**Universidade Federal de Pernambuco – Campus Agreste** 

**Núcleo de Tecnologia – Curso de Engenharia de Produção**

**Análise das Séries Temporais – Atividade 9**

**Alunos: Débora Nayanne da Silva e Vitor Ferreira Lins**

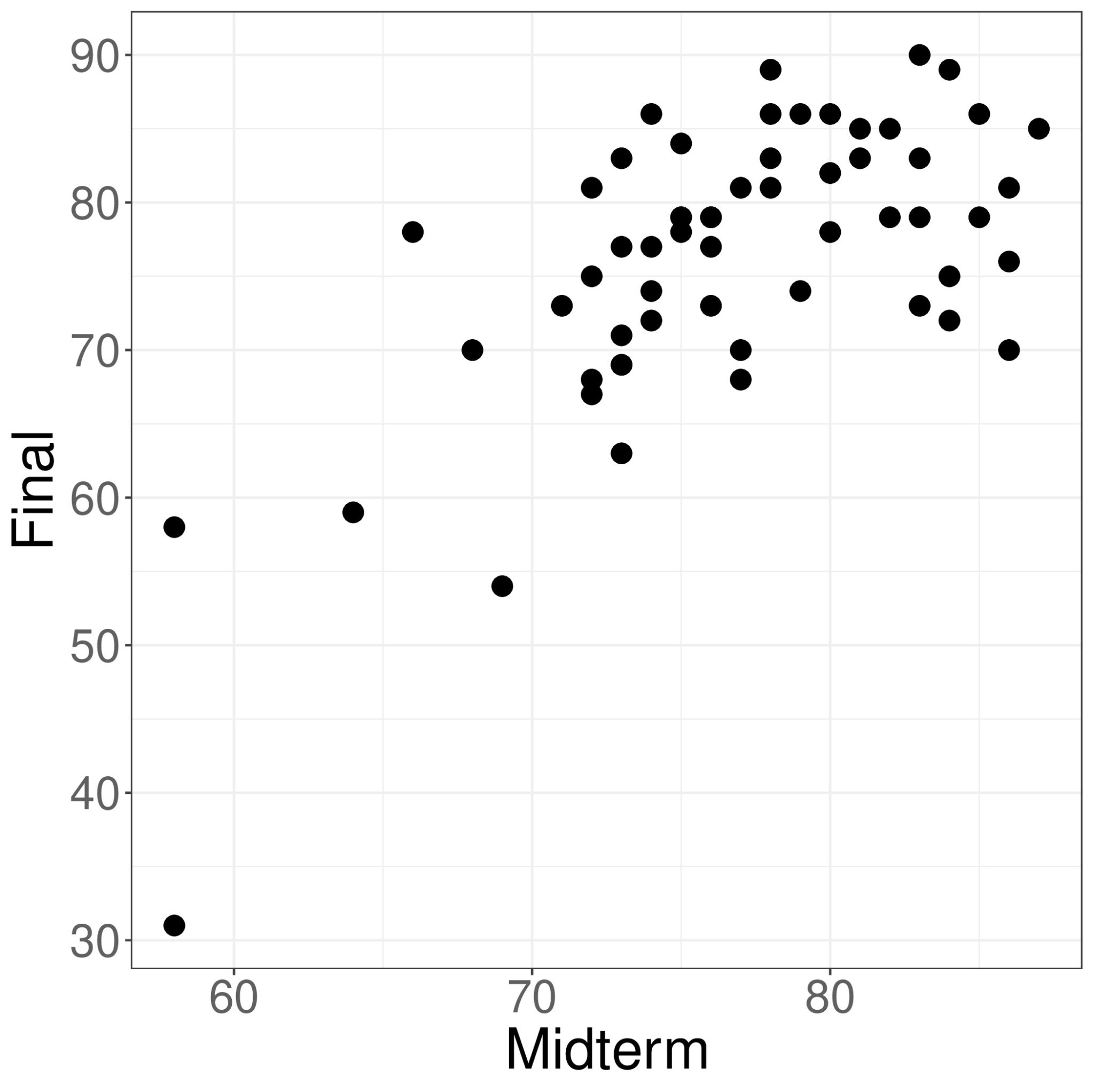
**Professores: Thyago Nepomuceno e Isloana Barros**

**1.**

* Erros não normais: é um problema não muito grave, depende do grau de não normalidade das variáveis. A não normalidade não causa viés na estimação dos parâmetros, mas impõe problemas aos desvios padrões e, desse modo, os intervalos de confiança dos testes de hipóteses podem não serem corretamente verificáveis. Alguns testes são utilizados para testar a normalidade dos dados, um dos mais frequentes é o teste de Kolmogorov-Smirnov. Em caso de não normalidade, alguns procedimentos podem ser adotados, como por exemplo, a transformação dos dados, esta, pode ser feita elevando os dados ao quadrado ou usando os seus logaritmos;
* Erros heterocedásticos: acontece quando a variância dos erros não é constante, os procedimentos de identificação e tratamento deste problema é basicamente o mesmo do problema de erros não normais, a heterocedasticidade não causa viés na estimação dos parâmetros, mas impõe problemas aos desvios padrões e, desse modo, os intervalos de confiança dos testes de hipóteses podem não serem corretamente verificáveis;
* Erros autocorrelacionados: em regressões, assume-se que os erros são independentes entre si, quando não, eles são autocorrelacionados. Normalmente isto ocorre ao tratar de dados em séries temporais, o erro do tempo t é correlacionado com o erro do tempo t-1. Para detectar este problema, usa-se testes de autocorrelação dos erros, uma solução é usar as primeiras diferenciações dos dados. Assim como os problemas dos pontos anteriores, erros autocorrelacionados não causam viés na estimação dos parâmetros.

**2.**

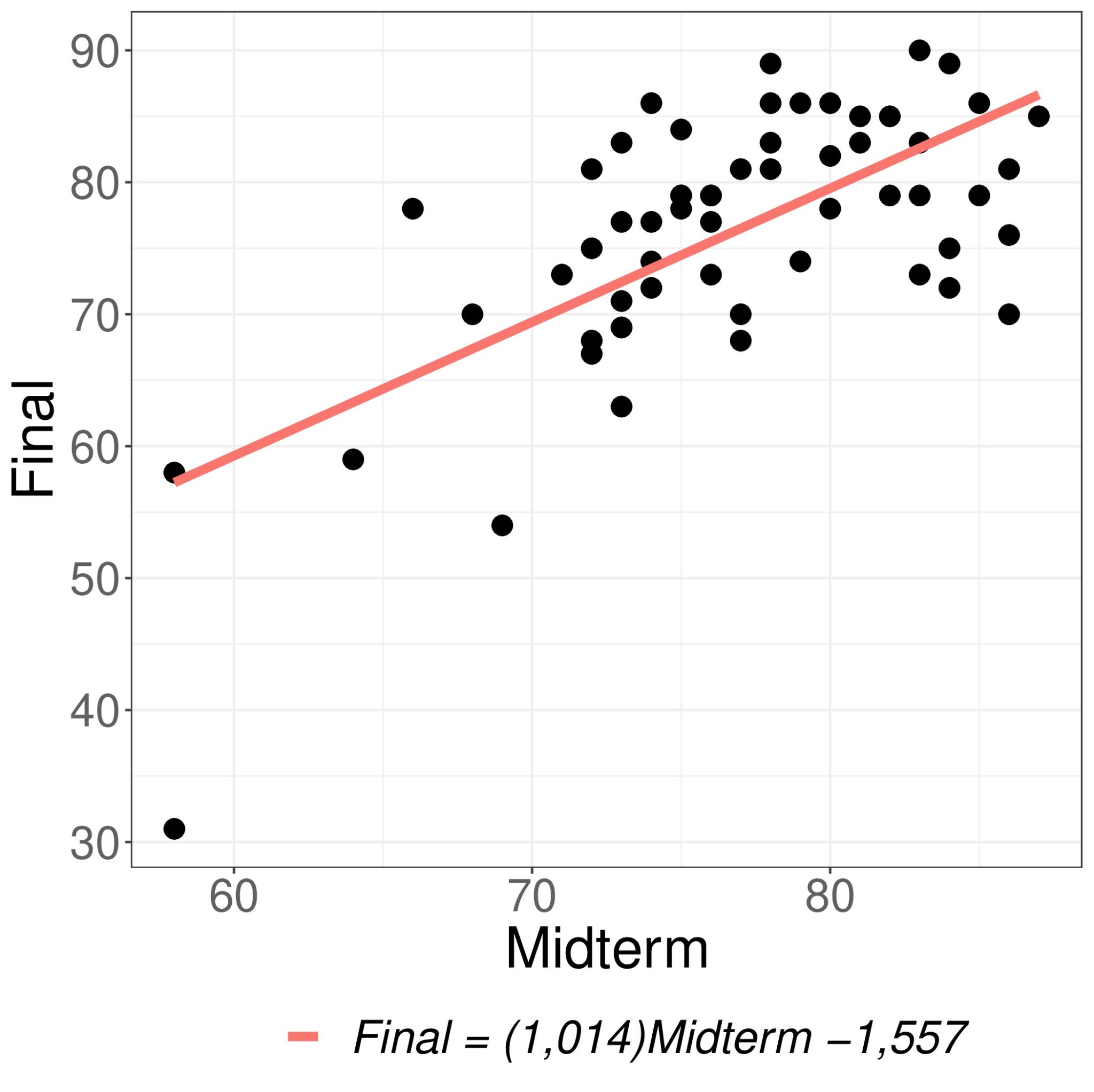
**a) GRÁFICO 1 - Diagrama de dispersão de Final Exam Score**



Olhando para os dados por este diagrama, percebe-se um comportamento linear entre as variáveis.

**b)** É aceitável um modelo linear, pois, o comportamento dos dados apresentam uma relação linear positiva (vide GRÁFICO 2).

**GRÁFICO 2 - Diagrama de dispersão Final Exam Score com linha de tendência**



**c)** O intercepto mostra que a cada 1 ponto no exame intermediário, esses indivíduos conseguem 1,01 pontos no exame final, ele tem significado, ou seja, a cada 1% a mais na nota do exame intermediário, espera-se 1,24% na nota no exame final.

**d)** O R2 diz que 43% dos resultados nos exames finais podem ser explicados pelo resultado no exame intermediário.

**e)** A estatística de F é igual a 42.22 (entre 1 e 56), assim, o modelo é muito significativo.

**f)** Não contém zero, isso significa que as variáveis são correlacionadas.

**g)** H0: diferenças de médias = 0

H1: diferenças de médias ≠ 0

Rejeita-se H0 ao nível de confiança de 95%. Graus de liberdade = 97.581; valor crítico = 0.7466.

**h)** O valor p é maior que 0,05, sendo rejeitado a níveis de confiança menores que 95%.

**i)** Os erros são muito dispersos, indicando uma não total normalidade dos erros.

**j)** Sim. Os erros se distanciam muito da linha de normalidade e do intervalo esperado.

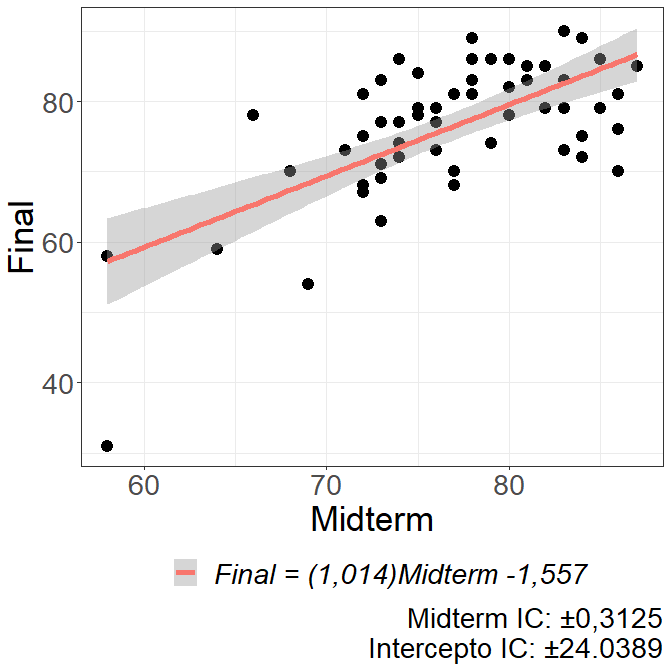
**k)** shapiro.test = 0.95326; p-value = 0.02575.

Percebe-se que o valor crítico é menor que 0,05, dessa forma, rejeita-se a hipótese de normalidade.

**l)** Não, só seria possível se os dados fossem temporais.

**m)**

**GRÁFICO 3 - Diagrama de dispersão Final Exam Score com linha de tendência e intervalo de confiança**



**2) a)** Calculando para fórmula:

(1)

O coeficiente delta () indica que a variável se trata de uma dummy, já que esta variável só pode assumir valores iguais a 0 ou 1. Os coeficientes obtidos foram:

**TABELA 1 - Valor dos coeficientes**

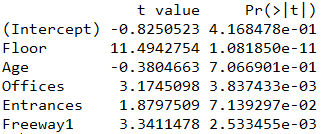
****

Todos os coeficientes correspondem, à expectativa, exceto em relação ao intercepto, mas isto pode ser explicado pelo fato de as variáveis “Entrances” e “Offices” não possuírem nenhum valor menor que 1, portanto o intercepto, especificamente neste modelo, deveria ser interpretado como a soma de alpha com beta 3 e beta 4 (), em vez de apenas .

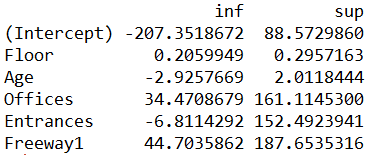
**2) b)** Apenas o intervalo de “Age” e do intercepto incluem o zero. Isto significa que todos os regressores são estatisticamente significantes, exceto “Age” e “Residuals”, existe a possibilidade de a remoção de algum dos regressores melhorar as estimativas do modelo.

**2) c)**

**TABELA 2 - Testes T**



**TABELA 3 - Intervalos de confiança**

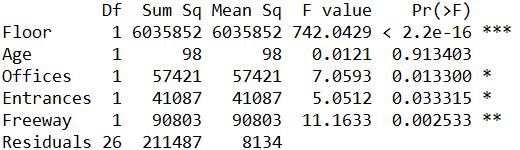


Com o nível de significância de 0,05, apenas “Floor”, “Offices” e “Freeway1” são aprovados como significantes, portanto, seus intervalos de confiança incluem o zero.

**2) d)** Para considerar um estimador como significante, o p-valor observado no teste T de Student deve ser menor que o nível de significância adotado, neste caso, deve ser menor que 0,05.

**2) e)** O R2 Ajustado encontrado foi 0,9608 e a avaliação da variância se encontra na tabela a seguir:

**TABELA 4 - Análise das Variâncias (ANOVA)**

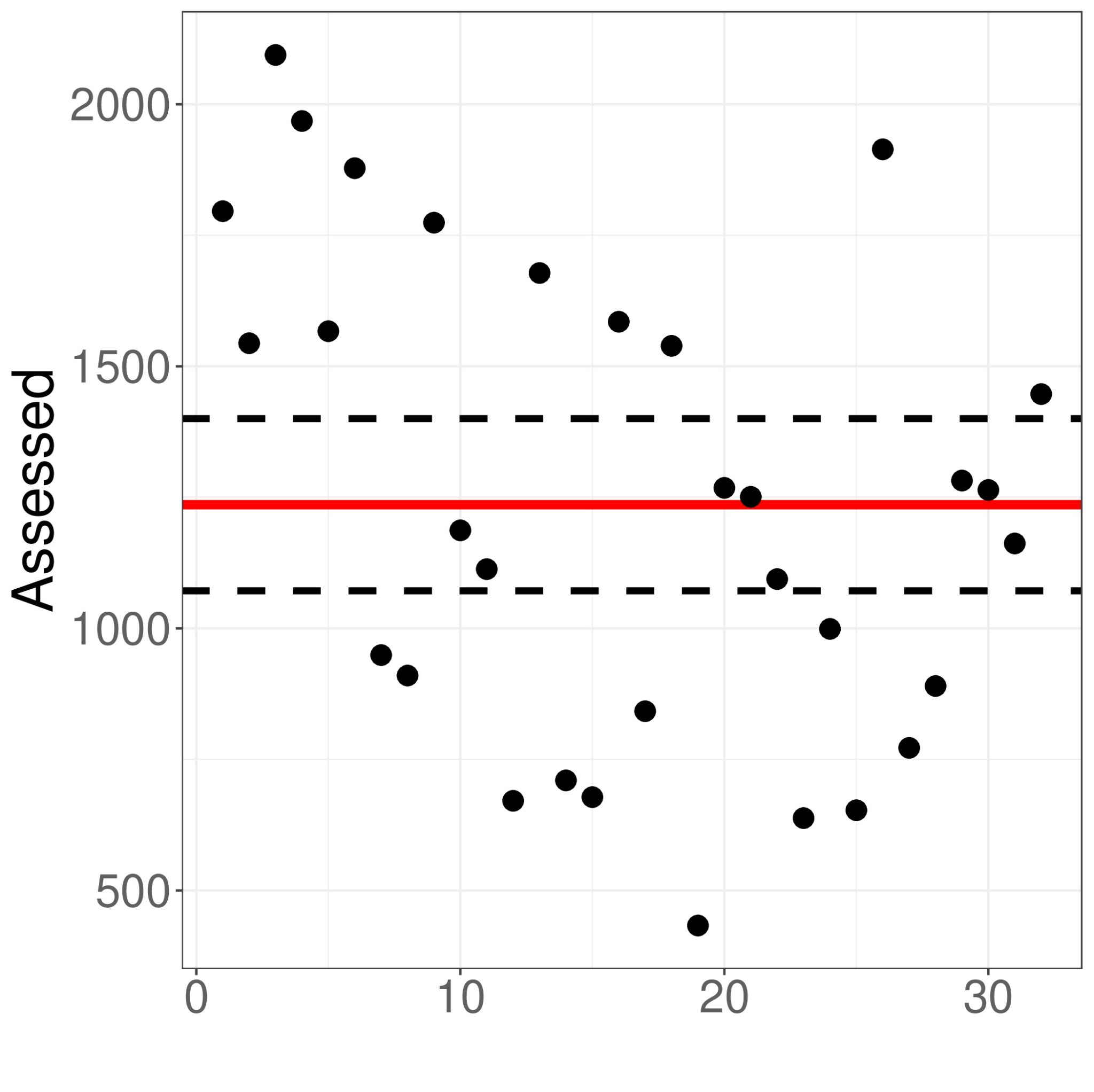


Após observar estas informações, é possível afirmar que o modelo se ajusta bem aos dados, embora o teste de variâncias encontre um possível problema na variável “Age” para o ajuste geral do modelo.

**2) f)** Ao utilizar o erro padrão, o intervalo de confiança observado não consegue acomodar a maioria dos dados, mesmo porque, sua interpretação diz respeito à diferença entre o ajuste do modelo amostral e o populacional, pois a medida de dispersão é ponderada pelo tamanho amostral.

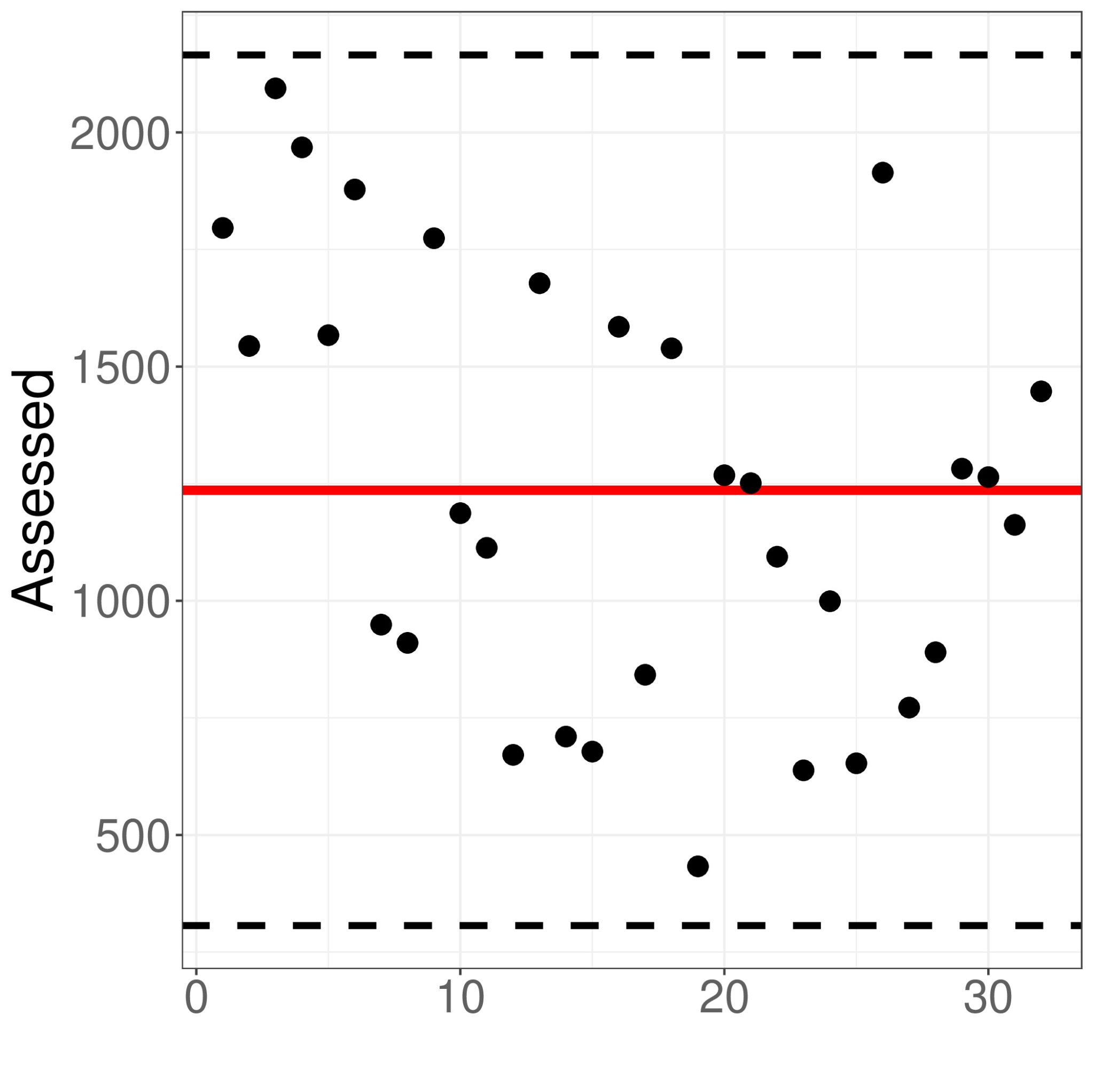
Neste caso, a interpretação do intervalo de confiança do erro padrão é de que a média populacional tem 95% de chance de estar contida na área marcada por linhas pontilhadas.

**GRÁFICO 4 - Intervalo de confiança do Erro Padrão**



Já ao usar o desvio padrão, é possível ver que os dados se encaixam bem ao intervalo de confiança, pois a interpretação teórica do desvio padrão diz respeito exclusivamente à distribuição dos dados. Ou seja, a interpretação é que 95% dos dados populacionais estão contidos na área marcada por linhas pontilhadas.

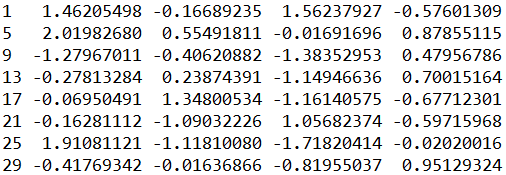
**GRÁFICO 5 - Intervalo de confiança do Desvio Padrão**



Deste ponto de vista, é possível afirmar que, ao assumir um nível de significância de 5%, a média simples pode ser representativa da população.

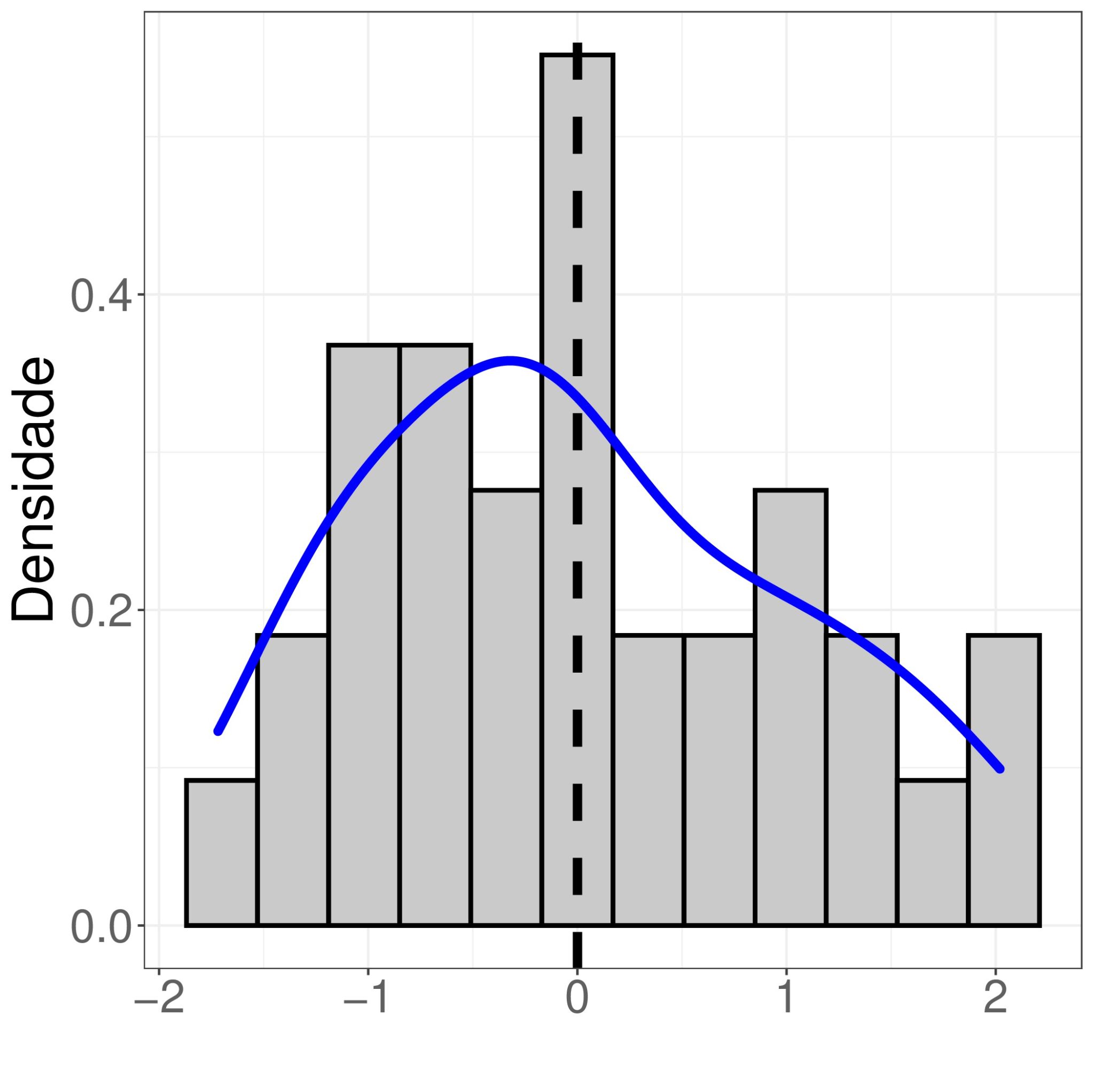
**2) g)** Todos os erros padronizados estão contidos no intervalo entre 2 e -2, exceto pelo quinto erro observado, que está a aproximadamente 2,02 erros padrão do ajuste do modelo, caracterizado como discrepante.

**TABELA 5 - Resíduos padronizados**



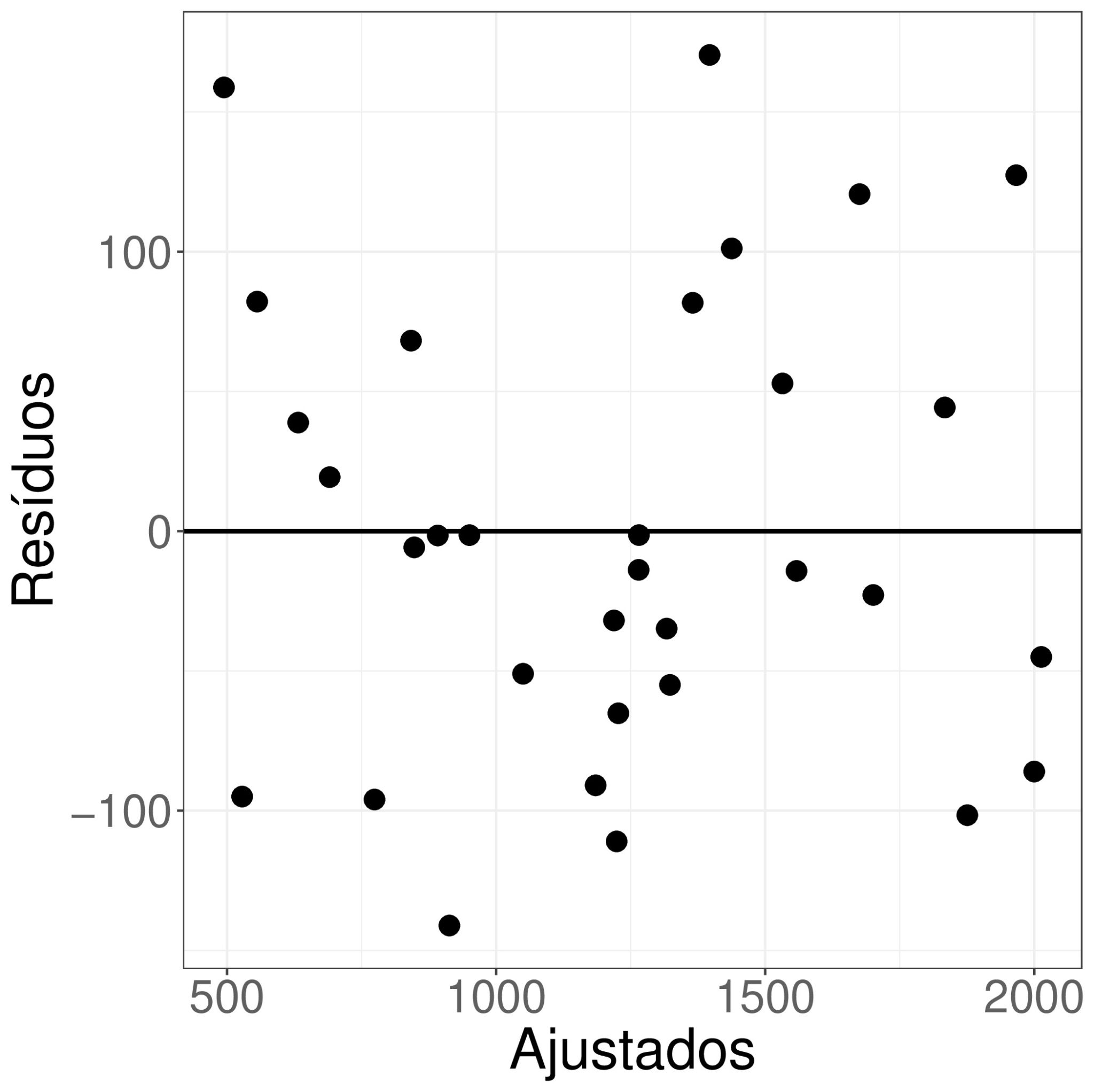
**2) h)** Ao observar o GRÁFICO 6, é possível perceber que os resíduos padronizados se ajustam de maneira semelhante, mas não suficientemente adequada à distribuição normal, uma vez que sua média não está alinhada com o topo da distribuição (vide linha azul de densidade), e que a simetria não é respeitada de maneira adequada.

**GRÁFICO 6 - Densidade e histograma**



**2) i)** O gráfico de Resíduos vs Valores ajustados (GRÁFICO 7) não indica uma presença relevante de heterocedasticidade no modelo.

**GRÁFICO 7 - Resíduos vs Valores Ajustados**



**2) j)** Não.