Universidade Federal de Mato Grosso

Instituto de Ciências Exatas e da Terra

Campus Universitário do Araguaia

Prova 2 de Estatística Geral

**Nome e RGA:**

**Resolva e envie a Prova 2 e envie no AVA em arquivo Word.**

**Tempo de duração da prova: 120 min**

**Questão 0:** Coloque o seu RGA na grade abaixo, um dígito para cada célula, e depois extraia os dígitos X, Y e Z. Um exemplo é dado abaixo.

Utilize esses **dígitos numéricos** para substituir as **letras X, Y e Z** nas questões da prova.

**Exemplo** com RGA 201811722006

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2 | 0 | 1 | 8 | 1 | 1 | 7 | 2 | 2 | 0 | 0 | 6 |
|  |  |  | X |  |  |  |  |  |  | Y | Z |

X = 8

Y = 0

Z = 6

**Use o Seu** RGA:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2 | 0 | 2 | 2 | 1 | 1 | 7 | 2 | 2 | 0 | 3 | 5 |
|  |  |  | X |  |  |  |  |  |  | Y | Z |

X = 2

Y = 3

Z = 5

Boa Prova!

**Questão 1:** Responda as questões abaixo (3,0):

a) Qual a relação entre população e amostra em estatística? ***Explique com as suas próprias palavras*** e dê um exemplo.

**Resposta**:

Em estatística, a população é o conjunto completo de todos os elementos ou indivíduos que possuem uma característica de interesse para um estudo. Já a amostra é um subconjunto menor dessa população, selecionado para representar suas características de forma prática e viável, já que muitas vezes é impossível ou impraticável estudar toda a população.

A relação entre os dois conceitos é que a amostra é usada para fazer inferências sobre a população. A ideia é que, se a amostra for bem escolhida (representativa e aleatória), as conclusões tiradas a partir dela podem ser generalizadas para a população com um certo grau de confiança. Por exemplo, se você quer saber a altura média dos brasileiros (população), pode medir a altura de um grupo menor de brasileiros (amostra) e usar esses dados para estimar a média da população.

**População**:

• É o **conjunto** **completo** deelementos (pessoas, objetos, eventos) que possuem alguma característica em comum e sobre o qual se quer tirar conclusões.

• Pode ser **finita** ou **infinita**.

• Exemplo: Todos os eleitores de um país.

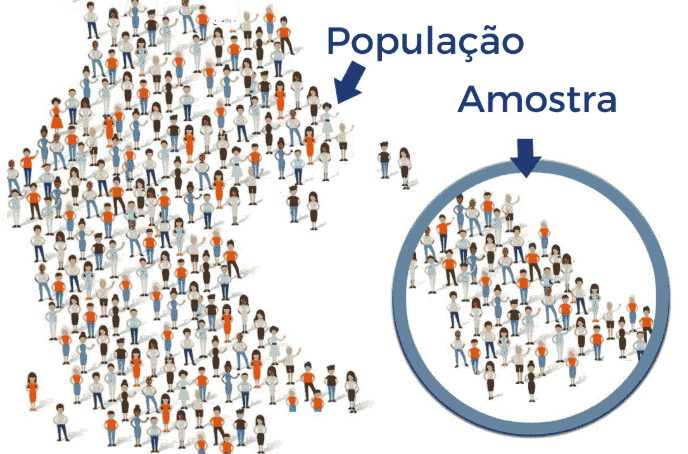
**Amostra**:

• É um **subconjunto** **representativo** da população.

• Utilizada quando não é viável estudar a população inteira (por tempo, custo, ou acessibilidade).

• Exemplo: 2.000 eleitores entrevistados em uma pesquisa eleitoral.

Exemplo: Imagine que você quer saber a opinião dos alunos de uma universidade (população, com 10.000 alunos) sobre um novo sistema de ensino. Em vez de entrevistar todos os 10.000 alunos, você seleciona aleatoriamente 500 alunos (amostra) e faz uma pesquisa com eles. Com base nas respostas dessa amostra, você pode estimar a opinião geral da universidade, desde que a amostra seja representativa.



b) Diga se a seguinte afirmação está correta ou não e explique a sua resposta: A **pergunta de uma pesquisa científica estatística** geralmentese refere à **população**, mesmo que os dados sejam coletados a partir de uma **amostra**.

**Resposta**:

A afirmação está correta.

Explicação: Em uma pesquisa científica estatística, o objetivo principal é obter informações ou tirar conclusões sobre uma população de interesse, como suas características, comportamentos ou tendências. No entanto, devido a limitações práticas (como tempo, custo ou acesso), geralmente é inviável coletar dados de todos os elementos da população. Por isso, os pesquisadores coletam dados de uma amostra, que é um subconjunto representativo da população.

A pergunta da pesquisa é formulada com foco na população porque o interesse está em entender ou descrever algo sobre esse grupo maior. A amostra serve como uma ferramenta para estimar ou inferir as características da população com base nos dados coletados. Por exemplo, se a pergunta de uma pesquisa é "Qual é a proporção de brasileiros que apoiam uma nova política?", o interesse está na população (todos os brasileiros), mesmo que os dados sejam coletados de uma amostra (um grupo menor de brasileiros entrevistados).

Portanto, a afirmação está correta, pois a pergunta da pesquisa visa a população, e a amostra é usada como um meio para responder a essa pergunta de forma prática e confiável, desde que a amostragem seja bem conduzida.

**Questão 2**: Um experimento foi conduzido para comparar a capacidade de enchimento de duas máquinas diferentes (A e B) de envasamento de defensivos. Dez frascos foram envasados em cada máquina, os volumes preenchidos em cada frasco foram medidos e os dados são apresentados na Tabela 1. Verifique se as máquinas enchem os frascos com o mesmo volume médio. **Copie os comandos e as saídas do programa *R* ao final da questão para que a mesma seja considerada para correção.** (3,5 pontos)

Tabela 1. Volume (mL) de enchimento de frascos de herbicidas por duas máquinas.

|  |  |
| --- | --- |
| Máquina | |
| ***A*** | ***B*** |
| 752 | 756 |
| 753 | 755 |
| 755 | 755 |
| 751 | 724 |
| 753 | 736 |
| 751 | 756 |
| 752 | 757 |
| 753 | 756 |
| 753 | 756 |
| 755 | 753 |

Para facilitar a análise, sugiro que se siga os 7 passos do teste de hipótese:

0. Entender a estrutura dos dados e escolher o modelo estatístico adequado

**Resposta**:

Temos dois grupos independentes (Máquina A e Máquina B) com 10 observações cada. Vamos utilizar um **teste t para duas amostras independentes** para comparar as médias.

1. Destacar a pergunta da pesquisa

**Resposta**:

As máquinas A e B enchem os frascos com o mesmo volume médio?

1. Definir as hipóteses a serem testadas e o nível de significância do teste

**Resposta**:

H0: *μA = μB (as médias são iguais)*

H1: *μA ≠ μB (as médias são diferentes)*

Nível de insignificância: α = 0,05

1. Estudar a amostra através de estatística descritiva

**Resposta**:

Vamos calcular as médias e desvios-padrão de cada grupo.

# Importando dados

dados <- read.csv(file.choose(),header=TRUE)

dados

# Estatística descritiva

# Médias

Media\_A <- mean(dados$A)

Media\_B <- mean(dados$B)

# Desvios Padrão

sd\_A <- sd(dados$A)

sd\_B <- sd(dados$B)

# Saídas

cat("Média A:", Media\_A, "\n") # Média A: 752.8

cat("Desvio Padrão A:", sd\_A, "\n") # Desvio Padrão A: 1.398412

cat("Média B:", Media\_B, "\n") # Média B: 750.4

cat("Desvio Padrão B:", sd\_B, "\n") # Desvio Padrão B: 11.16741

1. Calcular as estatísticas do teste e o valor de *p*

**Resposta**:

# (4) Calcular as estatísticas do teste e o valor de p

# 4.1 Teste de Variâncias (Homogeneidade)

var\_test <- var.test(dados$A,dados$B)

print(var\_test)

# F test to compare two variances

# data: dados$A and dados$B

# F = 0.015681, num df = 9, denom df = 9, p-value = 8.943e-07

# alternative hypothesis: true ratio of variances is not equal to 1

# 95 percent confidence interval:

# 0.00389486 0.06313034

# sample estimates:

# ratio of variances

# 0.01568068

# 4.2 Teste t para Amostras Independentes (Welch)

t\_test <- t.test(dados$A,dados$B, var.equal = FALSE)

print(t\_test)

# Welch Two Sample t-test

# data: dados$A and dados$B

# t = 0.67434, df = 9.2822, p-value = 0.5165

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

# 95 percent confidence interval:

# -5.613921 10.413921

# sample estimates:

# mean of x mean of y

# 752.8 750.4

1. Comparar o valor de *p* com o valor de *α* e concluir

**Resposta**:

p-value = 0.5165 > α = 0.05 → Não rejeitamos H0

1. Escrever a conclusão na forma de artigo científico ou técnico

**Resposta**:

**Conclusão**: Não há diferença significativa (p = 0.5165) no volume médio de enchimento entre as máquinas A e B.

No Passo 5, pode-se adotar a seguinte Regra de Decisão:

Se *p* > *α*, então não se rejeita H0.

Se *p* < *α*, então se rejeita H0 e aceita H1.

Obs.: Não esqueça de copiar os comandos e saídas do programa *R* ao final da resolução.

**Questão 3:** Os dados a seguir correspondem à variável renda familiar e gasto com alimentação (em unidades monetárias) de 25 famílias. Utilize o programa *R* para fazer a análise dos dados. **Copie os comandos e as saídas do programa *R* ao final da questão para que a mesma seja considerada para correção.** (3,5 pontos)

Tabela 1. Renda familiar e gasto com alimentação (em unidade monetária).

|  |  |
| --- | --- |
| Renda Familiar | Gasto com Alimentação |
| 3 | 1 |
| 5 | 2 |
| 10 | 6 |
| 10 | 7 |
| 20 | 10 |
| 20 | 12 |
| 20 | 15 |
| 30 | 8 |
| 40 | 10 |
| 50 | 20 |
| 60 | 20 |
| 70 | 25 |
| 70 | 30 |
| 80 | 25 |
| 100 | 40 |
| 100 | 35 |
| 100 | 40 |
| 120 | 30 |
| 120 | 40 |
| 140 | 40 |
| 150 | 50 |
| 180 | 40 |
| 182 | 42 |
| 203 | 63 |
| 205 | 65 |

(a) Faça um gráfico e ajuste uma reta de regressão para prever o gasto com alimentação (*variável depende – ordenada Y*) em função da renda familiar (*variável independente – abscissa X*).

**Resposta**:

# Importando dados

dados <- read.csv(file.choose(), header=TRUE)

dados

# (A) Faça um gráfico e ajuste uma reta de regressão para prever o gasto com alimentação (variável depende – ordenada Y)

# em função da renda familiar (variável independente – abscissa X).

#Gráfico de dispersão

plot(dados$Renda.Familiar, dados$Gasto.com.Alimentação,

main = "Renda Familiar vs. Gasto com Alimentação",

xlab = "Renda Familiar (u.m.)",

ylab = "Gasto com Alimentação (u.m.)",

pch = 19, col = "blue")

# Ajuste da reta de regressão

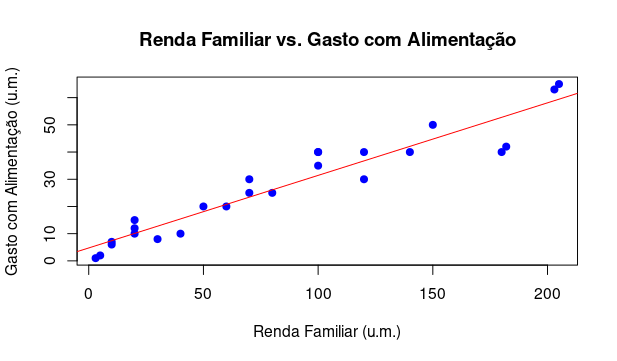
modelo <- lm(dados$Gasto.com.Alimentação ~ dados$Renda.Familiar)

abline(modelo, col = "red")

# Equação da reta

cat("Equação da reta: Gasto =", coef(modelo)[1], "+", coef(modelo)[2], "\* Renda\n")

# Equação da reta: Gasto = 4.764531 + 0.2667082 \* Renda



(b) Estime o coeficiente de correlação entre as variáveis e interprete.

**Resposta**:

# Coeficiente de correlação

correlacao <- cor(dados$Renda.Familiar, dados$Gasto.com.Alimentação)

cat("Coeficiente de correlação (r):", round(correlacao, 4), "\n")

# Coeficiente de correlação (r): 0.9514

# Interpretação

if (abs(correlacao) >= 0.7) {

cat("Interpretação: Correlação forte (|r| >= 0.7).\n")

} else if (abs(correlacao) >= 0.3) {

cat("Interpretação: Correlação moderada (0.3 <= |r| < 0.7).\n")

} else {

cat("Interpretação: Correlação fraca (|r| < 0.3).\n")

}

# Interpretação: Correlação forte (|r| >= 0.7).

(c) Qual o gasto com alimentação esperado para uma renda familiar de 155 unidades monetárias e para uma renda familiar de 321 unidades monetárias?

**Resposta**:

# Previsões

nova\_renda <- data.frame(renda = c(155, 321))

previsao <- predict(modelo, nova\_renda)

cat("Gasto esperado para Renda = 155 u.m.:", round(previsao[1], 2), "u.m.\n")

# Gasto esperado para Renda = 155 u.m.: 5.56 u.m.

cat("Gasto esperado para Renda = 321 u.m.:", round(previsao[2], 2), "u.m.\n")

# Gasto esperado para Renda = 321 u.m.: 6.1 u.m.

# Resumo do modelo (opcional)

summary(modelo)

# Call:

# lm(formula = dados$Gasto.com.Alimentação ~ dados$Renda.Familiar)

# Residuals:

# Min 1Q Median 3Q Max

# -12.7720 -4.0981 -0.0987 4.0937 8.5646

# Coefficients:

# Estimate Std. Error t value Pr(>|t|)

# (Intercept) 4.7645 1.8914 2.519 0.0192 \*

# dados$Renda.Familiar 0.2667 0.0180 14.813 2.97e-13 \*\*\*

# ---

# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

# Residual standard error: 5.736 on 23 degrees of freedom

# Multiple R-squared: 0.9051, Adjusted R-squared: 0.901

# F-statistic: 219.4 on 1 and 23 DF, p-value: 2.973e-13