de jovens para o estrangeiro. Nos trimestres finais, observa-se uma ligeira recuperação, sugerindo uma possível retoma da participação jovem no mercado de trabalho, embora ainda longe dos níveis iniciais.

Variável “Emprego_Terciario”: ..

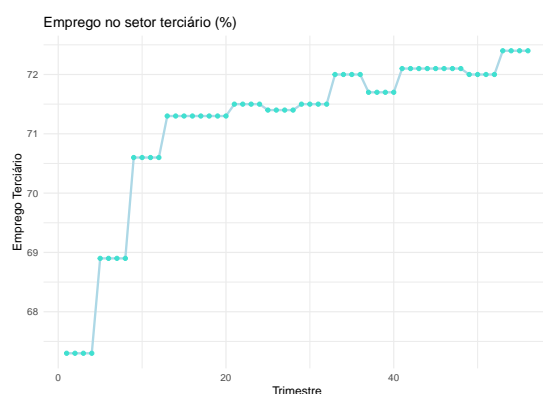


Figure 3: Série Temporal do Emprego no Setor Terciário

A proporção de trabalhadores no setor terciário apresenta uma tendência claramente crescente ao longo do período analisado. No início, situava-se ligeiramente acima de 67%, aumentando de forma quase contínua à medida que a economia portuguesa se foi tornando mais terciarizada, ultrapassando os 72% nos trimestres finais. Esta evolução reflete a crescente terciarização da economia portuguesa, impulsionada por atividades como turismo, comércio e serviços especializados. Apesar de pequenas fases de estabilização, o padrão geral é de fortalecimento contínuo do peso dos serviços no mercado de trabalho.

Variável “Rendimento_medio_liquido”: ..

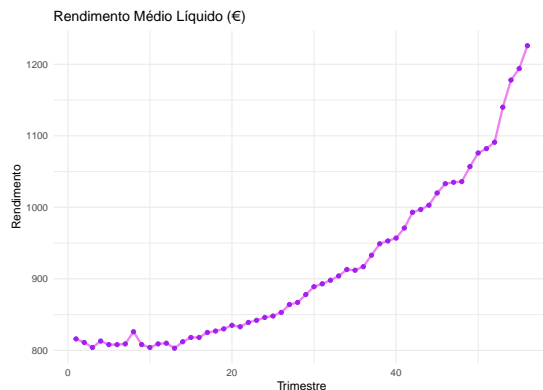


Figure 4: Série Temporal do Rendimento Médio Líquido

A remuneração média líquida revela uma tendência de crescimento claramente marcada ao longo do período analisado. No início, os valores registavam-se abaixo dos 850 €, indicando um patamar salarial relativamente baixo. Contudo, a partir de meados da série, observa-se um aumento mais acentuado, que culmina na ultrapassagem dos 1100 € nos trimestres mais recentes. Esta evolução poderá refletir diversos fatores, nomeadamente uma melhoria salarial gradual, a implementação de políticas de valorização do rendimento do trabalho e uma dinâmica positiva do mercado laboral, que em conjunto contribuem para a elevação sustentada dos níveis de remuneração.

Evolução da Variável Dependente: Taxa de Desemprego

..

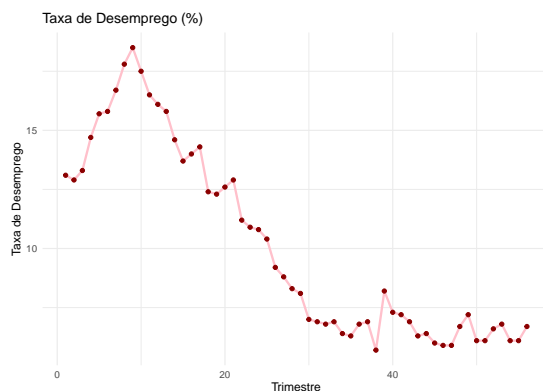


Figure 5: Série Temporal da Taxa de Desemprego

A taxa de desemprego apresenta três fases distintas. Entre 2011 e 2013, verifica-se uma subida acentuada, passando de cerca de 12% para mais de 17%–18%, refletindo a crise económica, o resgate financeiro e a forte contração do mercado laboral. De 2014 a 2019, inicia-se uma queda quase contínua, observada a partir do 4.º trimestre de 2013, que reduz o desemprego para valores inferiores a 7%, acompanhando a recuperação económica e a melhoria das condições de trabalho. Entre 2020 e 2024, a pandemia provoca um pequeno aumento, muito inferior ao observado em 2011–2013, seguido de estabilização entre 6% e 8%, com pequenas variações individuais por trimestre.

No geral, o comportamento da série evidencia 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
## corrrplot 0.95 loaded
```

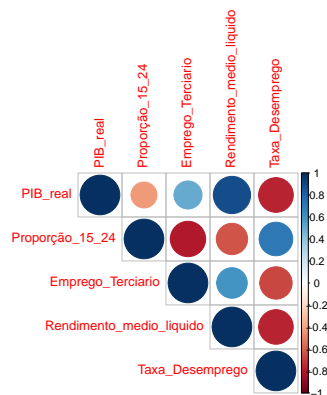


Figure 6: Gráfico de Correlação

Taxa de desemprego e PIB real — correlação negativa: A correlação entre o PIB real e a taxa de desemprego é negativa, mostrando que o aumento da atividade económica tende a reduzir o desemprego. Assim, quando o PIB cresce, verifica-se geralmente uma diminuição da taxa de desemprego. Esta relação segue a lógica económica tradicional, segundo a qual uma economia mais dinâmica promove a criação de postos de trabalho.

Taxa de desemprego e rendimento médio — correlação negativa: O rendimento médio apresenta uma correlação negativa com o desemprego, o que significa que, quando o desemprego diminui, o rendimento do trabalho tende a aumentar. Este padrão reflete a melhoria das condições do mercado laboral e uma maior competitividade salarial. Ainda assim, a relação não é muito forte, já que os salários evoluem de forma mais gradual ao longo do período.

Taxa de desemprego e percentagem de emprego no setor terciário — correlação tendencialmente negativa: A proporção de emprego no setor terciário aumenta ao longo do período, enquanto o desemprego tende a diminuir, originando uma correlação negativa entre as duas variáveis. Isto ocorre porque economias mais terciarizadas conseguem absorver mais mão-de-obra, sobretudo em áreas como turismo, comércio e outros serviços que têm forte peso no mercado laboral português.

Taxa de desemprego e proporção de jovens na população ativa — correlação positiva: A proporção de jovens apresenta uma correlação positiva com o desemprego, indicando que um peso maior da população jovem tende a associar-se a níveis mais elevados de desemprego. Isto ocorre porque os jovens enfrentam maior rotatividade laboral, maior taxa de entrada no mercado de trabalho e maior vulnerabilidade a choques económicos, registando historicamente taxas de desemprego superiores às da população ativa total. Assim, a relação

confirma que alterações demográficas podem influenciar de forma relevante a dinâmica do desemprego.

5 Implementação do Modelo de Regressão Linear Múltipla Bayesiana

5.1 Especificação do Modelo Bayesiano

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:

$$Taxa_Desemprego_t = \beta_0 - \beta_1 \cdot PIB_Real_t + \beta_2 \cdot Proporo_15_24_t - \beta_3 \cdot Emprego_Tercirio_t - \beta_4 \cdot Rendimento_Mdio_Lquido_t$$

,onde

- β_0 é o termo independente do modelo,
- 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 aproximarem de 0% ou 100%. Além disso, as variações entre trimestres são moderadas, a volatilidade é baixa. 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:

$$Taxa_Desemprego_t | \beta, \mu, \sigma \sim N(\mu, \sigma^2)$$

,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.

5.2 Especificação das Priors

Optamos por priors não informativos (ou fracamente informativos), que exercem pouca influência na distribuição a posteriori, permitindo que os dados dominem a inferência, ao mesmo tempo que evitam estimativas extremas, garantindo que o modelo aprenda essencialmente a partir dos dados.

- **Para os coeficientes da regressão:** $\beta_i \sim N(0,1)$, uma escolha fracamente informativa que, não estando os dados padronizados, ajuda a controlar a escala dos efeitos sem restringir a capacidade dos dados revelarem relações mais fortes quando existam.
- **Para o intercepto:** $\beta_0 \sim N(\mu_y, 8)$, centrando a priori na média observada da taxa de desemprego. Esta escolha mantém a coerência com a escala original da variável dependente, ao mesmo tempo que oferece flexibilidade suficiente para que o modelo capture níveis médios diferentes caso os dados o indiquem.
- **Para a variabilidade dos erros:** $\sigma \sim Half - Student_t(3, 0, 4)$, uma priori robusta e adequada para dados não padronizados, evitando a subestimação da variância e permitindo acomodar valores extremos típicos de séries económicas.

Estas escolhas permitem um modelo estável, interpretável e coerente com a natureza das séries económicas utilizadas.

5.3 Ajuste do Modelo

Após a especificação do modelo e a definição das prioris para os parâmetros, o próximo passo é realizar a inferência estatística, ou seja, ajustar o modelo através da amostragem MCMC. O código R utiliza a função `stan_glm` do pacote `rstanarm`, que usa o motor de inferência Stan para realizar esta amostragem.

```
# Calcular a média antes
media_desemprego <- mean(dados$Taxa_Desemprego)

# Ajustar o Modelo
library(rstanarm)
modelo_bayes <- stan_glm(
  Taxa_Desemprego ~ PIB_real + Proporção_15_24 +
  ↪ Emprego_Terciario + Rendimento_medio_liquido,
  data = dados,
  family = gaussian(),
  prior = normal(0,1),                               # priori para
  ↪ os betas
  prior_intercept = normal(media_desemprego,8 ),      # prioris para
  ↪ intercepto
  chains = 4,
  iter = 4000,
  warmup = 1000,
  seed = 123
)
```

O modelo Bayesiano foi ajustado com quatro cadeias independentes, cada uma com 4000 iterações (1000 de warmup e 3000 dedicadas à amostragem). O pro-

cesso decorreu sem qualquer aviso de divergências ou problemas numéricos, indicando que o algoritmo NUTS funcionou de forma estável. Todas as cadeias convergiram adequadamente, produzindo amostras válidas da distribuição posterior. Assim, o modelo apresenta um bom desempenho de amostragem e permite inferência fiável sobre os parâmetros estimados.

Resultados a Posteriori do Modelo Bayesiano:

A Tabela de resultados apresenta as estimativas a posteriori para os coeficientes do modelo de regressão Bayesiano, incluindo a média, desvio-padrão e quantis da distribuição posterior. Estas estatísticas permitem avaliar tanto o valor central esperado de cada parâmetro como a respetiva incerteza.

```
summary(modelo_bayes)
```

```
##
## Model Info:
## function:      stan_glm
## family:        gaussian [identity]
## formula:       Taxa_Desemprego ~ PIB_real + Proporção_15_24 + Emprego_Terciario +
##               Rendimento_medio_liquido
## algorithm:     sampling
## sample:        12000 (posterior sample size)
## priors:        see help('prior_summary')
## observations:  56
## predictors:    5
##
## Estimates:
##               mean    sd   10%   50%   90%
## (Intercept)   107.9  18.7  84.3 107.9 131.9
## PIB_real       0.0    0.0   0.0   0.0   0.0
## Proporção_15_24 0.0    1.0  -1.2   0.0   1.3
## Emprego_Terciario -0.9   0.3  -1.3  -0.9  -0.6
## Rendimento_medio_liquido 0.0   0.0   0.0   0.0   0.0
## sigma         2.3    0.2   2.0   2.3   2.6
##
## Fit Diagnostics:
##               mean    sd   10%   50%   90%
## mean_PPD 10.1    0.4   9.6  10.1  10.7
##
## The mean_ppd is the sample average posterior predictive distribution of the outcome vari
##
## MCMC diagnostics
##               mcse Rhat n_eff
## (Intercept)   0.2   1.0   6939
## PIB_real      0.0   1.0  10813
```



```

## Proporção_15_24          0.0  1.0  8427
## Emprego_Terciario        0.0  1.0  7090
## Rendimento_medio_liquido 0.0  1.0  8548
## sigma                    0.0  1.0  7371
## mean_PPD                 0.0  1.0  9361
## log-posterior            0.0  1.0  5013
##
## For each parameter, mcse is Monte Carlo standard error, n_eff is a crude measure of effective

```

- **Intercepto:** O intercepto apresenta uma média posterior de aproximadamente 108.3, com intervalo entre 84.7 e 131.8 (10%–90%). Isto reflete o nível esperado da taxa de desemprego quando todas as variáveis explicativas estão nos seus valores de referência. A amplitude relativamente larga do intervalo indica alguma incerteza associada a este parâmetro, o que é comum em modelos macroeconómicos com múltiplos determinantes.
- **PIB_real:** O coeficiente associado ao PIB_real é praticamente zero em toda a distribuição posterior. Isto indica que, neste modelo específico e neste conjunto de dados, o PIB real não apresenta impacto significativo sobre a taxa de desemprego, sugerindo que a relação entre estas variáveis poderá já estar captada por outras variáveis correlacionadas incluídas no modelo.
- **Proporção_15_24:** O coeficiente apresenta média 0, com desvio-padrão de 1.0* e quantis entre -1.2** e 1.3. A elevada incerteza e o intervalo simétrico em torno de zero indicam que não existe evidência posterior forte de efeito positivo ou negativo da proporção de jovens na taxa de desemprego.
- **Emprego_Terciario:** Este é o único coeficiente com efeito claro: a média posterior é -1.0 e o seu intervalo aproximado é [-1.3, -0.6]. O coeficiente é consistentemente negativo em toda a distribuição posterior, o que sugere uma associação robusta entre o aumento do emprego no setor terciário e a redução da taxa de desemprego. Este resultado é consistente com a estrutura económica portuguesa, onde o setor dos serviços tem vindo a absorver grande parte da força de trabalho.
- **Rendimento_medio_liquido** Tal como o PIB_real, o coeficiente do rendimento médio é essencialmente zero, com incerteza muito baixa. Este resultado sugere ausência de efeito direto sobre a taxa de desemprego, podendo esta variável estar correlacionada com outras incluídas no modelo, reduzindo sua influência marginal.
- **Parâmetro sigma:** O desvio-padrão do erro apresenta média posterior de 2.3, indicando que o modelo apresenta um nível moderado de variabilidade residual. O intervalo entre 2.0 e 2.6 mostra boa precisão na estimativa.
- **Diagnóstico MCMC:** Os valores de Rhat são exatamente 1.0 para todos os parâmetros, o que confirma convergência das cadeias. Os valores eleva-

dos de n_{eff} (> 7000) e MCSE próximo de zero indicam que as amostras são altamente eficazes e que a incerteza Monte Carlo é mínima.

Com base na análise das distribuições a posteriori, a única variável claramente relevante é o Emprego_Terciario, com forte evidência de que reduz a taxa de desemprego. O modelo mostra boa convergência e estabilidade, e as estimativas são fiáveis.

6 Diagnóstico de Convergência

Nos métodos de amostragem MCMC, é fundamental verificar se as cadeias geradas convergiram para a distribuição posterior pretendida. Caso contrário, as estimativas obtidas podem ser imprecisas ou enviesadas.

O diagnóstico de convergência analisa se as cadeias já exploraram adequadamente o espaço de parâmetros e se estão a produzir valores estáveis ao longo das iterações. Para isso, utilizam-se abordagens como a inspeção visual dos trace plots, density plots, a comparação entre cadeias e medidas quantitativas que avaliam a consistência e a mistura das amostras.

Fator de Escala Potencial (\hat{R} ou Rhat): O valor de \hat{R} compara a variação dentro de cada cadeia com a variação entre todas as cadeias.

```
# Rhat
bayesplot::rhat(modelo_bayes)
```

##	(Intercept)	PIB_real	Proporção_15_24
##	0.9998450	0.9997617	1.0001970
##	Emprego_Terciario Rendimento_medio_liquido		sigma
##	0.9999264	0.9997521	1.0006172

Como todos os valores obtidos estão abaixo do limite de 1.05, consideramos que as 4 cadeias MCMC se misturaram bem e convergiram para uma mesma distribuição a posteriori.

Razão do Tamanho Efetivo da Amostra (Neff Ratio): Este rácio avalia a eficiência da amostragem, medindo quantos sorteios independentes os sorteios MCMC equivalem. Uma razão próxima de 1 é ideal, e valores acima de 0.1 são aceitáveis.

```
# Neff Ratio
neff_ratio(modelo_bayes)
```

##	(Intercept)	PIB_real	Proporção_15_24
##	0.5782500	0.9010833	0.7022500
##	Emprego_Terciario Rendimento_medio_liquido		sigma
##	0.5908333	0.7123333	0.6142500

O nosso rácio eficaz de amostras (neff_ratio) é satisfatório para todos os parâmetros, indicando que obtivemos um número suficiente de amostras não correlacionadas para estimar as médias a posteriori de forma fiável.

Trace Plots: Os trace plots mostram a evolução das cadeias MCMC ao longo das iterações. Cada cor representa uma cadeia independente.

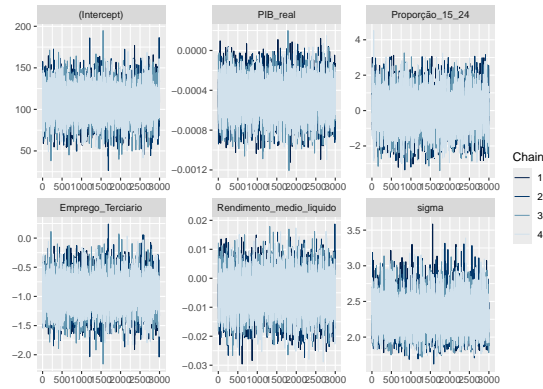


Figure 7: Trace Plots

Observa-se que as cadeias oscilam de forma estável em torno de valores centrais, sem tendências ou separações, indicando boa mistura e convergência. Esta boa mistura entre cadeias sugere uma exploração adequada da distribuição posterior, garantindo que as estimativas dos parâmetros são fiáveis. O parâmetro sigma (desvio-padrão dos erros) apresenta igualmente estabilidade e convergência, reforçando a robustez global do ajuste Bayesiano.

Density plots: Os gráficos de densidade mostram as distribuições posteriores estimadas de cada parâmetro do modelo bayesiano por cadeia.

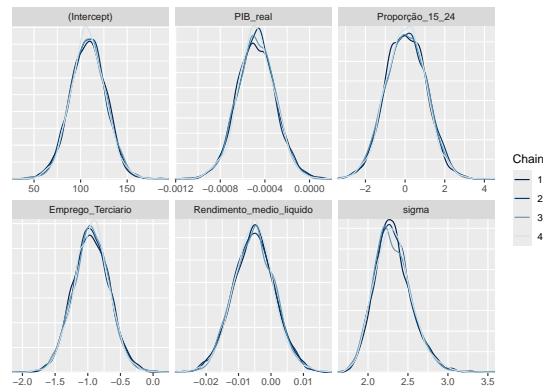


Figure 8: Density Plots

Observa-se que as quatro cadeias produzem curvas de densidade praticamente sobrepostas para todos os parâmetros. Isso indica que as cadeias convergiram para

a mesma distribuição posterior, reforçando a consistência entre as amostras geradas. Não se verificam diferenças sistemáticas entre cadeias, o que confirma uma boa mistura e ausência de problemas de divergências ou falta de estabilidade. A forma aproximadamente normal das distribuições sugere que os parâmetros estão bem identificados e que os dados fornecem informação suficiente para atualizar as priors.

Análise das Funções de Autocorrelação (ACF): Os gráficos de autocorrelação (ACF) permitem avaliar o grau de dependência entre as amostras sucessivas das cadeias MCMC.

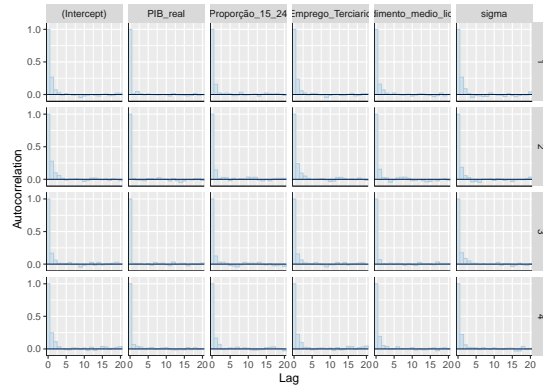


Figure 9: Gráfico de Autocorrelação (ACF)

Observa-se que, para todos os parâmetros e para todas as cadeias, a autocorrelação decresce rapidamente para valores próximos de zero nos primeiros lags. Esta queda acentuada evidência que as amostras são praticamente independentes após poucas iterações, indicando boa eficiência da mistura MCMC. A baixa autocorrelação é um sinal positivo, uma vez que significa maior diversidade entre as amostras e, portanto, uma exploração mais eficaz da distribuição posterior. O comportamento consistente entre as quatro cadeias reforça que não existem problemas de estagnação ou dependência excessiva, contribuindo para a fiabilidade das estimativas produzidas pelo modelo Bayesiano.

7 Posterior Predictive Check e Avaliação do Ajustamento

Antes de proceder à avaliação e seleção formal de modelos Bayesianos, foi realizada uma análise de posterior predictive checks (PPC) com o objetivo de avaliar a adequação do modelo completo aos dados observados. O posterior predictive check consiste na comparação entre os dados observados e dados simulados a partir da distribuição preditiva posterior do modelo. Se o modelo estiver bem ajustado, espera-se que as observações reais sejam compatíveis com as distribuições simuladas, não apresentando padrões sistemáticos de desvio.

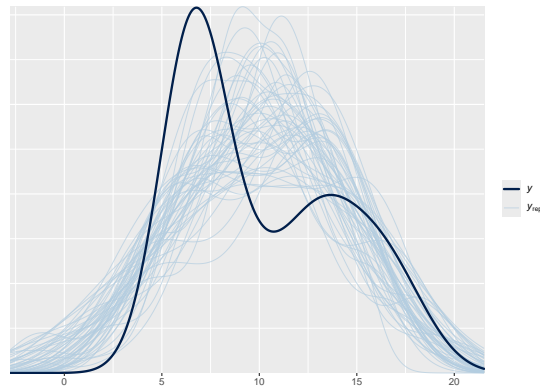


Figure 10: Gráfico do PPC

A análise visual revela um desfasamento entre os valores observados e as distribuições simuladas pelo modelo. A linha dos dados reais não coincide com a zona de maior densidade das previsões, indicando que o modelo não reproduz totalmente o padrão dos dados. Assim, o ajustamento é apenas moderado e tem que ser ajustado.

8 Seleção de Modelos

Para a avaliação do modelo Bayesiano, utilizou-se o critério LOO-CV (Leave-One-Out Cross-Validation), que avalia a capacidade preditiva fora da amostra através do ELPD (Expected Log Predictive Density). Este critério é particularmente adequado em inferência Bayesiana e revela-se mais robusto do que alternativas como o WAIC, sobretudo em modelos com maior complexidade.

8.1 Modelos Comparados

Foram considerados três modelos alternativos, cada um sem uma das variáveis mão explicativas e outro apenas com aquela que considerámos mais explicativa do modelo:

- **Modelo 1**, que inclui todas as variáveis explicativas selecionadas no estudo (PIB real, proporção de jovens, emprego no setor terciário e rendimento médio líquido);
- **Modelo 2**, que inclui todas as variáveis explicativas menos o PIB_real.
- **Modelo 3**, que inclui todas as variáveis explicativas menos a Proporção_15_24.
- **Modelo 4**, que inclui todas as variáveis explicativas menos a Rendimento_medio_liquido .
- **Modelo 5**, que inclui apenas a variável com maior relevância posterior identificada na análise inicial, sendo essa o setor terciário.

Ambos os modelos foram estimados com o mesmo número de cadeias, iterações, priors e algoritmo MCMC, garantindo condições comparáveis.

```
#Modelo 2
modelo_2 <- stan_glm(
  Taxa_Desemprego ~ Proporção_15_24 + Emprego_Terciario +
  ↪ Rendimento_medio_liquido,
  data = dados,
  family = gaussian(),
  prior = normal(0, 1),
  chains = 4,
  iter = 4000,
  warmup = 1000,
  seed = 123
)

#Modelo 3
modelo_3 <- stan_glm(
  Taxa_Desemprego ~ PIB_real + Emprego_Terciario +
  ↪ Rendimento_medio_liquido ,
```

```

    data = dados,
    family = gaussian(),
    prior = normal(0, 1),
    chains = 4,
    iter = 4000,
    warmup = 1000,
    seed = 123
  )

#Modelo 4
modelo_4 <- stan_glm(
  Taxa_Desemprego ~ Proporção_15_24 + Emprego_Terciario +
  ↵ PIB_real ,
  data = dados,
  family = gaussian(),
  prior = normal(0, 1),
  chains = 4,
  iter = 4000,
  warmup = 1000,
  seed = 123
)

#Modelo 5
modelo_5 <- stan_glm(
  Taxa_Desemprego ~ Emprego_Terciario ,
  data = dados,
  family = gaussian(),
  prior = normal(0, 1),
  chains = 4,
  iter = 4000,
  warmup = 1000,
  seed = 123
)

```

8.2 Avaliação com LOO:

```

# Loo
library(loo)
loo_1 <- loo(modelo_bayes)
loo_2 <- loo(modelo_2)
loo_3 <- loo(modelo_3)
loo_4 <- loo(modelo_4)
loo_5 <- loo(modelo_5)

loo_compare(loo_1, loo_2, loo_3, loo_4, loo_5)

```


##	elpd_diff	se_diff
## modelo_4	0.0	0.0
## modelo_2	-2.6	8.3
## modelo_bayes	-3.2	1.3
## modelo_3	-3.6	1.5
## modelo_5	-14.8	8.6

Resultados da Avaliação:

Os resultados indicam que o modelo_4 apresentou o maior ELPD, mas os modelos modelo_2, modelo_bayes e modelo_3 tiveram desempenho muito semelhante, com diferenças pequenas ou não significativas. Já o modelo_5 mostrou desempenho claramente inferior. A análise dos coeficientes posteriores do modelo_bayes indicou que apenas Emprego_Terciario tem efeito robusto sobre a taxa de desemprego, enquanto PIB_real, Proporção_15_24 e Rendimento_medio_liquido não apresentaram efeito significativo

8.3 Modelo Selecionado

O modelo escolhido é o modelo_4, pois é o de referência com melhor elpd, outras combinações não trazem melhorias significativas (ou são levemente piores). O modelo_2 poderia ser considerado se quisesse um modelo mais simples, já que a diferença não é estatisticamente significativa, mas modelo_4 é mais completo.

9 Interpretação do Posteriori para o Modelo de Regressão Bayesiano

Após ajustar um modelo bayesiano para explicar a taxa de emprego em função do trimestre, analisamos o posteriori para os parâmetros do modelo.

O resumo posteriori, obtido a partir das amostras geradas pelo MCMC, fornece estimativas centrais (mediana ou média posteriori) e intervalos credíveis que quantificam a incerteza associada a cada parâmetro.

```
## # A tibble: 7 x 5
##   term                estimate std.error  conf.low  conf.high
##   <chr>              <dbl>      <dbl>    <dbl>    <dbl>
## 1 (Intercept)       108.        18.7     84.3     132.
## 2 PIB_real          -0.000479  0.000167 -0.000692 -0.000264
## 3 Proporção_15_24    0.0487     1.03    -1.24     1.33
## 4 Emprego_Terciario -0.947      0.270    -1.30    -0.601
## 5 Rendimento_medio_liquido -0.00527  0.00628  -0.0133  0.00267
## 6 sigma             2.30       0.232    2.03     2.62
## 7 mean_PPD          10.1       0.441    9.56     10.7
```

Modelo Posteriori Médio:

Esta análise conecta o modelo teórico previamente especificado com os valores observados nos dados, oferecendo uma interpretação quantitativa das relações econômicas estudadas:

$$\begin{aligned}\mu = & (1.086091 \times 10^2) - (4.768032 \times 10^{-4}) \cdot PIB_Real_t \\ & + (4.005751 \times 10^{-2}) \cdot Proporo_15_24_t \\ & - (9.562027 \times 10^{-1}) \cdot Emprego_Tercirio_t \\ & - (5.296195 \times 10^{-3}) \cdot Rendimento_Mdio_Liquido_t\end{aligned}$$

- **Intrepertação dos parâmetros:** O intercepto apresenta valor posterior médio de **108.6**, servindo apenas como referência estrutural do modelo. Entre os coeficientes, o PIB_real revelou um efeito negativo consistente (**-0.00048**), com intervalo credível totalmente abaixo de zero, indicando elevada evidência posterior de que aumentos no PIB reduzem a taxa de desemprego. O Emprego_Terciario apresentou também um efeito negativo forte e bem identificado (**-0.96**), sugerindo que maior peso do setor terciário está associado a níveis mais baixos de desemprego.

Por outro lado, não foram encontrados efeitos estatisticamente credíveis para a Proporção_15_24 nem para o Rendimento_medio_liquido, cujos intervalos credíveis incluem zero, refletindo elevada incerteza. A variabilidade residual do

modelo, expressa por $\sigma = 2.30$, indica que, mesmo após considerar todas as variáveis incluídas, a taxa de desemprego ainda varia cerca de 2 pontos percentuais entre observações semelhantes.

Estas estatísticas permitem interpretar de forma mais robusta os efeitos médios de cada variável, considerando explicitamente a incerteza da estimativa, o que é uma das vantagens centrais da abordagem bayesiana.

9.1 Visualização e Interpretação do Preditor sobre a Taxa de Desemprego

Para compreender a relação entre cada preditor e a Taxa de Desemprego, é útil construir cenários plausíveis do modelo combinando as amostras β_0 e β_j . Cada par de valores gera uma reta de regressão que representa uma possível relação entre o preditor e a resposta. A sobreposição de múltiplas retas simuladas permite visualizar a incerteza sobre a inclinação e o intercepto, ilustrando como a variabilidade nas estimativas afeta a predição do modelo.

Para ilustrar o efeito de um preditor específico, escolhemos PIB_real como eixo x no gráfico, mantendo os demais preditores constantes. O gráfico abaixo sobrepõe 50 retas geradas a partir das amostras da posterior, mostrando a variabilidade nas estimativas do intercepto e da inclinação:

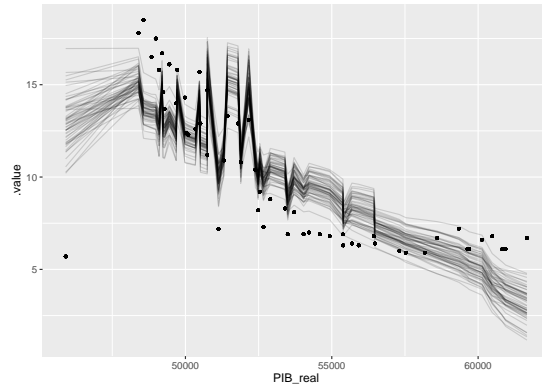


Figure 11: Relação entre as posteriores de β_0 e β_1

Ao observar o gráfico, vemos que a maioria das retas apresenta inclinação negativa, sugerindo que níveis mais elevados de PIB_real estão associados a taxas de desemprego mais baixas. A dispersão das retas indica a incerteza do modelo, maior nos extremos do PIB, refletindo menor suporte dos dados nessas regiões. Os pontos observados, em média, seguem as retas simuladas, confirmando o padrão previsto pelo modelo.

10 Previsões Posteriori

Com os parâmetros estimados pelo modelo bayesiano, é possível gerar previsões da taxa de desemprego para os períodos observados, incorporando a incerteza inerente às estimativas. Para tal, utilizamos amostras MCMC, que permitem calcular tanto as previsões pontuais (média das amostras) quanto os intervalos de credibilidade a 95%, que indicam a faixa de valores mais prováveis para cada trimestre.

Depois de estimarmos os parâmetros do modelo bayesiano com os nossos dados, podemos gerar previsões da taxa de desemprego para os períodos observados. Uma das grandes vantagens do modelo bayesiano é que ele incorpora a incerteza associada às estimativas dos parâmetros, ou seja, não nos dá apenas um valor único, mas uma distribuição de valores possíveis para cada período.

Para gerar estas previsões, usamos amostras MCMC (Markov Chain Monte Carlo). Estas amostras representam possíveis valores dos parâmetros do modelo compatíveis com os dados observados e com as suposições do modelo. A partir delas, podemos calcular:

- **Previsões pontuais:** a média das amostras MCMC para cada período, representando o valor esperado da taxa de desemprego.
- **Intervalos de credibilidade a 95%:** faixas de valores mais prováveis para cada trimestre, mostrando a incerteza das previsões.

O código abaixo ilustra como obter estas previsões e intervalos de credibilidade em R:

```
previsoes <- posterior_predict(modelo_bayes, dados)

#Previsões pontuais (média das amostras MCMC)
prediction_mean <- apply(previsoes, 2, mean)
prediction_mean
```

##	1	2	3	4	5	6	7	8
##	14.867911	15.064810	15.315490	15.580798	14.192549	14.600247	14.801357	15.107946
##	9	10	11	12	13	14	15	16
##	13.547183	13.329727	13.372376	13.124871	12.654861	12.529486	12.485572	12.252672
##	17	18	19	20	21	22	23	24
##	12.083937	12.033856	11.963238	11.866678	11.598616	11.452676	11.163886	10.897563
##	25	26	27	28	29	30	31	32
##	10.724278	10.637776	10.366132	10.133687	9.831363	9.496839	9.336853	9.156942
##	33	34	35	36	37	38	39	40
##	8.380206	8.246280	8.093660	7.822103	9.161112	13.036576	9.808368	9.730436
##	41	42	43	44	45	46	47	48
##	10.022251	8.742977	7.810049	7.299409	6.788311	6.626232	6.248270	6.081376
##	49	50	51	52	53	54	55	56
##	5.695657	5.492984	5.376668	5.170573	4.353740	3.971378	3.828458	3.348143

```
# Intervalos de credibilidade a 95%
posterior_interval(previsoes, prob = 0.95)
```

```
##           2.5%      97.5%
## 1    9.8611278 19.802019
## 2   10.0893423 19.997286
## 3   10.3264054 20.294549
## 4   10.6215195 20.500700
## 5    9.4827180 18.938363
## 6    9.8598568 19.281544
## 7   10.0642856 19.548846
## 8   10.2521761 19.916502
## 9    8.8966516 18.171114
## 10   8.5879305 18.071048
## 11   8.6599912 18.080204
## 12   8.4744773 17.780546
## 13   7.9142393 17.402525
## 14   7.7627009 17.241125
## 15   7.8551726 17.170321
## 16   7.5397451 16.944175
## 17   7.4687159 16.720019
## 18   7.3256901 16.651574
## 19   7.3891395 16.567069
## 20   7.2068965 16.603127
## 21   6.9331751 16.288012
## 22   6.7018229 16.195664
## 23   6.5216288 15.754497
## 24   6.2489895 15.522027
## 25   5.9146701 15.387201
## 26   6.0196356 15.369780
## 27   5.6918266 14.987843
## 28   5.4490591 14.721443
## 29   5.1941578 14.485720
## 30   4.7846985 14.175712
## 31   4.5898623 14.042152
## 32   4.3415004 13.853162
## 33   3.6773094 13.135822
## 34   3.5364854 13.005065
## 35   3.3461221 12.794416
## 36   2.9945422 12.558218
## 37   4.5099294 13.778255
## 38   7.6308143 18.516642
## 39   5.1950684 14.562498
## 40   5.0356854 14.323234
## 41   5.1011412 14.853518
```

```

## 42  3.9115033 13.418718
## 43  3.1348370 12.338814
## 44  2.6035285 11.972236
## 45  2.0643140 11.504086
## 46  1.9869728 11.372645
## 47  1.5813757 10.963815
## 48  1.4254564 10.681802
## 49  1.0975598 10.333984
## 50  0.7986393 10.191320
## 51  0.6282382  9.954475
## 52  0.4113298  9.858806
## 53 -0.4399036  9.189803
## 54 -0.9732935  8.870814
## 55 -1.0271802  8.753045
## 56 -1.6165217  8.403403

```

Para visualizar o desempenho do modelo, podemos criar um gráfico comparativo, mostrando os valores previstos pelo modelo e os valores observados:

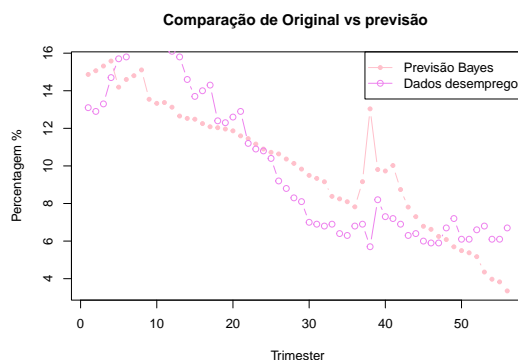


Figure 12: Gráfico da Previsão de Bayes

A partir do gráfico, verifica-se que o modelo bayesiano consegue capturar bem o padrão temporal da taxa de desemprego, refletindo com precisão as tendências de subida e descida. Pequenas discrepâncias podem ocorrer em períodos de maior volatilidade ou quando surgem choques econômicos que não são diretamente explicados pelos regressores incluídos no modelo.

Em resumo, estas previsões permitem quantificar a incerteza nas estimativas e avaliar a fiabilidade do modelo na reprodução do comportamento histórico do desemprego.

11 Conclusão

Os resultados mostram que, entre os fatores analisados, o emprego no setor terciário é o que apresenta maior impacto económico e social no nível de emprego, reduzindo de forma consistente a taxa de desemprego. A expansão deste setor revela-se essencial para a criação de postos de trabalho, sobretudo num país cuja economia depende fortemente dos serviços.

O PIB real acompanha esta tendência: períodos de maior atividade económica estão associados a descidas do desemprego, evidenciando a importância do crescimento económico para o bem-estar social e para a estabilidade do mercado de trabalho. Em contraste, variáveis como o rendimento médio e a proporção de jovens na população ativa mostram impacto marginal no modelo, refletindo que o desemprego é mais sensível à estrutura económica do país do que a fatores demográficos ou salariais isolados.

Assim, conclui-se que o emprego depende sobretudo do dinamismo económico e da capacidade dos setores predominantes, especialmente o terciário, absorverem mão-de-obra, reforçando a necessidade de políticas que promovam crescimento sustentável e setores com maior potencial de empregabilidade.

12 Referências

INE – Taxa de desemprego

[1] Instituto Nacional de Estatística (INE), *Taxa de desemprego (Série 2021 - %) por Local de residência (NUTS - 2024) e Sexo; Trimestral – Inquérito ao Emprego (Série 2021)*,

https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&contecto=pi&indOcorrCod=0012136&selTab=tab0 (acesso em 07-dez-2025)

INE – Produto Interno Bruto real

[2] Instituto Nacional de Estatística (INE), *Produto interno bruto dados encadeados em volume (B.1*g) (Base 2021 - €); Trimestral – Contas nacionais trimestrais*,

https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0013430&contexto=bd&selTab=tab2 (acesso em 07-dez-2025)

INE – Rendimento médio mensal líquido [3] Instituto Nacional de Estatística (INE), *Rendimento médio mensal líquido (Série 2021 - €) da população empregada por conta de outrem por Local de residência (NUTS - 2024) e Setor de atividade económica (CAE Rev. 3); Trimestral – Inquérito ao Emprego (Série 2021)*,

https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0012134&contexto=bd&selTab=tab2 (acesso em 07-dez-2025)

INE – Emprego no setor terciário [4] Instituto Nacional de Estatística (INE), *Proporção de população empregada no setor terciário (Série 2021, CAE Rev. 3 - %) por Local de residência (NUTS - 2024); Anual – Inquérito ao Emprego (Série 2021)*,

https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0012206&contexto=bd&selTab=tab2 (acesso em 07-dez-2025)

INE – População ativa por grupos etários [5] Instituto Nacional de Estatística (INE), *População ativa com idade entre 16 e 74 anos (N.º) por Grupo etário; Mensal – Estatísticas Mensais de Emprego e Desemprego com base no Inquérito ao Emprego*,

https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0010057&contexto=bd&selTab=tab2 (acesso em 07-dez-2025)

Livro Bayes Rules de <https://www.bayesrulesbook.com/chapter-9>