4.5 Teste de Comparação Entre Duas Médias

Supor o interesse em comparar duas populações com relação às suas médias. Ou seja, o interesse é verificar se uma variável qualitativa com dois níves tem efeito sobre uma variável quantitativa. Portanto, o objetivo será testar as seguintes hipóteses:

$$\left\{ \begin{array}{l} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{array} \right. \Rightarrow \left\{ \begin{array}{l} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \end{array} \right. \tag{4.14}$$

Para testar essas hipóteses, se a variável aleatória quantitativa seguir distribuição normal dentro dos dois níveis da variável qualitativa, quatro testes podem ser utilizados, como mostrado no esquema a seguir:

dependentes
(teste-t pareado)

$\nearrow$

2 amostras

$\searrow$ independentes

$\nearrow$ variâncias conhecidas (teste-z)

$\searrow$ variâncias desconhecidas

$\nearrow$ variâncias iguais (teste-t $\sigma_1^2 = \sigma_2^2$)

$\searrow$ variâncias diferentes (teste-t $\sigma_1^2 \neq \sigma_2^2$)

Portanto, para optar por um entre os quatro testes a serem utilizados, primeiramente é necessário verificar se as amostras obtidas pelo experimento são dependentes (pareadas) ou independentes (não pareadas). Duas amostras serão consideradas dependentes quando os indivíduos de ambas as amostras são os mesmos. Por exemplo, para avaliar o nível de tensão ocasionado por exames escolares, n alunos foram escolhidos e sua pulsação medida antes e depois do exame. Veja que, nesse caso, temos duas amostras, amostra da pulsação antes da realização de um exame e amostra da pulsação depois da realização do exame, no entanto os mesmos indivíduos foram mensurados antes e depois da realização do exame, logo, essas amostras são pareadas. Se as amostras forem dependentes o teste que será utilizado é o **teste-t pareado para comparação entre médias**.

Se as amostras forem independentes, o segundo passo é verificar se as variâncias de ambas as populações são conhecidas. Se essas variâncias populacionais forem conhecidas será utilizado o **teste-z para comparação entre médias**.

No caso de as amostras serem independentes com as variâncias populacionais desconhecidas, o terceiro passo é verificar se as variâncias das duas populações são iguais ou diferentes. Para verificar se essas variâncias são iguais ou diferentes deve-se realizar o **teste-F para comparação entre variâncias**, esse teste é utilizado para testar as seguintes hipóteses:

$$\left\{ \begin{array}{l} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{array} \right. \tag{4.15}$$

Após a realização do teste-F para comparação entre variâncias, se a hipótese nula não for rejeitada o teste a ser utilizado é o **teste-t para amostra independentes com variâncias populacionais iguais**. Caso a hipótese nula do teste-F for rejeitada o teste a ser utilizado é o **teste-t para amostra independentes com variâncias populacionais diferentes**.

A seguir alguns exemplos serão apresentados para ilustrar a aplicação desses testes.

4.5.1 Exemplo 1

Supor o interesse em verificar o efeito de região (central e oeste) no preço do metro quadrado praticado em um município. Para isso foram selecionados 20 imóveis da região central e 18 da região oeste, sendo o preço do metro quadrado avaliado para cada imóvel em reais. No entanto, o pesquisador tem a informação da variabilidade populacional entre o preço do metro quadrado praticado nas duas regiões, sendo elas $\sigma_{RC} = R\$71$ e $\sigma_{RO} = R\$82$ para as regiões central e oeste, respectivamente. Para a entrada do conjunto de dados no `Software R` a função `c` foi utilizada para a criação dos vetores contendo os valores do preço do metro quadrado praticado em ambas as regiões, como segue:

```
RC <- c(4120,4050,3960,3940,3890,3910,4090,4120,4040,4060,4030,3920,4060,3970,4030,4090,
        3960,3970,4000,4120)
RO <- c(3720,3490,3810,3540,3570,3770,3640,3660,3610,3740,3610,3590,3690,3740,3750,3800,
        3680,3640)
```

Seguindo os passos vistos na Seção 4.3 para a construção de testes de hipóteses, tem-se:

1. **Identificar H_0 e H_1**: Como o interesse é verificar se o preço do imóvel é afetado pela região em que ele se encontra, tem-se o interesse em testar as seguintes hipóteses,

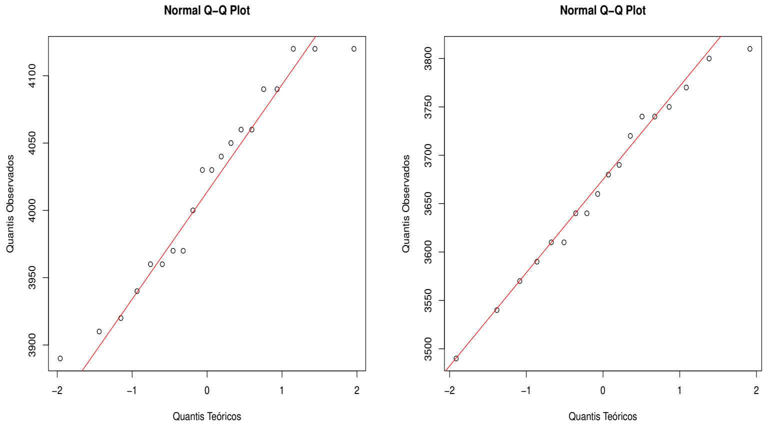
$$\begin{cases} H_0 : \mu_{RC} = \mu_{RO} \\ H_1 : \mu_{RC} \neq \mu_{RO} \end{cases}, \tag{4.16}$$

   em que, μ_{RC} é o preço médio do metro quadrado praticado em imóveis da região central e μ_{RO} é o preço médio do metro quadrado praticado em imóveis da região oeste.

2. **Escolher o teste estatístico**: Para testar esse tipo de hipóteses temos que optar por um entre os quatro testes de comparação entre médias que foram vistos na Seção 4.5. O primeiro passo é verificar se as variáveis são independentes ou não, como nesse caso os imóveis da região central são diferentes dos imóveis da região oeste, então as amostras são independentes. O segundo passo, então, é verificar se as variâncias populacionais são conhecidas, como nesse caso foi informado os valores dessas duas variabilidades ($\sigma_{RC} = R\$71$ e $\sigma_{RO} = R\$82$), as variâncias populacionais são conhecidas. Portanto, aqui será utilizado o **teste-z para comparação entre médias**.

3. **Fixar o nível de significância α**: Cometer o erro do Tipo I aqui é dizer que, de acordo com a amostra o preço médio do metro quadrado praticado em imóveis da região central é diferente do preço médio do metro quadrado praticado em imóveis da região oeste, quando na população os preços médios do metro quadrado praticados nas duas regiões é o mesmo. Supor que o grau de gravidade em cometer esse erro é brando, o nível de significância aqui será estipulado em $\alpha = 10\%$.

   Lembrado que esse teste tem como pressuposto que a variável aleatória quantitativa deve seguir distribuição normal para os dois grupos observados da variável qualitativa. Portanto, a partir dos gráficos da Figura 4.3 percebe-se que eles estão seguindo um comportamento linear acompanhando a linha vermelha teórica e como os p-valores do teste de Shapiro-Wilk em ambos os grupos, $p-valor_{RC} = 0,2804$ e $p-valor_{RO} = 0,8846$, é maior do que o nível de significância do teste ($\alpha = 10\%$), a hipótese de normalidade dos dados não é rejeitada (a obtenção do gráfico quantil-quantil e do teste de Shapiro-Wilk utilizando o `Software R` já foi abordada nesse texto). Portanto, o teste-z para comparação entre médias poderá ser utilizado para resolver esse problema.



(a) Região Central (b) Região Oeste

Figura 4.3: Gráficos quantil-quantil para verificar se o preço praticado em duas regiões de um município segue distribuição normal.

4. **Calcular os valores observados para o teste estatístico a partir dos dados amostrais**: Para realizar o **teste-z para comparação entre médias** utilizando o `Software R`, a função `z.test` será utilizada e a seguinte linha de comando deve ser executada (lembrando que para utilizar a função `z.test` a biblioteca `BSDA` deve estar instalada e deve ser carregada no `Software R`):

   z.test(RC,RO,sigma.x=71,sigma.y=82,conf.level=0.90)

   Os dois primeiros argumentos da função `z.test` são os vetores contendo os dados das variáveis aleatórias quantitativas em que se tem o interesse em comparar as suas médias; no argumento `sigma.x` é definido qual é o valor do desvio padrão populacional da variável definida no primeiro argumento da função `z.test` (como nesse caso o primeiro argumento colocado na função foi o vetor contendo os preços do metro quadrado praticados na região central, logo, o argumento `sigma.x` deve receber o valor $\sigma_{RC} = R\$71$); no argumento `sigma.y` é definido qual é o valor do desvio padrão populacional da variável definida no segundo argumento da função `z.test` (como nesse caso o segundo argumento colocado na função foi o vetor contendo os preços do metro quadrado praticados na região oeste, logo, o argumento `sigma.y` deve receber o valor $\sigma_{RC} = R\$82$); se a entrada dos dados na função `z.test` for feita de forma matricial (isso será visto mais a frente) então `sigma.x` e `sigma.y` são definidos, respectivemente, de acordo com a ordem alfabética ou numérica dos níveis de classificação da variável qualitativa; o argumento `conf.level` define o nível de confiabilidade estipulado para o cálculo do intervalo de confiança (como no Item 2 foi estipulado um nível de significância de 10% então o nível de confiabilidade é de 90%, pois a confiabilidade nada mais é do que $1 - \alpha$). Executando essa linha de comando o `Software R` irá retornar o seguinte resultado:

   Two-sample z-Test

   data: RC and RO

   z = 13.876, p-value < 2.2e-16

   alternative hypothesis: true difference in means is not equal to 0

90 percent confidence interval:
305.9143 388.1968
sample estimates:
mean of x mean of y
4016.500 3669.444

5. **Verificar se rejeita ou não a hipótese nula** H_0: Observando os valores acima, o p-valor é dado por $p - valor = 2,2\times^{-16}$, como o nível de significância estipulado foi de $\alpha = 0,10$, então $p - valor < \alpha$. Portanto, com 10% de significância, existem evidencias de que o preço médio do metro quadrado praticado nos imóveis da região central é diferente do preço médio do metro quadrado praticado nos imóveis da região oeste, ou seja, existe um efeito de região no preço do metro quadrado praticado nesse município.

   O p-valor é utilizado para verificar se rejeitamos ou não rejeitamos uma hipótese nula de interesse, no entanto ele não quantifica o quanto essa hipótese foi rejeitada. Por exemplo, a partir do teste foi concluido que o preço médio do metro quadrado dos imóveis das regiões central e oeste são diferentes, no entanto não se sabe o quanto diferentes esses preços praticados são. Para quantificar essa diferença o intervalo de confiança gerado acima pode ser observado, esse intervalo se trata do intervalo de confiança para a diferença entre as médias populacionais (como *default* a função `z.test` retorna o intervalo de confiança para a média do vetor estipulado no primeiro argumento menos a média do vetor estipulado no segundo argumento). Logo, temos que $IC\left(\mu_{RC} - \mu_{RO}, 90\%\right) = [305,91; 388,20]$, portanto, com 90% de confiabilidade, levantamos evidência de que o preço médio do metro quadrado praticado na região central é de $R\$305,91$ a $R\$388,20$ maior (sinal positivo do intervalo de confiança), se comparado com a região oeste. Outro fato importante que podemos concluir observando esse intervalo de confiança é que o valor zero não se encontra dentro dos limites gerados pelo intervalo, isso implica que $\mu_{RC} - \mu_{RO} \neq 0 \Rightarrow \mu_{RC} \neq \mu_{RO}$ (hipótese nula rejeitada).

4.5.2 Exemplo 2

Supor o interesse em verificar o efeito da doença fibrose cística no nível sérico de ferro. Para isso foram selecionadas 13 crianças que sofrem de fibrose cística e 11 crianças que não sofrem de fibrose cística (saudáveis), para ambos os grupos o nível sérico de ferro de cada criança foi medido em $\mu mol/l$. Para a entrada do conjunto de dados no `Software R` a função `c` foi utilizada para a criação dos vetores contendo os valores do o nível sérico de ferro medido para as crianças que sofrem de fibrose cística e para as saudáveis, como segue:

fibrose.cistica <- c(13.78,18.01,19.65,19.66,12.64,11.53,13.44,18.31,15.4,14.92,10.06,14.62,13.97)
saudaveis <- c(26.04,30.42,11.53,29.08,19.82,16.01,20.39,21.46,6.95,20.5,30.2)

Seguindo os passos vistos na Seção 4.3 para a construção de testes de hipóteses, tem-se:

1. **Identificar** H_0 **e** H_1: Como o interesse é verificar se o nível sérico de ferro é afetado pelo fato da criança apresentar fibrose cística, tem-se o interesse em testar as seguintes hipóteses,

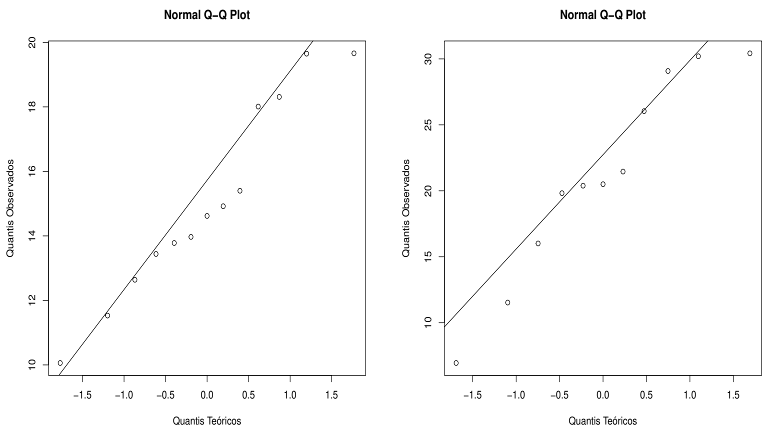
$$\begin{cases} H_0 : \mu_{fc} = \mu_s \\ H_1 : \mu_{fc} \neq \mu_s \end{cases}, \tag{4.17}$$

   em que, μ_{fc} é a média do nível sérico de ferro das crianças que sofrem de fibrose cística e μ_s a média do nível sérico de ferro das crianças saudáveis.

2. **Escolher o teste estatístico**: Para testar esse tipo de hipóteses temos que optar por um entre os quatro testes de comparação entre médias que foi visto na Seção 4.5. O primeiro passo é verificar se as variáveis são independentes ou não, como nesse caso as crianças que sofrem de fibrose cistica são diferentes das crianças saudáveis, então as amostras são independentes. O segundo passo, então, é verificar se as variâncias populacionais para ambos os grupos são conhecidas, como nesse caso não foram informados os valores dessas duas variabilidades populacionais, as variâncias populacionais são desconhecidas. O último passo, então, é realizar um teste-F para comparação entre variâncias (esse teste será realizado no Item 4). Se a hipótese nula não for rejeitada o teste a ser utilizado é o **teste-t para amostra independentes com variâncias populacionais iguais**. Caso a hipótese nula do teste-F for rejeitada o teste a ser utilizado é o **teste-t para amostra independentes com variâncias populacionais diferentes**.

3. **Fixar o nível de significância** α: Cometer o erro do Tipo I aqui é dizer que, de acordo com a amostra a média do nível sérico de ferro das crianças que sofrem de fibrose cística é diferente da média do nível sérico de ferro das crianças que não sofrem de fibrose cística, quando na população a média do nível sérico de ferro é a mesma para os dois grupos de crianças. Supor que o grau de gravidade em cometer esse erro é brando, o nível de significância aqui será estipulado em $\alpha = 10\%$.

   Lembrado que esse teste tem como pressuposto que a variável aleatória quantitativa deve seguir distribuição normal para os dois grupos observados da variável qualitativa. Portanto, a partir dos gráficos da Figura 4.4 percebe-se que eles estão seguindo um comportamento linear acompanhando a linha vermelha teórica e como os p-valores do teste de Shapiro-Wilk em ambos os grupos, $p-valor_{fc} = 0,5207$ e $p-valor_s = 0,4669$, é maior do que o nível de significância do teste ($\alpha = 10\%$), a hipótese de normalidade dos dados não é rejeitada. Portanto, o teste-F para comparação entre variâncias e os testes-t para amostra independentes com variâncias populacionais iguais e para variâncias populacionais diferentes podem ser utilizados para resolver esse problema.



(a) Fibrose Cística (b) Saudáveis

Figura 4.4: Gráficos quantil-quantil para verificar se o nível sérico de ferro segue distribuição normal para crianças que sofrem de fibrose cística e para crianças saudáveis.

4. **Calcular os valores observados para o teste estatístico a partir dos dados amostrais**: Primeiramente o teste-F será realizado para verificar se as variâncias populacionais são

iguais ou diferentes. Para realizar o teste-F para comparação entre variâncias utilizando o `Software R`, a função `var.test` será utilizada e a seguinte linha de comando deve ser executada:

var.test(fibrose.cistica,saudaveis)$p.value

Os dois primeiros argumentos da função `var.test` são os vetores contendo os dados das variáveis aleatórias quantitativas em que se tem o interesse em comparar as suas variâncias; como, nesse caso, o interesse é verificar se as duas variâncias são diferentes ou iguais apenas para selecionar o teste adequado para a comparação entre médias (hipóteses principais), necessitamos somente do p-valor do teste-F, por esse motivo após a função `var.test` foi utilizado a sintaxe `$p.value`. Ao executar essa linha de comando o `Software R` irá retornar o valor 0.003995031, que é o p-valor do Teste-F para comparação entre variâncias populacionais. Portanto, como $p - valor < \alpha$ ($\alpha = 10\%$), com 10% de significância rejeita-se a hipótese nula de igualdade entre variâncias.

Como a igualdade entre as variâncias populacionais foi rejeitada, então o teste que será utilizado aqui é o **teste-t para amostra independentes com variâncias populacionais *diferentes***, para realizar esse teste utilizando o `Software R`, a função `t.test` será utilizada e a seguinte linha de comando deve ser executada:

t.test(fibrose.cistica,saudaveis,var.equal=FALSE,conf.level=0.90)

Os dois primeiros argumentos da função `t.test` são os vetores contendo os dados das variáveis aleatórias quantitativas em que se tem o interesse em comparar as suas médias; no argumento `var.equal` é definido se as variâncias populacionais são iguais ou diferentes, esse argumento recebe valor "FALSE" se as variâncias populacionais são diferentes e valor "TRUE" se as variâncias populacionais são iguais; o argumento `conf.level` define o nível de confiabilidade estipulado para o cálculo do intervalo de confiança. Executando essa linha de comando o `Software R` irá retornar o seguinte resultado:

Welch Two Sample t-test

data: fibrose.cistica and saudaveis

t = -2.4777, df = 12.687, p-value = 0.02813

alternative hypothesis: true difference in means is not equal to 0

90 percent confidence interval:

-10.384245 -1.717993

sample estimates:

mean of x mean of y

15.07615 21.12727

5. **Verificar se rejeita ou não a hipótese nula** H_0: Observando os valores acima, o p-valor é dado por $p - valor = 0,02813$, como o nível de significância estipulado foi de $\alpha = 0,10$, então $p - valor < \alpha$. Portanto, com 10% de significância, existem evidências de que a média do nível sérico de ferro das crianças que sofrem de fibrose cística é diferente do nível sérico médio de ferro das crianças que não sofrem de fibrose cística, ou seja, existe um efeito da doença fibrose cística no nível sérico de ferro das crianças.

Como o `Software R` retorna o intervalo de confiança para a média do vetor estipulado no primeiro argumento menos a média do vetor estipulado no segundo argumento da função `t.test`, logo, temos que $IC\left(\mu_{fc} - \mu_s, 90\%\right) = [-10,38; -1,72]$. Portanto, com 90% de confiabilidade, levantamos evidências de que o nível sérico médio de ferro das crianças que sofrem de fibrose cística é de $1,72\mu mol/l$ a $10,38\mu mol/l$ menor (sinal negativo do intervalo de confiança), se comparado com as crianças saudáveis. Outro fato importante que podemos concluir observando esse intervalo de confiança é que o valor zero não se encontra dentro dos limites gerados pelo intervalo, isso implica que $\mu_{fs} - \mu_s \neq 0 \Rightarrow \mu_{fz} \neq \mu_s$ (hipótese nula rejeitada).

4.5.3 Exemplo 3

Supor o interesse em verificar o efeito da hidroclorotiazida (medicamento) na pressão sistólica de indíviduos. Para isso foi ministrado placebo (preparação neutra quanto a efeitos farmacológicos) em 11 indíviduos e a pressão sistólica foi medida em $mmHg$. Após algum tempo, foi ministrado a hidroclorotiazida nesses mesmos 11 indíviduos e a pressão sistólica também foi medida em $mmHg$. Para a entrada do conjunto de dados no `Software R` a função `c` foi utilizada para a criação dos vetores contendo os valores da pressão sistólica de indíviduos antes e depois da intervenção (hidroclorotiazida), como segue:

```
Placebo <- c(211,210,210,203,196,190,191,177,173,170,163)
Hidroclorotiazida <- c(181,172,196,191,167,161,178,160,149,119,156)
```

Seguindo os passos vistos na Seção 4.3 para a construção de testes de hipóteses, tem-se:

1. **Identificar** H_0 **e** H_1: Como o interesse é verificar se a pressão sistólica é afetada pela hidroclorotiazida, tem-se o interesse em testar as seguintes hipóteses,

$$\begin{cases} H_0 : \mu_p = \mu_h \\ H_1 : \mu_p \neq \mu_h \end{cases}, \tag{4.18}$$

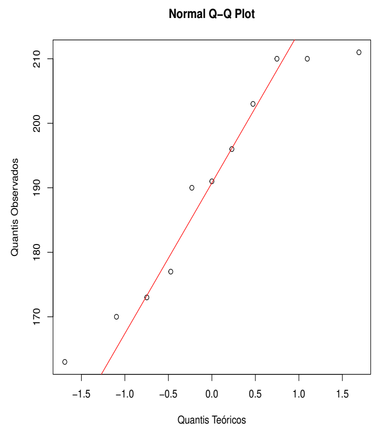
   em que, μ_p é a pressão sistólica média dos indivíduos que tomaram placebo e μ_h a pressão sistólica média dos indivíduos que tomaram hidroclorotiazida.

2. **Escolher o teste estatístico**: Para testar esse tipo de hipóteses temos que optar por um entre os quatro testes de comparação entre médias que foi visto na Seção 4.5. O primeiro passo é verificar se as variáveis são independentes ou não, como nesse caso os indivíduos que tomaram o placebo são os mesmos indivíduos que tomaram a hidroclorotiazida, então as amostras são dependentes. Portanto, aqui será utilizado o **teste-t pareado para comparação entre médias**.

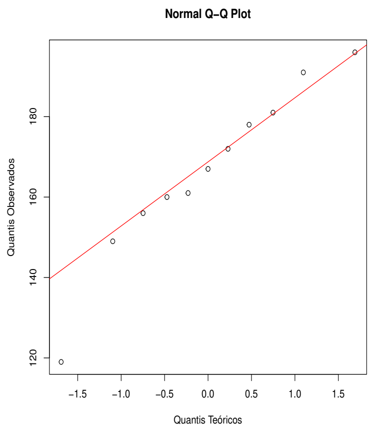
3. **Fixar o nível de significância** α: Cometer o erro do Tipo I aqui é dizer que, de acordo com a amostra a pressão sistólica média dos indivíduos que tomaram placebo é diferente da pressão sistólica média dos indivíduos que tomaram hidroclorotiazida, quando na população a pressão sistólica média é a mesma para os dois grupos de indivíduos. Supor que o grau de gravidade em cometer esse erro é grave, o nível de significância aqui será estipulado em $\alpha = 1\%$.

   Lembrado que esse teste tem como pressuposto que a variável aleatória quantitativa deve seguir distribuição normal para os dois grupos observados da variável qualitativa. Portanto, a partir dos gráficos da Figura 4.5 percebe-se que eles estão seguindo um comportamento linear acompanhando a linha vermelha teórica e como os p-valores do teste de

Shapiro-Wilk em ambos os grupos, $p - valor_p = 0,2719$ e $p - valor_h = 0,6507$, é maior do que o nível de significância do teste ($\alpha = 1\%$), a hipótese de normalidade dos dados não é rejeitada. Portanto, o teste-t pareado para comparação entre médias pode ser utilizado para resolver esse problema.



(a) Placebo



(b) Hidroclorotiazida

Figura 4.5: Gráficos quantil-quantil para verificar se a pressão sistólica segue distribuição normal para indivíduos que tomaram placebo e hidroclorotiazida.

4. **Calcular os valores observados para o teste estatístico a partir dos dados amostrais**: Para realizar o **teste-t pareado para comparação entre médias** utilizando o `Software R`, a função `t.test` será utilizada e a seguinte linha de comando deve ser executada:

   t.test(Placebo,Hidroclorotiazida,paired=TRUE,conf.level=0.99)

   Os dois primeiros argumentos da função `t.test` são os vetores contendo os dados das variáveis aleatórias quantitativas em que se tem o interesse em comparar as suas médias; no argumento `paired` é definido se as amostras são dependentes, como *default* esse argumento recebe o valor "FALSE" (amostras independentes), como as amostras nesse caso são dependentes é necessário modificao o *default* para o valor "TRUE"; o argumento `conf.level` define o nível de confiabilidade estipulado para o cálculo do intervalo de confiança. Executando essa linha de comando o `Software R` irá retornar o seguinte resultado:

   Paired t-test

   data: Placebo and Hidroclorotiazida

   t = 6.08, df = 10, p-value = 0.0001188

   alternative hypothesis: true difference in means is not equal to 0

   99 percent confidence interval:

   11.48967 36.51033

   sample estimates:

   mean of the differences

   24

5. **Verificar se rejeita ou não a hipótese nula** H_0: Observando os valores acima, o p-valor é dado por $p - valor = 0,0001188$, como o nível de significância estipulado foi de $\alpha = 0,01$, então $p - valor < \alpha$. Portanto, com 1% de significância, existem evidências de que a pressão sistólica média dos indivíduos que tomaram placebo é diferente da pressão sistólica média dos indivíduos que tomaram hidroclorotiazida, ou seja, existe um efeito da hidroclorotiazida na pressão sistólica dos indivíduos.

   Como o `Software R` retorna o intervalo de confiança para a média do vetor estipulado no primeiro argumento menos a média do vetor estipulado no segundo argumento da função `t.test`, logo, temos que $IC\left(\mu_p - \mu_h, 99\%\right) = [11,49; 36,51]$. Portanto, com 99% de confiabilidade, levantamos evidências de que a pressão sistólica média dos indivíduos que tomaram placebo é de $11,49mmHG$ a $36,51mmHg$ maior (sinal positivo do intervalo de confiança), se comparado com os indivíduos que tomaram hidroclorotiazida. Outro fato importante que podemos concluir observando esse intervalo de confiança é que o valor zero não se encontra dentro dos limites gerados pelo intervalo, isso implica que $\mu_p - \mu_h \neq 0 \Rightarrow \mu_p \neq \mu_h$ (hipótese nula rejeitada).

4.5.4 Exemplo 4

Considerando o conjunto de dados "Dieta.xlsx" (disponível na plataforma Moodle), supor o interesse em verificar o efeito das dietas (Dieta A e B) na proporção de perda de peso dos indivíduos. Seguindo os passos vistos na Seção 4.3 para a construção de testes de hipóteses, tem-se:

1. **Identificar H_0 e H_1**: Como o interesse é verificar se a proporção de perda de peso é afetada pelo tipo de dieta, tem-se o interesse em testar as seguintes hipóteses,

$$\begin{cases} H_0 : \mu_A = \mu_B \\ H_1 : \mu_A \neq \mu_B \end{cases}, \tag{4.19}$$

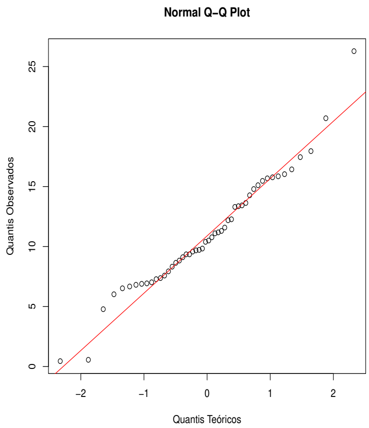
   em que, μ_A é a média da proporção de perda de peso dos indivíduos submetidos a Dieta A e μ_B a média da proporção de perda de peso dos indivíduos submetidos a Dieta B.

2. **Escolher o teste estatístico**: Para testar esse tipo de hipóteses temos que optar por um entre os quatro testes de comparação entre médias que foram vistos na Seção 4.5. O primeiro passo é verificar se as variáveis são independentes ou não, como, nesse caso, os indivíduos que foram submetidos a Dieta A são diferentes dos indivíduos submetidos a Dieta B, então as amostras são independentes. O segundo passo, então, é verificar se as variâncias populacionais para ambos os grupos são conhecidas, como, nesse caso, não foram informados os valores dessas duas variabilidades populacionais, as variâncias populacionais são desconhecidas. O último passo, então, é realizar um teste-F para comparação entre variâncias (esse teste será realizado no Item 4). Se a hipótese nula não for rejeitada o teste a ser utilizado é o **teste-t para amostra independentes com variâncias populacionais iguais**. Caso a hipótese nula do teste-F for rejeitada o teste a ser utilizado é o **teste-t para amostra independentes com variâncias populacionais diferentes**.

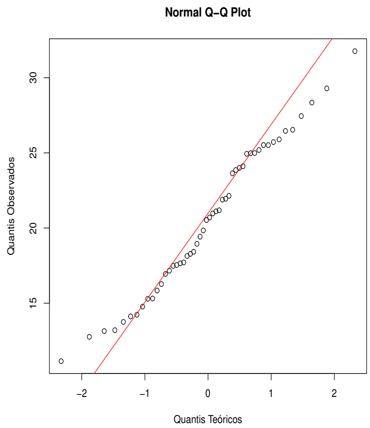
3. **Fixar o nível de significância** α: Cometer o erro do Tipo I aqui é dizer que, de acordo com a amostra a média da proporção de perda de peso dos indivíduos submetidos a Dieta A é diferente da média da proporção de perda de peso dos indivíduos submetidos a Dieta B, quando na população a média da proporção de perda de peso é a mesma para os dois

grupos de indivíduos. Supor que o grau de gravidade em cometer esse erro é brando, o nível de significância aqui será estipulado em $\alpha = 10\%$.

Lembrado que esse teste tem como pressuposto que a variável aleatória quantitativa deve seguir distribuição normal para os dois grupos observados da variável qualitativa. Portanto, a partir dos gráficos da Figura 4.6 percebe-se que eles estão seguindo um comportamento linear acompanhando a linha vermelha teórica e como os p-valores do teste de Shapiro-Wilk em ambos os grupos, $p-valor_A = 0,1545$ e $p-valor_B = 0,3682$, é maior do que o nível de significância do teste ($\alpha = 10\%$), a hipótese de normalidade dos dados não é rejeitada. Portanto, o teste-F para comparação entre variâncias e os testes-t para amostra independentes com variâncias populacionais iguais e para variâncias populacionais diferentes podem ser utilizados para resolver esse problema.



(a) Dieta A



(b) Dieta B

Figura 4.6: Gráficos quantil-quantil para verificar se a proporção de perda de peso segue distribuição normal para indivíduos submetidos a Dieta A e B.

4. **Calcular os valores observados para o teste estatístico a partir dos dados amostrais**: Primeiramente o teste-F será realizado para verificar se as variâncias populacionais são iguais ou diferentes. Para realizar o teste-F para comparação entre variâncias utilizando o `Software R`, a função `var.test` será utilizada e a seguinte linha de comando deve ser executada (aqui foi criado o objeto "dados" contendo a matriz de dados):

   var.test(dados\$Perda∼dados\$Dieta)\$p.value

   Veja que agora o primeiro argumento da função `var.test` é separado pelo símbolo $\sim$, antes do símbolo $\sim$ é definido o vetor com as observações da variável quantitativa que se tem interesse em realizar o teste de comparação entre variâncias, após o símbolo $\sim$ é definido o vetor com as observações da variável qualitativa que representa os grupos que serão comparados; como nesse caso o interesse é verificar se as duas variâncias são diferentes ou iguais apenas para selecionar o teste adequado para a comparação entre médias (hipóteses principais), necessitamos apenas do p-valor do teste-F, por esse motivo após a função `var.test` foi utilizado a sintaxe `$p.value`. Ao executar essa linha de comando o `Software R` irá retornar o valor 0.7360251, que é o p-valor do Teste-F para comparação entre variâncias populacionais. Portanto, como $p - valor > \alpha$ ($\alpha = 10\%$), com 10% de significância não rejeita-se a hipótese nula de igualdade entre variâncias.

Como a igualdade entre as variâncias populacionais não foi rejeitada, então o teste que será utilizado aqui é o **teste-t para amostra independentes com variâncias populacionais *iguais***, para realizar esse teste utilizando o `Software R`, a função `t.test` será utilizada e a seguinte linha de comando deve ser executada (aqui foi criado o objeto "dados" contendo a matriz de dados):

t.test(dados$Perda dados$Dieta,var.equal=TRUE,conf.level=0.90)

Veja que agora o primeiro argumento da função `t.test` é separado pelo símbolo $\sim$, antes do símbolo $\sim$ é definido o vetor com as observações da variável quantitativa que se tem interesse em realizar o teste de comparação entre média, após o símbolo $\sim$ é definido o vetor com as observações da variável qualitativa que representa os grupos que serão comparados; no argumento `var.equal` é definido se as variâncias populacionais são iguais ou diferentes, esse argumento recebe valor "FALSE" se as variâncias populacionais são diferentes e valor "TRUE" se as variâncias populacionais são iguais; o argumento `conf.level` define o nível de confiabilidade estipulado para o cálculo do intervalo de confiança. Executando essa linha de comando o `Software R` irá retornar o seguinte resultado:

Two Sample t-test

data: dados$Perda by dados$Dieta

t = -9.7569, df = 98, p-value = 4.079e-16

alternative hypothesis: true difference in means is not equal to 0

90 percent confidence interval:

-11.094829 -7.867571

sample estimates:

mean in group A mean in group B

11.0398 20.5210

5. **Verificar se rejeita ou não a hipótese nula** H_0: Observando os valores acima, o p-valor é dado por $p - valor = 4.079 \times 10^{-16}$, como o nível de significância estipulado foi de $\alpha = 0,10$, então $p - valor < \alpha$. Portanto, com 10% de significância, existem evidências de que a média da proporção de perda de peso dos indivíduos submetidos a Dieta A é diferente da média da proporção de perda de peso dos indivíduos submetidos a Dieta B, ou seja, existe um efeito do tipo de dieta na proporção de perda de peso dos indivíduos.

   Como a entrada de dados na função `t.test` foi feita de forma matricial, o `Software R` retorna o intervalo de confiança para a média da proporção de perda de peso dos indivíduos submetidos a Dieta A menos a média da proporção de perda de peso dos indivíduos submetidos a Dieta B, pois ele segue a ordem alfabética ou numérica da classificação da variável qualitativa. Logo, temos que $IC\,(\mu_A - \mu_B, 90\%) = [-11,09; -7,87]$. Portanto, com 90% de confiabilidade, levantamos evidências de que a média da proporção de perda de peso dos indivíduos submetidos a Dieta A é de $7,87\mu mol/l$ a $11,09\mu mol/l$ menor (sinal negativo do intervalo de confiança), se comparado com os indivíduos submetidos a Dieta B. Outro fato importante que podemos concluir observando esse intervalo de confiança é que o valor zero não se encontra dentro dos limites gerados pelo intervalo, isso implica que $\mu_A - \mu_B \neq 0 \Rightarrow \mu_A \neq \mu_B$ (hipótese nula rejeitada).