ESTATÍSTICA APLICADA I | 21041

Atividades Formativas (Exercícios e Resoluções)

Catarina S. Nunes
Secção de Matemática, Departamento de Ciências e Tecnologia
Universidade Aberta

UAb

ATIVIDADE FORMATIVA 1

Temas:

- Introdução à Inferência Estatística
- Estimação Pontual
- Estimação por Intervalos de Confiança

Objetivos

- Conceitos das principais distribuições discretas e contínuas;
- Objetivos da Inferência Estatística
- Estimadores pontuais e suas propriedades
- Método da Máxima Verosimilhança
- Principais distribuições amostrais
- Intervalos de confiança para: propoções; diferença de proporções; médias; diferença de médias; variância e desvio padrão; razão de variâncias
- Implementação em **R** dos conceitos anteriores

Descrição da Atividade (Exercícios a Resolver)

1. Considere dois universos normais:

$$X_1 \sim N\left(\mu_1; \sigma_1\right)$$

$$X_2 \sim N\left(\mu_2; \sigma_2\right)$$

e amostras aleatórias independentes de cada um dos universos (de tamanhos n_1 e n_2 respetivamente).
Deduza a distribuição amostral da estatística $T = \bar{X}_1 + \bar{X}_2$.
Que poderia concluir se fosse desconhecida a distribuição dos universos?

2. Para o parâmetro θ de certa população, foram indicados dois estimadores: $\hat{\theta}_1$ e $\hat{\theta}_2$. Diga qual preferiria, sabendo que:

$$E\left[\hat{\theta}_1\right] = \frac{n+1}{n}\theta \quad VAR\left[\hat{\theta}_1\right] = \frac{k}{n}$$

$$E\left[\hat{\theta}_2\right] = \frac{n+1}{n}\theta \quad VAR\left[\hat{\theta}_2\right] = \frac{k}{n+3}$$

onde k é uma constante e n o tamanho da amostra.

3. Seja a estatística:

$$T = \frac{(n-1)X_1 + X_n}{n}$$

definida com base numa amostra aleatória de dimensão n, recolhida de uma população normal.

3.1 Verifique se T constitui um estimador não enviesado ou centrado para a média da população.

3.2 Será T um estimador consistente para aquele parâmetro da população?

4. Derive o estimador de máxima verosimilhança para o parâmetro p de uma população com distribuição de Bernoulli, calculados a partir de uma amostra de dimensão N.

5. Encontre os estimadores de máxima verosimilhança para os parâmetros μ e σ^2 de uma população normal, calculados a partir de uma amostra de dimensão N.

6. A variável aleatória X segue uma distribuição com função densidade de probabilidade dada por

$$f(x) = \begin{cases} \frac{(x-\alpha)}{2}, & se \;\; \alpha < x < \alpha + 2 \\ 0, & no \;\; caso \;\; contrário \end{cases}$$

onde α é um parâmetro desconhecido.
Recorrendo ao método de máxima verosimilhança, estime o parâmetro α a partir da seguinte amostra aleatória, constituída por 10 observações

3.5 ; 4.3 ; 2.8 ; 4.5 ; 2.9 ; 3.3 ; 3.8 ; 2.9 ; 4.0 ; 3.9

7. O tempo que uma máquina leva a executar certa operação em cada peça produzida é sujeita a variações. Para verificar se as condições de funcionamento estão dentro das normas, registou-se 12 vezes o referido tempo.
Os resultados (em segundos) foram os seguintes:

29	33	36	35
36	40	32	37
31	35	30	36

Construa um intervalo de confiança a 95% para o tempo médio de execução da tarefa pela máquina em análise.
NOTA: suponha que se pode aceitar que o tempo de execução da tarefa segue uma distribuição aproximadamente normal.

8. Um fabricante produz peças de diâmetro especificado em 100 mm. Querendo estimar o verdadeiro diâmetro num grande lote, selecionou 25 peças ao acaso, que depois de medidas forneceram os seguintes valores:

$$\sum x_i = 2530\, mm \qquad \sum (x_i - \bar{x})^2 = 384\, mm^2$$

8.1 Identifique o universo a ser objeto de estudo.

8.2 Apresente um estimativa para o diâmetro médio do lote.

8.3 Faça o mesmo através de um intervalo com um grau de confiança de 99%. Que vantagens resultam da identificação de um intervalo de confiança, em relação à estimativa pontual na alínea 6.2?

8.4 Quantas peças deveriam ser incluídas na amostra, se se pretendesse aumentar a precisão do intervalo - reduzir a sua amplitude para 3 mm?

9. Pretende-se avaliar a proporção p de indivíduos com uma certa doença numa população com um erro que não exceda $\pm 2\%$ para um coeficiente de confiança de 95%.
Qual a dimensão da amostra a recolher, se:

9.1 nada se sabe sobre p;

9.2 se admitir que p é da ordem de grandeza 10%.

Exercícios Computacionais em R

10. Elabore uma rotina e implemente-a no R, para gerar 100 números pseudo-aleatórios com distribuição:

10.1 $X \sim Poisson(\lambda = 2)$

10.2 $X \sim U\,[0, 1]$ (Distribuição Uniforme)

10.3 $X \sim N(5, 5)$ (Distribuição Normal)

11. Considere a seguinte amostra:

109 118 119 121 121 134 111 125 137 121 133 111 118 102 112 121 109 117 114
105 132 122 132 134 109 112 115 121 122 120 128 119 116 114 111 102 120 122

Determine um intervalo de confiança para a média da população de onde esta amostra foi retirada. Considere $\alpha = 1\%$.

12. Uma empresa de transportes consome em média, 1000 litros de biodiesel, por dia. O consumo diário durante a última semana foi de 995, 1015 , 985 , 1004 , 907, 1002, 976 litros.
Determine um I.C. a 99% para variância do consumo diário.

UAb

ATIVIDADE FORMATIVA 2

Temas:

- Testes de Hipóteses Paramétricos
- Testes de Hipóteses Não Paramétricos

Objetivos

- Definição de hipóteses estatísticas
- Testes de hipóteses para proporções e diferenças de proporções
- Testes de hipóteses para médias e diferenças entre médias
- Teste de hipóteses para a variâncias e a razão de variâncias
- Implementação em **R** de testes paramétricos
- Erro de tipo I e erro de tipo II
- Teste de Qui-quadrado para a independência
- Testes de ajustamento de distribuições
- Testes não paramétricos para comparação de amostras independentes e dependentes

Descrição da Atividade (Exercícios a Resolver)

1. O departamento de controlo de qualidade de uma empresa conserveira, especificou que o peso líquido médio por embalagem de atum deve ser 500 gramas.
A experiência passada indica que os pesos são normalmente distribuídos com desvio-padrão σ=15 gramas.
Se numa amostra de 20 embalagens de atum for encontrado um peso líquido médio de 495 gramas, constitui isso prova suficiente de que o verdadeiro peso médio é inferior ao estabelecido? (α=0.05)

2. A despesa semanal em alimentação de um agregado familiar pertencente à classe alta de rendimentos tem desvio padrão σ=170 euros. Crê-se que a despesa semanal média é de 1500 euros; sendo 1420 euros a hipótese alternativa, e fixando o nível de significância α=0.05, com base numa amostra aleatória de tamanho n, obteve-se a probabilidade de erro de tipo II β=0.10 aproximadamente.
Determine o tamanho da amostra.

3. Num processo de fabricação de placas de vidro, produzem-se bolhas que se distribuem aleatoriamente pelas placas. Com base na abundante informação recolhida pelo Departamento de Qualidade, a densidade média das bolhas estima-se, até há pouco tempo, em 0.4 bolhas/m^2.

Recentemente fez-se uma tentativa para melhorar o processo produtivo, em particular no que toca ao aparecimento daqueles defeitos. Depois de serem introduzidas as alterações no processo, recolheu-se uma amostra constituída por 15 placas de 4.5 m^2 e registou-se o número de bolhas em cada uma delas. Os resultados foram os seguintes:

1, 0, 3, 0, 2, 3, 1, 0, 1, 1, 2, 0, 1, 2, 1

Verifique, ao nível de significância de 5%, se a densidade esperada de bolhas diminuiu.

4. O presidente de um clube de futebol pretende avançar com um projeto que tem suscitado opiniões controversas. Ele afirma em sua defesa que mais de 50% dos sócios concordam com o projeto. O que se deve concluir da afirmação do presidente, sabendo que em 80 sócios escolhidos ao acaso, 47 se manifestaram a favor? (Utilize α=0.10)

5. No laboratório JP Análises, foi recentemente desenvolvido um método mais barato e mais rápido do que o convencional para efetuar certas análises de rotina.

Em testes realizados com o objetivo de comparar a fiabilidade do novo método com a do convencional, repetiu-se sucessivamente a mesma análise, tendo-se obtido os seguintes resultados:

	Nº de Observações	$\sum_n (x_n - \bar{x})^2$
Método convencional	24	30.5
Método novo	30	55.7

5.1 Poder-se-á concluir que o novo método é menos fiável do que o método convencional?

5.2 Será que a média de 3 observações obtidas com o novo método é mais precisa do que uma observação recolhida com base no método convencional?

Efetue os testes adequados ao nível de significância de 5%.

6. A fábrica FunBallons tem uma máquina que produz balões com uma resistência à compressão com valor esperado de 5.18 kg/cm^2 e variância de 0.0625 $(kg/cm^2)^2$.

Os últimos 12 balões produzidos foram recolhidos e ensaiados, tendo-se obtido para a resistência média à compressão o valor de 4.95 kg/cm^2. Admite-se que a resistência à compressão segue uma distribuição normal.

6.1 Poder-se-á afirmar, ao nível de significância de 5%, que os balões produzidos recentemente são menos resistentes do que o habitual?

6.2 Qual a potência do teste efetuado em **6.1**, admitindo que o valor esperado da resistência à compressão dos balões produzidos recentemente é de 4.90 kg/cm^2.

7. Testes de Hipóteses Não Paramétricos

7.1 Explique como aplicaria o Teste dos sinais para comparar duas amostras dependentes.

7.2 Explique como aplicaria o Teste de Mann-Whitney para comparar duas amostras independentes.

8. Admita-se que foi conduzida uma experiência no âmbito da qual se procurou testar se existe alguma relação entre a qualidade da secagem de máquinas de lavar roupa de um certo tipo e a velocidade de rotação a que se eleva o tambor da roupa na fase de centrifugação. Os resultados desta experiência, efetuada com base no comportamento de 90 máquinas (amostra aleatória), estão representados na tabela seguinte:

		Qualidade da secagem				
		Medíocre	Suficiente	Boa	Muito Boa	Total
Velocidade de rotação	600rmp	12	8	7	3	30
	900rpm	9	10	7	4	30
	1200rpm	2	9	8	11	30
	Total	23	27	22	18	90

Verifique se as duas variáveis em questão são ou não relacionadas.

9. Uma companhia área registou qual o número de passageiros que, tendo efetuado reserva para um determinado destino, acabava por não fazer o *check in*. Para 100 voos escolhidos aleatoriamente, os resultados foram os seguintes:

Nº de ausências	Nº de voos
0	21
1	36
2	23
3	13
4	4
5	2
6	1

Teste, ao nível de significância de 5%, se a distribuição do número de ausências por voo segue uma distribuição de Poisson.

Exercícios Computacionais em R

10. Considere a seguinte amostra aleatória, relativa aos tempos de visualização de um indíviduo (em minutos), de dois canais de televisão em 10 dias consecutivos:

Canal 1: 35 47 51 42 25 43 59 72 49 61
Canal 2: 32 49 54 39 39 44 65 43 65 55

Teste a hipótese de igualdade de médias populacionais, admitindo a igualdade de variâncias populacionais.

11. Uma empresa de transportes consome em média, 1000 litros de biodiesel, por dia. O consumo diário durante a última semana foi de 995, 1015 , 985 , 1004 , 907, 1002, 976 litros.
Teste ao nível de significância de 5% que a média do consumo diário é igual a 990 litros.

12. Dois grupos de individuos foram vacinados contra uma doença com duas vacinas experimentais. Após um período de tempo préviamente fixado, fez-se a determinação de um índice que representa a abundância de anticorpos por mm^3 de sangue, sendo os valores obtidos nos dois grupos, os que estão registados na tabela seguinte (admita a normalidade dos dados):

Grupo 1 (x)	94.8	89.4	96.7	93.5	96.4	85.3	87.4	89.5	85.7								
Grupo 2 (y)	83.4	85.6	86.6	84.4	83.4	86.7	83.4	82.7	84.8	83.8	84.9	84.4	85.3	84.8	82.3	81.5	84.3

12.1 Teste ao nível de significância de 1% se $\sigma_x^2 = \sigma_y^2$.

12.2 Averigue se a vacina administrada ao grupo 1, produz piores resultados do que a administrada ao grupo 2. $(\alpha = 5\%)$.

UAb

ATIVIDADE FORMATIVA 3

Temas:

- Introdução à Inferência Estatística
- Estimação Pontual
- Estimação por Intervalos de Confiança
- Testes de Hipóteses Paramétricos
- Testes de Hipóteses Não Paramétricos
- Análise de Variância a 1 fator

Objetivos

Esta atividade formativa é de revisões sobre todos os temas da unidade curricular, nesse sentido os objetivos são os de todos os temas estudados ao longo do semestre. Pretende-se que esta atividade seja uma preparação adicional para as provas finais.

Descrição da Atividade (Exercícios a Resolver)

1. Uma variável aleatória X representa o número de avarias de um dispositivo eletrónico durante um mês. A variável X segue uma lei de Poisson de parâmetro λ desconhecido. Foram indicados dois estimadores para λ:

$$\hat{\lambda} = \frac{X_1+X_2+\ldots+X_n}{n} \quad \tilde{\lambda} = \frac{X_1+X_n}{2}$$

com população: $X \sim Poisson(x;\lambda)$ e amostra: $(X_1, X_2, \ldots, X_n)$. Compare os dois estimadores propostos quanto ao não enviesamento e eficiência.

2. Os valores que se apresentam seguidamente representam os atrasos de comboios, expressos em minutos, verificados numa determinada estação:

2.23, 7.74, 1.03, 0.08, 11.14, 12.72, 0.42, 5.17

Admitindo que os atrasos seguem uma distribuição uniforme $U(0,\theta)$ e que a amostra é aleatória, calcule a estimativa de máxima verosimilhança do parâmetro θ.
Recorde que sendo $X \sim U(a,b)$, se tem

$$f(x) = \begin{cases} 1/(b-a) & se\ a \leq x \leq b \\ 0, & se\ x < a\ ou\ se\ x > b \end{cases}$$

3. No âmbito do estudo de uma determinada operação de montagem, recolheram-se 25 observações do tempo necessário para a sua realização. A variância amostral obtida foi de $(0.3\ horas)^2$.
Construa intervalos de confiança a 90, 95 e 99% para a variância dos tempos de montagem, indicando as hipóteses subjacentes à construção destes intervalos. Admita que se trata de uma distribuição normal.

4. O fabricante do chá "SleepGood"reivindicou que o chá era muito eficaz nos casos de insónia. Recolhida uma amostra de 250 pessoas, verificou-se que o chá ajudou 190.

4.1 Qual o intervalo de confiança a 95% para a proporção de eficácia do chá, em termos populacionais?

4.2 Que dimensão deveria ter a amostra de pessoas a recolher, para que a proporção p de eficácia do chá se situe num intervalo de amplitude 0.10 com um nível de confiança de 99%?

5. O gerente de um banco está interessado em analisar a diferença entre os valores esperados dos saldos das contas à ordem de duas das suas agências. De cada uma delas foi recolhida uma amostra aleatória de saldos, tendo-se registado os seguintes resultados:

Agência A: dimensão da amostra: $N_A = 17$

$$\bar{x}_A = 48.2\ euros$$
$$s_A = 2.74\ euros$$

Agência B: dimensão da amostra: $N_B = 13$

$$\bar{x}_B = 41.25\ euros$$
$$s_B = 3.91\ euros$$

Calcule os intervalos de confiança a 90% e a 95% para a diferença entre os valores esperados dos saldos à ordem das agências A e B.

6. A PrecisionQ é uma empresa que tenciona importar um grande lote de instrumentos de precisão para posterior distribuição em Portugal.
Os fabricantes garantem que o respetivo peso médio é de 100 gramas. Sendo no entanto, o peso uma característica importante na qualidade do produto, resolveu-se testar a garantia do fabricante. Para tal, o departamento técnico da PrecisionQ obteve uma amostra de 15 instrumentos, donde resultaram os seguintes valores:

$$\sum_{i=1}^{15} x_i = 1344\ \ g \qquad \sum_{i=1}^{15} (x_i - \bar{x})^2 = 3150\ \ g^2$$

Admitindo que o peso é normalmente distribuído, diga qual a conclusão a tirar (com α=0.01).

7. Numa sondagem 60 das 200 pessoas inquiridas revelaram-se conhecedoras de determinado produto.
Após uma campanha publicitária foi feito novo inquérito a 300 pessoas, das quais 111 se revelaram conhecedoras do produto.
Pode-se considerar que, devido à campanha de publicidade, o referido produto se tornou mais conhecido? (utilize α=0.01)

8. Com o objetivo de analisar o tipo de audiência de um programa de televisão.

8.1 Obtiveram-se as idades de 100 telespetadores, selecionados aleatoriamente, cujo desvio padrão é $s = 2.95$.
Convencionando-se que a audiência é considerada heterogénea se a variância das idades ultrapassar os 6 anos, que conclui para $\alpha = 0.05$?

8.2 Recolhemos uma amostra de 200 telespetadores, dos quais 110 eram do sexo feminino.
Podemos concluir para $\alpha = 0.01$ que o programa é menos cativante para as pessoas do sexo masculino?

9. Os dados da tabela seguinte referem-se a fornecimentos de morangos feitos por 3 agricultores à empresa IoguLeite:

	morangos rejeitados	morangos aceitáveis	morangos bons
Agricultor A	12	23	89
Agricultor B	8	12	62
Agricultor C	21	30	119

Teste, para um nível de significância α=0.05, se a qualidade dos morangos varia com os fornecedores.

10. A empresa Delta estava interessada em delegar numa terceira empresa, o serviço relativo ao transporte dos artigos por ela produzidos, desde o local de produção até ao porto de mar mais próximo. Para o efeito, procedeu a um estudo de mercado em relação ás empresas Alfa, Beta e Gama, no qual a decisão pela escolha da empresa a selecionar entre as três, iria basear-se no menor tempo médio gasto para transportar os produtos. O estudo de mercado revelou que os tempos gastos pelas três empresas de transporte seguem uma distribução normal com a mesma variância. Os dados seguintes são relativos a três amostras aleatórias de seis observações, dos tempos gastos (em horas) pelas três empresas no referido transporte:

$$Alfa: \ \sum_{i=1}^{6} x_{\alpha i} = 202.99998; \ \sum_{i=1}^{6} (x_{\alpha i} - \bar{x}_\alpha)^2 = 0.13335$$

$$Beta: \ \sum_{i=1}^{6} x_{\beta i} = 215.5002; \ s_\beta = \sqrt{0.03767}$$

$$Gama: \ \sum_{i=1}^{6} x_{\gamma i} = 221.2998; \ s_\beta^2 = 0.02567$$

Justificando adequadamente e retirando todas as conclusões possíveis, compare os tempos médios gastos pelas três empresas. Com base nesta comparação o que aconselharia à Delta? (utilize um nível de significância de 0.05)

11. Um determinado ministério está preocupado com os aumentos dos custos verificados no decurso de projetos de restauração de edifícios, nas cidades A, B, C e D. Portanto, decidiu analisar os custos associados a diferentes projetos, calculando para cada um deles a razão entre o custo final incorrido e o custo inicialmente previsto. Os resultados foram:

Cidade	Custo incorrido/Custo previsto					
A	1.0	0.8	1.9	1.1	2.7	
B	1.7	2.5	3.0	2.2	3.7	1.9
C	1.0	1.3	3.2	1.4	1.3	2.0
D	3.8	2.8	1.9	3.0	2.5	

O fator ”cidade”tem efeito significativo sobre o agravamento dos custos? (Utilize um nível de confiança α=0.05.)

12. Considere a seguinte tabela de ANOVA obtida a partir de uma experiência cujo objetivo consiste na comparação do efeito de certos tratamentos aplicados a culturas de milho e onde foram usadas amostras de igual dimensão:

Origem de variação	Graus de liberdade	Soma de Quadrados	Quadrados Médios	Razão de variância
Entre Tratamentos		246		
Resíduo	30		22	—
Total	34		—	—

12.1 Preencha a tabela.

12.2 Determine o número de tratamentos em estudo.

12.3 Determine a dimensão das amostras usadas.

12.4 Conclua acerca da existência ou não de diferenças significativas entre tratamentos, ao nível de significância de 5%.

UAb

ATIVIDADE FORMATIVA 1 - Proposta de Resolução

1. $\bar{X}_1 \sim N\left(\mu_1; \frac{\sigma_1}{\sqrt{n_1}}\right)$ e $\bar{X}_2 \sim N\left(\mu_2; \frac{\sigma_2}{\sqrt{n_2}}\right)$
pelo Teorema da aditividade da Normal temos:
$\left(\bar{X}_1 + \bar{X}_2\right) \sim N\left(\mu_{\bar{X}_1+\bar{X}_2}; \sigma_{\bar{X}_1+\bar{X}_2}\right)$
$E\left[\bar{X}_1 + \bar{X}_2\right] = E\left[\bar{X}_1\right] + E\left[\bar{X}_2\right] = \mu_1 + \mu_2$
$VAR\left[\bar{X}_1 + \bar{X}_2\right] = VAR\left[\bar{X}_1\right] + VAR\left[\bar{X}_2\right] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$
já que as amostras são independentes e portanto também X_1 e X_2. Ficará então:
$T = \bar{X}_1 + \bar{X}_2 \sim N\left(\mu_1 + \mu_2; \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$
Se for desconhecida a distribuição dos universos, desde que as amostras sejam grandes, pelo teorema do limite central a distribuição amostral de T será aproximadamente normal.

2. Qualquer dos estimadores é enviesado. Como $E\left[\hat{\theta}_1\right] = E\left[\hat{\theta}_2\right] \neq \theta$, será de preferir o mais eficiente.
$VAR\left[\hat{\theta}_1\right] = \frac{k}{n} > VAR\left[\hat{\theta}_2\right] = \frac{k}{n+3}$ deve preferir-se $\hat{\theta}_2$.

3. .

3.1 T é um estimador não enviesado para μ se e só se $E\left[T\right] = \mu$
$E[T] = E\left[\frac{(n-1)X_1+X_n}{n}\right] = \frac{1}{n}\left[(n-1)E\left[X_1\right] + E\left[X_n\right]\right] = \frac{1}{n}\left[(n-1)\mu + \mu\right] = \mu$
T é um estimador não enviesado para μ

3.2 Se $n \overset{lim}{\rightarrow} \infty\ EQM(T) = 0$ então T é um estimador consistente em média quadrática.
$EQM(T) = VAR(T) + [enviesamento(T)]^2$
como T é um estimador não enviesado para μ o enviesamento é zero, e:
$n \overset{lim}{\rightarrow} \infty\ EQM(T) = n \overset{lim}{\rightarrow} \infty\ VAR(T) =$
$n \overset{lim}{\rightarrow} \infty\ VAR\left[\frac{(n-1)X_1+X_n}{n}\right] =$
$n \overset{lim}{\rightarrow} \infty\ \frac{1}{n^2} VAR\left[(n-1)X_1 + X_n\right] =$
$n \overset{lim}{\rightarrow} \infty\ \frac{1}{n^2}\left[(n-1)^2 VAR(X_1) + VAR(X_n)\right] =$
$n \overset{lim}{\rightarrow} \infty\ \frac{1}{n^2}\left[(n^2-2n+1)\sigma^2 + \sigma^2\right] = \sigma^2$
T não é um estimador consistente em média quadrática para μ

4. $\hat{p}_{ML} = \frac{\sum x_i}{N}$

5. $\hat{\mu}_{ML} = \bar{X}$ e $\hat{\sigma^2}_{ML} = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \mu\right)^2$

6. $\hat{\alpha} = 2.5$

7. $[32.15; 36.19]$

8. .

8.1 O lote a fornecer ao cliente

8.2 101.2 mm

8.3 $[98.96; 103.44]$

8.4 $n \approx 48$

9. .

9.1 $n \geq 2401$

9.2 $n \geq 865$

10. .

10.1 $> rpois(100, 2)$

10.2 $> runif(100, min = 0, max = 1)$

10.3 $> rnorm(100, 5, 5)$

11. $> dados < -c(109, 118, 119, 121, 121, 134, 111, 125, 137, 121, 133, 111, 118, 102, 112, 121, 109, 117, 114, 105, 132, 122, 132, 134, 109, 112, 115, 121, 122, 120, 128, 119, 116, 114, 111, 102, 120, 122)$

$> length(dados)$

$[1]38$

$> liminf < -mean(dados) - qt(0.995, 37) * sd(dados)/sqrt(38)$

$> limsup < -mean(dados) + qt(0.995, 37) * sd(dados)/sqrt(38)$

$> limsup$

$[1]122.5675$

$> liminf$

$[1]114.7483$

R: $[114.7483; 122.5675]$

12. $X < -c(995, 1015, 985, 1004, 907, 1002, 976)$

$> c((length(X)-1)*var(X)/qchisq(0.995, df = 6), (length(X)-1)*var(X)/qchisq(0.005, df = 6))$

$[1]420.4167 \;\; 11539.7444$

UAb

ATIVIDADE FORMATIVA 2 - Proposta de Resolução

1. X - v.a. peso em gramas de embalagens de atum. $X \sim N(\mu; \sigma = 15)$.
Amostra: $n = 20$ e $\bar{X} = 495$.
Teste de hipóteses para μ, população normal e variância (σ^2) conhecida:

$$H_0 : \mu \geq 500 \quad (\mu \geq \mu_0)$$

$$H_1 : \mu < 500 \quad (\mu < \mu_0) \qquad (\alpha = 0.05)$$

Estatística de teste:

$$T = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

Tabela da distribuição normal reduzida, valor crítico $-z_{1-\alpha} = -z_{0.95} = -1.645$.
Região de Aceitação (RA): $[-1.645; +\infty[$
Região de Rejeição ou ou Região Critica (RC): $]-\infty; -1.645[$
Decisão: Rejeitar H_0 se $T \in RC : T < -1.645$ ou Não Rejeitar H_0 se $T \in RA : T \geq -1.645$.
Valor de Teste:

$$T_{obs} = \frac{495 - 500}{15/\sqrt{20}} = -1.491 \geq -1.645$$

Valor de teste $T_{obs} = -1.491 \in RA \Rightarrow$ não se rejeita H_0. Não existe prova que o peso médio das embalagens de atum tenha decrescido.

2. X - v.a. despesa semanal em alimentação por agregado familiar pertencente à classe alta de rendimentos, em euros, com $\sigma = 170$ euros.
Amostra: n ?
Teste de hipóteses para μ, população normal e variância (σ^2) conhecida:

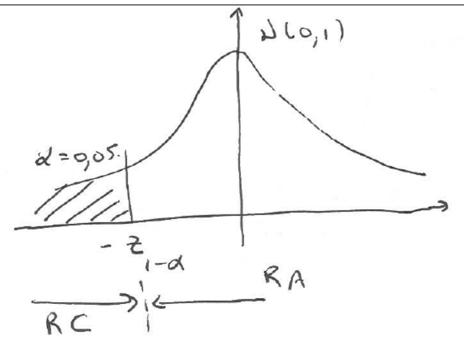
$$H_0 : \mu = 1500 \quad (\mu = \mu_0)$$

$$H_1 : \mu = 1420 \quad (\mu = \mu_1) \qquad (\alpha = 0.05)$$

n ? tal que $\beta = 0.10$.
O teste utilizado (supondo n grande) é

$$T = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

Semelhante ao exercício anterior, pela tabela da distribuição normal reduzida $-z_{0.95} = -1.645$, com as mesmas RA e RC.



$$P\left[T \in RC | H_0 \quad Verdadeira\right] = \alpha$$

n será tal que satisfaça:

$$\alpha = P\left[Rejeitar \;\; H_0 | H_0 \quad Verdadeira\right] = 0.05$$

e

$$\beta = P\left[Nao \quad Rejeitar \;\; H_0 | H_0 \quad Falsa\right] = 0.10$$

desenvolvendo

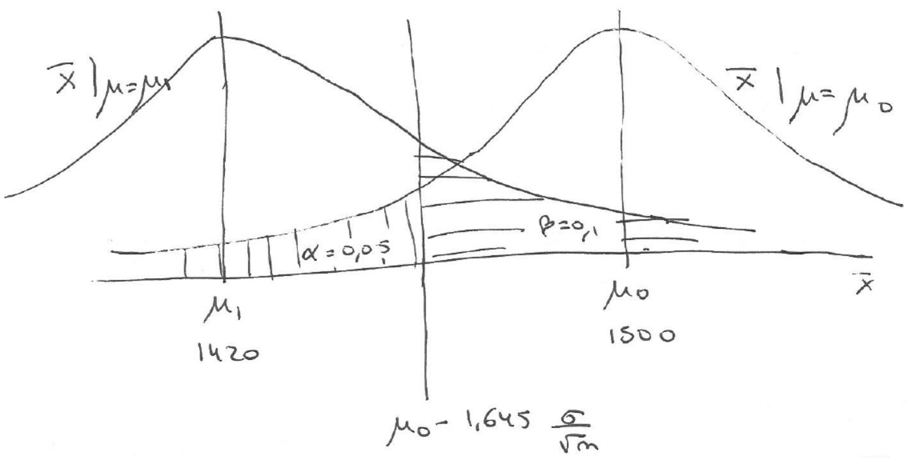
$$\alpha = 0.05 = P\left[\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \leq -1.645|_{\mu=\mu_0}\right]$$
$$= P\left[\bar{X} \leq \mu_0 - 1.645\frac{\sigma}{\sqrt{n}}|_{\mu=\mu_0}\right]$$

A fronteira da RC em termos de $\bar{X}$ é: $\mu_0 - 1.645\frac{\sigma}{\sqrt{n}}$.

$$\beta = 0.10 = P\left[\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > -1.645|_{\mu=\mu_1}\right]$$
$$= P\left[\bar{X} > \mu_0 - 1.645\frac{\sigma}{\sqrt{n}}|_{\mu=\mu_1}\right]$$
$$= P\left[Z > \frac{(\mu_0 - 1.645\sigma/\sqrt{n}) - \mu_1}{\sigma/\sqrt{n}}\right]$$
$$= P\left[Z > \frac{(1500 - 1.645\ 170/\sqrt{n}) - 1420}{170/\sqrt{n}}\right]$$

Com $P(Z < z) = 0.10$ para $z = 1.282$ (vendo na tabela da distribuição normal reduzida). Virá:

$$\frac{(1500 - 1.645\ 170/\sqrt{n}) - 1420}{170/\sqrt{n}} = 1.282 \Rightarrow n = 39$$



3. Y - v. a. número de bolhas em cada placa de $4.5m^2$ e segue uma distribuição de Poisson $Y \sim Poisson(\lambda)$.
Antes de melhorar o processo, o número médio de bolhas é:

$$\lambda = 0.4\,\frac{bolhas}{m^2} \times 4.5m^2 = 1.8\,\frac{bolhas}{placa}$$

Teste de hipóteses unilateral à esquerda para $\mu_y = E[Y]$:

$$H_0 : \mu_y = \lambda = 1.8$$

$$H_1 : \mu_y < \lambda = 1.8 \qquad (\alpha = 0.05)$$

Após as alterações no processo, obteve-se:

$$\bar{y} = \frac{1+0+3+\cdots+1+2+1}{15} = \frac{18}{15} = 1.2\frac{bolhas}{placa}$$

Admitiu-se que $Y \sim Poisson(\lambda)$, então $E[Y] = \mu_y = \lambda$ e $VAR[Y] = \lambda$. Com $N = 15$ vamos considerar suficiente para se aproximar pela distribuição Normal:

$$\bar{Y} \sim N\left(\mu_y = \lambda, \sigma^2 = \frac{\lambda}{N}\right)$$

A estatística de teste:

$$Z = \frac{\bar{Y} - \lambda}{\sqrt{\frac{\lambda}{N}}}$$

que quando H_0 é verdadeira $Z \sim N(0,1)$ - aproximadamente.
Valor de Teste:

$$Z_{obs} = \frac{1.2 - 1.8}{\sqrt{\frac{1.8}{15}}} = -1.732$$

Para $\alpha = 0.05$ da tabela da distribuição Normal reduzida temos $-z_{1-\alpha} = -z_{0.95} = -1.645$.
Decisão:
$-1.732 < -1.645 \Rightarrow H_0$ é rejeitada ao nível de significância de 5%.
Valor de teste $Z_{obs} = -1.732 \notin RA \Rightarrow$ rejeita-se H_0. Confirma-se a diminuição da densidade de defeitos (bolhas).

4. p - a proporção dos sócios que concordam com o projecto - Pop. Bernoulli.
Amostra: $n = 80$ e $\sum x_i = 47$.
Teste de hipóteses para p:

$$H_0 : p \leq 0.5$$

$$H_1 : p > 0.5 \qquad (\alpha = 0.10)$$

Será de sustentar a opinião do presidente se se conseguir rejeitar H_0.
Estatística de teste (amostra grande $n > 30$):

$$Z = \frac{\bar{X} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$$

Neste caso $p_0 = 0.5$.

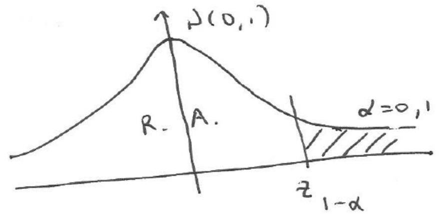


Tabela da distribuição Normal $z_{1-\alpha} = z_{0.90} = 1.282$.

$$RA : \,]-\infty; 1.282\,] \quad e \quad RC : \,]\,1.282; +\infty\,[$$

Valor de Teste:

$$Z_{obs} = \frac{\bar{X} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.5875 - 0.5}{\sqrt{\frac{0.5(0.5)}{80}}} = 1.75$$

com $\bar{X} = \sum x_i/n = 47/80 = 0.5875$.
Decisão:
$Z_{obs} = 1.75 > 1.282 \Leftrightarrow Z \in RC \Rightarrow$ rejeita-se H_0.

Com $\alpha = 0.1$ é de sustentar a opinião do presidente.

5. X_c - v.a resultados obtidos nas análises com o método comvencional.
X_N - v.a. resultados obtidos nas análises com o método novo.

5.1 Como

$$s^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{n-1}$$

temos

$$s_c^2 = \frac{30.5}{24-1} = 1.33 \Rightarrow s_c = 1.15$$

$$s_N^2 = \frac{55.7}{30-1} = 1.92 \Rightarrow s_N = 1.39$$

Admite-se que os resultados das análises seguem distribuições normais

$$X_c \sim N(\mu_c; \sigma_c)$$

$$X_N \sim N(\mu_N; \sigma_N)$$

A fiabilidade do método $\Rightarrow$ repetibilidade $\Rightarrow$ dispersão dos resultados que são obtidos em condições idênticas $\Rightarrow$ trata-se de comparação de variâncias de duas populações normais.
Testes de Hipóteses:

$$H_0 : \sigma_N^2 = \sigma_c^2 \quad \left(\frac{\sigma_N^2}{\sigma_c^2} = 1\right)$$

$$H_1 : \sigma_N^2 > \sigma_c^2 \quad \left(\frac{\sigma_N^2}{\sigma_c^2} > 1\right)$$

Se H_0 for verdadeira, a estatística de teste

$$F = \frac{s_n^2}{s_c^2} \sim F_{n_N - 1; n_c - 1}$$

RC, sendo

$$H_1 : \frac{\sigma_N^2}{\sigma_c^2} > 1 \Rightarrow \{F \in \mathbb{R} : F > F_{n_N-1;n_c-1;1-\alpha}\} \Leftrightarrow \left]\, F_{n_N-1;n_c-1;1-\alpha}; +\infty \right[$$

Tabela F de Snedcor com $n_N - 1 = 29$ e $n_c - 1 = 23$ graus de liberdade e $\alpha = 0.05$, temos $F_{29,23,0.95} \cong 1.96$, e a $RC : \left] 1.96, +\infty \right[$.
Valor de Teste:

$$F_{obs} = \frac{1.92}{1.33} = 1.44 < 1.96 \Rightarrow \notin RC$$

Valor de teste $F_{obs} = 1.44 \in RA \Rightarrow$ não se rejeita H_0. Ao nível de confiança de 95% não se pode concluir que o método novo é menos fiável do que o método convencional.

5.2 Seja Z a média de 3 observações obtidas com o novo método:

$$VAR(Z) = \sigma_z^2 = \frac{\sigma_N^2}{3}$$

Considerando hipóteses idênticas às de 5.1

$$H_0 : \sigma_c^2 = \sigma_z^2 = \frac{\sigma_N^2}{3}$$

$$H_1 : \sigma_c^2 > \sigma_z^2 = \frac{\sigma_N^2}{3}$$

Se H_0 for verdadeira, a estatística de teste

$$F = \frac{s_c^2}{s_z^2} = \frac{s_c^2}{\frac{\sigma_N^2}{3}} \sim F_{n_c-1;n_N-1}$$

Região Critica (RC): $F_{23;29;0.95} \cong 1.91$ (Tabela F Snedcor) $RC : \left] 1.91; +\infty \right[$.
Valor de Teste Valor de Teste:

$$F_{obs} = \frac{1.33}{\frac{1.92}{3}} = 2.08 > 1.91 \Rightarrow \in RC$$

Valor de teste $F_{obs} = 2.08 \notin RA \Rightarrow$ rejeita-se H_0. Utilizando a média de 3 observações obtidas a partir do novo método é mais preciso do que recorrer a observações individuais obtidas a partir do método convencional.

6. X - resistência ‘compressão dos balões. $X \sim N(\mu, \sigma^2)$.
$\mu = 5.18 Kg/cm^2$
$\sigma^2 = 0.0625\,(Kg/cm^2)^2 \Rightarrow \sigma = 0.25 Kg/cm^2$
$n = 12$ e $\bar{x} = 4.9 Kg/cm^2$

6.1 Teste de hipóteses

$$H_0 : \mu = 5.18$$

$$H_1 : \mu < 5.18$$

Estatística de teste

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

$RC:]-\infty; -z_{1-\alpha}[\Leftrightarrow]-\infty; -z_{0.95}[\Leftrightarrow]-\infty; -1.645[$
Valor de teste:

$$Z_{obs} = \frac{4.95 - 5.18}{0.25/\sqrt{12}} = -3.18 < -1.645 \Rightarrow \in RC$$

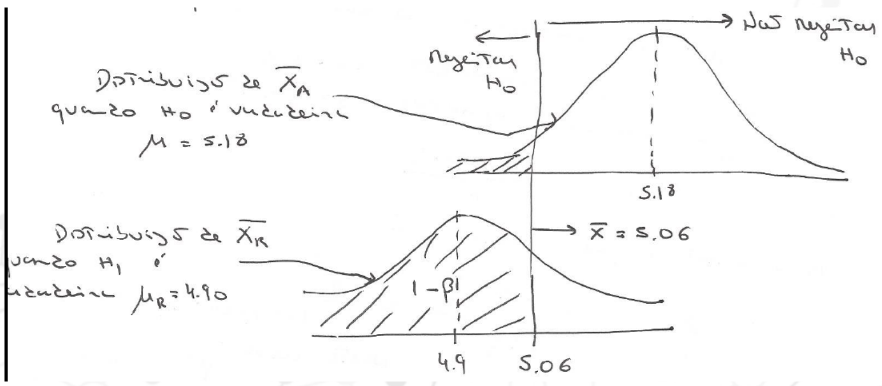
Valor de teste $Z_{obs} = -3.18 \notin RA \Rightarrow$ rejeita-se H_0. Conclui-se que os balões produzidos atualmente são menos resistentes do que o habitual.

6.2 X_A - resistência dos balões produzidos anteriormente e $\mu_A = 5.18$
X_R - resistência dos balões produzidos recentemente e $\mu_R = 4.90$
Admite-se que a variância não se altera $\sigma_A^2 = \sigma_R^2 = \sigma^2 = 0.0625$
Erro de tipo II - não rejeitar H_0 quando H_1 é verdadeira. Probabilidade de erro de tipo II = β.
A potência do teste = probabilidade de rejeitar H_0 quando H_0 é falsa =$1-\beta$.
O valor critico de $-z_{0.95} = -1.645$ (obtido em 6.1), separa a região critica (RC) da região de aceitação (RA) de H_0.
O valor critico na distribuição de $\bar{X}$ é dado por:

$$\frac{\bar{X} - 5.18}{\frac{0.25}{\sqrt{12}}} = -1.645 \Rightarrow \bar{X} = 5.06$$

Sabendo que H_0 é falsa e que $H_1: \mu_R = 4.90$ a potência do teste corresponde a:

$$1 - \beta = P\left[\bar{X}_R \leq 5.06 | H_1\right]$$



A probabilidade de $\bar{X}_R$ ocorrer na RC de H_0 é dada por:

$$P\left[\bar{X}_R \leq 5.06\right] = P\left(Z = \frac{\bar{X}_R - 4.90}{\frac{0.25}{\sqrt{12}}} \leq \frac{5.06 - 4.90}{\frac{0.25}{\sqrt{12}}}\right)$$

sendo $Z \sim N(0,1)$.

$$\Leftrightarrow P(Z \leq 2.217) = 0.98669 = 1 - \beta$$

Note que o valor 0.98669 é lido na tabela da distribuição Normal Reduzida.
Portanto a potência do teste $(1-\beta)$ é de 98.669 %.

7. Testes de Hipóteses Não Paramétricos

7.1 Páginas 139 a 141 do livro adotado.

7.2 Páginas 144 a 147 do livro adotado.

8. Teste de Qui-quadrado baseado na Tabela de Contingência.

H_0: As variáveis são independentes
H_1: As variáveis não são independentes
O_{ij} - valores observados $i = 1, 2, 3$
E_{ij} - valores esperados $j = 1, 2, 3, 4$

$$E_{ij} = \frac{O_{i\cdot}O_{\cdot j}}{n}$$

A estatística de teste é

$$X^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{(r-1)(s-1);1-\alpha}$$

neste caso $(r-1)(s-1) = 6$ com $\alpha = 0.05$.
H_0: A qualidade da secagem e a velocidade de centrifugação são independentes
H_1: A qualidade de secagem e a velocidade de centrifugação não são independentes
Regra de decisão:
Valor critico da distribuição $\chi^2_{6;0.95} = 12.59$ (Tabela da distribuição Qui-Quadrado).
Calculo da estatística de teste, utilizando a tabela de valores esperados:

		Qualidade da secagem				
		Medíocre	Suficiente	Boa	Muito Boa	Total
Velocidade	600rmp	7.67	9	7.33	6	30
de rotação	900rpm	7.67	0	7.33	6	30
	1200rpm	7.67	9	7.33	6	30
	Total	23	27	22	18	90

$$X^2_{obs} = \frac{(12-7.67)^2}{7.67} + \frac{(8-9)^2}{9} + \frac{(7-7.33)^2}{7.33} + \frac{(3-6)^2}{6} + \ldots + \frac{(11-6)^2}{6} = 13.516$$

Decisão: $X^2_{obs} = 13.516 > 12.59$ (valor crítico) $\Rightarrow$ rejeita-se H_0. Fica estabelecido (ao nível de significância de 5 %) que as variáveis "'qualidade de secagem"' e "'velocidade de rotação"' são independentes.

9. A amostra é aleatória e tem dimensão suficiente para efectuar o teste Qui-quadrado.

Estimativa do parâmetro λ da distribuição de Poisson:

$$\hat{\lambda} = \frac{(21 \times 0 + 36 \times 1 + 23 \times 2 + 13 \times 3 + 4 \times 4 + 2 \times 5 + 1 \times 6)}{100} = 1.5$$

Hipóteses a testar:
H_0: a distribuição do número de ausências por voo é de Poisson(1.5)
H_1: a distribuição do número de ausências por voo não é de Poisson(1.5)
Seja: O_i: frequência observada na classe i
E_i: a frequência esperada na classe i
A estatística de teste é:

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Que quando H_0 é verdadeira $X^2 \sim \chi^2_{(k-1)-R}$ onde k- número de classes e R- número de parâmetros da distribuição populacional estimados a partir da amostra

Nº de ausências	Freq. Observada O_i	Freq. Esperada $E_i \mid H_0(\lambda = 1.5)$	$\frac{(O_i - E_i)^2}{E_i}$
0	21	23.3	0.076
1	36	33.5	0.187
2	23	25.1	0.1756
3	13	12.6	0.013
4	4	4.7	
$5 \;\}4/5/6$	$2 \;\}7$	$1.4 \;\}6.5$	0.038
6	1	0.4	

$$\sum O_i = 100 \quad \sum E_i = 100 \quad X^2_{obs} = 0.490$$

A seguinte regra prática deve ser aplicada para utilizar este teste com confiança:
• Dimensão da amostra não inferior a 30
• Frequência esperada em cada classe não inferior a 5 ($E_i \geq 5$)
Dado que a frequência esperada em algumas classes ($i = 4, 5, 6$) é inferior a 5 ($E_i < 5$), é conveninente agregar classes adjacentes, de forma a obter uma nova categoria que satisfaça a condição.
E portanto temos agora:
$k = 5; R = 1 \Rightarrow X^2 \sim \chi^2_3$
Para $\alpha = 0.05$ o valor critico na tabela de χ^2 com 3 graus de liberdade é $\chi^2_{3;0.95} = 7.81$.
Decisão: $X^2_{obs} = 0.49 < 7.81$ (valor crítico) $\Rightarrow$ não rejeitar H_0. Para um nível de significância de 5 % , é plausível que a distribuição em causa seja de Poisson ($\lambda = 1.5$)

10. .

$> canal1 <- c(35, 47, 51, 42, 25, 43, 59, 72, 49, 61)$
$> canal2 <- c(32, 49, 54, 39, 39, 44, 65, 43, 65, 55)$
$> t.test(canal1, canal2, var.equal = T)$

Two Sample t-test

data: canal1 and canal2
t = -0.0181, df = 18, p-value = 0.9858
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-11.72363 11.52363
sample estimates:
mean of x mean of y
48.4 48.5

Conclusão: Não rejeitamos a hipótese de igualdade entre médias ao nível de significância de 5% , $p - value = 0.9858 > 0.05$.

11. .

$> X <- c(995, 1015, 985, 1004, 907, 1002, 976)$
$> t.test(X, mu = 990)$

One Sample t-test

data: X
t = -0.4823, df = 6, p-value = 0.6467
alternative hypothesis: true mean is not equal to 990
95 percent confidence interval:
950.0877 1016.7695
sample estimates:
mean of x
983.4286

Podemos concluir que sendo $p - value = 0.6467 > 0.05$, então com um risco de 5%, não devemos rejeitar a hipótese de que a média do consumo diário de biodiesel é 990 litros.

12. .

$> grupo1$
$[1]$ 94.8 89.4 96.7 93.5 96.4 85.3 87.4 89.5 85.7
$> grupo2$
$[1]$ 83.4 85.6 86.6 84.4 83.4 86.7 83.4 82.7 84.8 83.8 84.9 84.4 85.3 84.8 82.3
$[16]$ 81.5 84.3

12.1 .

$> var.test(grupo1, grupo2, conf.level = 0.99)$

F test to compare two variances

data: grupo1 and grupo2
F = 10.1496, num df = 8, denom df = 16, p-value = 0.0001108
alternative hypothesis: true ratio of variances is not equal to 1
99 percent confidence interval:
2.245156 68.644639
sample estimates:
ratio of variances
10.14959

Conclusão: Verificando-se que o valor de prova $p = 0.000108 < 0.01$, ao nível de significância de 1%, a hipótese nula deve ser rejeitada, ou seja, a variância populacional do grupo 1 não é igual à variância populacional do grupo 2.

12.2 .

$> t.test(grupo1, grupo2, var.equal = T, alternative = "greater")$

Two Sample t-test

data: grupo1 and grupo2
t = 5.753, df = 24, p-value = 3.137e-06
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:

4.717117 Inf
sample estimates:
mean of x mean of y
90.96667 84.25294

Conclusão: Rejeitamos a hipótese de igualdade de médias ao nível de significância de 5%. Não rejeitamos a hipótese alternativa onde se afirma que a média do grupo 1 é menor (produz piores resultados) do que a média do grupo 2.

UAb

ATIVIDADE FORMATIVA 3 - Proposta de Resolução

1. Não enviesamento:

$$E[\hat{\lambda}] = E\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] = \frac{1}{n}E\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}n\lambda = \lambda$$

$$E[\tilde{\lambda}] = E\left[\frac{X_1 + X_n}{2}\right] = \frac{1}{2}E[X_1 + X_n] = \frac{1}{2}(\lambda + \lambda) = \lambda$$

Como $E\left[\hat{\lambda}\right] = E\left[\tilde{\lambda}\right] = \lambda$ ambos os estimadores são não enviesados ou centrados para λ.
Eficiência:

$$VAR[\hat{\lambda}] = VAR\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] = \frac{1}{n^2}VAR\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}n\lambda = \lambda/n$$

$$VAR[\tilde{\lambda}] = VAR\left[\frac{X_1 + X_n}{2}\right] = \frac{1}{4}VAR[X_1 + X_n] = \frac{1}{4}(\lambda + \lambda) = \lambda/2$$

$\frac{VAR[\hat{\lambda}]}{VAR[\tilde{\lambda}]} = 2/n$, conclui-se que se $n = 2$ os estimadores são igualmente eficientes, se $n > 2$ então $\hat{\lambda}$ é mais eficiente do que $\tilde{\lambda}$.
NOTA: Se $X \sim Poisson(x; \lambda)$ então $E\,[X] = \lambda$ e $VAR\,[X] = \lambda$.

2. X - v.a. "atrasos de comboio"$X \sim U(0, \theta)$

$$f(x) = \begin{cases} 1/\theta & 0 \leq x \leq \theta \\ 0 & x < 0 \;\; ou \;\; x > \theta \end{cases}$$

As oito observações

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
2.23	7.74	1.03	0.08	11.14	12.72	0.42	5.17

A função de verosimilhança da amostra é:

$$L\left(x_1, x_2, \ldots, x_8|\theta\right) = \prod_{i=1}^{8} \frac{1}{\theta} = \frac{1}{\theta^8}$$

e

$$L(\theta) = \begin{cases} \left(\frac{1}{\theta}\right)^8 & se \;\; \theta \geq max(x_i) \\ 0 & se \;\; \theta < max(x_i) \end{cases}$$

Note que há restrições em θ, uma vez que $\theta \geq x_i$ para todo o $i = 1, \ldots, 8$. Assim

$$\theta \geq max(x_1, x_2, \ldots, x_8) \;\; e \;\; ln(L) = -8ln(\theta)$$

Derivando

$$\frac{dln(L)}{d\theta} = -\frac{8}{\theta} < 0 \;\; para \;\; \theta \geq max(x_1, x_2, \ldots, x_8)$$

Não há qualquer ponto estacionário, ou seja, maximizar L é equivalente a minimizar θ, satisfazendo a condição $\theta \geq max(x_i)$.
Neste caso a $\hat{\theta}_{ML} = max\,(x_1, \ldots . x_8) = 12.72$ minutos.

3. X - v.a. tempo necessário para a realização da operação de montagem.
$N = 25$, $s^2 = 0.3^2 = 0.09horas^2$ e $X \sim N(\mu, \sigma^2)$.
A variável fulcral é:

$$(N-1)\frac{s^2}{\sigma^2} \sim \chi^2_{N-1}$$

O intervalo de confiança (IC) para a variância σ^2 de uma população normal com um grau de confiança $1 - \alpha$ é dado por:

$$\left[\frac{(n-1)s^2}{\chi^2_{N-1;\alpha/2}}; \frac{(n-1)s^2}{\chi^2_{N-1;1-\alpha/2}}\right]$$

A partir da tabela da distribuição χ^2 temos:

$1-\alpha$	$\chi^2_{24;1-\alpha/2}$	$\chi^2_{24;\alpha/2}$
0.9	13.85	36.4
0.95	12.40	39.4
0.99	9.89	45.6

Neste caso temos $N = 25$ e $s^2 = 0.09$
•$IC(90\%) = [0.059; 0.156]$
•$IC(95\%) = [0.055; 0.174]$
•$IC(99\%) = [0.047; 0.218]$
Note-se que os intervalos não são centrados em $s^2 = \hat{\sigma}^2 = 0.09$, porque a distribuição χ^2 não é simétrica.

4. p- proporção de eficácia do chá; $n = 250$, $\hat{p} = 0,76$

4.1 IC:

$$\left[\hat{p} - z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}; \hat{p} + z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right]$$

$\alpha = 0.05$ e $z_{1-\alpha/2} = z_{0.975} = 1.96$ (tabela da distribuição normal reduzida).
IC: $[0.707; 0.813]$

4.2

$$z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} = \frac{h}{2}$$

$\hat{p} = 0.76 \Leftrightarrow 1 - \hat{p} = 0.24$
Para 99% $\Rightarrow$ $\alpha = 0.01 \Rightarrow z_{1-\alpha/2} = z_{0.995} = 2.58$ (tabela da distribuição normal reduzida).

$$z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} = \frac{h}{2}$$

$$\Leftrightarrow n = \frac{\left(z_{1-\alpha/2}\right)^2 p(1-p)}{(h/2)^2}$$

$$\Leftrightarrow n = \frac{(2.58)^2 \times 0.76 \times 0.24}{(0.05)^2} = 486$$

A amostra deveria ser de 486 pessoas.

5. X_A - saldo de uma conta à ordem da agência A em euros. X_B - saldo de uma conta à ordem da agência B em euros.
$N_A = 17 \;\; \bar{x}_A = 48.2 \;\; s_A = 2.74$
$N_B = 13 \;\; \bar{x}_B = 41.5 \;\; s_B = 3.91$
Admite-se que X_A e X_B seguem distribuições normais independentes, com variâncias não necessariamente iguais:

$$\bar{X}_A - \bar{X}_B \sim N\left(\mu_A - \mu_B; \frac{\sigma_A^2}{N_a} + \frac{\sigma_B^2}{N_B}\right)$$

$$Z = \frac{\left(\bar{X}_A - \bar{X}_B\right) - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{N_A} + \frac{\sigma_B^2}{N_B}}} \sim N(0,1)$$

como σ_A e σ_B são desconhecidos e não são válidas as aproximações $s_A \cong \sigma_A$ e $s_B \cong \sigma_B$ (porque as amostras são de pequena dimensão)

$$\frac{\left(\bar{X}_A - \bar{X}_B\right) - (\mu_A - \mu_B)}{\sqrt{\frac{s_A^2}{N_A} + \frac{s_B^2}{N_B}}} \sim t_v$$

onde v (número de graus de liberdade) é dado por:

$$v = \frac{\left(\frac{s_A^2}{N_A} + \frac{s_B^2}{N_B}\right)^2}{\sqrt{\frac{(s_A^2/N_A)^2}{N_A-1} + \frac{(s_B^2/N_B)^2}{N_B-1}}} = \frac{2.617}{\sqrt{0.0122 + 0.1151}} = 7.335 \cong 7$$

É recomendado adoptar o número inteiro imediatamente inferior.
O intervalo de confiança para $\mu_A - \mu_B$ a 90% $(\alpha = 0.1)$ é:

$$\left[\left(\bar{X}_A - \bar{X}_B\right) - t_{7;1-\alpha/2}\sqrt{\frac{s_A^2}{N_A} + \frac{s_B^2}{N_B}}; \left(\bar{X}_A - \bar{X}_B\right) + t_{7;1-\alpha/2}\sqrt{\frac{s_A^2}{N_A} + \frac{s_B^2}{N_B}}\right]$$

$$= [6.7 - 1.895 \times 1.272; 6.7 + 1.895 \times 1.272] = [4.29; 9.11]$$

em que $t_{7;0.95} = 1.895$ (tabela da distribuição t-student).
A 95% $(\alpha = 0.05)$ temos $t_{1-\alpha/2} = t_{7;0.975} = 2.365$ e o IC é $[3.69; 9.71]$.

6. Teste para a média (população normal com variância desconhecida).
X - peso em gramas dos instrumentos, $X \sim N(\mu, \sigma^2)$.
Amostra: $n = 15 \quad \sum x_i = 1344g \quad \sum (x_i - \bar{x})^2 = 3150g^2$.
Teste de hipóteses:

$$H_o : \mu = 100 \;\; (\mu = \mu_0)$$

$$H_1 : \mu \neq\neq 100 \;\; (\mu \neq \mu_0)$$

com $\alpha = 0.01$.
Estatística de teste:

$$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$$

Região crítica (RC):

$$t_{n-1;1-\alpha/2} = t_{14;0.995} = 2.977$$

$$RC: \,]-\infty; -2.977\,[\cup]\,2.977; +\infty\,[$$

Valor de teste:

$$T_{obs} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{89.6 - 100}{15/\sqrt{15}} = -2.685$$

onde $\bar{x} = \sum x_i/n = 1344/15 = 89.6$ e $s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{3150}{14} = 225 \Rightarrow s = 15$.
Decisão:
$T_{obs} = -2,685 > -2,977 \notin RC \Rightarrow$ Não rejeitar H_0. Não existe razão suficientemente forte para duvidar da garantia do fabricante.

7. Teste para a diferença de proporções.
p_1 - proporção de conhecedores do produto antes da campanha
p_2 - proporção de conhecedores do produto após a campanha
Amostra 1: $n_1 = 200 \quad \bar{x}_1 = 60/200 = 0.3$
Amostra 2: $n_2 = 300 \quad \bar{x}_2 = 111/300 = 0.37$
Teste de hipóteses:

$$H_0 : p_2 - p_1 = 0 \quad (p_2 = p_1)$$

$$H_1 : p_2 - p_1 > 0 \quad (p_2 > p_1)$$

Estamos a tentar perceber se o produto se tornou mais conhecido após a campanha $(p_2 > p_1)$, com $\alpha = 0.01$.
Estatística de teste:

$$Z = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

sob H_0 pode escrever-se

$$Z = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$$

Quando H_0 é verdadeira e n_1 e n_2 são grandes. É preciso estimar p utilizando um estimador ponderado:

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{60 + 111}{200 + 300} = 0.342$$

RC:

$$-z_{1-\alpha} = -z_{0.99} = -2.326$$

e o teste é bilateral à esquerda:

$$RC: \,]-\infty; -2.326\,[$$

Valor de teste:

$$T_{obs} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.3 - 0.37}{\sqrt{0.342 \times 0.658\left(\frac{1}{200}+\frac{1}{300}\right)}} = -1.62$$

Decisão:
$T_{obs} = -1.62 > -2.326 \notin RC \Rightarrow$ Não rejeitar H_0. A diferença de 7 pontos encontrada amostralmente não é significativamente diferente de zero $\Rightarrow$ a campanha publicitária não deve ter produzido efeito.

8. .

8.1 Teste para a variância.
X - idades dos telespetadores supondo $X \sim N(\mu, \sigma^2)$
$n = 100 \quad s = 2.95 \Leftrightarrow s^2 = 8.70 \quad \alpha = 0.05$
Testar as hipoteses:

$$H_0 : \sigma^2 = 6 \quad versus \quad H_1 : \sigma^2 > 6$$

Estatística de teste:

$$X^2 = \frac{ns^2}{\sigma^2} \sim \chi^2_{n-1}$$

RC (utilizando a tabela da distribuição Qui-Quadrado):

$$\chi^2_{n-1;1-\alpha} = \chi^2_{99;0.95} = 123.2$$

$$RC :]\,123.2; +\infty\,[$$

Valor de teste:

$$X^2_{obs} = \frac{ns^2}{\sigma_0^2} = \frac{100 \times 8.70}{6} = 145$$

Decisão:
$X^2_{obs} = 145 > 123.2 \in RC \Rightarrow$ Rejeitar H_0. A população pode ser considerada heterogénea.

8.2 Teste para a proporção.
p - proporção de telespetadores do sexo feminino.
Testar hipoteses:

$$H_0 : p = 0.5 \quad versus \quad H_1 : p > 0.5$$

com $\alpha = 0.01$, $n = 200$ e $\hat{p} = 100/200 = 0.55$.
Estatística de teste:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

RC (utilizando a tabela da distribuição normal reduzida):

$$z_{1-\alpha} = z_{0.99} = 2.236$$

$$RC :]\,2.236; +\infty\,[$$

Valor de teste:

$$Z_{obs} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.55 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{200}}} = 1.414$$

Decisão:
$Z_{obs} = 1.414 < 2.326 \notin RC \Rightarrow$ Não rejeito H_0. Não podemos concluir que o programa seja menos cativante para as pessoas do sexo masculino.

9. Teste Qui-Quadrado para a Independência. Hipóteses:
H_0 : Não existe relação entre a qualidade dos morangos e a ”identidade”dos agricultores (as variáveis são independentes)
H_1 : Existe relação entre a qualidade dos morangos e a ”identidade”dos agricultores (as variáveis não são independentes)
O_{ij} - valores observados; E_{ij} - valores esperados com $i = 1, \ldots, r; j = 1, \ldots, s$, e :

$$E_{ij} = \frac{O_{i\bullet} \times O_{\bullet j}}{n} \quad i = 1, \ldots, r; j = 1, \ldots, s$$

neste caso $r = s = 3$
Estatística de teste;

$$X^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{(r-1)(s-1);1-\alpha}$$

com $\alpha = 0.05$, pela tabela $\chi^2_{4;0.95} = 9.49$
Tabela dos valores esperados:

	Rejeitados	Aceitáveis	Bons	Total
A	13.52	21.47	89.09	128
B	8.98	14.31	59.39	82
C	18.44	29.27	121.40	170
Total	41	65	270	n=376

Valor de teste:

$$X^2 = \frac{(12 - 13.52)^2}{13.52} + \frac{(23 - 21.47)^2}{21.47} + \cdots + \frac{(119 - 121.40)^2}{121.40} = 1.6$$

Decisão:
Valor de teste $= X^2 = 1.6 < 9.49 \Rightarrow$ Não rejeitar H_0. É possivel que as duas características sejam independentes.

10. ANOVA - Condições de aplicabilidade: independência (amostras aleatórias); normalidade; homocedasticidade.
Hipóteses:
$H_0 : \mu_1 = \mu_2 = \mu_3$ versus $H_1 : \mu_i \neq \mu_j$ para pelo menos um par (i, j)
Tabela Auxiliar:

	Alfa	Beta	Gama	
n_i	6	6	6	n=18
$x_{i\bullet}$	203	215.5	221.3	$\sum_i x_{i\bullet} = x_{\bullet\bullet}$=639.9
$\sum_j x_{ij}^2$	6868.3	7740.23	8162.41	$A = \sum_i \sum_j x_{ij}^2$=22770.94
$\frac{x_{i\bullet}^2}{n_i}$	6868.168	7740.042	8163.282	$B = \sum_i \frac{x_{i\bullet}^2}{n_i}$=22770.49
$\bar{x}_{i\bullet}$	33.833	35.917	36.883	$C = \frac{x_{\bullet\bullet}^2}{n}$=22741.336

$SQA = B - C = 29.1544$
$SQE = A - B = 0.45$
$SQT = A - C = 29.60444$
Tabela ANOVA:

O.V.	G.L.	S.Q.	Q.M.	R.V.
Entre amostras	2	29.154	14.577	485.907
Erro	15	0.45	0.03	
Total	17	29.604		

Valor Crítico: $F_{k-1;n-k;1-\alpha} = F_{2;15;0.95} = 3.68$
Decisão: $F_0 = 485.907 > 3.68 \Rightarrow$ Rejeitar $H_0 \Rightarrow$ Existem diferenças entre os tempos médios.
Teste de comparações múltiplas de Tukey para verificar entre que valores médios estão as diferenças significativas.
Hipóteses:
$H_0 : \mu_i = \mu_j (i \neq j = 1, 2, 3)$ versus $H_1 : \mu_i \neq \mu_j$
Intervalo de confiança genérico e estatística de teste:

$$\mu_i - \mu_j = |\bar{x}_{i\bullet} - \bar{x}_{j\bullet}| \pm q_{k;n-k;\alpha}\sqrt{\frac{SQE}{r(n-k)}}$$

Neste caso temos $k = 3$, $n = 18$, $r = 6$ (nº de amostra por fator) e $\sqrt{\frac{QME}{n_i(n-k)}} = \sqrt{\frac{0.03}{6}} = 0.07$.
Valor crítico $q_{k;n-k;\alpha} = q_{3;15;0.05} = 3.67$ (tabela de Tukey).
Tabela das diferenças de pares de médias em valores absolutos:

$\lvert\bar{y_i}_\bullet - \bar{y_j}_\bullet\rvert$	$\bar{y_1}_\bullet = 33.83$	$\bar{y_2}_\bullet = 35.92$	$\bar{y_3}_\bullet = 36.83$
$\bar{y_1}_\bullet = 33.83$		2.08*	3.05*
$\bar{y_2}_\bullet = 35.62$			0.97*

Todos os valores * são valores superiores a $3.67 * 0.07 = 0.26$ então $\forall_{i,j}\, |\bar{y_i}_\bullet - \bar{y_j}_\bullet| > 0.26 \Rightarrow$ rejeitar $H_0 : \mu_i = \mu_j$.
Os pesos médios são todos significativamente diferentes. A empresa Alfa, apresenta a melhor média amostral, assim parece ser razoável ser ela a escolhida.

11. ANOVA - Condições de aplicabilidade: independência (amostras aleatórias); normalidade; homocedasticidade.
Hipóteses ($\alpha = 0.05$):
$H_0 : \mu_A = \mu_B = \mu_C = \mu_D = \mu$ versus $H_1 : \mu_i \neq \mu_j$ para pelo menos um par (i, j)
Cálculos Auxiliar:
$\bar{x}_A = (1 + 0.8 + 1.9 + 1.1 + 2.7)/5 = 1.5$
$\bar{x}_B = 2.5 \quad \bar{x}_C = 1.7 \quad \bar{x}_D = 2.8$
$\sum_i \sum_j (x_{ij} - \bar{x})^2 = SQT = 16.619$ onde $\bar{x} = \sum_i \sum_j x_{ij}/n = 2.12$
$SQA = \sum_{i=1}^{k} n_i x_{i\bullet}^2 - n\bar{x}^2 = 6.159$
$SQE = SQT - SQA = 10.460$
Tabela ANOVA:

O.V.	G.L.	S.Q.	Q.M.	R.V.
Entre amostras	3	6.159	2.053	3.533
Erro	18	10.460	0.581	
Total	21	16.619		

Valor Crítico: $F_{k-1;n-k;1-\alpha} = F_{3;18;0.95} = 3.16$
Decisão: $F_0 = 3.533 > 3.16 \Rightarrow$ Rejeitar $H_0 \Rightarrow$ Existem diferenças entre cidades no que diz respeito ao agravamento dos custos.

12. .

12.1 Cálculos auxiliares:

$GL_{total} = GL_{Trat} + GL_{Res} = 34 - 30 = 4$

$QM_{Res} = \frac{SQ_{Res}}{GL_{Res}} \Rightarrow SQ_{Res} = 22 \times 30 = 660$

$SQ_{total} = SQ_{Trat} + SQ_{Res} = 246 + 660 = 906$

$QM_{Trat} = \frac{SQ_{Trat}}{GL_{Trat}} = \frac{246}{4} = 61.5$

$RV = \frac{QM_{Trat}}{QM_{Res}} = \frac{61.5}{22} = 2.795$

Tabela ANOVA:

Origem de variação	Graus de liberdade	Soma de Quadrados	Quadrados Médios	Razão de variância
Entre Tratamentos	4	246	61.5	2.795
Resíduo	30	660	22	—
Total	34	906	—	—

12.2 O número de graus de liberdade para os tratamentos é dado pelo número de tratamentos menos 1 $= k - 1 \Rightarrow k = GL_{Trat} + 1 = 5$

12.3 Número de observações $= n = GL_{total} + 1 = 35$, como temos 5 tratamentos e amostras de igual dimensão $\Leftrightarrow \frac{35}{5} = 7$ amostras por tratamento

12.4 Valor Crítico: $F_{k-1;n-k;1-\alpha} = F_{4;30;0.95} = 2.69$. Decisão: $F_0 = 2.795 > 2.69 \Rightarrow$ Rejeitar $H_0 \Rightarrow$ Existem diferenças significativas entre tratamentos.

FIM