Atividade 3 - Respostas

**Questão 1.a - Elabore e interprete o teste do qui-quadrado de independência para verificar se há associação estatisticamente significativa entre o turno escolar e a percepção dos pais quanto à qualidade do ensino. Utilize um nível de significância de 5%.**

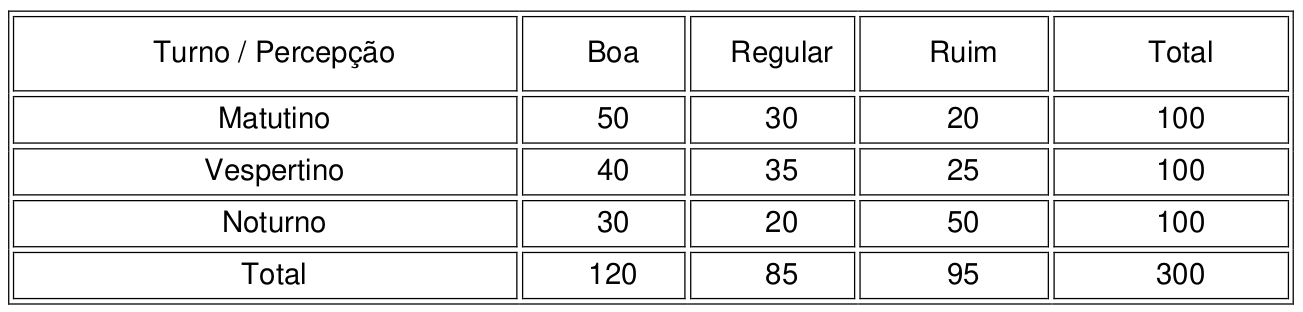
A análise de dados pode ser utilizada para investigar a associação entre o turno escolar (matutino, vespertino ou noturno) e a percepção dos pais quanto à qualidade do ensino (boa, regular ou ruim) por meio do teste do qui-quadrado de independência.

Este teste é uma ferramenta estatística que permite verificar se há uma relação estatisticamente significativa entre duas ou mais variáveis de classificação.

O teste do qui-quadrado de independência formula as seguintes hipóteses:

* **Hipótese Nula (H0)**: As variáveis “turno escolar” e “percepção dos pais sobre a qualidade do ensino” são independentes, ou seja, não há associação estatisticamente significativa entre elas.
* **Hipótese Alternativa (H1)**: As variáveis não são independentes (estão associadas), ou seja, há uma associação estatisticamente significativa entre as variáveis “turno escolar” e “percepção dos pais sobre a qualidade do ensino”.

A tabela de contingência com os resultados da pesquisa realizada com 300 responsáveis por estudantes da rede pública é a apresentada abaixo:

As células internas da tabela representam as frequências observadas. O teste compara essas frequências observadas com as frequências esperadas, que seriam os valores esperados caso não houvesse associação entre as variáveis. As frequências esperadas são calculadas pela fórmula: (Total da linha x Total da coluna) / Total de observações.

A interpretação é baseada no p-valor comparado ao nível de significância () estabelecido, que neste caso é de 5% (0.05). O teste de significância, por exemplo, é determinado pela comparação do p-valor com o nível de significância. Se o p-valor for menor que o nível de significância, a hipótese nula é rejeitada.

Para os dados fornecidos, o R calculou o valor da estatística de qui-quadrado e o p-valor correspondente, além dos graus de liberdade. No caso, para uma tabela de contingência R x C, os graus de liberdade são (R-1) x (C-1).

O cálculo do teste do qui-quadrado utilizando um nível de significância de 5% e do p-valor pode ser verificado no script1.R (em anexo).

O valor do p-valor de 4.115756e-05 foi extraído da execução do script1.R. Como o p-valor é menor que 0.05, então rejeitamos a hipótese nula (H0). Isso significa que há evidências estatísticas suficientes para concluir que existe uma associação significativa entre o turno escolar e a percepção dos pais sobre a qualidade do ensino. Em outras palavras, a percepção da qualidade do ensino depende do turno em que o estudante está matriculado.

**Questão 1.b - Calcule o coeficiente de contingência e interprete seu valor à luz da força da associação entre as variáveis.**

Para calcular o coeficiente de contingência e interpretar seu valor à luz da força da associação entre as variáveis “turno escolar” e “percepção dos pais sobre a qualidade do ensino” foram utilizados os dados da tabela de contingência fornecida e a linguagem R.

O coeficiente de contingência de Pearson (**C**) e o coeficiente de contingência corrigido de *Tschuprow* (**T**) são medidas da força da associação entre o turno escolar e a percepção dos pais quanto à qualidade do ensino.

O **C** é uma medida de associação utilizada para variáveis qualitativas. Ele quantifica o grau de relacionamento, associação ou dependência entre as classificações em uma tabela de contingência. O **C** varia de 0 a um valor máximo que depende do número de linhas e colunas da tabela, mas que nunca atinge 1. Quanto maior o **C**, maior o grau de associação entre as variáveis. Já o **T**, que pode atingir o valor máximo de 1, se o número de linhas for igual ao número de colunas e houver uma dependência total.

O valor do **C** de 0,28 e **T** de 0,41 foram extraídos da execução do script1.R (em anexo). O coeficiente C de *Pearson* indica uma associação fraca a moderada entre as variáveis. Já o coeficiente T de *Tschuprow* permite comparações melhores, indicando a existência de uma associação moderada entre as variáveis, sugerindo que há uma relação relevante entre o turno e a percepção dos pais.

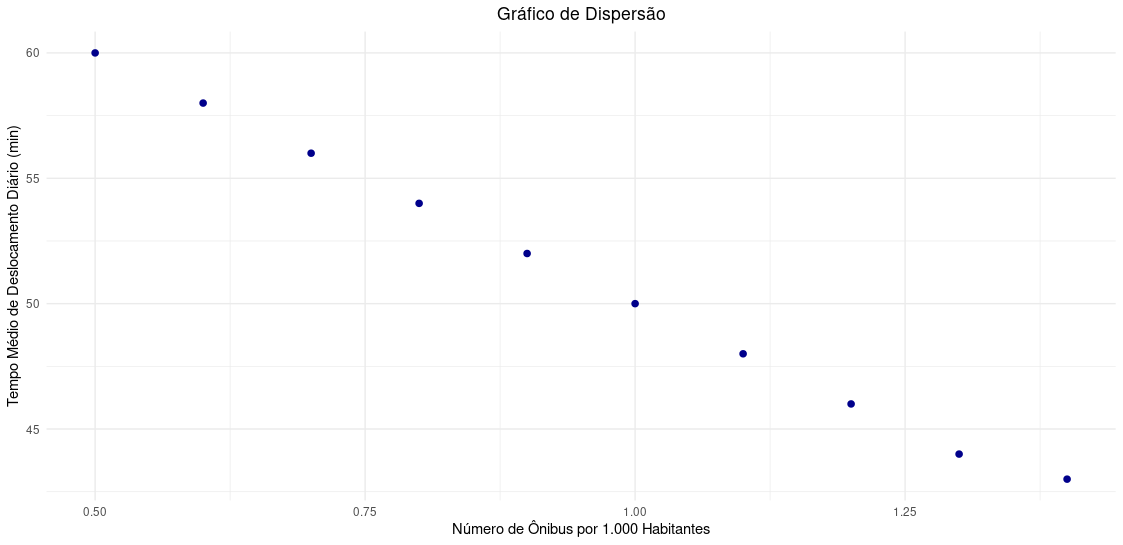
**Questão 1.c - Com base nos resultados obtidos, discuta como essas informações podem subsidiar políticas públicas voltadas para a melhoria da qualidade do ensino nos diferentes turnos escolares.**

Os resultados obtidos evidenciam uma associação estatisticamente significativa entre o turno escolar e a percepção dos pais quanto à qualidade do ensino, com destaque para a maior insatisfação no turno noturno. Essas informações podem subsidiar políticas públicas ao direcionar esforços para identificar e corrigir desigualdades estruturais, pedagógicas e de gestão específicas de cada turno, promovendo ações mais eficazes e direcionadas. Investimentos em formação docente, recursos pedagógicos, infraestrutura e acompanhamento educacional no turno noturno, por exemplo, podem melhorar a percepção e, consequentemente, a qualidade do ensino oferecido, promovendo maior equidade no sistema educacional.

**Questão 2.a - Construa um gráfico de dispersão com os dados apresentados. Qual o tipo de associação (positiva, negativa ou inexistente) parece estar presente entre as variáveis?**

Um gráfico de dispersão é uma ferramenta gráfica utilizada para apresentar a relação entre duas variáveis quantitativas. A visualização dos pontos no gráfico permite identificar o tipo de associação entre elas: positiva (pontos tendem a subir da esquerda para a direita), negativa (pontos tendem a descer da esquerda para a direita) ou inexistente (sem padrão aparente).

O gráfico de dispersão apresentado abaixo foi obtido a partir da execução do script2.R (em anexo).



Ao observar o gráfico de dispersão, percebe-se que, à medida que o "Número de Ônibus por 1.000 Habitantes" aumenta (movendo-se para a direita no eixo X), o "Tempo Médio de Deslocamento Diário (min)" tende a diminuir (movendo-se para baixo no eixo Y). Isso sugere uma associação negativa entre as variáveis. Ou seja, cidades com mais ônibus por habitante tendem a ter um tempo médio de deslocamento menor.

**Questão 2.b - Calcule o coeficiente de correlação linear de Pearson entre o número de ônibus por habitante e o tempo médio de deslocamento. Interprete o valor obtido.**

O coeficiente de correlação linear de *Pearson* é uma medida estatística que quantifica a força e a direção da relação linear entre duas variáveis quantitativas. O valor varia de -1 a 1, onde:

• Um valor próximo de 1 indica uma forte correlação linear positiva.

• Um valor próximo de -1 indica uma forte correlação linear negativa.

• Um valor próximo de 0 indica uma correlação linear fraca ou inexistente.

O valor do coeficiente de correlação linear de *Pearson* foi obtido por meio da execução do script2.R (em anexo), resultando no valor de -0,99.

**Questão 2.c - Discuta de que forma os resultados encontrados podem orientar a formulação de políticas públicas para melhorar a mobilidade urbana nas cidades analisadas.**

Os resultados encontrados, que demonstram uma associação linear negativa muito forte entre o número de ônibus por habitante e o tempo médio de deslocamento, fornecem uma base robusta para a formulação de políticas públicas de mobilidade urbana, alinhando-se ao objetivo do departamento de avaliar a relação entre a oferta de transporte coletivo e a redução no tempo de deslocamento. Essa evidência sugere que cidades com maior oferta de ônibus tendem a apresentar um tempo médio de deslocamento significativamente menor, indicando a importância de priorizar investimentos na expansão e otimização do transporte público coletivo. Ao utilizar esses dados, o departamento pode subsidiar decisões estratégicas e justificar a alocação de recursos escassos em ações que visam diretamente a melhoria da mobilidade urbana e a eficiência do gasto público, com base em evidências que minimizam erros e otimizam processos.