NOTA: Todos os resultados obtidos aqui foram utilizando a linguagem de programação python com o pacote estatístico **scipy.stats**. Os gráficos foram gerados utilizando a biblioteca **matplotlib**. Todos os códigos estão disponíveis em: **http://bit.ly/me-iago**

Exercício 5

Sejam as variáveis aleatórias:
X: Número médio de vezes de brincadeiras iniciadas com macaco macho.
Y: Número médio de vezes de brincadeiras iniciadas com macaco fêmea.

Como trata-se de duas amostras de tamanho 10, que é menor que 30, é necessário realizar teste de normalidade nas amostras para verificar se estas podem ser modeladas por uma distribuição normal. As hipóteses para esse teste são:

H_0 : A amostra provém de uma população normal
H_1 : A amostra não provém de uma população normal

Pode-se utilizar o teste de Shapiro em ambas amostras a fim de verificar a normalidade. Os testes consideraram o nível de significância como $\alpha = 0.005$. Os seguintes resultados foram encontrados:

Amostra	**p-valor**
X (machos)	0.7962
Y (fêmeas)	0.6585

Portanto, considerando significância estatística de $\alpha = 0.005$, não há evidência nessas amostras de dados para se rejeitar a hipótese nula sobre a normalidade da população de de macacos machos e fêmeas.

Podemos extrair as informações de valores de médias, desvios padrões e tamanhos amostrais para ambas amostras:

Amostra	μ_A	S_A	n_A
X (machos)	3.7667	0.4734	6
Y (fêmeas)	1.8950	0.3424	6

Onde $A \in \{X, Y\}$

Como as observações de cada amostras são separados entre macacos machos e fêmeas, consideramos que as amostras são independentes. Sendo assim, é necessário calcular a variância combinada das duas amostras. Porém, precisamos saber primeiramente se podem ser consideradas iguais $(S_x \simeq S_y)$. Isso acontecerá se a proporção da divisão dos seus quadrados forem menores que 4:

$$S_x^2/S_y^2 = 1.9116 < 4 \qquad S_y^2/S_x^2 = 0.5231 < 4$$

Sendo assim, podemos considerar $S_x \simeq S_y$. Agora, podemos calcular a variância combinada:

$$S_{comb} = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{(n_x - 1) + (n_y - 1)} = 0.4131$$

Seja $\overline{D}$ a variável aleatória da diferença entre as médias dos números de brincadeiras com machos e fêmeas, de tal forma que:

$$\overline{D}_{obs} = \mu_x - \mu_y = 1.8771$$

Conforme discutido anteriormente, consideramos que os dados são uma amostra pequena. Sendo assim, $\overline{D}$ é uma aproximação por t-student com $n_1 + n_2 - 2 = 10$ graus de liberdade.

Finalmente, deseja-se testar, com significância de $\alpha = 0.005$:

$$H_0 : \mu_D = \mu_x - \mu_y = 0$$
$$H_a : \mu_D = \mu_x - \mu_y \neq 0$$

Utilizando o python, encontra-se as seguintes saídas:

$\overline{D}_{obs} = 1.8771$	# Diferença das médias de X e Y
$df = n_x + n_y - 2 = 10$	# Graus de liberdade
$S_{comb} = 0.4131$	# Variância combinada
$t = 7.8472$	# Valor de $\overline{D}_{obs}$ em t-student
$p_{val} = 1.3937e - 05$	# p-valor para o t encontrado anteriormente

- **Conclusão**:

Como o p-valor obtido é menor que o valor de significância considerada, a hipótese nula, H_0, é rejeitada. Isto é, existe evidência na amostra para acreditar que existem diferença nos resultados de brincadeiras iniciadas por macacos machos e fêmeas com coeficiente de confiança de 95%.

Exercício 6

Sejam as variáveis aleatórias:
X: Capacidade aeróbica de americanos
Y: Capacidade aeróbica de peruanos.

Como trata-se de duas amostras de tamanho menor que 30, é necessário realizar teste de normalidade nas amostras para verificar se estas podem ser modeladas por uma distribuição normal. As hipóteses para esse teste são:

H_0 : A amostra provém de uma população normal
H_1 : A amostra não provém de uma população normal

Pode-se utilizar o teste de Shapiro em ambas amostras a fim de verificar a normalidade. Os testes consideraram o nível de significância como 5%. Os seguintes resultados foram encontrados:

Amostra	**p-valor**
X (americanos)	0.4529
Y (peruanos)	0.0267

Portanto, considerando significância estatística de $\alpha = 0.005$, não há evidência nessas amostras de dados para se rejeitar a hipótese nula sobre a normalidade da população de americanos, X. Porém, a rejeita-se a hipótese nula de peruanos, Y. Para fins didáticos de resolução desse exercício, considera-se que ambas amostras provém de populações normais.

Podemos extrair as informações de valores de médias, desvios padrões e tamanhos amostrais para ambas amostras:

Amostra	μ_A	S_A	n_A
X (americanos)	38.1	4.4335	10
Y (peruanos)	46.8	5.4378	20

Onde $A \in \{X, Y\}$

Como as observações de cada amostras são separados entre macacos machos e fêmeas, consideramos que as amostras são independentes. Sendo assim, é necessário calcular a variância combinada das duas amostras. Porém, precisamos saber primeiramente se podem ser consideradas iguais $(S_x \simeq S_y)$. Isso acontecerá se a proporção da divisão dos seus quadrados forem menores que 4:

$$S_x^2/S_y^2 = 0.6647 < 4 \qquad S_y^2/S_x^2 = 1.4550 < 4$$

Sendo assim, podemos considerar $S_x \simeq S_y$.

Finalmente, deseja-se testar, com $\alpha = 0.005$ significância:

$$H_0 : \mu_x - \mu_y = 0$$
$$H_1 : \mu_x - \mu_y \neq 0$$

Como temos ambas amostras pequenas, n < 30, utiliza-se um teste t com a tabela t-student e com $n_x + n_y - 2 = 28$ graus de liberdade. O teste será bilateral.

Utilizando o python, encontra-se as seguintes saídas:

$\overline{D}_{obs} = -8.7$ # Diferença das médias observada
$df = 28$ # Graus de liberdade
$S_{comb} = 5.1343$ # S combinado
$t = -4.3751$ #valor da tabela
$p\_valor = 0.0002$

- **Conclusão**:

Como o p-valor obtido é menor que o valor de significância considerada, a hipótese nula, H_0, é rejeitada. Isto é, existe evidência na amostra para acreditar que há diferença entre as médias de capacidade aeróbica de americanos e peruanos com coeficiente de confiança de 95%.

Exercício 7

Sejam as variáveis aleatórias:
X: Número de intervalos de 1 minuto que registraram uma dada quantidade de carros passando.

Avaliamos se X pode ser modelada por um modelo de Poisson.

As classes de frequências observadas menores que 5, são agrupadas com outras classes próximas a esta para criar uma classe com frequência mais alta. Pela tabela, as observações são:

Número de carros	0	1	2	3	4	5
Frequência	8	13	10	4	2	1

Agrupando as classes, conforme mencionado, temos:

Número de mortos	0	1	2	$\geq$3
Frequência	8	13	10	7

É necessário estimar o parâmetro λ para a distribuição de Poisson. Esse valor pode ser estimado com a média amostral. Então, temos:

$$\lambda = \mu_x = 1.5263$$

Assim, temos $n_{classes} - 1 - 1 = 2$ graus de liberdade. A aproximação do parâmetro λ fez perder um grau de liberdade.

Construindo o vetor de frequências observadas e esperadas pela distribuição de Poisson com o valor $\lambda = 1.5263$.

Número de mortos	0	1	2	$\geq$3
Frequência observada (o)	8	13	10	7
Frequência esperada(e)	8.259	12.605	9.612	7.516

Finalmente, calculamos o valor qui-quadrado dado por:

$$Q^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$
$$Q^2 = 0.0709$$

Com esse valor, verificaremos o teste de hipótese abaixo, onde deseja-se testar com significância de $\alpha = 0.05$.

$$H_0 : X \sim Po(\lambda)$$
$$H_a : X \neq Po(\lambda)$$

Utilizando o python, encontra-se as seguintes saídas:

$$q_{critico} = 5.9915$$

$$p\_valor = 0.9652$$

- **Conclusão:**

Com um $p\_valor = 0.9652 > \alpha = 0.05$, não existe evidências para rejeitar a hipótese nula, H_0, Sendo assim, podemos acreditar que população proveniente da amostra X sobre a quantidade de carros passando em 1 minuto, pode ser aproximada por uma distribuição de Poisson com $\lambda = 1.5263$ com confiança de 95%.

Exercício 8

Seja a variável aleatória:
X: Frequência registrada para um dado número k de soldados mortos por coices de cavalo no período informado.

Avaliamos se X pode ser modelada por um modelo de Poisson.

As classes de frequências observadas menores que 5, são agrupadas com outras classes próximas a esta para criar uma classe com frequência mais alta. Pela tabela, as observações são:

Número de mortos	0	1	2	3	4
Frequência	109	65	22	3	1

Agrupando as classes, conforme mencionado, temos:

Número de mortos	0	1	$\geq$ 2
Frequência	109	65	26

É necessário estimar o parâmetro λ para a distribuição de Poisson. Esse valor pode ser estimado com a média amostral. Então, temos:

$$\lambda = \mu_x = 0.61$$

Assim, temos $n_{classes} - 1 - 1 = 1$ graus de liberdade. A aproximação do parâmetro λ fez perder um grau de liberdade.

Construindo o vetor de frequências observadas e esperadas pela distribuição de Poisson com o valor $\lambda = 0.61$.

Número de mortos	0	1	$\geq$2
Frequência observada (o)	8	65	26
Frequência esperada(e)	108.670	66.289	25.041

Finalmente, calculamos o valor qui-quadrado dado por:

$$Q^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$
$$Q^2 = 0.063$$

Com esse valor, verificaremos o teste de hipótese abaixo, onde deseja-se testar com significância de $\alpha = 0.05$.

$$H_0 : X \sim Po(\lambda)$$
$$H_a : X \neq Po(\lambda)$$

Utilizando o python, encontra-se as seguintes saídas:

$$q_{critico} = 3.842$$

$$p\_valor = 0.8021$$

- **Conclusão:**

Com um $p\_valor = 0.8021 > \alpha = 0.05$, não existe evidências para rejeitar a hipótese nula, H_0, Sendo assim, podemos acreditar que população proveniente da amostra X sobre a quantidade k de soldados mortos por coices de cavalo no período informado, pode ser aproximada por uma distribuição de Poisson com $\lambda = 0.61$ com confiança de 95%.

Exercício 9

Sejam as variáveis aleatórias X definidas pelo enunciado, deseja-se testar com 5% de confiança, se o modelo normal se adequa aos dados com o seguinte teste de significância:

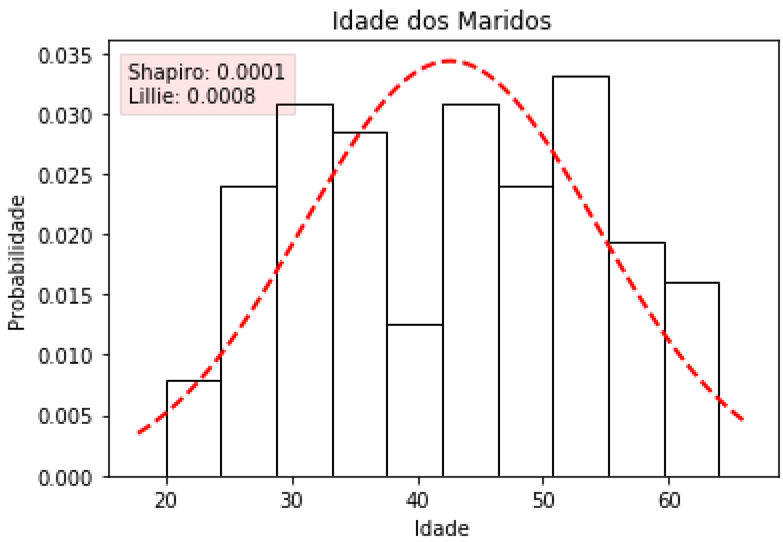
$H_0 :$ A amostra provém de uma população normal
$H_1 :$ A amostra não provém de uma população normal

Utilizaremos os testes de Shapiro-Wilk e Lilliefors comparando o p-valor com nível de significância de $\alpha = 0.005$.

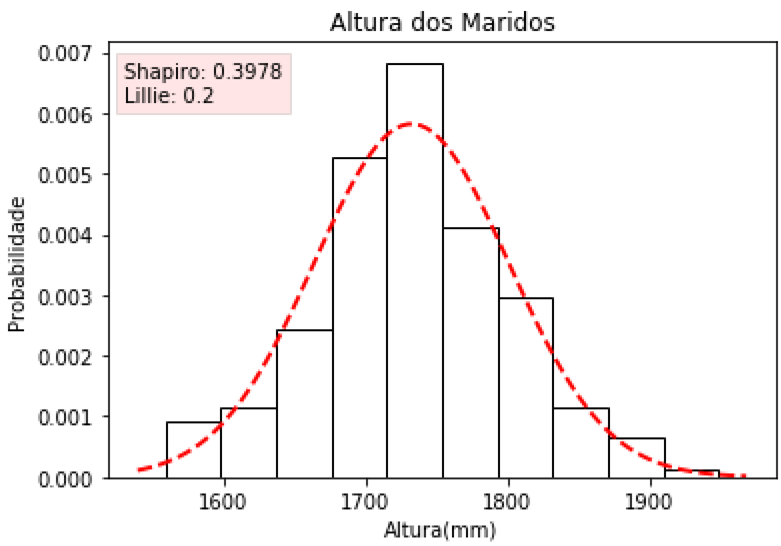
As linhas em vermelho nos resultados demonstram o ajuste de um modelo normal considerando os valores de média e desvio padrão de cada um dos dados. Podemos perceber que quando o modelo normal ajusta os dados, a curva em vermelho ajusta-se bem a densidade de probabilidades dos dados.

Lembrando que os dados que apresentavam os valores iguais a 0 (zero) foram removidos, conforme solicitado.

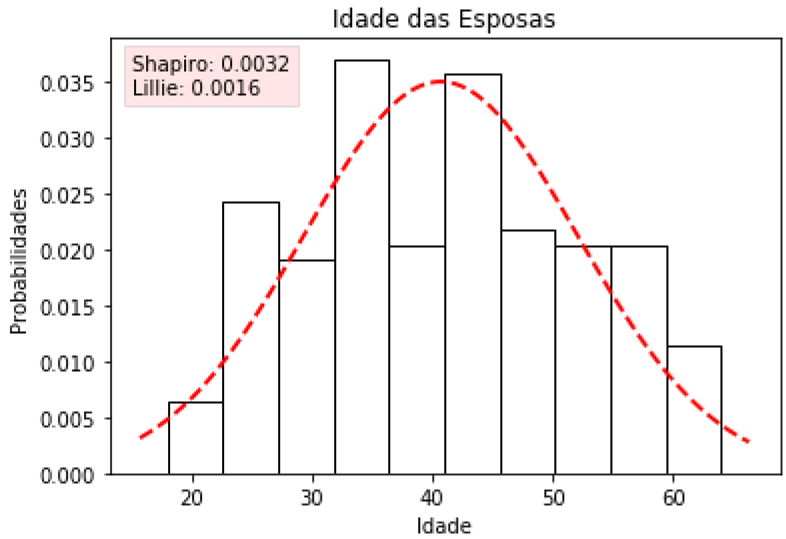
Idade dos Maridos: Ambos resultados de p-valor encontrados com o Shapiro-Wilk e Lilliefors foram menores do que o valor de α. Portanto, há evidências para rejeitar a hipótese nula e podemos inferir que a população da Idade dos Maridos não é ajustada por um modelo normal. Visualmente, percebemos que a normal em vermelho não se ajusta aos dados.



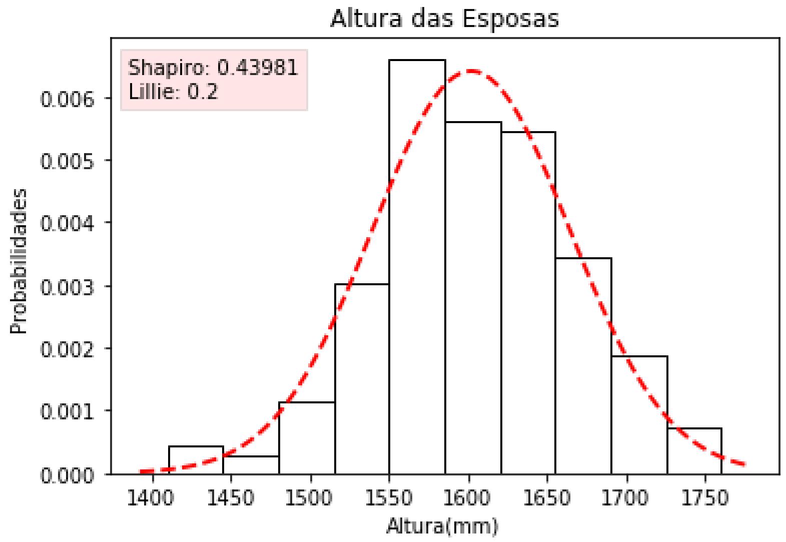
Altura dos Maridos: Ambos resultados de p-valor encontrados com o Shapiro-Wilk e Lilliefors foram maiores do que o valor de α. Portanto, não há evidências para rejeitar a hipótese nula e podemos inferir que a população da Altura dos Maridos é ajustada por um modelo normal. Visualmente, percebemos que a normal em vermelho se ajusta bem aos dados.



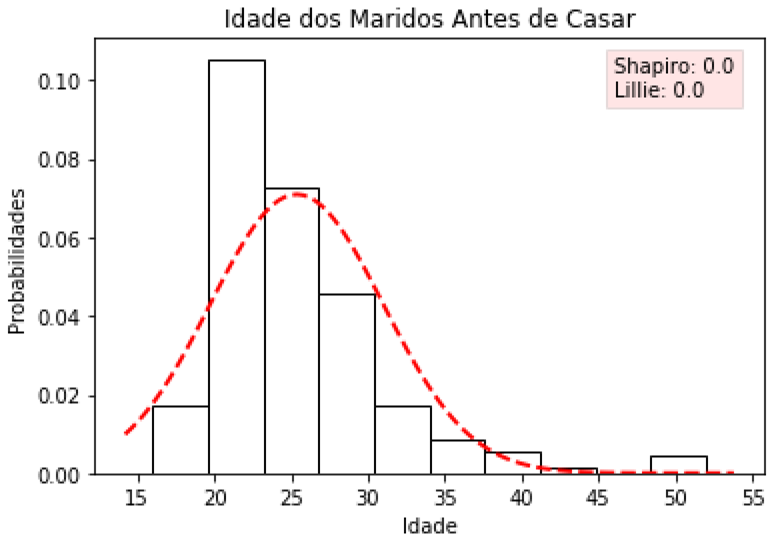
Idade das Esposas: Ambos resultados de p-valor encontrados com o Shapiro-Wilk e Lilliefors foram menores do que o valor de α. Portanto, há evidências para rejeitar a hipótese nula e podemos inferir que a população da Idade das Esposas não é ajustada por um modelo normal. Visualmente, percebemos que a normal em vermelho não se ajusta aos dados.



Altura das Esposas: Ambos resultados de p-valor encontrados com o Shapiro-Wilk e Lilliefors foram maiores do que o valor de α. Portanto, não há evidências para rejeitar a hipótese nula e podemos inferir que a população da Altura das Esposas é ajustada por um modelo normal. Visualmente, percebemos que a normal em vermelho se ajusta bem aos dados.



Idade dos Maridos Antes de Casar: Ambos resultados de p-valor encontrados com o Shapiro-Wilk e Lilliefors foram menores do que o valor de α. Portanto, há evidências para rejeitar a hipótese nula e podemos inferir que a população da Idade dos Maridos não é ajustada por um modelo normal. Visualmente, percebemos que a normal em vermelho não se ajusta aos dados.



*** Utilizou-se o teste KS, mas o resultado de p_valor obtido em todos os testes acima foram iguais a 0. Dessa forma, não apresentou-se os resultados deste teste nos gráficos.**

Exercício 10

Sejam a variável aleatória:
X: Medidas de velocidade da luz obtidas pelo equipamento de Michelson

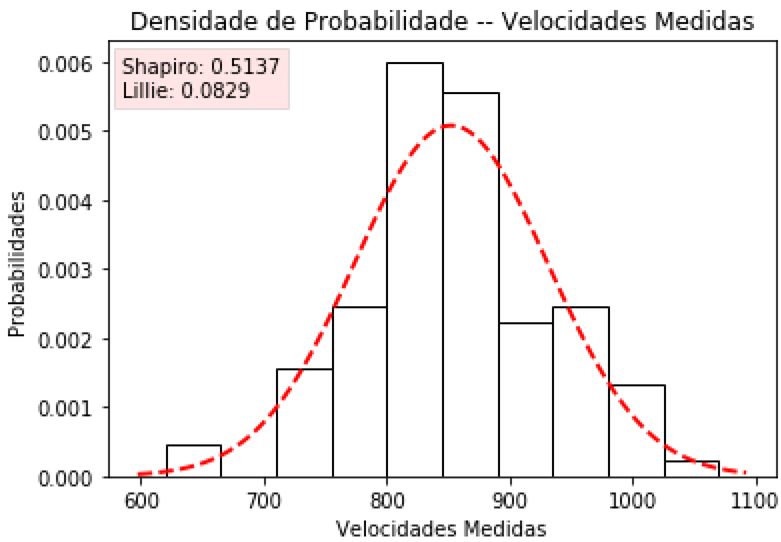
Deseja-se testar com 5% de significância, se o modelo normal se adequa aos dados com o seguinte teste de hipóteses:

H_0 : A amostra provém de uma população normal
H_1 : A amostra não provém de uma população normal

Utilizaremos os testes de Shapiro-Wilk e Lilliefors comparando o p-valor com nível de significância de $\alpha = 0.005$.

As linhas em vermelho nos resultados demonstram o ajuste de um modelo normal considerando os valores de média e desvio padrão de cada um dos dados.

Podemos perceber que quando o modelo normal ajusta os dados, a curva em vermelho ajusta-se bem a densidade de probabilidades dos dados.



Ambos resultados de p-valor encontrados com o Shapiro-Wilk e Lilliefors foram maiores do que o valor de α. Portanto, não há evidências para rejeitar a hipótese nula e podemos inferir que a população da Altura das Esposas é ajustada por um modelo normal. Visualmente, percebemos que a normal em vermelho se ajusta bem aos dados.

*** Utilizou-se o teste KS, mas o resultado de p-valor obtido neste teste foi igual a 0. Dessa forma, não apresentou-se o resultado deste teste nos gráficos.**