

ENTREGA EM 12/06/2018

Considere que o arquivo com os dados usados na Lista de Exercícios 2, referente aos alunos da universidade TYU seja a POPULAÇÃO da sua pesquisa. Desta população você irá retirar amostras aleatórias para resolver os problemas a seguir. O relatório com as respostas deverá ser entregue IMPRESSO com todas as tabelas, medidas e gráficos julgados convenientes, produzidos através do R, Excel, GPower ou qualquer outro aplicativo. As amostras coletadas devem ser enviadas por e-mail ao professor, devidamente identificadas com os nomes dos alunos e nomes das variáveis. Recomenda-se que antes de qualquer procedimento **remova todas as linhas com dados perdidos**.

Primeira parte – Distribuição Amostral da Média

1) Usando qualquer aplicativo que considere apropriado¹ proceda a retirada de 1000 amostras aleatórias da variável **Renda** (primeiramente exclua as linhas com dados perdidos), com os seguintes tamanhos: 4, 16, 64 e 256 elementos. Com base nos resultados encontrados responda os itens a seguir:

- Pelo teorema central do limite admite-se que o valor esperado da média amostral seja a média populacional que se pretende estimar, e eles serão tão mais próximos à medida que aumenta o tamanho da amostra. Isso é confirmado pelos resultados obtidos nas amostras? JUSTIFIQUE.
- Além do que foi dito acima, admite-se que o desvio padrão das médias amostrais será igual ao desvio padrão populacional dividido pela raiz quadrada do tamanho da amostra, e eles serão tão mais próximos à medida que aumenta o tamanho da amostra. As amostras retiradas confirmam essa afirmação? JUSTIFIQUE.
- Outra característica muito importante do teorema central do limite é que a distribuição amostral da média será cada vez mais próxima de uma distribuição normal, à medida que o tamanho da amostra aumenta, independentemente da forma da distribuição da variável na população. Sabe-se que a variável **Renda** não tem distribuição normal na população. Os resultados das amostras corroboram esta afirmação (usem os gráficos apropriados²). JUSTIFIQUE.
- Obtenha os intervalos de 95% de confiança para as 1000 amostras de cada tamanho. Construa gráficos apropriados para mostrá-los. Analisando os gráficos, o que você pode concluir sobre a precisão dos intervalos à medida que aumenta o tamanho de amostra? JUSTIFIQUE.

Segunda parte – Análise da variável Nota no ENEM

2) Há uma grande preocupação em estimar o comportamento da variável **Nota no ENEM** dos alunos, de maneira a caracterizar melhor seu o perfil. Alguém da universidade sugeriu que você retirasse uma amostra de 25 alunos e registrasse os valores de Nota no ENEM³. **Considere que a variância populacional de Nota no ENEM é DESCONHECIDA**. Com base nos resultados encontrados responda os itens a seguir (lembre-se de excluir inicialmente as linhas com dados perdidos).

- Construa um gráfico de probabilidade normal para os valores da amostra.
 - É possível considerar que os dados provêm de uma população com distribuição normal? JUSTIFIQUE.

¹ Pode ser feito através do suplemento de dados do Excel. Maiores detalhes no vídeo:

<https://www.youtube.com/watch?v=zkoMzbuYudI&t=1s>

² Podem ser feitos através do Microsoft Excel. Maiores detalhes no vídeo:

<https://www.youtube.com/watch?v=HRhKR30en70>

³ A amostragem também pode ser realizada através do Microsoft Excel. Maiores detalhes no vídeo:

https://www.youtube.com/watch?v=g4XFcUv_5NE

- a.2 – Com base na resposta da letra a.1 você recomendaria a utilização de técnicas como Intervalo de confiança e teste paramétrico de média? JUSTIFIQUE.
- b) Independente dos resultados da letra a), encontre o intervalo de 95% de confiança para a média populacional da Idade dos hóspedes. Interprete o resultado.
- c) Independente dos resultados da letra a), qual seria o tamanho mínimo de amostra necessário para obter um intervalo de 95% de confiança para a média populacional da Nota no ENEM, com uma precisão de 5 pontos? A amostra coletada é suficiente? JUSTIFIQUE.
- d) Independente dos resultados da letra a), a direção da TYU acredita que a média de Nota no ENEM dos alunos é maior do que 450 pontos, por pesquisas anteriores. Aplicando o teste estatístico apropriado, os dados confirmam isso, a 5% de significância? JUSTIFIQUE.
- e) Analise o intervalo de confiança da letra b). Os limites encontrados corroboram os resultados encontrados na letra d)? JUSTIFIQUE.
- f) Independente dos resultados da letra a), suponha 5% de significância e o desvio padrão amostral como boa estimativa do desvio padrão populacional. Se a média real da Nota no ENEM fosse de 470 pontos, qual seria o poder do teste? Você acha o valor aceitável? JUSTIFIQUE.
- g) Independente dos resultados da letra a), qual deveria ser o tamanho mínimo de amostra para detectar com 95% de probabilidade que a média da Nota no ENEM dos alunos é igual a 470 pontos. Suponha 5% de significância e o desvio padrão amostral como boa estimativa do desvio padrão populacional. A amostra coletada é suficiente? JUSTIFIQUE.

Terceira parte – Análise da variável Turno

3) Há interesse em avaliar a **Turno** dos alunos. Retire uma amostra aleatória de 200 elementos (novamente, primeiramente exclua as linhas com dados perdidos). Registre a proporção amostral de alunos do turno Integral e responda os itens a seguir:

- a) É possível aproximar a distribuição amostral da proporção por uma normal para cálculo do intervalo de confiança da proporção de alunos do turno Integral? JUSTIFIQUE.
- b) Independente dos resultados da letra a), encontre o intervalo de 95% de confiança para a proporção populacional de alunos do turno Integral. Interprete o resultado.
- c) Independente dos resultados da letra a), qual seria o tamanho mínimo de amostra para obter um intervalo de 95% para a proporção populacional de alunos do turno Integral com uma precisão de 5%. A amostra coletada é suficiente? JUSTIFIQUE.
- d) Independente dos resultados da letra a), qual seria o tamanho mínimo de amostra para obter um intervalo de 95% para a proporção populacional de alunos do turno Integral com uma precisão de 5%, se você não pudesse ter tirado a amostra piloto?
- e) Um colega seu acredita que a proporção de alunos do turno Integral é menor do que 35% (0,35) e você não.
- e.1 – É possível aproximar a distribuição amostral da proporção por uma normal para a realização de teste de hipóteses da proporção de alunos do turno Integral? JUSTIFIQUE.
- f) Independente dos resultados do item e.1, aplicando o teste de hipóteses apropriado a 5% de significância o seu colega está certo? JUSTIFIQUE.
- g) Analise o intervalo de confiança da letra b). Os limites encontrados corroboram os resultados encontrados na letra f)? JUSTIFIQUE.
- h) Independente dos resultados do item e.1, se a proporção real de alunos do turno Integral fosse igual a 33% (0,33) qual seria o poder do teste? Você acha o valor aceitável? JUSTIFIQUE.
- i) Independente dos resultados do item e.1, qual deveria ser o tamanho mínimo de amostra para detectar com 95% de probabilidade que a proporção populacional de alunos do turno Integral é igual a 0,33. Use 5% de significância. A amostra coletada é suficiente? JUSTIFIQUE.

Quarta parte – Análise da variável Renda em função da variável Opinião

4) Há interesse em avaliar a **Renda** dos alunos em função da sua **Opinião** sobre o curso. Mas contrariando a prática usual resolveu-se agrupar os valores de Opinião, criando uma nova variável OpiniãoC, na qual os valores Muito insatisfeito, Insatisfeito e Indiferente passarão a ser INSATISFEITO e os valores Satisfeito e Muito satisfeito passarão a ser SATISFEITO.

- a) Crie a nova variável OpiniãoC, recodificando a variável Opinião⁴ (novamente, primeiramente exclua as linhas com dados perdidos).
- b) Classifique o arquivo de dados em função dos valores de OpiniãoC⁵.
- c) Retire uma amostra aleatória de 15 rendas de alunos com OpiniãoC INSATISFEITO.
- d) Retire uma amostra aleatória de 15 rendas de alunos com OpiniãoC SATISFEITO.
- e) SUPONHA que as rendas são provenientes de uma distribuição normal, e que as variâncias populacionais são DESCONHECIDAS.

e.1 – Obtenha os intervalos de 95% de confiança para as médias das rendas com OpiniãoC INSATISFEITO e SATISFEITO. Interprete os resultados.

e.2 – Com base nos resultados do item e.1 há evidência de que as médias das rendas são diferentes entre alunos com OpiniãoC INSATISFEITO e SATISFEITO? JUSTIFIQUE.

e.3 – Aplique o teste de hipóteses apropriado e verifique se há indicação de diferença entre as variâncias de rendas de alunos com OpiniãoC INSATISFEITO e SATISFEITO. Use 5% de significância.

e.4 – Com base nos resultados do item e.3, aplique o teste de hipóteses apropriado e verifique se há evidência de que as médias das rendas são diferentes entre alunos com OpiniãoC INSATISFEITO e SATISFEITO. Use 5% de significância.

e.5 – Compare os resultados dos itens e.2 e e.4. Eles chegam à mesma conclusão sobre a diferença entre as médias? JUSTIFIQUE.

e.6 – Calcule o poder do teste para as seguintes diferenças entre as médias das rendas (em desvios padrões) dos alunos com OpiniãoC INSATISFEITO e SATISFEITO:

-3 -2,5 -2 -1,5 -1 -0,5 0,5 1,0 1,5 2 2,5 3

Use 5% de significância. Use como estimativa do desvio padrão populacional o desvio agrupado calculado a partir dos desvios padrões e tamanho de cada subgrupo no teste de diferença de médias do item e.4.

e.7 – Repita o item e.6 para tamanhos de amostra 20, 30 e 60 elementos em cada grupo, supondo que o desvio padrão agrupado permanece o mesmo). Construa um gráfico com os valores de média e poder do teste para os diferentes tamanhos de amostra (curva característica de operação). Analise os resultados e recomende qual seria o tamanho de amostra mais apropriado conciliando poder do teste e economia na amostragem. JUSTIFIQUE.

Professor Marcelo Menezes Reis, 24/04/2018

⁴ Veja como realizar isso no Microsoft Excel no vídeo: https://www.youtube.com/watch?v=rNj_SP399c4

⁵ Veja como realizar isso no Microsoft Excel no vídeo: https://www.youtube.com/watch?v=4Bz8U_E2_4o