Econometria 1

Pedro Henrique Rocha Mendes*

Lista 1

1) O que diz a Lei das Expectativas Iteradas?
   Lei das Expectativas Iteradas: $E(Y) = E[E(Y|X)]$, ou seja, a expectativa de Y é igual à expectativa de sua expectativa condicional. Assim, se temos $E(Y|X)$ como uma função de X e calculamos o valor esperado para a distribuição de valores de X, obtemos a expectativa de Y.[1]

2) Demonstre os resultados a seguir, partindo, para isso, das definições dos valores populacionais das estatísticas variância ($V[X]$) e covariância ($cov(X, Y)$). E considerando que $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} x_i$ e $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$:

a. $Var(X) = E(X^2) - E(X)^2$

$$\begin{aligned}
Var(X) &= E\{(X - E(X))^2\} \\
&= E[X^2 - 2X \underbrace{E(X)}_{\text{constante}} + \underbrace{E(X)^2}_{\text{constante}}] \\
&= E(X^2) - 2E(X)E(X) + E(X)^2 \\
&= E(X^2) - \cancel{2E(X)^2} + \cancel{E(X)^2} \\
&= E(X^2) - E(X)^2
\end{aligned}$$

b. $cov(X, Y) = E(XY) - E(X)E(Y)$

$$\begin{aligned}
cov(X, Y) &= E\{[X - E(X)][Y - E(Y)]\} \\
&= E[XY - X \underbrace{E(Y)}_{\text{constante}} - Y \underbrace{E(X)}_{\text{constante}} + \underbrace{E(X)E(Y)}_{\text{constante}}] \\
&= E(XY) - \cancel{2E(X)E(Y)} + \cancel{E(X)E(Y)} \\
&= E(XY) - E(X)E(Y)
\end{aligned}$$

*RA: 11201811516

[1] GUJARATI, Damodar N ; PORTER, Dawn C, Econometria Básica, 4a edição, **São Paulo, Editora Campus**, 2006.

c. $Var(X+Y) = Var(X) + Var(Y) + 2cov(X,Y)$

$$\begin{aligned} Var(X+Y) &= E\{[X+Y-E(X+Y)]^2\} \\ &= E\{[X+Y-E(X)-E(Y)]^2\} \end{aligned}$$

Para $\alpha \equiv X - E(X), \beta \equiv Y - E(Y)$:

$$\begin{aligned} &= E[(\alpha+\beta)^2] \\ &= E[(\alpha^2 + 2\alpha\beta + \beta^2)] \\ &= E(\alpha^2) + 2E(\alpha\beta) + E(\beta^2) \\ &= \underbrace{E\{[X-E(X)]^2\}}_{Var(X)} + 2\underbrace{E\{[X-E(X)][Y-E(Y)]\}}_{cov(X,Y)} + \underbrace{E\{[Y-E(Y)]^2\}}_{Var(Y)} \end{aligned}$$

d. $E[(X-\bar{X}) + (Y-\bar{Y})|X,Y] = 0$

$$\begin{aligned} E[(X-\bar{X}) + (Y-\bar{Y})|X,Y] &= E(X-\bar{X}|X,Y) + E(Y-\bar{Y}|X,Y) \\ &= E(X|X,Y) - E(\bar{X}|X,Y) + E(Y|X,Y) - E(\bar{Y}|X,Y) \\ &= E(X|X,Y) + E(Y|X,Y) \\ &- E\left(\frac{1}{N}\sum_{i=1}^{N} x_i \middle| X,Y\right) - E\left(\frac{1}{N}\sum_{i=1}^{N} y_i \middle| X,Y\right) \\ &= E(X|X,Y) + E(Y|X,Y) \\ &- \frac{1}{N}E\left(\sum_{i=1}^{N} x_i \middle| X,Y\right) - \frac{1}{N}E\left(\sum_{i=1}^{N} y_i \middle| X,Y\right) \\ &= E(X|X,Y) + E(Y|X,Y) \\ &- N^{-1}E(x_1 + \cdots + x_n|X,Y) - N^{-1}E(y_1 + \cdots + y_n|X,Y) \\ &= E(X|X,Y) + E(Y|X,Y) \\ &- N^{-1}[E(x_1|X,Y) + \cdots + E(x_n|X,Y)] \\ &- N^{-1}[E(y_1|X,Y) + \cdots + E(y_n|X,Y)] \\ &= E(X|X,Y) + E(Y|X,Y) \\ &- \cancel{N^{-1}N}E(X|X,Y) - \cancel{N^{-1}N}E(Y|X,Y) \\ &= E(X|X,Y) + E(Y|X,Y) - E(X|X,Y) - E(Y|X,Y) \\ &= 0 \end{aligned}$$

e. $\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{N}(x_i - \bar{x})\,y_i$

$$
\begin{aligned}
\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y}) &= \sum_{i=1}^{N}(x_iy_i - x_i\bar{y} - y_i\bar{x} + \bar{x}\bar{y}) \\
&= \sum_{i=1}^{N}[y_i(x_i-\bar{x}) - x_i\bar{y} + \bar{x}\bar{y}] \\
&= \sum_{i=1}^{N} y_i(x_i-\bar{x}) - \sum_{i=1}^{N} x_i\bar{y} + \sum_{i=1}^{N}\bar{x}\bar{y} \\
&= \sum_{i=1}^{N} y_i(x_i-\bar{x}) - \bar{y}\underbrace{\sum_{i=1}^{N} x_i}_{N\bar{x}} + N\bar{x}\bar{y} \\
&= \sum_{i=1}^{N} y_i(x_i-\bar{x}) \cancel{- N\bar{x}\bar{y}} \cancel{+ N\bar{x}\bar{y}} \\
&= \sum_{i=1}^{N} y_i(x_i-\bar{x})
\end{aligned}
$$

3) A variável aleatória X tem a seguinte função densidade de probabilidade:

$$
f(x) = \begin{cases} c(1-x) \text{ se } 0 < x \leq 1 \\ c(x+1) \text{ se } 1 < x \leq 2 \\ cx \text{ se } 2 < x \leq 3 \\ 0 \text{ caso contrário} \end{cases}
$$

a. Qual o valor de c?

$$
\int_{-\infty}^{0} 0dx + \int_0^1 c(1-x)dx + \int_1^2 c(x+1)dx + \int_2^3 cxdx + \int_3^{\infty} 0dx = 1
$$
$$
3c - \tfrac{5c}{2} = 1
$$
$$
c = \tfrac{2}{11}
$$

b. Qual a função distribuição cumulativa de X?

$$
F(x) = \int_{-\infty}^{x} f(t)dt
$$

- $\int_{-\infty}^{x} 0dt = 0$
- $\int_0^x \frac{2}{11}(1-t)dt = \frac{2x-x^2}{11}$
- $\int_1^x \frac{2}{11}(t+1)dt = \frac{x^2+2x-3}{11}$
- $\int_2^x \frac{2}{11}tdt = \frac{2}{11}\left(\frac{x^2}{2}-2\right)$
- $\int_3^x 0dt = 0$

$$F(x) = \begin{cases} 0, \text{ se } x < 0 \\ \frac{2x-x^2}{11}, \text{ se } 0 < x \leq 1 \\ \frac{x^2+2x-3}{11}, \text{ se } 1 < x \leq 2 \\ \frac{2}{11}\left(\frac{x^2}{2} - 2\right), \text{ se } 2 < x \leq 3 \\ 1, \text{ se } x > 3 \end{cases}$$

c. Calcule $E(X)$ e $Var(X)$.

$$\begin{aligned} E(X) &= \int_{D_x} xf(x)dx \\ &= \int_{-\infty}^{0} 0xdx + \int_{0}^{1} \frac{2}{11}x(1-x)dx + \int_{1}^{2} \frac{2}{11}x(x+1)dx + \\ &\int_{2}^{3} \frac{2}{11}x^2dx + \int_{3}^{\infty} 0xdx \\ &= 0 - \frac{1}{33} + \frac{23}{33} + \frac{38}{33} + 0 \\ &= \frac{62}{33} \end{aligned}$$

$$\begin{aligned} Var(X) &= \int_{D_x} (x - E(x))^2 f(x)dx = E(X^2) - E(X)^2 \\ E(X^2) &= \int_{-\infty}^{0} 0x^2dx + \int_{0}^{1} \frac{2}{11}x^2(1-x)dx + \int_{1}^{2} \frac{2}{11}x^2(x+1)dx + \\ &\int_{2}^{3} \frac{2}{11}x^3dx + \int_{3}^{\infty} 0x^2dx \\ &= 0 - \frac{1}{66} + \frac{73}{66} + \frac{65}{22} + 0 \\ &= \frac{269}{66} \\ Var(X) &= \frac{269}{66} - \left(\frac{62}{33}\right)^2 \\ &= \frac{1189}{2178} \approx 0{,}546 \end{aligned}$$

4) É possível dividir uma amostra estatística em subamostras de igual tamanho, selecionadas aleatoriamente. Isso foi feito com os dados da POF 2008 do IBGE, gerando-se nove subamostras, cada uma com 1.217 observações. Dentro de cada subamostra, foi calculada a média para a variável "renda per capita". Posteriormente, calculou-se a variância das médias subamostrais, obtendo-se um valor de 22,17. Explique porque este valor se mostra consideravelmente inferior ao valor da variância da amostra, i.e., trata-se da variância calculada na amostra como um todo, sem divisão em subamostras, o qual corresponde a 35.621,24. Considere, para isso, a tabela abaixo.

Tabela 1: Médias e variâncias para a renda per capita dentro das subamostras

Subamostra	Média	Variância
1	267,79	36.338,93
2	270,28	34.318,42
3	273,24	36.996,46
4	281,04	36.909,41
5	273,78	36.756,27
6	270,82	34.368,63
7	263,76	35.114,03
8	269,96	35.175,32
9	270,68	34.670,53

A média das médias subamostrais, por se basear numa medida de tendência central, perde as informações que a variância subamostral nos dá. Assim, as médias subamostrais vão se espalhar menos ao redor da média amostral. Isso implica em uma perda relevante de informação, pois a variância das médias subamostrais não está representada na variância da média da amostra, podendo levar a conclusões incorretas a respeito da distribuição da renda *per capita*.

5. Seja $X_1, X_2, \ldots, X_N$ uma sequência de variáveis aleatórias (VAs) independentes e identicamente distribuídas (i. i. d.) com média e variância populacionais dadas, respectivamente, por μ e σ^2, i. e., $E[X_i] = \mu$ e $V[X_i] = \sigma^2$, $i = 1, \ldots, N$. Responda as perguntas abaixo:

a. Verifique se a propriedade de ausência de viés na estimação da média populacional é atendida pelo estimador $\tilde{X} = \frac{1}{N-1}\sum_{i=1}^{N} X_i$.

- $B(\tilde{X}) = E(\tilde{X}) - \mu$
- estimador não enviesado $\leftrightarrow B(\tilde{X}) = 0$

$$
\begin{aligned}
E(\tilde{X}) - \mu &= 0 \\
E(\tilde{X}) &= \mu \\
E\left(\frac{1}{N-1}\sum_{i=1}^{N} X_i\right) &= \mu \\
\frac{1}{N-1}E\left(\sum_{i=1}^{N} X_i\right) &= \mu \\
\frac{1}{N-1}[E(X_1 + \cdots + X_N)] &= \mu \\
\frac{1}{N-1}[E(X_1) + \cdots + E(X_N)] &= \mu \\
\frac{1}{N-1}(\mu + ... + \mu) &= \mu \\
\frac{N}{N-1}\mu &= \mu
\end{aligned}
$$

O estimador $\tilde{X}$ é enviesado em $\frac{N}{N-1}$.

b. Obtenha a variância populacional do estimador do item anterior e verifique se tal estimador é eficiente (i.e., apresenta menor variância populacional) relativamente a um segundo estimador para a média populacional correspondente à $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$.

$$
\begin{aligned}
Var(\tilde{X}) &= Var\left(\frac{1}{N-1}\sum_{i=1}^{N} X_i\right) \\
&= \left(\frac{1}{N-1}\right)^2 Var\left(\sum_{i=1}^{N} X_i\right) \\
&= \left(\frac{1}{N-1}\right)^2 Var(X_1 + \cdots + X_N) \\
&= \left(\frac{1}{N-1}\right)^2 Var(X_1) + \cdots + Var(X_N) \\
&= \left(\frac{1}{N-1}\right)^2 \sigma^2 + \cdots + \sigma^2 \\
&= [(N-1)^{-1}]^2 N\sigma^2 \\
&= \frac{N}{(N-1)^2}\sigma^2 \ (i)
\end{aligned}
$$

$$
\begin{aligned}
Var(\bar{X}) &= Var\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) \\
&= \left(\frac{1}{N}\right)^2 Var\left(\sum_{i=1}^{N} X_i\right) \\
&= \left(\frac{1}{N}\right)^2 Var\left(X_1 + \cdots + X_N\right) \\
&= \left(\frac{1}{N}\right)^2 [Var(X_1) + \cdots + Var(X_N)] \\
&= \left(\frac{1}{N}\right)^2 (\sigma^2 + \cdots + \sigma^2) \\
&= \left(\frac{1}{N}\right)^2 N\sigma^2 \\
&= (N^{-1})^2 N\sigma^2 \\
&= \frac{\sigma^2}{N}\ (ii)
\end{aligned}
$$

Apesar de $\tilde{X}$ não ser um candidato a estimador eficiente por ser enviesado (demonstrado no item a.), ele também não apresenta a menor variância populacional, pois sua variância, encontrada em (i), é maior que a de $\bar{X}$ encontrada em (ii) sempre que N > 1. Como N necessariamente precisa ser maior que 1, $\bar{X}$ é o estimador eficiente para a média populacional.

6. O governo do Estado de São Paulo implementou um programa de qualificação para trabalhadores vítimas de desemprego tecnológico no setor rural. Um exemplo é o da introdução de máquinas colheitadeiras em substituição à colheita manual em plantios de cana-de-açúcar. Você foi contratado para determinar se os trabalhadores que passaram por este programa de qualificação tiveram sua remuneração aumentada significativamente. O indicador de impacto do programa, calculado para cada trabalhador, é a diferença de remuneração antes e depois do treinamento, sendo representado por $W_i, i = 1, \ldots, N$. Este se distribui normalmente com $W_i \sim N(\mu, \sigma^2), i = 1, \ldots, N$. É tomada uma amostra de $N = 100$ trabalhadores e obtida a estimativa pontual para o valor populacional do impacto médio, μ. O valor da estimativa pontual é de $\bar{W} = N^{-1}\sum_{i=1}^{N} W_i = 100$, o desvio padrão estimado, $s = \sqrt{N^{-1}\sum_{i=1}^{N}(W_i - \bar{W})^2} = 640$. Neste caso, o valor populacional do desvio padrão é desconhecido e, portanto, a estatística do teste é $T = \frac{\bar{W}-\mu_0}{s/\sqrt{N}} \sim t_{N-1}$, uma VA com distribuição t de Student com $N-1$ graus de liberdade. O símbolo μ_0 representa o valor da média populacional de W_i definido pela hipótese nula, zero, no caso, i. e., $\mu_0 = 0$.

a. Obtenha os valores críticos para o teste de hipóteses bicaudal. Para isso você pode utilizar a tabela da distribuição t ao final dos livros-texto ou empregar a função `qt()` do R .

- $H_0 : \mu_0 = 0$
- $H_1 : \mu_0 \neq 0$

$$T = \frac{100}{640/\sqrt{100}} \sim t_{99}$$

```
# Estatística de teste
t <- 100/(640/sqrt(100))
t
```

```
## [1] 1.5625
```

```
# Intervalo de confiança
c(qt(0.025, df = 99, lower.tail = T), qt(0.975, df = 99))
```

```
## [1] -1.984217  1.984217
```

b. Obtenha o p-valor do teste (o que pode ser feito com base nas tabelas ao final dos livros-texto ou utilizando a função ‘pt()‘ do R).

```
# pvalor bicaudal
pvalor <- t |> pt(df = 99, lower.tail = F)*2*100
pvalor
```

```
## [1] 12.13613
```

c. Qual é o resultado do teste? Explique com detalhe como, com base nos resultados dos itens anteriores e na estimativa pontual, é possível concluir acerca da existência de um impacto relevante ou não do programa de qualificação.
   Segundo o teste de hipóteses, a estatística de teste caiu no intervalo de não-rejeição da hipótese nula. Além disso, o valor-p da estatística de teste é de 12.1361336, acima dos 5% de significância. Logo, não se pode afirmar que os trabalhadores que passaram por este programa de qualificação tiveram sua remuneração aumentada.

Referências

GUJARATI, Damodar N ; PORTER, Dawn C. Econometria Básica, 4a edição. **São Paulo, Editora Campus**, 2006.