

Lab 3/Lista 2

Gabarito

Usando os dados em “gujarati.dta” e o do file para o laboratório 3 no Moodle, responda os seguintes exercícios. Para este lab, a lista deve ser entregue com o do file onde se deve explicar os procedimentos executados para responder a cada pergunta. Cada estudante deve entregar um arquivo PDF com as respostas dos exercícios e um do file na quinta-feira, 15 de setembro.

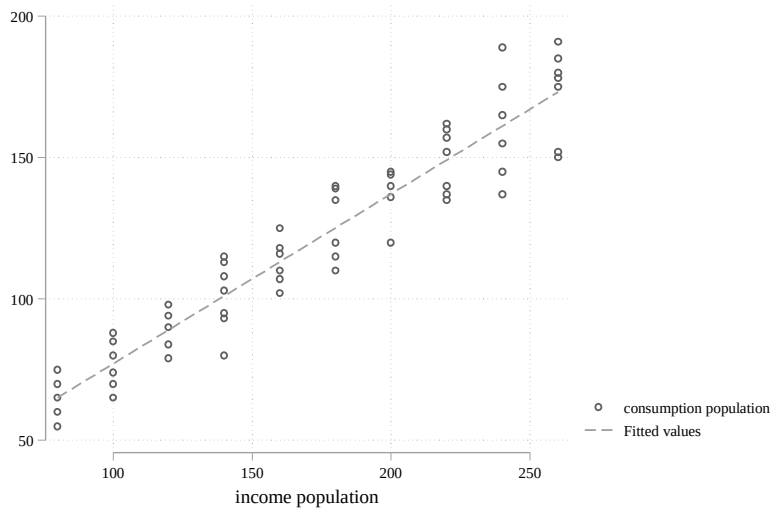
1. Faça uma análise comparando os resultados da regressão bivariada que analisa como a renda influencia o consumo na população (N=60)(Modelo 1) com os resultados do mesmo modelo em:
 - a) duas amostras aleatórias (n=10) (Modelos 2 e 3); e,
 - b) duas amostras aleatórias (n=30) (Modelos 4 e 5).

Baseado nos resultados observados, qual tipo de amostragem é preferível (n=10 ou n=30),. Justifique sua resposta.

The population regression function is:

$$CONSUMPTION = 17.10 + 0.60 \text{ INCOME} + \varepsilon$$

Note: Be sure you can correctly interpret of the slope and intercept.



a) duas amostras aleatórias (n=10) (Modelos 2 e 3); e,

Table 1.

	(1) consumption	(2) consumption	(3) consumption
income	0.600*** [0.549, 0.652]	0.543** [0.293, 0.793]	0.597*** [0.440, 0.753]
_cons	17.10*** [7.662, 26.54]	24.72 [-24.53, 73.98]	16.73 [-10.48, 43.93]
N	60	10	10

95% confidence intervals in brackets

* p<0.05, ** p<0.01, *** p<0.001

As we can see from Table 1, each random sample includes the “true” value of our population. In this sense, there is no evidence of bias. However, as compared to the population, the confidence intervals for the intercept and income coefficient are much wider. This is not surprising given the small n.

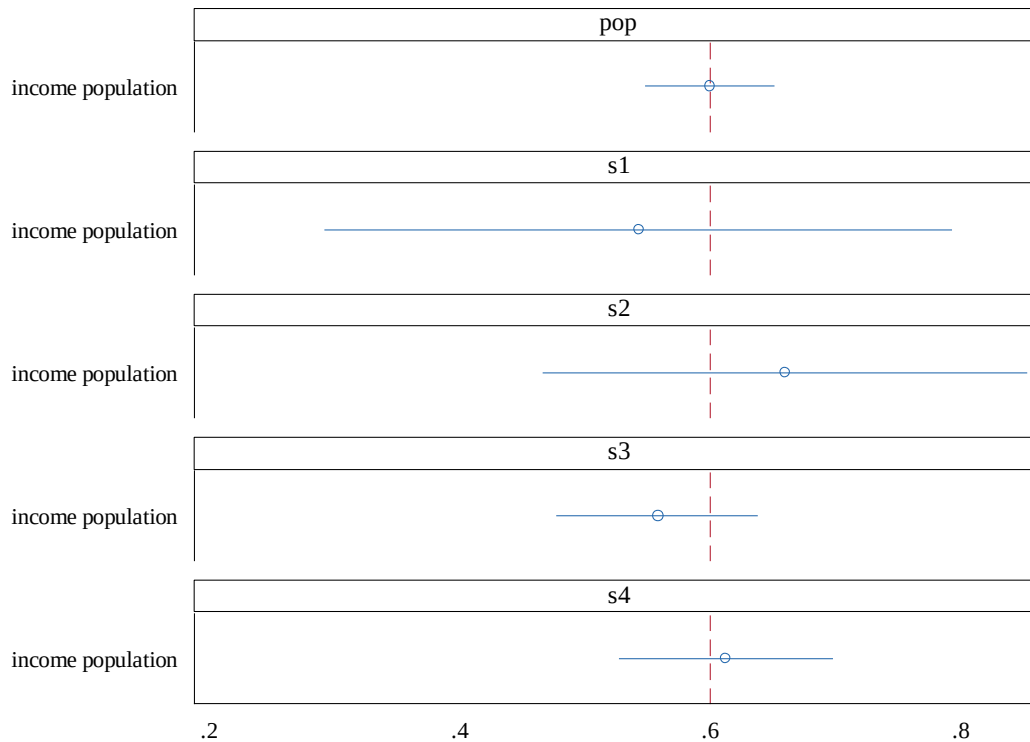
b) duas amostras aleatórias (n=30) (Modelos 4 e 5).

Table 2

	(1) consumption	(2) consumption	(3) consumption
income	0.600*** [0.549, 0.652]	0.558*** [0.478, 0.638]	0.612*** [0.519, 0.705]
_cons	17.10*** [7.662, 26.54]	22.99** [8.697, 37.29]	13.64 [-4.818, 32.10]
N	60	30	30

As we can see from Table 2, each random sample includes the “true” value of our population. As compared to Table 1, the standard error of the parameter estimates has been reduced and the confidence intervals are narrower. In this sense, a relatively larger random sample is preferred. Figure 1 summarizes this relationship succinctly as we see the plot of the 95% confidence intervals in all 5 models. The true population parameter is included in the confidence interval for each sample regression estimate, but the confidence intervals are much wider when the sample size is smaller.

Figure 1. Plot of Coefficient Estimates for Population and Random Samples



2. Faça uma análise comparando os
 - a) Os resultados reportados na pergunta 1 para uma amostra aleatória (Modelo 1, n=10)
 - b) Os resultados obtidos do mesmo modelo aplicado a duas amostras onde você estratifica a amostra por renda e logo dentro de cada strata, você aleatoriamente escolhe uma amostra de esse grupo (n=10, Modelo 6).

Baseado nos resultados observados, houve alguma consequência sobre nossas inferências sobre a população?

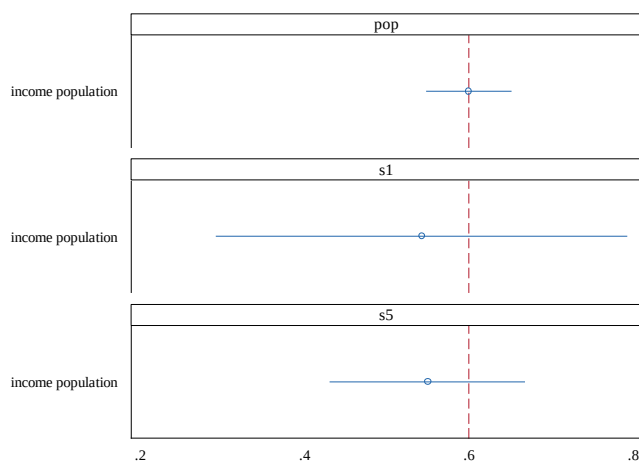
	(1) consumption	(2) consumption	(3) consumption
income	0.600*** [0.549, 0.652]	0.543** [0.293, 0.793]	0.499*** [0.345, 0.654]
_cons	17.10*** [7.662, 26.54]	24.72 [-24.53, 73.98]	32.70* [4.940, 60.47]
N	60	10	10

95% confidence intervals in brackets

* p<0.05, ** p<0.01, *** p<0.001

We still correctly estimate the effect of income on consumption, but the standard errors are still larger than the population.

Figure 2. Plot of Coefficient Estimates for Population and Random Samples



3. Agora teste se os achados dependem da distribuição de renda e consumo da população. Para implementar esta simulação, altere as características da população e refaça todos os modelos estimados na pergunta 1. Por exemplo, você poderia alterar as

características da população de modo a fazer a amostra mais parecida com a população brasileira (que tem mais pobres e menos ricos). Reporte seus achados.