Estatística Aplicada II

Gabarito - Lista 5

Problema 6.1.

Queremos encontrar o intervalo de confiança de 95% para o desvio-padrão da amostra obtida, para isso utilizaremos a distribuição X^2.

Em um primeiro momento, devemos transformar o desvio-padrão em variância - distribuição X^2 nos permite apenas analisar a variância de uma amostra -, dessa forma:

$$s^2 = 20^2 = 400$$

Com esse valor, podemos realizar outra transformação que é demonstrado seguir uma distribuição X^2: $\frac{(n-1)\,s^2}{\sigma^2}$, onde n é o número de observações e σ^2 representa a variância populacional.

Como esse valor segue uma distribuição X^2, podemos construir um intervalo de confiança para ele utilizando os valores críticos de ambos os lados da distribuição. Temos que o valor crítico que representa a área de 2,5% à esquerda da distribuição com 40-1 = 39 graus de liberdade é, aproximadamente 24,433; e que o valor crítico à direita é, aproximadamente 59,342. Dessa forma temos o seguinte intervalo:

$$[24,433 \le \frac{(n-1)\,s^2}{\sigma^2} \le 59,342]$$

Rearranjando os termos para que obtenhamos um intervalo de confiança para a variância populacional:

$$[\frac{24,433}{(n-1)\,s^2} \le \frac{1}{\sigma^2} \le \frac{59,342}{(n-1)\,s^2}] = [\frac{(n-1)\,s^2}{59,342} \le \sigma^2 \le \frac{(n-1)\,s^2}{24,433}] = [\frac{39 * 400}{59,342} \le \sigma^2 \le \frac{39 * 400}{24,433}]$$

Portanto, temos o intervalo de confiança para a variância populacional:

$$[262,883 \le \sigma^2 \le 638,481]$$

Para obtermos o intervalo de confiança para o desvio-padrão, basta extrair a raiz quadrada em ambos os limites, portanto obtemos:

$$[16,214 \leq \sigma \leq 25,268]$$

Problema 6.2.

Seguindo a mesma lógica do exercício anterior, utilizaremos a distribuição X^2 para encontrar o intervalo de confiança de 99% para o desvio-padrão.

Inicialmente transformemos o desvio-padrão em variância:

$$s^2 = 15^2 = 225$$

Depois disso, com a transformação $\frac{(n-1)\,s^2}{\sigma^2}$, podemos construir um intervalo com os valores críticos aproximados que representam uma área de 0,5% em cada uma das extremidades da distribuição com 70-1 = 69 graus de liberdade:

$$[43,275 \leq \frac{(n-1)\,s^2}{\sigma^2} \leq 104,215]$$

Com isso podemos manipular a expressão a fim de encontrar um intervalo de confiança para a variância populacional:

$$[\frac{43,275}{(n-1)\,s^2} \leq \frac{1}{\sigma^2} \leq \frac{104,215}{(n-1)\,s^2}] = [\frac{(n-1)\,s^2}{104,215} \leq \sigma^2 \leq \frac{(n-1)\,s^2}{43,275}] = [\frac{69 * 225}{104,215} \leq \sigma^2 \leq \frac{69 * 225}{43,275}]$$

Dessa forma, obtemos o intervalo para a variância populacional:

$$[148,971 \leq \sigma^2 \leq 358,752]$$

Extraindo a raiz quadrada, temos o intervalo de confiança para o desvio-padrão:

$$[12,205 \leq \sigma \leq 18,941]$$

Problema 6.10.

Para analisarmos se há alguma relação entre o tipo de carro e sua condição de tráfego podemos realizar um teste de hipótese - usando a distribuição X^2 - do tipo centrado em uma tabela de contingência.

Primeiramente, vamos definir nossas hipóteses nula e alternativa. Como a hipótese nula deve ser bem definida para que possamos calcular os valores esperados de cada categoria, temos:

H_0: Não há associação entre o tipo de automóvel e as suas condições para trafegar

H_1: Há associação entre o tipo de automóvel e suas condições para trafegar

Em posse das hipóteses nula e alternativa, façamos uma tabela com todos os dados:

	Condições Satisfatórias	**Condições Insatisfatórias**	**Total**
Automóveis Privados	114	30	144
Automóveis de Empresas	84	24	108
Peruas	36	12	48
Caminhões	44	20	64
Ônibus	36	12	48
Total	314	98	412

Tabela 1: Condições de tráfego de cada tipo de automóvel

Como nossa hipótese nula postula que não há qualquer relação entre as categorias e sua condição para trafegar, espera-se que as proporções dos carros de cada categoria nas divisões entre condições satisfatórios e insatisfatórios seriam as mesmas de sua proporção na amostra total de veículos - por exemplo, se a proporção de carros privados na amostra é $\frac{144}{412}$, esperaríamos que o número de carros com condições satisfatórias seria $314 * \frac{144}{412}$.

De forma mais geral, temos que o valor esperado para cada célula da tabela é dado por:

$$Valor\ Esperado\ =\ \frac{Total\ da\ Linha * Total\ da\ Coluna}{Total\ Geral}$$

Sendo assim, podemos construir uma nova tabela contendo o valor esperado:

	Condições Satisfatórias	**Condições Insatisfatórias**	**Total**
Automóveis Privados	114 (109,748)	30 (34,252)	144
Automóveis de Empresas	84 (82,311)	24 (25,689)	108
Peruas	36 (36,583)	12 (11,417)	48
Caminhões	44 (48,777)	20 (15,223)	64
Ônibus	36 (36,583)	12 (11,417)	48
Total	314	98	412

Tabela 2: Condições de tráfego de cada tipo de automóvel com valores esperados entre parênteses

Sabemos também que podemos construir um valor que segue uma distribuição X^2 da seguinte maneira:

$$X^2 = \Sigma \frac{(O-E)^2}{E}$$

Onde O representa o valor observado e E o valor esperado. Dessa forma, podemos obter a estatística X^2:

$$X^2 = 2,883$$

Agora devemos verificar se essa estatística do teste encontra-se ou não na região de rejeição, para isso devemos encontrar o valor crítico de X^2. Note que o termo $(O\text{-}E)^2$ do cálculo da estatística do teste não permite que se cometa erros à esquerda na distribuição X^2, de forma que devemos realizar um teste unicaudal e, portanto, o valor crítico será o valor que corresponde à 95% da área da distribuição X^2.

Temos que o número de graus de liberdade para esse teste é dado por:

$$\nu = (l - 1) * (c - 1) = (5 - 1) * (2 - 1) = 4 * 1 = 4$$

Onde l corresponde ao número de linhas na tabela de contingência e c corresponde ao número de colunas.

Verificando a tabela, temos que o valor crítico para 4 graus de liberdade, é 9,488. Como $2,883 < 9,488$, temos que a estatística do teste não se encontra na região de rejeição e, portanto, a uma significância de 5% não há qualquer relação entre o tipo de automóvel e suas condições para tráfego.

Problema 6.11.

Para realizarmos um teste de igualdade entre as duas variâncias podemos utilizar a distribuição F.

O que queremos testar é a hipótese de que as duas variâncias populacionais são iguais, em notação matemática temos:

H_0: $\sigma_1^2 = \sigma_2^2$

H_1: $\sigma_1^2 \neq \sigma_2^2$

Mas essas hipóteses equivalem às seguintes:

H_0: $\frac{\sigma_1^2}{\sigma_2^2} = 1$

H_1: $\frac{\sigma_1^2}{\sigma_2^2} \neq 1$

É demonstrável que o quociente $\frac{s_1^2}{s_2^2}$ segue uma distribuição F, portanto é possível montarmos a seguinte estatística de teste:

$$F = \frac{s_1^2}{s_2^2}$$

Então, seja $s_1^2 = 55$ e $s_2^2 = 48$, temos que a estatística F é igual a $\frac{55}{48} = 1,146$. Devemos checar, agora, se esse valor da estatística de teste se encontra na região de rejeição. A distribuição F possui dois graus de liberdade $\nu_1 = n_1-1$ e $\nu_2 = n_2-1$, onde n_1 e n_2 são os números de observações de cada amostra. Nesse caso temos $\nu_1 = 24$ e $\nu_2 = 29$. Além disso, sabemos que o teste deve ser bicaudal, mas como utilizamos o maior valor no numerador, só precisamos obter o valor crítico para o 0,5% à direita da distribuição. de forma que o valor crítico para se obter uma significância de 1% deve ser obtido através da tabela que corte a cauda da direita a um nível de 0,005.

Olhando a tabela obtemos um valor crítico para F de 2,7594. Como $1,146 < 2,7594$, temos que a estatística de teste não se encontra na região de rejeição, logo, não rejeitamos a hipótese nula.

Problema 6.14.

Para testarmos se há alguma diferença entre as vendas dos três tipos de loja usaremos a técnica de Análise de Variância (ANOVA). Essa técnica compara a soma dos quadrados entre grupos (BSS) - soma dos quadrados das diferenças entre as médias de cada fator (multiplicado por quantas observações temos naquele fator - com a soma dos quadrados dentro dos grupos (WSS) - soma dos quadrados das diferenças entre a observação e a média de seu fator.

Considerando as seguintes hipóteses:

H_0: $\mu_{Supermercado} = \mu_{Pequeno\ Mercado} = \mu_{Quiosque}$

H_1: Pelo menos uma média é diferente

A estatística F será dada pela seguinte expressão:

$$F = \frac{BSS/(k-1)}{WSS/(n-k)}$$

Onde (k-1) e (n-k) são os graus de liberdade da distribuição F - k representa o número de grupos presentes na análise e n representa o número total de observações.

Calculemos então BSS e WSS:

$$BSS = 4 * (\textit{média do supermercado} - \textit{média geral})^2 + 4 * (\textit{média de pequenos mercados} - \textit{média geral})^2 + 3 * (\textit{média de quiosques} - \textit{média geral})^2$$

$$\Rightarrow BSS = 4 * (329 - 284,272)^2 + 4 * (267,25 - 284,272)^2 + 3 * (249 - 284,272)^2 = 12.893,712$$

$$WSS = (355-329)^2 + (251-329)^2 + (408-329)^2 + (302-329)^2 + (288-267,25)^2 + (257-267,25)^2 + (225-267,25)^2 + (299-267,25)^2 + (155-249)^2 + (352-249)^2 + (240-249)^2$$

$$\Rightarrow WSS = 36.584,75$$

Como temos k=3 e n=11, a estatística do teste será:

$$F = \frac{12.893,712/2}{36.584,75/8} = 1,4097$$

Olhando na tabela da distribuição F a um nível de significância de 5% com graus de liberdade 2 e 8, encontramos o valor crítico 4,4590. Como $1,4097 < 4,4590$, a estatística de teste não se encontra na região de rejeição, logo não se rejeita H_0 e, portanto, a um nível de significância de 5% não há indícios de uma diferença significativa nas vendas.