Estatística Inferencial para o Direito.

Autor: Felipe Navarro Romero Miguel (navarromiguel787@gmail.com (mailto:navarromiguel787@gmail.com)).

Dez. 2020.

ANOVA em um sentido

Introdução

Objetivo

Assim como o documentos anteriores, esse notebook tem por objetivo apresentar temáticas e técnicas de estatística inferencial para usos práticos no mundo do Direito. No caso, o interesse é abordar a comparação de mais de duas amostras utilizando teste paramétrico F da análise de variância (ANOVA). Pressupõem-se que se conheça os documentos anteriores ou, pelo menos, alguns temas já trabalhados sobre estatística como definição de teste paramétrico, distribuição normal, valor p, poder do teste e teste Shapiro-Wilk, teorema central do limite e tamanho de amostra.

Sobre os notebooks anteriores (mesma pasta)

No primeiro notebook (teste t para amostras pareadas) buscamos realizar um teste t de student para testar a hipótese de se existem diferença no arbitramento de honorários em câmaras de um mesmo tribunal (ou tribunais distintos). No segundo (Teste Wilcoxon para amostras pareadas), realizamos o teste Wilcoxon para amostras pareadas para avaliar se havia diferenças nas decisões entre duas instâncias no que concerne à valorização do dano moral. Em ambos os casos se trabalhou com simulações e se realizou teste de normalidade Shapiro-Wilk.

Pergunta e procedimento

Procuramos responder a pergunta: "Se as médias de preços cobrados por três empresas de certo seguimento de uma mesma cidade são significativamente diferentes". Esse tipo de questão é relevante porque pode indicar eventual cartel (consistindo em conduta contra a livre concorrência). No caso vamos utilizar dados aleatórios que seguem uma distribuição normal representando três empresas distintas (todavia, os procedimentos aqui propostos podem ser usados para números bem maiores de grupos distintos).

Resumo e explicação do Procedimento

(I) Geração de amostra de casos (simulando a obtenção de amostra aleatória): vamos criar amostras para três empresas distintas.

(II) Realização de teste de normalidade e de variância: é necessário realizar um teste de normalidade e de variância porque, caso as distribuições não sejam normais, ou tenham variância homogênea (isto é, mesma medida de dispersão), não devemos realizar os testes abaixo de comparação entre as amostras/grupos, devendo procurar por outros procedimentos (outros tipos de testes).

(III) Execução do teste F para a comparação das amostras: porque a distribuição das amostras de interesse é normal e há homogeneidade de variâncias podemos fazer um teste de hipótese paramétrico que passe desses pressupostos.

Última observação

Ao invés de vírgulas, a separação decimal se dará por ponto, dado que assim funciona a linguagem original do Python 3.

Bibliotecas

In [1]:
```
import numpy as np
import matplotlib.pyplot as plt
import random
import pandas as pd
from scipy import stats
from scipy.stats import norm
from scipy.stats import f_oneway
from scipy.stats import bartlett
from statsmodels.stats.multicomp import pairwise_tukeyhsd
from statsmodels.stats.power import FTestAnovaPower
```

Teste F

Já falamos sobre outros tipos de teste de hipótese para duas amostras pareadas, agora devemos considerar uma situação nova: a comparação de três ou mais amostras (cada uma representativa de um grupo distinto) tendo em vista verificar se elas têm médias/medianas idênticas (o que indica que elas vêm de uma mesma população). Em tese, é possível se realizar testes de hipóteses de duas em duas dessas amostras, mas essa forma de proceder, além de mais trabalhosa, pode ser inviável em se tratando de múltiplas amostras (por exemplo 10 ou 20). Nestas situações, é preferível se utilizar de um teste alternativo conhecido como teste F que faz parte da "Análise de Variância de fator único" (One-Way ANOVA).

A ANOVA consiste em uma técnica de análise de variância entre amostras que se utiliza do teste F para comparar as médias entre amostras distintas, podendo ser utilizado para situações nas quais se busca comparar mais de duas amostras distintas. Ela tem por hipótese nula (H_o) de que a média entre todas as amostras é idêntica ($\mu_1 = \mu_2 = ... = \mu_n$) e por hipótese alternativa (H_1) que pelo menos uma dessas médias é distinta das demais.

O teste F, quando empregado a análise de variância entre múltiplas amostras pressupõe: distribuição normal das populações; independência entre as amostras; e variância homogênea dentro de cada amostra. A violação desses pressupostos, pincipalmente homogeneidade de variância aumenta o risco de se cometer erro de tipo I pelo teste F conforme demonstrado em simulações (ROGAN, KESELMAN, 1977; DELACRE et. al. 2019). Por uma questão de simplicidade, vamos nos utilizar do teste F partindo de populações que sigam os seus pressupostos, até porque trata-se do teste mais "famoso" ligado a ANOVA de um fator (todavia, existem outros tipos de testes que buscam ser alternativas mais robustas ao teste F).

O teste F não é nada mais do que a divisão da *variância entre as amostras* pela *variância dentro das amostras*. Por uma questão de simplicidade não vamos refazer aqui a tabela da análise de variância, optando por ir direto para a fórmula de F - para os interessados recomendamos ver a excelente apresentação e explicação da tabela de ANOVA para mais de duas amostras em Bussab e Morettin (2017, p 449-452) -. Segundo KIM (2017, p. 24), a fórmula do teste F é:

$$F = \frac{Intergroup\ variance}{Intragroup\ variance} = \frac{\frac{\sum_{i=1}^{K} n(\bar{Y}_i - \bar{Y})^2}{K-1}}{\frac{\sum_{ij=1}^{n} (\bar{Y}_{ij} - \bar{Y})^2}{N-K}}$$

Em que "$\bar{Y}_i$ é a média do grupo i; n_i é o número de observações do grupo do grupo i; $\bar{Y}$ é a média geral; K é o número de grupos; Y_{ij} é a observação j do valor do grupo i e N é o número total de valores observados" ("Here, $\bar{Y}_i$ is the mean of the group i; n_i is the number of observations of the group i; $\bar{Y}$ is the overall mean; K is the number of groups; Y_{ij} is the jth observational value of group i; and N is the number of all observational values.") (KIM, 2017, p. 24).

Calculado F, compara-se o valor obtido com a tabela F para se tomar uma decisão quanto a hipótese nula (rejeitando-se essa quando F obtido for maior que F crítico).

Explicando F gráficamente

Sendo F a razão entre a variância entre amostras e a a variância dentro das amostras, não é difícil de se entender as consequências do teste. Isto é, quanto maior a variância intragrupo maior o valor de F encontrado, e o exato oposto se aplica a variância entre grupos.

Isso pode ser visualizado gráficamente: em cima temos um gráfico com alta variancia entre amostras e baixa variança intramostras; em baixo temos um gráfico de três distribuições normais com baixa variancia entre amostras e alta variancia intramostras. Como se nota, no primeiro gráfico, é mais fácil encontrar um valor de F estatísticamente significativo que no segundo.

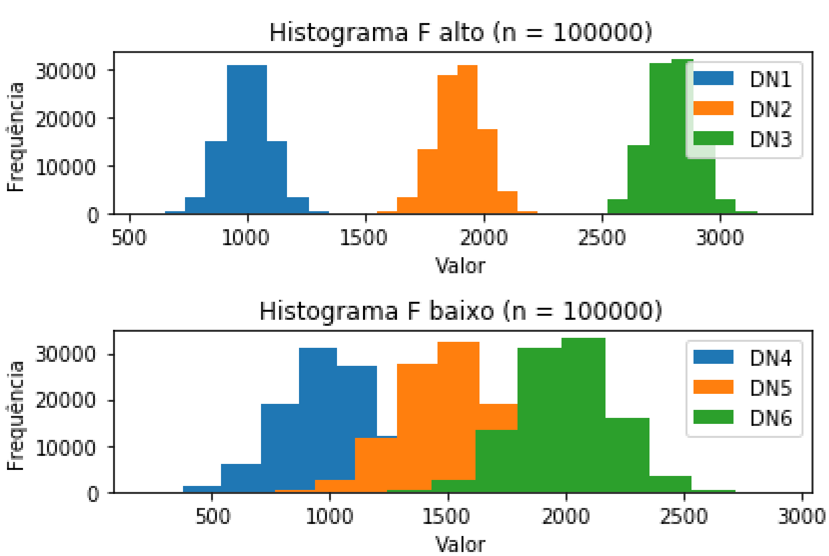
In [2]:
```python
# Criando distribuições normais de alta variancia entre amostras
DN1 = np.random.normal(loc= 1000, scale = 100,
                       size = 100000)
DN2 = np.random.normal(loc= 1900, scale = 100,
                       size = 100000)
DN3 = np.random.normal(loc= 2800, scale = 100,
                       size = 100000)

# Gráfico de alta variancia entre amostras
fig, axes = plt.subplots(2,1)
axes[0].hist(DN1)
axes[0].hist(DN2)
axes[0].hist(DN3)
axes[0].set_title('Histograma F alto (n = 100000)')
axes[0].set_xlabel('Valor')
axes[0].set_ylabel('Frequência')
axes[0].legend(("DN1", "DN2", "DN3"), loc = 1)

# Criando distribuições normais de alta variancia intramostras
DN4 = np.random.normal(loc=1000, scale = 200,
                       size = 100000)
DN5 = np.random.normal(loc=1500, scale = 200,
                       size = 100000)
DN6 = np.random.normal(loc= 2000, scale = 200,
                       size = 100000)


# Gráfico de alta variancia intramostras
axes[1].hist(DN4)
axes[1].hist(DN5)
axes[1].hist(DN6)
axes[1].set_title('Histograma F baixo (n = 100000)')
axes[1].set_xlabel('Valor')
axes[1].set_ylabel('Frequência')
axes[1].legend(("DN4", "DN5", "DN6"), loc = 1)

plt.tight_layout()
```



Teste de Tukey

Sendo F encontrado superior ao F crítico, podemos rejeitar a hipótese de igualdade entre as médias em questão. Todavia, ainda há um problema de saber quais médias são distintas entre si. Para resolver isso, realizamos um teste de comparação de cada par de médias possível, conhecido como teste de Tukey. Trata-se de teste desenvolvido por John W. Tukey para obter mais informações das diferenças entre as médias após a análise de variância (nesse sentido, trata-se de teste "post-hoc").

Por uma questão de simplicidade não vamos entrar nos detalhes do teste aqui, mas recomendamos a leitura do trabalho original de Tukey (1949) para os interessados.

Porque teste Tukey e não teste t?

É preferível se utilizar do teste de Tukey para essa comparação ao invés de vários testes t porque o primeiro diminui a chance de cometermos erros de tipo I (quanto mais testes t nós realizarmos para comparar médias, maiores as chances de cometermos erros de tipo I) (Tukey, 1949, p. 100). Ressalta-se, todavia, que o teste de Tukey requer as mesmas condições da ANOVA (LEE; LEE, 2018, p. 355).

Realizando o teste

Gerando amostras aleatórias normais

In [3]:
```
E1 = np.random.normal(loc=1000, scale = 400,
                      size = 30)
E2 = np.random.normal(loc=1000, scale = 400,
                      size = 30)
E3 = np.random.normal(loc=2000, scale = 400,
                      size = 30)
```

Testando a normalidade

In [4]:
```
print(stats.shapiro(E1).pvalue)
print(stats.shapiro(E2).pvalue)
print(stats.shapiro(E3).pvalue)
```

```
0.038139041513204575
0.4213254153728485
0.38374707102775574
```

Conforme os resultados do teste Wilk-Shapiro ($p > 0.05$ para todas as amostras), conclui-se que as amostras seguem uma distribuição normal.

Testando a homogeneidade da variancia

Para compararmos a homogeneidade das variâncias realizamos o teste de Bartlett, desenvolvido por Maurice Stevenson Bartlett (1937).

In [5]:
```
bartlett(E1, E2, E3)
```

Out[5]:
```
BartlettResult(statistic=0.4078523939439561, pvalue=0.8155225570004445)
```

Pelos resultados do teste (p > 0.05), devemos aceitar a hipótese nula de que as variancias são iguais entre as amostras.

Realizando o teste F

In [6]:
```
f_oneway(E1,E2,E3)
```

Out[6]:
```
F_onewayResult(statistic=72.96579782218028, pvalue=2.481144246547845e-19)
```

Pelos resultados do teste (considerando p < 0.05), devemos rejeitar a hipótese nula de que inexiste diferença do valor da normal entre as amostras.

Realizando o teste de Tukey

In [7]:
```
# Unificando as amostras/grupos em um dataframe (ET) para realizar o teste Tukey
ET = pd.DataFrame({'E1': (E1), 'E2': (E2), 'E3': (E3)})
ET = ET.melt()
#Realizando o teste Tukey
print(pairwise_tukeyhsd(endog = ET['value'], groups = ET['variable'], alpha = 0.0
```

```
  Multiple Comparison of Means - Tukey HSD, FWER=0.05
========================================================
group1 group2  meandiff p-adj   lower     upper   reject
--------------------------------------------------------
    E1     E2  -43.9421   0.9 -293.2203   205.336  False
    E1     E3 1071.0224 0.001  821.7443 1320.3005   True
    E2     E3 1114.9645 0.001  865.6864 1364.2427   True
--------------------------------------------------------
```

Como podemos observar pelos resultados supra, existe diferenças estatisticamente significativas entre as empresas E1 e E3, E2 e E3, mas não entre E1 e E2, o que indica que E3 possui uma média substancialmente distinta das demais. O valor real da diferença, para um nível de confiança de 95%, se encontra entre os limites inferior ("lower") e superior ("upper") conforme definido na tabela.

Últimas considerações

A violação das pressuposições do teste F pode afetar sua performance em controlar por erros de tipo I; nessas situações, pode-se buscar alternativas ao referido teste (isto é, procurar por testes potencialmente mais robustos a depender da situação). Para evitar complicar em demasia o textos, optamos por não entrar nas alternativas ao teste F, todavia, para o interessado no assunto recomendamos os texto Lix et. al. (1996). Além disso, a homogeneidade de variância também pode ser um problema na aplicação do teste Tukey, devendo se avaliar outras alternativas nessa situação (LEE; LEE, 2018, p. 355).

Referências

BUSSAB, Wilton de O.; MORETTIN, Pedro A.. *Estatística básica*. 9ª ed. São Paulo: Saraiva, 2017.

BARTLETT, M.. Properties of Sufficiency and Statistical Tests. Proceedings of the Royal Society of London. *Series A, Mathematical and Physical Sciences*, vol. 160, n. 901, 1937, P. 268-282. Disponível em: http://www.jstor.org/stable/96803 (http://www.jstor.org/stable/96803). Acesso em: 14 Dez. 2020.

DELACRE, M., LEYS, C., MORA, Y. L., LAKENS, D. Taking Parametric Assumptions Seriously: Arguments for the Use of Welch's F-test instead of the Classical F-test in One-Way ANOVA. *International Review of Social Psychology*, vol. 32, n. 1, 2019, p. 1-12. Disponível em: https://www.rips-irsp.com/article/10.5334/irsp.198/ (https://www.rips-irsp.com/article/10.5334/irsp.198/).Acesso em: 13 Dez. 2020.

ROGAN, J., KESELMAN, H. Is the ANOVA F-Test Robust to Variance Heterogeneity When Sample Sizes Are Equal?: An Investigation Via a Coefficient of Variation. *American Educational Research Journal*, vol. 14, n. 4, 1977, p. 493-498. Acesso em: 13 Dez. 2020, Disponível em: http://www.jstor.org/stable/1162346 (http://www.jstor.org/stable/1162346).

KIM, Tae Kyun. Understanding one-way ANOVA using conceptual figures. *Korean Journal of Anesthesiology*, v. 70, n. 1, p. 22-26, 2017. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5296382/pdf/kjae-70-22.pdf (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5296382/pdf/kjae-70-22.pdf). Acesso em: 09 Dez. 2020.

LEE, S., LEE, D. K.. What is the proper way to apply the multiple comparison test?. *Korean journal of anesthesiology*, vol. 71, n. 5, 2018, 353–360. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6193594/ (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6193594/). Acesso em: 14 Dez. 2020.

LIX, L., KESELMAN, J., KESELMAN, H.. Consequences of Assumption Violations Revisited: A Quantitative Review of Alternatives to the One-Way Analysis of Variance "F" Test. *Review of Educational Research*, vol. 66, n. 4, 1996, 579-619. Disponível em: http://www.jstor.org/stable/1170654 (http://www.jstor.org/stable/1170654). Acesso em: 13 Dez. 2020.

TUKEY, John W. Comparing Individual Means in the Analysis of Variance. *Biometrics*, Vol. 5, n. 2, 1949, p. 99-114. Disponível em: https://www.jstor.org/stable/3001913?origin=crossref&seq=1 (https://www.jstor.org/stable/3001913?origin=crossref&seq=1). Acesso em: 11 Dez. 2020.