VERIFICAÇÃO DA ADEQUAÇÃO DO MODELO DE ANÁLISE DE VARIÂNCIA
ANÁLISE DE RESÍDUOS

Conforme foi apresentado anteriormente, o modelo de análise de variância assume que as observações são independentes e normalmente distribuídas com a mesma variância para cada tratamento.

A validade da suposição de independência pode ser verificada
por meio do gráfico de resíduos contra o tempo
(ordem de coleta das observações).

GRÁFICO DE RESÍDUOS CONTRA O TEMPO

Se neste gráfico os resíduos estiverem situados, aproximadamente, em torno de uma faixa horizontal centrada em $e_i = 0$, então será obtida uma indicação da validade da suposições de independência.

Por outro lado, a presença de seqüências de resíduos positivos e negativos ou padrões de alternância de sinais podem indicar que as observações ***não são independentes***.

RESÍDUO

Um resíduo é definido como:

$$\boxed{e_{ij} = x_{ij} - \bar{x}_i}$$

Isto é, um resíduo é uma diferença entre uma observação e a média do tratamento correspondente.

Configurações especiais no gráfico de resíduos contra a ordem de obtenção dos dados indicam que as observações não são independentes.

Note que a violação da suposição de independência dos erros ε_{ij} pode exercer sérios efeitos sobre a validade das inferências realizadas por meio da análise de variância.

Como esse é um problema difícil de ser corrigido, é importante tentar impedir a sua ocorrência. Geralmente, o emprego de uma aleatorização adequada para a coleta dos dados faz com que a condição de independência não seja violada.

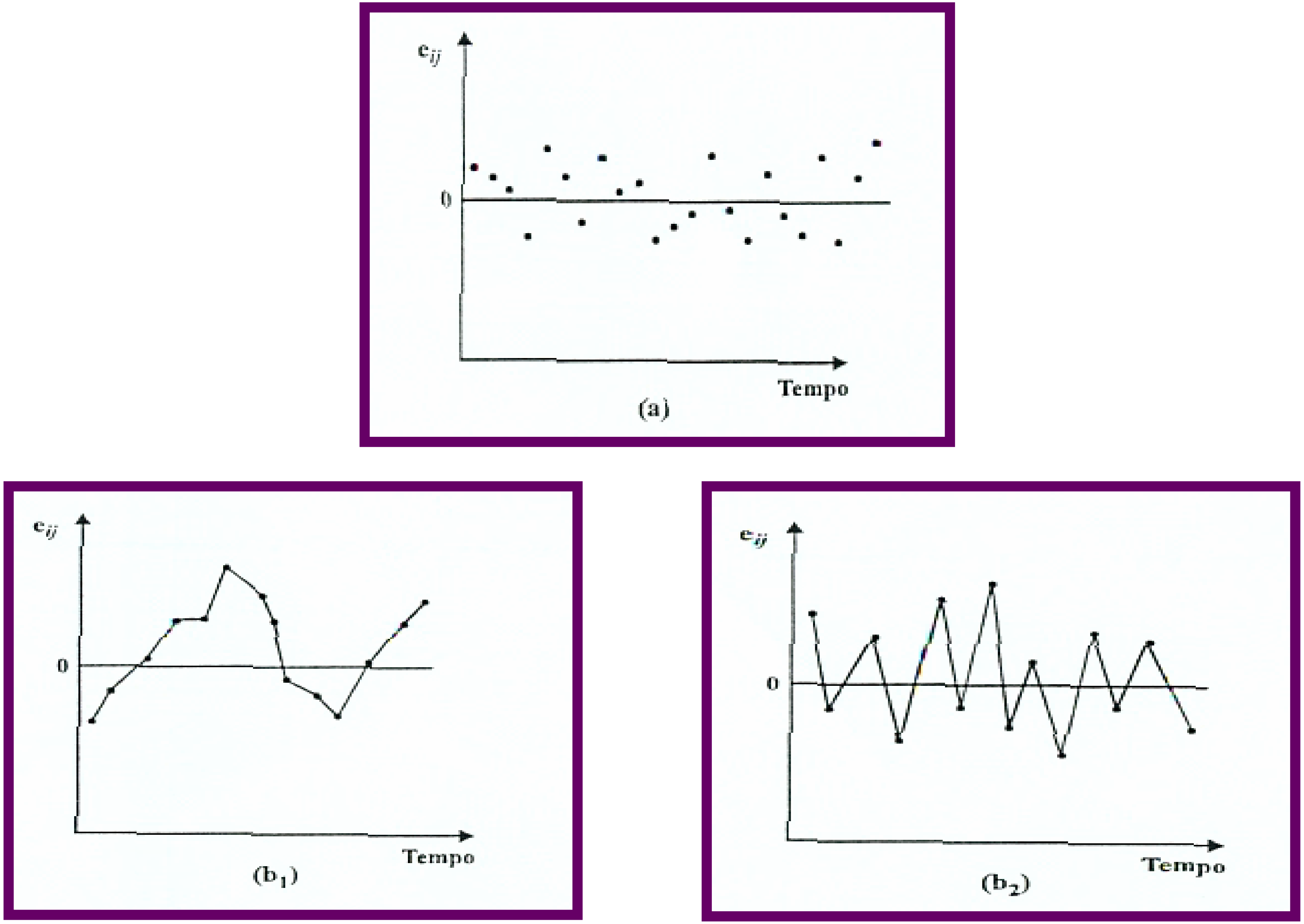


Figura 1: Exemplos de Gráficos de Resíduos contra o Tempo Indicando: (a) Validade da Suposição de Independência.
(b) Violação da Suposição de Independência.

GRÁFICO DE RESÍDUOS CONTRA AS MÉDIAS DOS TRATAMENTOS

Para avaliar a validade da suposição de igualdade de variâncias em todos os níveis do fator, devemos traçar o gráfico dos resíduos contra médias e analisar a dispersão dos resíduos. Se a suposição é válida, esta dispersão não deve depender do valor de $\overline{x}_i$ a seguir:

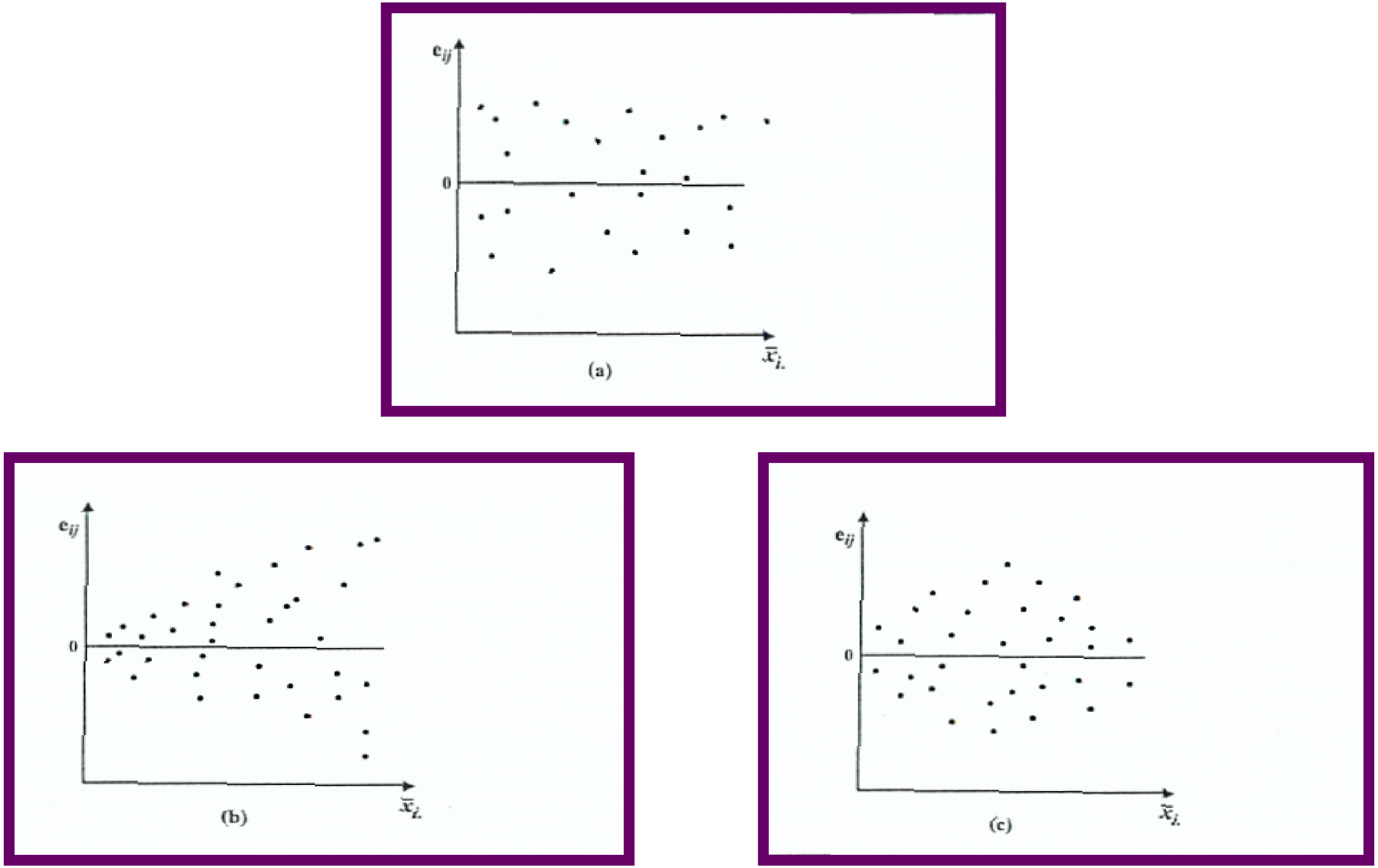


Figura 2 : Padrões para os Gráficos de Resíduos contra as Médias: (a) Satisfatório, (b) Funil, (c) Laço Duplo

Se a faixa de dispersão no gráfico de resíduos contra as médias dos tratamentos depender do valor de $\overline{x}_i$, terá sido obtida uma indicação de que a suposição de igualdade de variâncias não é valida.

A abordagem mais usual para lidar com situações onde a variação não é constante, consiste em utilizar *transformações para estabilizar a variância* e então aplicar as técnicas já ensinadas aos dados transformados.

É importante notar que, neste caso, as conclusões da análise de variância se aplicam aos dados transformados e devem ser estendidas com cuidado aos dados originais.

A análise de variância com amostras de mesmo tamanho, o teste F será somente ligeiramente afetado, caso a suposições de igualdade de variâncias seja violada.

GRÁFICO DE PROBABILIDADE NORMAL

A validade da suposição de normalidade pode ser verificada por meio de um gráfico de probabilidade normal para os resíduos. Neste gráfico cada resíduo é representado em função de seu valor esperado, o qual é calculado supondo que os resíduos seguem uma distribuição normal.

Será considerado que a suposições de normalidade é válida se os pontos do gráfico estiverem localizados, aproximadamente, ao longo de uma linha reta. Na visualização da linha reta, devem ser enfatizados os valores centrais do gráfico e não es extremos.

Se no gráfico de probabilidade normal os pontos não estiverem localizados, aproximadamente, ao longo de uma linha reta, terá sido obtida uma indicação de que a suposição de normalidade não é válida.

Para determinar os valores esperados dos resíduos ordenados, supondo que a condição de normalidade seja válida, serão utilizados os seguintes resultados:

- O valor esperado dos erros ε_{ij} é zero para o modelo de análise de variância considerado.

- O desvio padrão dos erros ε_{ij} é estimado por

$$\boxed{\sqrt{QM_{Res}}}$$

É possível demonstrar que, para uma variável aleatória com distribuição normal de média zero e desvio padrão estimado por $\boxed{\sqrt{QM_{Res}}}$, é uma boa aproximação para o valor esperado da *i*-ésima menor observação (E_i), em uma amostra aleatória de tamanho *n*, é :

$$E_i = \sqrt{QMR}\left[z\left(\frac{i - 0{,}375}{n + 0{,}25} \right) \right]$$

onde *z (A)* representa o 100(*A*)º percentil da distribuição normal padronizada, isto é,

$P[z \leq z(A)] = A$ se z ~ N(0,1).

Como a visualização da aproximação dos pontos por uma linha reta é subjetiva, também devemos calcular o coeficiente de correlação linear (r) entre os resíduos (e_i) e seus valores esperados sob a suposição de normalidade (E_i).

A expressão para o cálculo do coeficiente de correlação linear é dada por:

$$r = \frac{\sum_{i=1}^{n}(E_i - \bar{E})(e_i - \bar{e})}{\left[\sum_{i=1}^{n}(e_i - \bar{e})^2 \sum_{i=1}^{n}(E_i - \bar{E})^2\right]^{1/2}}$$

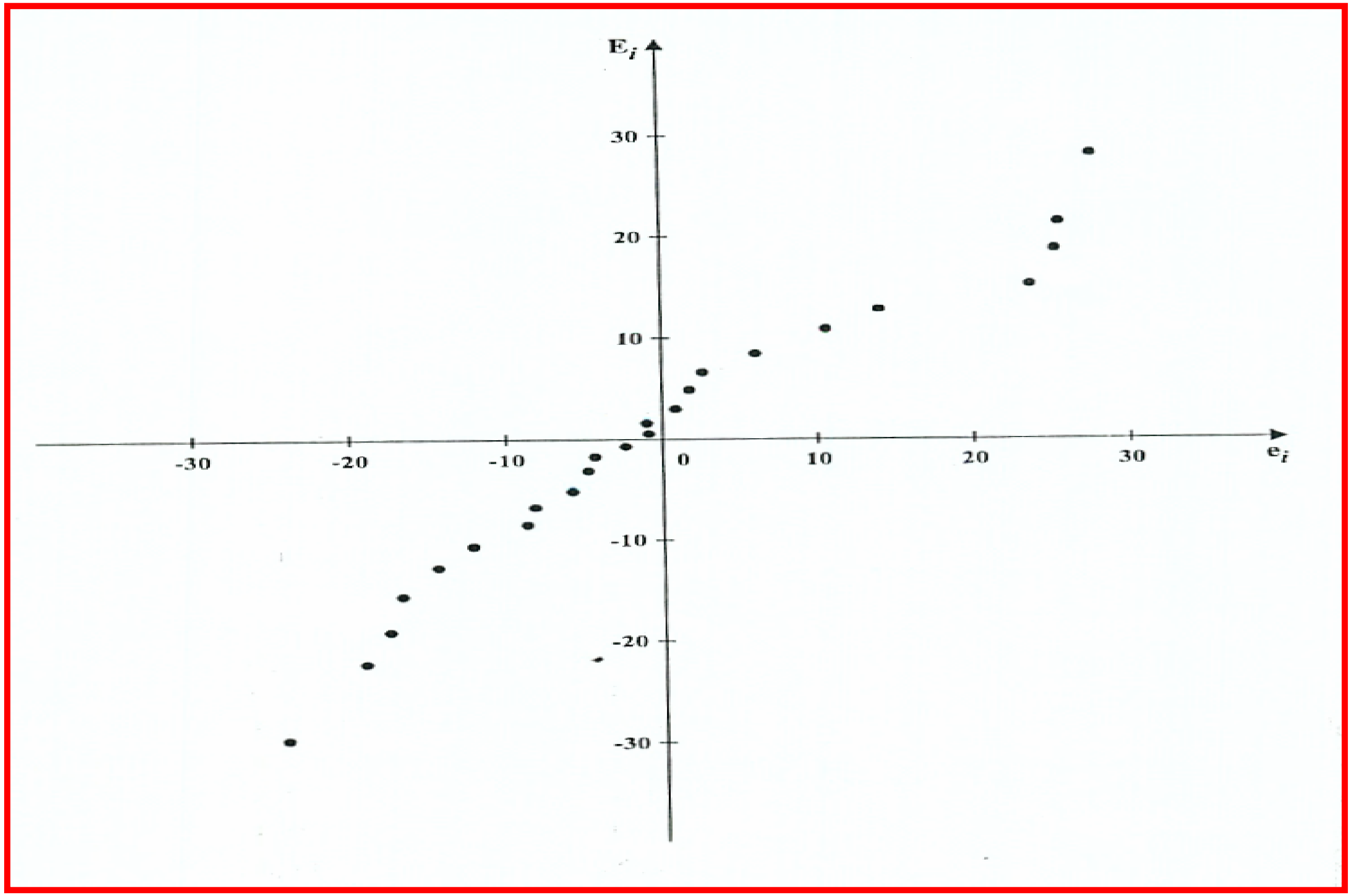
onde:

$$\bar{e} = \frac{1}{n}\sum_{i=1}^{n} e_i \qquad \bar{E} = \frac{1}{n}\sum_{i=1}^{n} E_i$$

Teoricamente $\boxed{\overline{e} = \overline{E} = 0}$, mas devido a aproximações realizadas durante os cálculos dos valores para e_i e E_i, estas médias podem não ser iguais a zero.

Um valor próximo de 1 para o coeficiente de correlação linear, é uma indicação da existência de um forte relacionamento linear entre os resíduos e seus valores esperados sob a suposição de normalidade, o que significa dizer que a suposição de normalidade pode ser considerada satisfeita.

Gráfico de Probabilidade Normal para os resíduos



É importante destacar que, se a distribuição do erro ε não segue uma distribuição normal, apresentando pequenos desvios em relação a esta distribuição, este fato não exerce grandes efeitos sobre o teste F, os intervalos de confiança para as médias e o método de Duncan de comparações múltiplas.

As violações das suposições de normalidade e variância constante geralmente ocorrem simultaneamente.

Felizmente, na maioria dos casos, a mesma transformação utilizada para estabelecer a variância também faz com que a distribuição dos erros se aproxime da distribuição normal.

Portanto, se a ocorrência destas violações for constatada, em primeiro lugar deve ser empregada uma transformação para estabilizar a variância e a seguir os resíduos devem ser analisados com o objetivo e verificar se a suposição de normalidade ainda continua sendo violada.

Verificação da Adequação do Modelo de Análise de Variância

- Resíduos $e_{ij} = x_{ij} - \bar{x}_i$
- Gráficos de Resíduos

Gráfico	Para avaliar a validade da suposição de
Resíduos contra o tempo	Independência
Resíduos contra médias $\overline{X}_i$	Variância constante
Prob. Normal para os resíduos	Normalidade

Teste de normalidade baseado no coeficiente de correlação linear.