**Atividade Pré-Avaliação: 2ª Prova**

O conjunto de dados anexado corresponde a um ensaio clínico realizado pela "Radiation Therapy Oncology Group" dos Estados Unidos. O estudo é composto por pacientes com câncer na boca e garganta. Os pacientes que entraram no estudo foram aleatoriamente atribuídos a um de dois grupos de tratamento, terapia somente com radiação (padrão) ou a terapia de radiação em conjunto com um agente quimioterápico (teste). Cada pacientes foi observado até vir a óbito, e o tempo de sobrevivência do diagnóstico da doença até o óbito foi calculada em dias (Tempo). Além disso, verificou-se o grau em que a célula do tumor se assemelha a célula hospedeira (Grau), o local em que o tumor estava alojado ao iniciar o tratamento (Local) e a idade do paciente no dia do diagnóstico da doença (Idade). Abaixo segue a descrição das variáveis observadas nesse ensaio clínico:

1) Tratamento: 1=padrão, 2=teste;

2) Tempo: Tempo de sobrevivência a partir do diagnóstico da doença (em dias);

3) Grau: 1=bem diferenciado, 2=moderadamente diferenciado, 3=pouco diferenciado;

4) Local: 1=arco faucial, 2=fossa tonsilar, 4=faringe;

5) Idade do paciente no dia do diagnóstico da doença (em anos).

Utilizando essas informações testar hipóteses de tal forma que seja possível verificar se:

a) O tempo de sobrevivência (Tempo) é afetado pelo tipo de tratamento (Tratamento);

b) O tempo de sobrevivência (Tempo) é afetado pelo grau em que a célula do tumor se assemelha a célula hospedeira (Grau);

c) O tempo de sobrevivência (Tempo) é afetado pelo local em que o tumor estava alojado ao iniciar o tratamento (Local);

d) O tempo de sobrevivência (Tempo) está correlacionado com a  idade do paciente no dia do diagnóstico da doença (Idade);

e) O grau em que a célula do tumor se assemelha a célula hospedeira (Grau) está associada ao local em que o tumor estava alojado ao iniciar o tratamento (Local).

**OBS: Para todos os itens definir as hipóteses que serão testadas (H0 e H1); citar o teste paramétrico e sua versão não-paramétrica necessário para resolver o problema e explicar o motivo da escolha; estipular um nível de significância explicando o motivo da escolha; calcular todos os P-valores (interpretar o resultado); se necessário calcular o intervalo de confiança (interpretar o resultado); fazer uma conclusão final. (Os gráficos de normalidade e de análise resídual devem ser mostrados).**

**RESOLUÇÃO**

* **ITEM (A)**

**Para verificarmos se o tempo de sobrevivência é afetado pelo tipo de tratamento utilizamos o teste de hipótese:**

Onde  **é a média do tempo de sobrevivência para pacientes que usaram o tratamento tipo 1 (padrão - tratamento com radiação) e sendo a média do tempo de sobrevivência para pacientes que usaram o tratamento tipo 2 (teste - tratamento com radiação em conjunto com um agente quimioterápico). Tendo em mente que estamos realizando uma análise de variáveis quantitativas *vs* qualitativas, utilizaremos o teste paramétrico teste-t para variâncias desconhecidas, pois os tempos de tratamento dos pacientes são independentes entre si, visto que o mesmo paciente não foi submetido a ambos os tipos de tratamento e as variâncias populacionais são desconhecidas. A versão não paramétrica para o teste-t é o teste** Wilcoxon-Mann-Whitney.

Para definirmos (nível de significância) é necessário analisar os possíveis impactos de um erro do tipo 1 (rejeitar  **sendo ela verdadeira). Neste caso, cogitando a rejeição de é possível que haja uma redução no tempo de vida como consequência do uso de agentes quimioterápicos, sabendo que isso implicaria em mortes prematuras, se trata de um erro grave, logo é pertinente utilizar** .

Para que seja utilizado o teste paramétrico é necessário que as variáveis em seus respectivos grupos sigam distribuição normal, logo nós aplicamos o gráfico de dispersão de quartis e o teste de Shapiro Wilk para os tipos de tratamento.

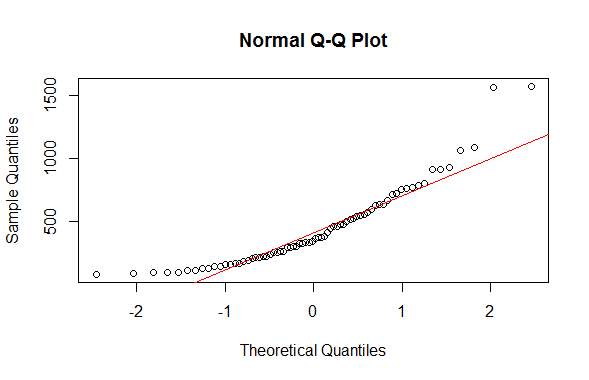


Figura 1 - Gráfico de dispersão de quartis do tempo de vida (dias) para o tratamento do tipo 1, a reta em vermelho representa um modelo de dispersão ideal.

Aplicando o teste de Shapiro Wilk encontramos o p-valor = . Comparando o p-valor com , rejeitamos a  **(**p-valor < ). Portanto é necessário que seja utilizado o teste Wilcoxon-Mann-Whitney.

**(CASO var.test VERDADEIRO)** Realizando o teste-t obtemos o p-valor maior que o nível de significância, portanto é aceita a  **e pode-se dizer com 5% de significância que as médias populacionais para o tempo de cada tratamento são diferentes. Com 95% de confiabilidade e a diferença entre as médias populacionais está entre 26,58 e 202,14 dias.**

**Realizando a versão não paramétrica do teste-t obtemos um p-valor = 0,08 (p-valor >** ), portanto as médias do tempo de vida para ambos os tratamentos são iguais, com isso, é possível afirmar que o tipo de tratamento não interfere no tempo de vida.

* **ITEM (B)**

**Para verificarmos se o tempo de sobrevivência é afetado pelo grau em que o tumor e as células hospedeiras se assemelham utilizamos o teste de hipótese:**

Onde  **é a média do tempo de sobrevivência para pacientes que possuíam semelhança de** grau 1 (bem diferenciado),  **para grau 2** (moderadamente diferenciado) e  **para o grau** 3 (pouco diferenciado)**. Tendo em mente que estamos realizando uma análise de variáveis quantitativas *vs* qualitativas para mais de dois grupos, utilizaremos o teste ANOVA *One Way*. A versão não paramétrica da ANOVA *One Way* é o teste** Kruskal-Wallis.

Para definirmos (nível de significância) é necessário analisar os possíveis impactos de um erro do tipo 1 (rejeitar  **sendo ela verdadeira). Neste caso, cogitando a rejeição de , na pior das hipóteses, seria adotado um tipo de tratamento específico para cada nível de compatibilidade, logo, tenderia a melhora do paciente mesmo que a compatibilidade não intervisse em seu tempo de vida. Com isso, é possível afirmar que se trata de um erro brando, entretanto, por ser uma análise clínica, será utilizado um** .

Para que seja utilizado o teste ANOVA *One Way* é necessário fazer uma análise residual e constatar que os resíduos seguem distribuição normal, são homocedásticas e independentes. Ao verificarmos a primeira exigência obtivemos o seguinte gráfico:

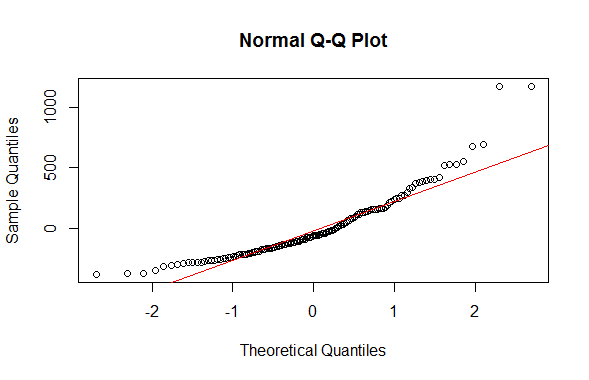


Figura 2 - Gráfico de dispersão de quartis dos resíduos sabendo que reta em vermelho representa um modelo de dispersão ideal.

Aplicando o teste de Shapiro Wilk encontramos o p-valor = . Comparando o p-valor com , rejeitamos a  **(**p-valor < ). Portanto é necessário que seja utilizada o teste de Kruskal-Wallis.

**(CASO ANOVA VERDADEIRO) Após analisarmos a normalidade dos resíduos, a independência e a homocedasticidade realizamos o teste de Turkey como pós teste para constatarmos ... .**

**Realizando a versão não paramétrica do ANOVA *One Way* obtivemos um p-valor = 0.97 (p-valor > )** portanto as médias do tempo de vida para todos os graus de compatibilidade são iguais, com isso, é possível afirmar que o grau de compatibilidade entre as células hospedeiras e o câncer não interfere no tempo de vida.

* **ITEM (C)**

**Para verificarmos se o tempo de sobrevivência é afetado pelo local do tumor utilizamos o teste de hipótese:**

Onde  **é a média do tempo de sobrevivência para pacientes que possuíam tumor no local 1** (arco faucial),  **para o local 2** (fossa tonsilar) e  **para o local 4** (faringe)**. Tendo em mente que estamos realizando uma análise de variáveis quantitativas *vs* qualitativas para mais de dois grupos, utilizaremos o teste ANOVA *One Way*. A versão não paramétrica da ANOVA *One Way* é o teste** Kruskal-Wallis.

Para definirmos (nível de significância) é necessário analisar os possíveis impactos de um erro do tipo 1 (rejeitar  **sendo ela verdadeira). Neste caso, cogitando a rejeição de , na pior das hipóteses, seria adotado um tipo de tratamento específico para cada local diferente onde o tumor está instalado, logo, tenderia a melhora do paciente mesmo que o local do câncer não intervisse em seu tempo de vida. Com isso, é possível afirmar que se trata de um erro brando, entretanto, por ser uma análise clínica, será utilizado um** .

Para que seja utilizado o teste ANOVA *One Way* é necessário fazer uma análise residual e constatar que os resíduos seguem distribuição normal, são homocedásticas e independentes. Ao verificarmos a primeira exigência obtivemos o seguinte gráfico:

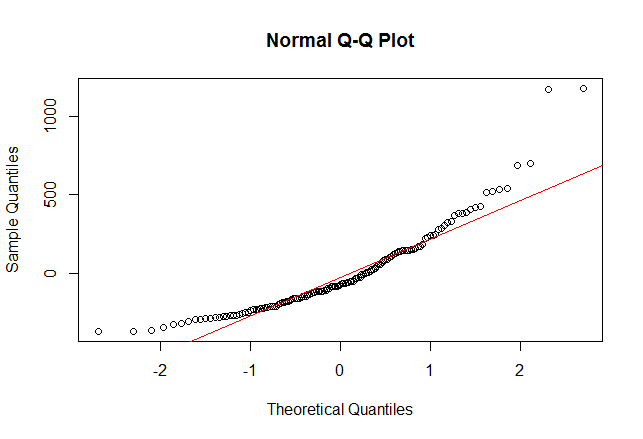


Figura 3 - Gráfico de dispersão de quartis dos resíduos sabendo que reta em vermelho representa um modelo de dispersão ideal.

Aplicando o teste de Shapiro Wilk encontramos o p-valor = . Comparando o p-valor com , rejeitamos a  **(**p-valor < ). Portanto é necessário que seja utilizada o teste de Kruskal-Wallis.

**(CASO ANOVA VERDADEIRO) Após analisarmos a normalidade dos resíduos, a independência e a homocedasticidade realizamos o teste de Turkey como pós teste para constatarmos ... .**

**Realizando a versão não paramétrica do ANOVA *One Way* obtivemos um p-valor = 0.97 (p-valor >** ) portanto as médias do tempo de vida para todos os os locais de câncer são iguais, com isso, é possível afirmar que o local onde o câncer está instalado não interfere no tempo de vida.

* **ITEM (D)**

**Para verificarmos se o tempo de sobrevivência está correlacionado com a idade do paciente utilizamos os testes de hipótese:**

Onde **é o coeficiente de correlação entre o tempo de vida e a idade do paciente. Tendo em mente que estamos realizando uma análise de variáveis quantitativas *vs* quantitativas é pertinente utilizar um dos testes de correlação entre variáveis: Pearson, Spearman ou Kendall.**

Para definirmos (nível de significância) é necessário analisar os possíveis impactos de um erro do tipo 1 (rejeitar  **sendo ela verdadeira). Neste caso, cogitando a rejeição de é possível que haja uma redução no tempo de vida com dependência direta da idade do paciente, logo, se ignorado este fator, pode acarretar mortes prematuras, então, por se tratar de um erro grave, utilizaremos = 0,01. O teste paramétrico para este caso é o teste de Pearson e, em contrapartida, os não paramétricos são os testes de Kendall e Spearman.**

Para que seja utilizado o teste de Pearsoné necessário fazer uma análise residual e constatar que os resíduos seguem distribuição normal, são homocedásticas e independentes. Ao verificarmos a primeira exigência obtivemos o seguinte gráfico:

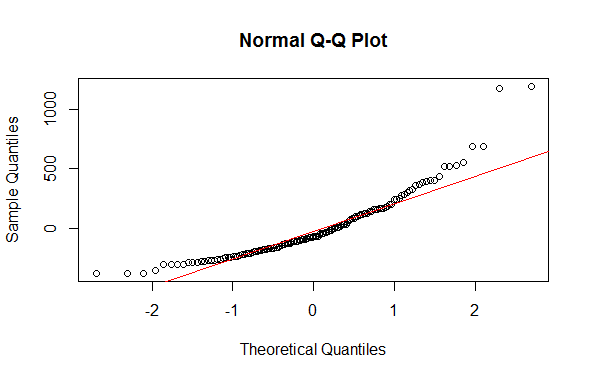


Figura 4 - Gráfico de dispersão de quartis dos resíduos sabendo que reta em vermelho representa um modelo de dispersão ideal.

Aplicando o teste de Shapiro Wilk encontramos o p-valor = . Comparando o p-valor com , rejeitamos a  **(**p-valor < ). Portanto, já que se trata de um caso com 0,01, iremos utilizar o teste de Kendall.

**(CASO Resíduos VERDADEIRO) Após analisarmos a normalidade dos resíduos, a independência e a homocedasticidade realizamos o teste de Turkey como pós teste para constatarmos ... .**

**Realizando o teste de Kendall obtivemos um p-valor = 0.34 (p-valor >** ) portanto é possível afirmar que não há correlação entre o tempo de vida do paciente e sua idade.

* **ITEM (E)**

**Para verificarmos se há associação entre o grau de compatibilidade das células hospedeira e o tumor com o local onde o tumor está instalado utilizamos os seguintes testes de hipótese:**

**Tendo em mente que estamos realizando uma análise de variáveis qualitativas *vs* qualitativas é pertinente utilizar os testes de regressão logística como teste paramétrico ou o teste do Chi-quadrado como teste não paramétrico.**

Para definirmos (nível de significância) é necessário analisar os possíveis impactos de um erro do tipo 1 (rejeitar  **sendo ela verdadeira). Neste caso, cogitando a rejeição de** , **na pior das hipóteses, seria adotado um tipo de tratamento específico para cada nível de compatibilidade relacionando-o com o local do tumor, logo, tenderia a melhora do paciente mesmo que a compatibilidade não tivesse relação com o local em que o tumor está instalado. Com isso, é possível afirmar que se trata de um erro brando, entretanto, por ser uma análise clínica, será utilizado um** .

Por conta da finalidade deste curso de Probabilidade e Estatística não será utilizado o teste de regressão logística por conta de não ser abordado conteúdo suficientemente aprofundado para a análise de casos para utilização do mesmo, com isso, será utilizado sua versão não paramétrica: o teste do Chi-quadrado.

**Realizando o teste do Chi-quadrado obtivemos um p-valor = 0.02 (p-valor <** ), portanto é possível afirmar que **há associação entre o grau de compatibilidade das células hospedeira e o tumor com o local onde o tumor está instalado.**