



EFFECTOS DE UN DISEÑO MUESTRAL COMPLEJO SOBRE  
LAS ESTIMACIONES DE LOS MODELOS DE REGRESIÓN  
BETA

Tesis para optar al grado de magíster en estadística

Estudiante:  
Diego Leonardo Oyarzún Trujillo

Profesor Guía:  
Carlos Felipe Henríquez Roldán, PhD

Valparaíso, enero 2018

El trabajo de seminario presenta un estudio comparativo entre distintos modelos de regresión, que pueden o no incluir las características de un diseño muestral complejo y la distribución beta de la variable respuesta. El objetivo principal de estudio, es determinar si existen efectos sobre las estimaciones de un modelo de regresión cuya variable respuesta posea una distribución beta, bajo un diseño muestral complejo. Para realizar un análisis de forma empírica, se ha utilizado la base de datos desarrollada a partir del diseño muestral complejo de la encuesta de caracterización socio-económica nacional en su versión 2017 a cargo del Ministerio de Desarrollo Social de Chile. Para ello fue necesario revisar tanto su diseño muestral como el libro de códigos, con el objetivo de comprender las posibles limitaciones provenientes del diseño muestral. Para poder evaluar los efectos sobre las estimaciones, se realiza una comparación entre cuatro modelos de regresión distintos, que parte desde un análisis que se denominará ingenuo, un modelo de regresión simple sin la inclusión de factores de expansión, hasta un análisis de regresión beta que considera el uso de factores de expansión. Considerando un análisis ingenuo, los resultados del análisis indican que, existe un efecto significativo sobre los tamaños de las estimaciones de los parámetros, atribuible a la distribución beta de la variable respuesta. Y por otro lado, también se logra observar un efecto significativo sobre las estimaciones de los errores y los IC, atribuible a la inclusión de las características de un diseño muestral complejo.

# Índice general

<b>Resumen</b>	<b>II</b>
<b>Abreviaturas</b>	<b>III</b>
<b>Lista de figuras</b>	<b>V</b>
<b>Lista de tablas</b>	<b>VI</b>
<b>Introducción</b>	<b>1</b>
<b>1. Estado del arte</b>	<b>3</b>
1.1. Muestreo Complejo . . . . .	3
1.1.1. Efectos de diseño . . . . .	6
1.1.2. Factores de expansión . . . . .	9
1.1.3. Estimación de varianzas . . . . .	13
1.2. Modelo de regresión lineal bajo MC . . . . .	16
1.3. Diseño muestral CASEN 2017 . . . . .	23
1.4. Modelos de regresión con respuesta beta . . . . .	33
1.4.1. Estimación de parámetros . . . . .	35
1.4.2. Estimación de varianzas . . . . .	39
1.4.3. Diagnóstico del modelo . . . . .	40
1.4.4. Inferencia . . . . .	42
<b>2. Metodología</b>	<b>43</b>
2.1. Objetivo general . . . . .	43
2.2. Objetivos específicos . . . . .	43
2.3. Pregunta de investigación . . . . .	43
2.4. Hipótesis . . . . .	44
2.5. Metodología de trabajo . . . . .	44
2.5.1. Consideraciones previas al análisis . . . . .	44
2.5.2. Orden del proceso de análisis . . . . .	45
2.5.3. Métodos para la medición de los resultados . . . . .	47
<b>3. Análisis y resultados</b>	<b>49</b>
3.1. Descripción de las variables utilizadas para la construcción de las variables respuesta . . . . .	49
3.2. Proceso de construcción de las variables respuesta . . . . .	52

3.2.1. Construcción de la variable proporción del ingreso total del hogar que se utiliza para gastos de arriendo corregida (parrtoc). . . . .	53
3.2.2. Construcción de la variable proporción del ingreso total del hogar que se debe a subsidios corregida (psubtoc). . . . .	56
3.3. Descripción de las covariables consideradas . . . . .	61
3.3.1. Covariables categóricas . . . . .	62
3.3.2. Covariables numéricas . . . . .	64
3.4. Análisis de regresión . . . . .	67
3.4.1. Análisis en torno a la covariable categórica sexo . . . . .	67
3.4.2. Análisis en torno a la covariable categórica región . . . . .	74
3.4.3. Análisis en torno a la covariable categórica zona . . . . .	85
3.4.4. Análisis en torno a la covariable numérica metros cuadrados que posee la vivienda (v12) . . . . .	91
3.4.5. Análisis en torno a la covariable numérica cantidad de hogares que existen en la vivienda (v28) . . . . .	97
3.4.6. Análisis en torno a la covariable numérica nivel educacional (educ)	103
3.5. Resumen general del análisis . . . . .	108
<b>4. Conclusiones</b>	<b>111</b>
<b>5. Anexos</b>	<b>115</b>

# Índice de figuras

1.1. Influencia de las características de un diseño muestral complejo sobre la estimación de los errores estándar. . . . .	6
1.2. Proceso de construcción e identificación de estratos y unidades primarias de muestreo en el diseño muestral de la encuesta CASEN 2017. . . . .	26
1.3. Esquema de construcción de los FE en el diseño muestral de la encuesta CASEN 2017 . . . . .	32
1.4. Histogramas de la distribución beta con distintos valores en sus parámetros, obtenidos a partir de simulaciones en el <i>software</i> estadístico Stata. . . . .	34
1.5. Histogramas de la distribución beta con distintos valores en sus parámetros, obtenidos a partir de simulaciones en el <i>software</i> estadístico Stata. . . . .	35
3.1. Comparación entre las distribuciones de las variables proporción del ingreso total del hogar que se utiliza para gastos de arriendo ( <i>parrto</i> ) y su versión corregida ( <i>parrtoc</i> ) . . . . .	55
3.2. Comparación entre las distribuciones de las variables proporción del ingreso total del hogar que se debe a ingresos por subsidios ( <i>psubto</i> ) y su versión corregida ( <i>psubtoc</i> ) . . . . .	59
3.3. Histogramas de dos distribuciones beta con distintos valores en sus parámetros, obtenidos a partir de simulaciones en el <i>software</i> estadístico Stata. . . . .	60
3.4. Resumen de las condiciones y el proceso realizado para la elección de las variables en la base de datos CASEN 2017 que serán parte de la creación de las variables respuesta proporción de los ingresos totales que corresponden a ingresos por subsidios corregida ( <i>psubtoc</i> ) y la variable respuesta proporción de los ingresos totales que se gastan en concepto de arriendo corregida ( <i>parrtoc</i> ) . . . . .	61

# Índice de cuadros

3.1. Valores descriptivos de la variable total de ingresos del hogar corregido (ytotcorh). . . . .	50
3.2. Valores descriptivos para la variable gastos en arriendo (v19). . . . .	50
3.3. Valores descriptivos de la variable ingresos por concepto de subsidios monetarios (ysub). . . . .	51
3.4. Valores descriptivos de la variable proporción del ingreso total del hogar que se utiliza para gastos de arriendo (parrto). . . . .	53
3.5. Valores descriptivos de la variable proporción del ingreso total del hogar que se utiliza para gastos de arriendo corregida (parrtoc). . . . .	54
3.6. Comparación de medias entre las variables proporción de los ingresos totales que se gastan por concepto de arriendo (parrto) y su versión corregida (parrtoc), teniendo en cuenta el diseño muestral complejo de la encuesta CASEN 2017. . . . .	54
3.7. Valores descriptivos de la variable proporción del ingreso total del hogar que corresponde a ingresos por subsidios (psubto). . . . .	56
3.8. Valores descriptivos de la variable proporción del ingreso total del hogar que corresponde a ingresos por subsidios corregida (psubtoc). . . . .	57
3.9. Comparación de medias entre las variables, proporción de los ingresos totales que corresponden a ingresos por subsidios (psubto) y su versión corregida (psubtoc), teniendo en cuenta el diseño muestral complejo de la encuesta CASEN 2017. . . . .	58
3.10. Valores descriptivos de la covariable sexo. . . . .	62
3.11. Valores descriptivos de la covariable región. . . . .	63
3.12. Valores descriptivos de la covariable zona. . . . .	64
3.13. Valores descriptivos de la covariable metros cuadrados que posee la vivienda (v12). . . . .	65
3.14. Valores descriptivos de la covariable número de hogares existentes en la vivienda (v28). . . . .	65
3.15. Valores descriptivos de la covariable nivel educacional (educ). . . . .	66
3.16. Comparación entre los distintos modelos de regresión que consideran la variable respuesta proporción del ingreso total que corresponde a ingresos por subsidios corregida (psubtoc) y la covariable sexo. . . . .	67
3.17. Comparación entre los distintos modelos de regresión que consideran la variable respuesta proporción del ingreso total que se gasta en arriendo corregida (parrtoc) y la covariable sexo. . . . .	70

3.18. Comparación de los resultados obtenidos en los análisis de regresión para las variables respuesta proporción de ingresos totales del hogar que corresponden a subsidios corregida (psubtoc) y proporción de los ingresos totales del hogar que se gastan en arriendo corregida (parrtoc) con respecto a la covariable sexo. . . . .	74
3.19. Características principales de los modelos de regresión que consideran la variable respuesta proporción del ingreso total que corresponde a ingresos por subsidios corregida (psubtoc) y la covariable región. . . . .	74
3.20. Resultados de la aplicación del modelo de RLS sin FE para la variable respuesta proporción de los ingresos totales que corresponden a ingresos por subsidios corregida (psubtoc) y la covariable región. . . . .	75
3.21. Resultados de los modelos de regresión entre la variable proporción de los ingresos totales del hogar que corresponden a ingresos por subsidios corregida (psubtoc) y la covariable región, para la estimación del parámetro constante . . . . .	76
3.22. Resultados de los modelos de regresión entre la variable proporción de los ingresos totales del hogar que corresponden a ingresos por subsidios corregida (psubtoc) y la covariable región, para la estimación del valor único (3) . . . . .	77
3.23. Resultados de los modelos de regresión entre la variable proporción de los ingresos totales del hogar que corresponden a ingresos por subsidios corregida (psubtoc) y la covariable región, para la estimación del valor único (6) . . . . .	78
3.24. Resultados de los modelos de regresión entre la variable proporción de los ingresos totales del hogar que corresponden a ingresos por subsidios corregida (psubtoc) y la covariable región, para la estimación del valor único (12) . . . . .	79
3.25. Características principales de los modelos de regresión que consideran la variable respuesta proporción del ingreso total que se gasta en arriendo corregida (parrtoc) y la covariable región. . . . .	80
3.26. Resultados de los modelos de regresión entre la variable proporción de los ingresos totales del hogar que se gastan por concepto de arriendo corregida (parrtoc) y la covariable región, para la estimación del parámetro constante. . . . .	81
3.27. Resultados de los modelos de regresión entre la variable proporción de los ingresos totales del hogar que se gastan por concepto de arriendo corregida (parrtoc) y la covariable región, para la estimación del valor único (3). . . . .	82
3.28. Resultados de los modelos de regresión entre la variable proporción de los ingresos totales del hogar que se gastan por concepto de arriendo corregida (parrtoc) y la covariable región, para la estimación del valor único (6). . . . .	83
3.29. Resultados de los modelos de regresión entre la variable proporción de los ingresos totales del hogar que se gastan por concepto de arriendo corregida (parrtoc) y la covariable región, para la estimación del valor único (12). . . . .	84

3.30. Comparación entre los distintos modelos de regresión que consideran la variable respuesta proporción del ingreso total que corresponde a ingresos por subsidios corregida (psubtoc) y la covariable zona. . . . .	85
3.31. Comparación entre los distintos modelos de regresión que consideran la variable respuesta proporción del ingreso total que se gasta en arriendo corregida (parrtoc) y la covariable zona. . . . .	88
3.32. Comparación de los resultados obtenidos en los análisis de regresión para las variables respuesta proporción de ingresos totales del hogar que corresponden a subsidios corregida (psubtoc) y proporción de los ingresos totales del hogar que se gastan en arriendo corregida (parrtoc) y la covariable zona. . . . .	91
3.33. Comparación entre los distintos modelos de regresión que consideran la variable respuesta proporción del ingreso total que corresponde a ingresos por subsidios corregida (psubtoc) y la covariable metros cuadrados que posee la vivienda (v12). . . . .	92
3.34. Comparación entre los distintos modelos de regresión que consideran la variable respuesta proporción del ingreso total que se gasta en arriendo corregida (parrtoc) y la covariable metros cuadrados que posee la vivienda (v12). . . . .	94
3.35. Comparación de los resultados obtenidos en los análisis de regresión para las variables respuesta proporción de ingresos totales del hogar que corresponden a subsidios corregida (psubtoc) y proporción de los ingresos totales del hogar que se gastan en arriendo corregida (parrtoc) y la covariable cantidad de metros cuadrados de la vivienda (v12). . . . .	97
3.36. Comparación entre los distintos modelos de regresión que consideran la variable respuesta proporción del ingreso total que corresponde a ingresos por subsidios corregida (psubtoc) y la covariable cantidad de hogares que existen en la vivienda (v28). . . . .	98
3.37. Comparación entre los distintos modelos de regresión que consideran la variable respuesta proporción del ingreso total que se gasta en arriendo corregida (parrtoc) y la covariable cantidad de hogares que existen en la vivienda (v28). . . . .	100
3.38. Comparación de los resultados obtenidos de los análisis de regresión para las variables respuesta proporción de ingresos totales del hogar que corresponden a subsidios corregida (psubtoc) y proporción de los ingresos totales del hogar que se gastan en arriendo corregida (parrtoc) y la covariable cantidad hogares existentes en la vivienda (v28). . . . .	102
3.39. Comparación entre los distintos modelos de regresión que consideran la variable respuesta proporción del ingreso total que corresponde a ingresos por subsidios corregida (psubtoc) y la covariable nivel educacional (educ). . . . .	103
3.40. Comparación entre los distintos modelos de regresión que consideran la variable respuesta proporción del ingreso total que se gasta en arriendo corregida (parrtoc) y la covariable nivel educacional (educ). . . . .	106

3.41. Comparación de los resultados obtenidos en los análisis de regresión para las variables respuesta proporción de ingresos totales del hogar que corresponden a subsidios corregida (psubtoc) y proporción de los ingresos totales del hogar que se gastan en arriendo corregida (parrtoc) y la covariable nivel educacional (educ). . . . .	108
5.1. Resultados de la aplicación del modelo de RLS con FE para la variable respuesta proporción de los ingresos totales que corresponden a ingresos por subsidios corregida (psubtoc) y la covariable región. . . . .	116
5.2. Resultados de la aplicación del modelo de RB sin FE para la variable respuesta proporción de los ingresos totales que corresponden a ingresos por subsidios corregida (psubtoc) y la covariable región. . . . .	117
5.3. Resultados de la aplicación del modelo de RB con FE para la variable respuesta proporción de los ingresos totales que corresponden a ingresos por subsidios corregida (psubtoc) y la covariable región. . . . .	118
5.4. Resultados de la aplicación del modelo de RLS sin FE para la variable respuesta proporción de los ingresos totales que se gastan en arriendo corregida (parrtoc) y la covariable región. . . . .	119
5.5. Resultados de la aplicación del modelo de RLS con FE para la variable respuesta proporción de los ingresos totales que se gastan en arriendo corregida (parrtoc) y la covariable región. . . . .	120
5.6. Resultados de la aplicación del modelo de RB sin FE para la variable respuesta proporción de los ingresos totales que se gastan en arriendo corregida (parrtoc) y la covariable región. . . . .	121
5.7. Resultados de la aplicación del modelo de RB con FE para la variable respuesta proporción de los ingresos totales que se gastan en arriendo corregida (parrtoc) y la covariable región. . . . .	122