

Trabalho4

higor lucas de Araujo Freitas

2024-09-16

Parte 1 - Inferência sobre a média de uma população

1 - Inferência sobre a Média de Tempo de Entrega de uma Empresa de Logística Uma empresa de logística afirma que o tempo médio de entrega de suas encomendas é de no máximo 48 horas. Você, como analista de dados, deseja testar essa afirmação para verificar se o tempo médio de entrega real está de acordo com essa alegação.

Você coleta uma amostra aleatória de 35 encomendas recentes e registra os tempos de entrega (em horas): 49, 47, 46, 50, 48, 45, 49, 47, 46, 50, 51, 45, 47, 49, 50, 46, 48, 47, 46, 49, 50, 45, 47, 46, 48, 49, 45, 46, 50, 48, 47, 49, 45, 50, 46.

Realize um teste t para avaliar se o tempo médio de entrega é significativamente maior do que 48 horas, assumindo um nível de significância de 0,05.

tarefas: 1) Formular as Hipóteses: Hipótese nula (H_0): A media de entrega de encomendas são no maximo 48 horas ($\mu=48$). Hipótese alternativa (H_1): Tempo medio de entrega não é menor que 45 horas 2) Executar o Teste t One-Sided: Use a função `t.test` em R para testar a hipótese de que o tempo médio de entrega excede 48 horas.

3) Interpretação dos Resultados

```
# Instalar o pacote necessário se não estiver instalado
if (!require(ggplot2)) install.packages("ggplot2")
```

```
## Carregando pacotes exigidos: ggplot2
```

```
# Parâmetros da amostra
```

```
dados <- c(49, 47, 46, 50, 48, 45, 49, 47, 46, 50, 51, 45, 47, 49, 50, 46, 48, 47, 46, 49, 50, 45, 47, 46, 48, 49, 45, 46, 50, 48, 47, 49, 45, 50, 46)
desvio_padrao_amstral <- sd(dados)
media_amostra_horas <- mean(dados)
tamanho_amostra <- 35
media_horas <- 48
```

```
# Gerar uma amostra aleatória que reflète a média e o desvio padrão amostrais
```

```
set.seed(123) # Para garantir reprodutibilidade
```

```
amostra <- rnorm(n = tamanho_amostra, mean = media_amostra_horas, sd = desvio_padrao_amstral)
```

```
# Realizar o teste t
```

```
resultado_teste <- t.test(x = amostra, mu = media_horas, alternative = "two.sided", conf.level = 0.95)
```

```
# Exibir os resultados
```

```
print(resultado_teste)
```

```
##
## One Sample t-test
##
## data: amostra
## t = -1.1409, df = 34, p-value = 0.2619
## alternative hypothesis: true mean is not equal to 48
## 95 percent confidence interval:
##  47.07720 48.25923
## sample estimates:
## mean of x
##  47.66822
```

Interpretação

De acordo com os resultados para testar se a media da amostra(47.6682) é significativamente diferente de 48 horas

Estatística t

A estatística t obtida foi -1.1409. Isso representa a diferença entre a média observada (47.66822) e a média hipotética (48), padronizada pela variação dos dados

Graus de liberdade (df)

Os graus de liberdade (df) são 34 (tamanho da amostra - 1). Reflete a informação disponível para estimar a variabilidade.

P-value

O p-value é 0.2619. Este valor indica a probabilidade de observar uma diferença igual ou mais extrema que a observada, caso a média verdadeira seja realmente 48. Um valor de p acima de 0.05 (nível de significância comum) indica que não há evidências suficientes para rejeitar a hipótese nula.

Hipótese Alternativa

A hipótese alternativa é que a média verdadeira é diferente de 48 (“true mean is not equal to 48”). Isso foi determinado pelo teste bilateral realizado.

Intervalo de Confiança de 95%

intervalo de confiança de 95% para a média é [47.07720, 48.25923]. Isso significa que, com 95% de confiança, o valor verdadeiro da média populacional está dentro desse intervalo.

Estimativa da Amostra

A média da amostra estimada foi 47.66822.

Conclusão

Com um p-value de 0.2619, que é maior do que o nível de significância padrão de 0.05, não há evidências suficientes para rejeitar a hipótese nula. Portanto, não podemos concluir que a média verdadeira é significativamente diferente de 48 com base nos dados fornecidos. O intervalo de confiança também inclui o valor 48, o que reforça essa conclusão.

=====

2 - Inferência sobre a Média de Tempo de Entrega de Uma Transportadora Uma transportadora promete aos seus clientes que o tempo médio de entrega de encomendas é de no máximo 5 dias. Suspeita-se que essa promessa não está sendo cumprida, e você foi contratado para analisar se o tempo médio de entrega realmente difere de 5 dias.

Você selecionou uma amostra aleatória de 50 encomendas entregues no último mês e registrou os seguintes tempos de entrega (em dias): `tempos_entrega <- c(6, 5, 7, 5, 6, 4, 6, 5, 7, 4, 5, 6, 6, 5, 4, 6, 5, 6, 7, 5, 5, 6, 5, 6, 4, 5, 6, 5, 4, 6, 5, 7, 5, 6, 5, 4, 6, 5, 7, 4, 5, 6, 5, 6, 4, 5, 6, 5, 4, 6)`

Use um teste t para avaliar se o tempo médio de entrega das encomendas é significativamente diferente de 5 dias. Assuma um nível de significância de 0,05.

tarefas: 1) Formular as Hipóteses:

Hipótese nula (H0): O tempo médio de entrega é 5 dias ou menos Hipótese alternativa (H1): O tempo médio de entrega é maior que 5 dias

2) Executar o Teste t One-Sided: Use a função `t.test` em R para testar a hipótese de que o tempo médio de entrega excede 48 horas.

3) Interpretação dos Resultados

```
# Instalar pacotes necessários, se ainda não estiverem instalados
if (!require(ggplot2)) install.packages("ggplot2")
library(ggplot2)

# 1. Gerar a Amostra
# Dados da amostra
tempos_entrega <- c(6, 5, 7, 5, 6, 4, 6, 5, 7, 4, 5, 6, 6, 5, 4, 6, 5, 6, 7, 5, 5, 6, 5, 6, 4, 5, 6, 5,
# Testar a Não-Normalidade com o Teste de Shapiro-Wilk
teste_normalidade <- shapiro.test(tempos_entrega)
print("AQUI")
```

```
## [1] "AQUI"
```

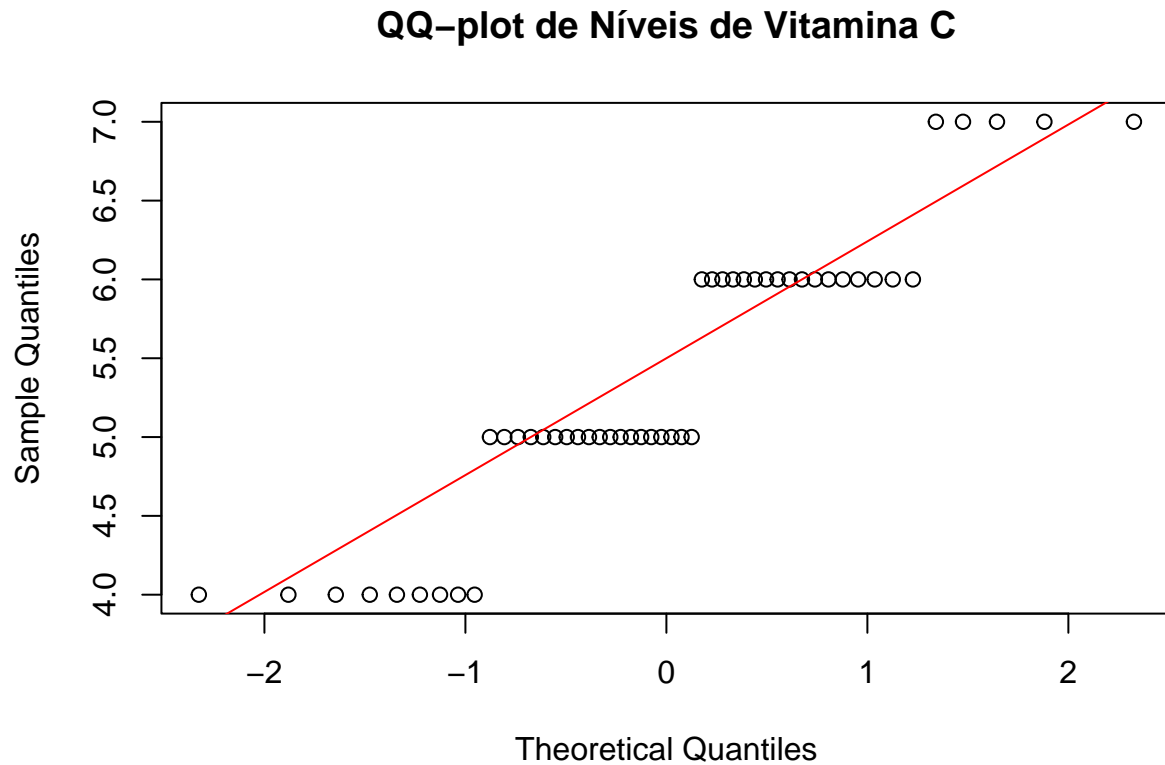
```
print(teste_normalidade)
```

```
##
## Shapiro-Wilk normality test
##
## data:  tempos_entrega
## W = 0.87784, p-value = 9.63e-05
```

```
if (teste_normalidade$p.value < 0.05) {
  print("A distribuição da amostra não é normal. ")
} else {
  print("A distribuição da amostra não apresentou evidência de desvio da normalidade.")
}
```

```
## [1] "A distribuição da amostra não é normal. "
```

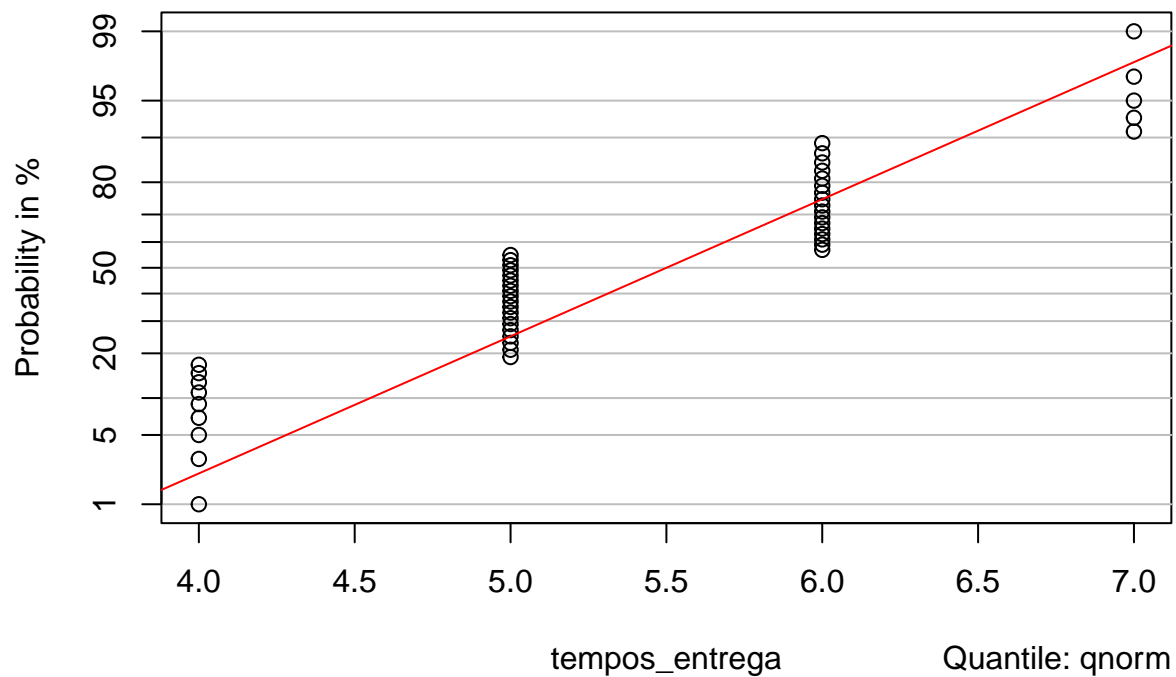
```
# Desenhar um QQ-plot com a reta normal em vermelho  
qqnorm(tempos_entrega, main = "QQ-plot de Níveis de Vitamina C")  
qqline(tempos_entrega, col = "red")
```



```
# Desenhar um Normal Probability Plot  
if (!require(e1071)) install.packages("e1071")
```

```
## Carregando pacotes exigidos: e1071
```

```
library(e1071)  
probplot(tempos_entrega)
```



```
# Realizar o teste t one-sided
resultado_teste <- t.test(tempos_entrega, mu = 5, alternative = "greater", conf.level = 0.95)

# Exibir os resultados
print(resultado_teste)
```

```
##
## One Sample t-test
##
## data:  tempos_entrega
## t = 2.8345, df = 49, p-value = 0.003325
## alternative hypothesis: true mean is greater than 5
## 95 percent confidence interval:
##  5.147067      Inf
## sample estimates:
## mean of x
##      5.36
```

Interpretação

A hipótese alternativa é que o tempo médio é superior a 5 dias

Estatística t

A estatística t calculada foi 2.8345. Esse valor indica quantos desvios padrão a média amostral (5.36) está acima da média hipotética (5). Como o valor t é positivo e maior que 2, isso sugere uma diferença considerável entre a média da amostra e o valor hipotético.

Graus de liberdade (df)

Os graus de liberdade (df) foram 49. Em um teste t de uma amostra, o número de graus de liberdade é igual ao tamanho da amostra menos 1

P-value

O p-value obtido foi 0.003325. Este valor representa a probabilidade de observar uma diferença igual ou maior do que a observada, caso a média verdadeira seja de fato 5 dias. Como o p-value é menor que 0.05, há evidências estatisticamente significativas para rejeitar a hipótese nula.

Hipótese Alternativa

A hipótese alternativa testada foi que o tempo médio de entrega é maior do que 5 dias. Isso está indicado no resultado pelo uso de `alternative = "greater"`

Intervalo de Confiança de 95%

O intervalo de confiança unilateral de 95% para a média foi de [5.147067, Inf]. Isso significa que, com 95% de confiança, a média verdadeira dos tempos de entrega está acima de 5.147 dias, sem um limite superior definido (indicado pelo "Inf").

Estimativa da Amostra

A média amostral foi 5.36 dias. Esse é o valor médio dos tempos de entrega para as 50 encomendas observadas.

Conclusão

Com um p-value de 0.003325, que é significativamente menor do que o nível de significância de 0.05, rejeitamos a hipótese nula. Isso sugere que o tempo médio de entrega das encomendas da transportadora é significativamente maior do que 5 dias.

Parte 2 - Inferência sobre a média de duas populações

3 - Comparação do Desempenho de Duas Marcas de Baterias com amostras independentes Uma empresa de eletrônicos deseja comparar o desempenho de duas marcas de baterias, A e B, para decidir qual delas incluir em seus produtos. O desempenho é medido pelo número de horas que as baterias duram em um teste padronizado. São selecionadas amostras aleatórias de 30 baterias de cada marca e testadas. Os dados gerados abaixo são fictícios e representam o número de horas de duração para cada bateria nas amostras:

Use um teste t para duas amostras independentes para avaliar se existe uma diferença significativa na duração média entre as baterias das marcas A e B. Assuma um nível de significância de 0,05.

- 1) Formular as Hipóteses: Hipótese nula (H_0): Não há diferença significativa na duração média entre as duas marcas ($\mu_A = \mu_B$)

Hipótese alternativa (H_1): Existe uma diferença significativa na duração média entre as duas marcas ($\mu_A \neq \mu_B$)

- 2) Executar o Teste t para Duas Amostras Independentes: Use a função `t.test` em R para comparar as médias das duas marcas.

3) Interpretar o resultado

```
# Configuração para garantir a reprodutibilidade
set.seed(123)

duracao_marca_A <- rnorm(30, mean = 50, sd = 5) # Duração média de 50 horas, desvio padrão de 5
duracao_marca_B <- rnorm(30, mean = 52, sd = 5) # Duração média de 52 horas, desvio padrão de 5

# Realizar o teste t para duas amostras independentes
resultado_teste <- t.test(duracao_marca_A, duracao_marca_B, alternative = "two.sided", conf.level = 0.95)

# Exibir os resultados
print(resultado_teste)
```

```
##
##  Welch Two Sample t-test
##
## data:  duracao_marca_A and duracao_marca_B
## t = -2.659, df = 56.559, p-value = 0.01018
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.4827131 -0.7717079
## sample estimates:
## mean of x mean of y
##  49.76448  52.89169
```

Interpretação

Estatística t

A estatística t obtida foi -2.659, o que indica a magnitude e direção da diferença entre as médias das duas amostras. O valor negativo sugere que a média da marca A é menor que a média da marca B.

Graus de liberdade (df)

O teste foi realizado com aproximadamente 56.559 graus de liberdade. ### P-value O P-value é 0.01018, que é menor do que o nível de significância convencional de 0.05. Isso indica que há evidências estatisticamente significativas para rejeitar a hipótese nula de que não há diferença entre as médias das duas marcas ### Intervalo de Confiança de 95% O intervalo de confiança para a diferença verdadeira nas médias estende-se de -5.4827131 a -0.7717079. Isso significa que, com 95% de confiança, a verdadeira diferença entre as médias das marcas A e B está entre -5.48 e -0.77. Como o intervalo é totalmente negativo, isso sugere que a duração da marca A é significativamente menor que a da marca B.

Estimativas das Médias:

As médias estimadas são 49.76448 horas para o grupo baterias A e 52.89169 horas para o grupo baterias B. Esses valores indicam que, em média, a duração do produto da marca B é maior que o da marca A.

Conclusão

Com base no P-value de 0.01018 e no intervalo de confiança que inclui o zero, podemos rejeitar a hipótese nula. Há evidências suficientes para concluir que há uma diferença significativa nas durações médias entre as marcas A e B. Especificamente, a marca A tem uma duração significativamente menor que a marca B.

4 -

Avaliação do Efeito de um Suplemento Nutricional no Tempo de Recuperação Uma empresa de nutrição esportiva desenvolveu um novo suplemento que alega reduzir o tempo de recuperação após exercícios intensos. Para testar essa alegação, um estudo foi realizado com voluntários que foram submetidos a uma sessão de exercícios padronizada, medindo o tempo de recuperação sem suplemento e depois de um dia com a ingestão do suplemento.

Seriam selecionados aleatoriamente 30 voluntários para o estudo. Os tempos de recuperação (em horas) seriam registrados antes do uso do suplemento e em outro dia com a ingestão do suplemento. Na verdade, vamos gerar dados fictícios para as amostras de tempos de recuperação antes e depois para 30 pessoas.

Use um teste t pareado one-sided para avaliar se a ingestão do suplemento resultou em uma redução significativa no tempo de recuperação após exercícios intensos. Assuma um nível de significância de 0,05.

- 1) Formular as Hipóteses: Hipótese nula (H_0): O suplemento não reduz o tempo de recuperação ($\mu_{\text{depois}} \geq \mu_{\text{antes}}$)

Hipótese alternativa (H_1): O suplemento reduz o tempo de recuperação ($\mu_{\text{depois}} < \mu_{\text{antes}}$)

- 2) Executar o Teste t Pareado One-Sided: Utilize a função `t.test` em R para comparar as médias dos tempos de recuperação antes e depois da ingestão do suplemento.
- 3) Interprete o resultado.

```
# Configuração para garantir a reprodutibilidade
set.seed(123)
# Gerar dados fictícios para as amostras
recuperacao_antes <- rnorm(30, mean = 480, sd = 10)
recuperacao_depois <- recuperacao_antes - rnorm(30, mean = 0.01, sd = 1)
# Espera-se uma redução no tempo de recuperação porque para cada elemento em recuperacao_antes é subtraído um valor aleatório de recuperacao_depois

# Realizar o teste t one-sided para duas amostras pareadas
resultado_teste <- t.test(recuperacao_depois, recuperacao_antes, alternative = "greater", paired = TRUE)

# Exibir os resultados
print(resultado_teste)
```

```
##
## Paired t-test
##
## data: recuperacao_depois and recuperacao_antes
## t = -1.2352, df = 29, p-value = 0.8867
```



```
## alternative hypothesis: true mean difference is greater than 0
## 95 percent confidence interval:
##  -0.447409      Inf
## sample estimates:
## mean difference
##      -0.1883383
```

Interpretação

Estatística t

A estatística t foi -1.2352. Um valor negativo indica que a média da diferença entre a recuperação depois e antes é negativa, ou seja, a recuperação antes foi maior que a recuperação depois.

Graus de liberdade (df)

Os graus de liberdade são 29

P-value

O p-value foi 0.8867, que é muito maior do que o nível de significância padrão de 0.05. Isso significa que não há evidências suficientes para rejeitar a hipótese nula. Portanto, não podemos concluir que há uma diferença significativa entre a recuperação antes e depois.

Intervalo de Confiança de 95%

O intervalo de confiança para a diferença média entre os grupos é [-0.447409, Inf]. Como o intervalo inclui zero e o limite inferior é negativo, isso significa que a diferença pode ser zero ou negativa, sugerindo que a recuperação antes pode ser maior ou igual à recuperação depois.

Estimativas das Médias:

A diferença média observada entre os dois momentos foi de -0.1883383, indicando que, em média, a recuperação antes foi ligeiramente maior do que a recuperação depois.

Conclusão

Com um p-value de 0.8867, que é bem acima de 0.05, não há evidências estatísticas para rejeitar a hipótese nula. Isso significa que não há suporte suficiente para afirmar que a recuperação depois foi significativamente maior do que a recuperação antes. A diferença observada é pequena e não significativa.