

# Trabajo práctico final de Biometría II: Desigualdades en salud

Grupo 6: Violeta López, Camila Vázquez Cañas, Federico Picado, Inés Beckerman

10/11/2020

## Introducción

Las enfermedades crónicas no transmisibles (ENT), como las enfermedades cardiovasculares, enfermedades respiratorias crónicas, diabetes y cáncer, son un problema de salud cada vez mayor en todo el mundo. En nuestro país representan el 80% del total de las causas de muertes y el 76% de los años de vida ajustados por discapacidad (WHO 2014). Estas enfermedades suelen compartir factores de riesgo en común, como son la alimentación inadecuada, el sedentarismo y el tabaquismo (WHO 2005).

Una enfermedad suele no tener una única causa y los determinantes que la producen son múltiples y pertenecen a distintos niveles de organización.

En los últimos años se ha enfatizado la relevancia de los factores sociales como determinantes de múltiples enfermedades, a través de múltiples factores de riesgo (Phelan et al. 2010; Marmot and Allen 2014). Existe fuerte evidencia de que la mayoría de los factores de riesgo de ENT están estructurados a lo largo del gradiente socioeconómico (Di Cesare et al. 2013).

Comprender cómo las desigualdades en salud se distribuyen espacialmente y si se reducen o no en el tiempo es un objetivo fundamental de la salud pública; monitorear su distribución y evolución es necesario para evaluar y dirigir las posibles acciones para enfrentarlas.

Con el objetivo de comprender cómo las desigualdades en salud se distribuyen espacialmente y si se reducen o no en el tiempo, la Comisión de la OMS sobre los Determinantes Sociales de la Salud realizó un llamado a la acción para el monitoreo rutinario de las desigualdades en salud (WHO 2009). Este monitoreo permite hacer un seguimiento sistemático de las desigualdades sanitarias y de los determinantes sociales de la salud, así como evaluar los efectos de las políticas e intervenciones en la equidad sanitaria.

En Argentina se implementó en 2005 la primera edición de la Encuesta Nacional de Factores de Riesgo (ENFR), que es representativa a nivel nacional y provincial, y que desde entonces se realiza con una periodicidad de 4 años. La ENFR proporciona información confiable sobre la prevalencia de factores de riesgo en ENT y su evolución en el tiempo.

Por todo lo expuesto, el objetivo es estudiar la asociación entre conductas saludables y nivel socioeconómico en mayores de 18 años, a partir de información recolectada por la ENFR 2018 (ENFR, INDEC 2018).

En particular, en este trabajo se estudiará la asociación entre el consumo de frutas/verduras y el nivel socioeconómico a escala individual, hogar y provincial.

## Metodología

La ENFR se realizó entre septiembre y noviembre de 2018, sobre 29224 personas mayores de 18 años de toda la República Argentina, residentes en viviendas particulares de áreas urbanas de 5000 y más habitantes.

El cuestionario de cada individuo incluyó preguntas sobre aspectos socioeconómicos (tanto a nivel individual como a nivel del hogar y del núcleo familiar) como aspectos sobre hábitos saludables y de factores de riesgo a nivel individual.

En particular, en este Trabajo se analizó la relación entre **el consumo de al menos 5 porciones diarias de frutas o verduras y 5 variables socioeconómicas a distintos niveles:**

- A nivel individual:

1) **Cobertura de salud.**

2 niveles: 1) Con obra social, prepaga o servicio de emergencia médica; 2) Solo cobertura pública.

2) **Nivel de instrucción.**

2 niveles: 1) Secundario incompleto o menos; 2) Secundario completo o más.

- A nivel del hogar:

1) **Vivienda deficitaria.**

2 niveles: 1) Sí; 2) No

2) **Quintil de ingresos del jefe de hogar.**

5 niveles, correspondientes a los 5 quintiles.

- A nivel de la provincia:

1) **NBI (porcentaje de viviendas con necesidades básicas insatisfechas en cada provincia).**

3 niveles, 1) Bajo; 2) Medio; 3) Alto.

Los datos de NBI fueron obtenidos a partir del Informe Censal de 2010. Allí, el NBI para cada provincia fue calculado como el porcentaje de hogares con necesidades básicas insatisfechas.

Para la agrupación entre bajo, medio y alto, se dividieron los valores de NBI provinciales en terciles, de modo que quedaran entre 7 y 9 provincias por categoría. La categoría más baja abarcó a valores de NBI de hasta 9%, la media, de entre 9% y 15%, y la alta, mayor a 15%.

Además, se consideraron **3 variables de control** para abarcar parte de la heterogeneidad en las observaciones. Estas variables son:

- 1) Edad, variable cuantitativa cuyo rango original va de 18 a 104 años.
- 2) Género, 2 niveles, 1) Hombre; 2) Mujer.
- 3) Provincia de residencia, 24 niveles, todas las provincias más la Ciudad Autónoma de Buenos Aires.

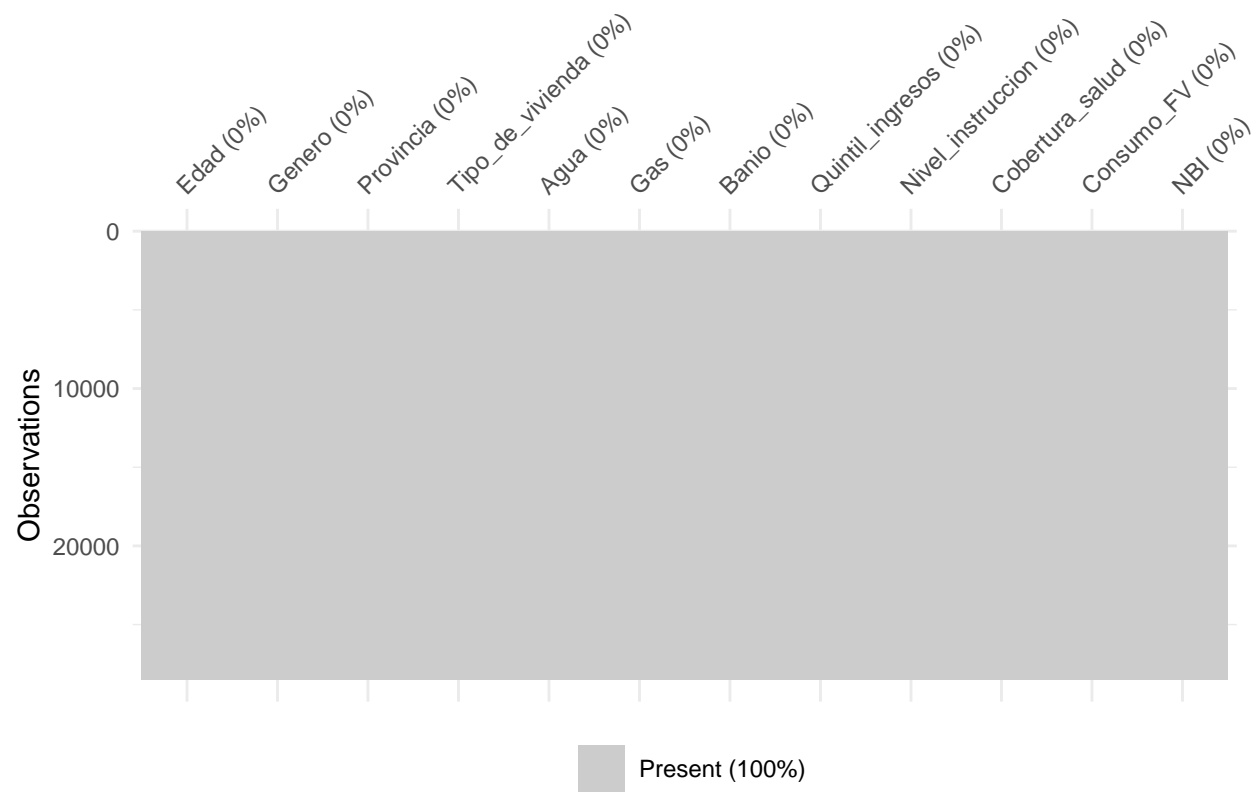
**Detalles de la elección y construcción de las variables en el ANEXO I.**

Para el análisis, como la variable respuesta es una variable dicotómica con una potencial distribución de probabilidades de Bernoulli, se ajustó a una regresión logística. Para medir magnitud de efecto de las variables explicativas (VE), se estimaron los odds ratios (OR) con sus respectivos intervalos de confianza del 95% ajustados por las otras VE. Las predicciones sobre cambios en la probabilidad en función de cada VE se calcularon a escala de la variable respuesta. En este trabajo no se puso a prueba la interacción de las variables, bajo el criterio de que se trata de un estudio observacional y que no encontramos motivos fuertes que nos permitiesen plantear una hipótesis al respecto. Todos los análisis se efectuaron utilizando Rstudio V. 3.6.3.

# Resultados

## Estadística descriptiva

## [1] "NULL"



Presencia/ausencia de datos faltantes para cada variable.

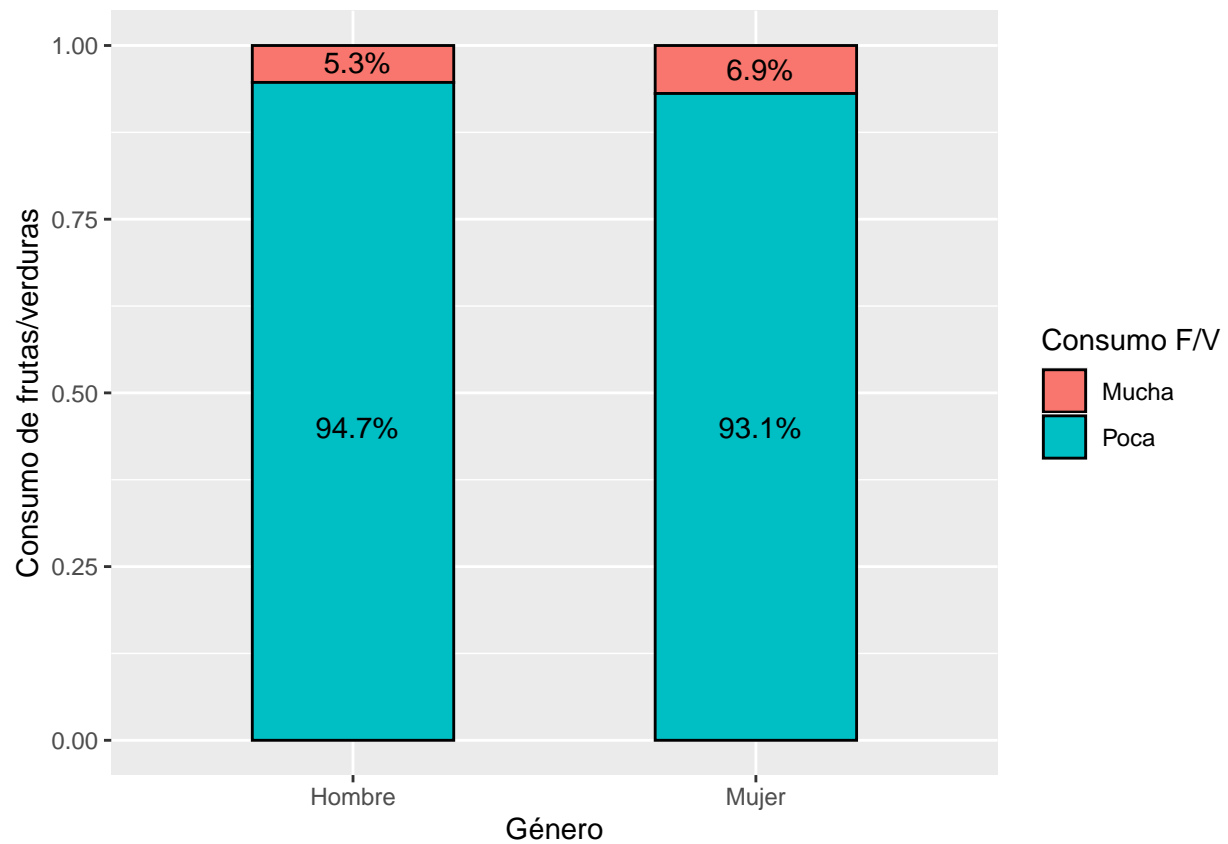
No hay datos faltantes.

Figura 1:

## Gráficos descriptivos

### *Variables de control*

#### Consumo según género



*Figura 2:* Gráfico de barras de proporción de consumo de muchas o pocas porciones diarias de frutas/verduras según el género (“hombre” o “mujer”).

Observamos que para ambos géneros, la gran mayoría de las personas consume pocas frutas o verduras diarias. Las mujeres parecen comer un poco más de frutas/verduras que los hombres.

## Consumo segun provincia

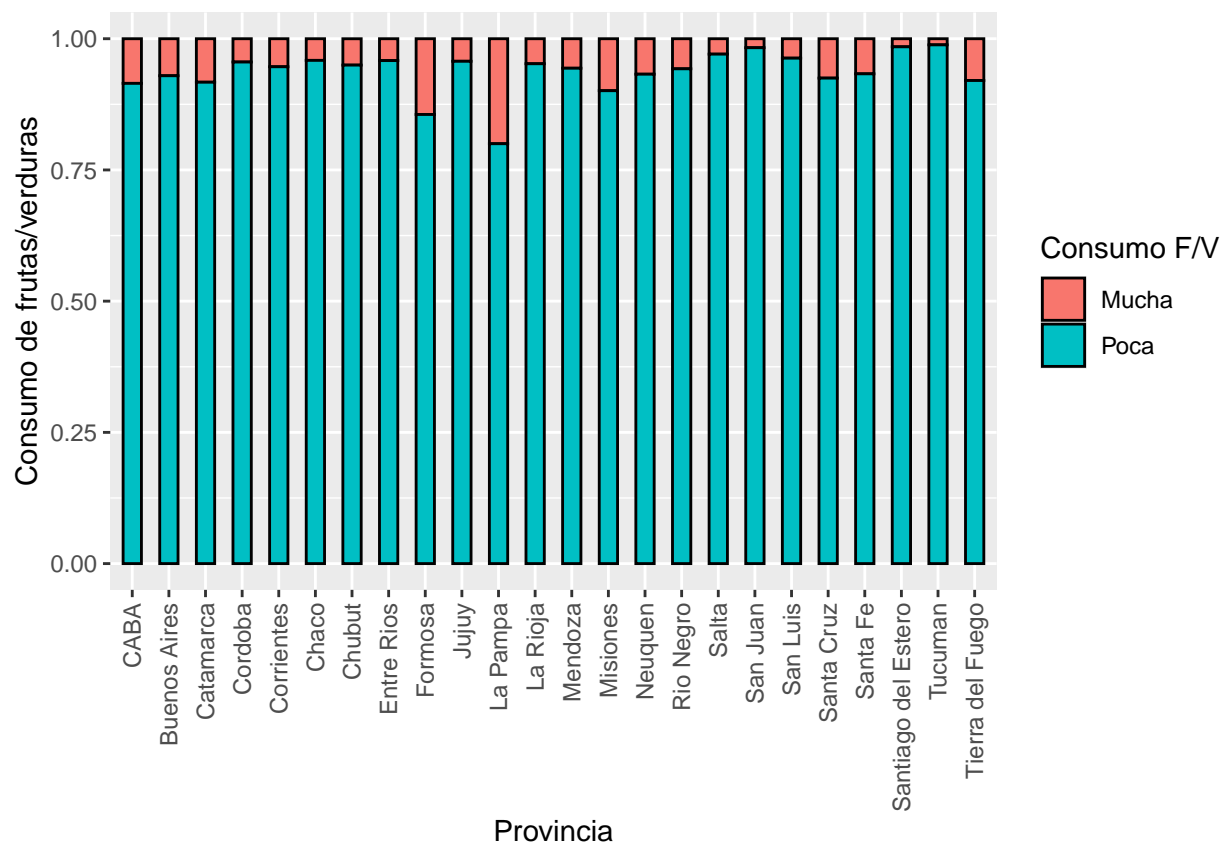
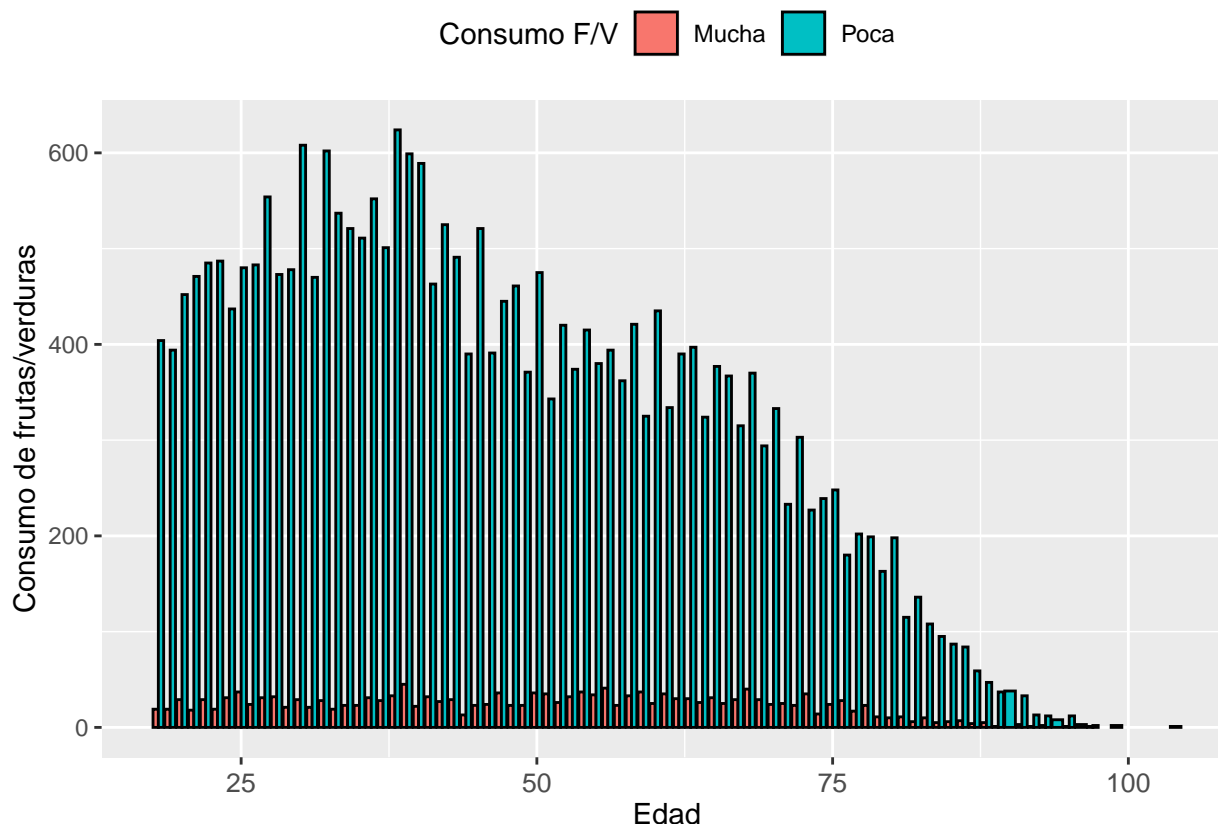


Figura 3: Gráfico de barras de proporción de consumo de muchas o pocas porciones diarias de frutas/verduras según la provincia de residencia.

En todas las provincias, entre un 80% y un 99% de las personas aproximadamente consumen menos de 5 porciones diarias de frutas o verduras. En algunas provincias, este porcentaje parece ser bastante menor que el promedio (por ejemplo, en La Pampa y Formosa) y en otras, bastante mayor (como en Santiago del Estero, Tucumán o San Juan).

## Consumo según edad



*Figura 4:* Gráfico de barras de frecuencia de consumo de muchas o pocas porciones diarias de frutas/verduras según la edad (18 a 104 años). Las barras azules representan poco consumo y las barras rojas mucho consumo.

Al analizar la altura de cada barra, se observa la distribución de edades dentro de la muestra. La distribución es asimétrica, con una moda de 38 años y una media de 46,5 años.

En rojo, observamos la cantidad de gente de cada edad que consume 5 o más porciones diarias de frutas o verduras. La proporción es baja en todas las edades. Más adelante analizaremos si hay diferencias significativas entre distintas edades.

### Tabla de medias de edad según el consumo de frutas/verduras

##	Consumo_FV	Edad.m	Edad.s
## 1	Mucha	50.06772	18.18864
## 2	Poca	46.34076	18.01619

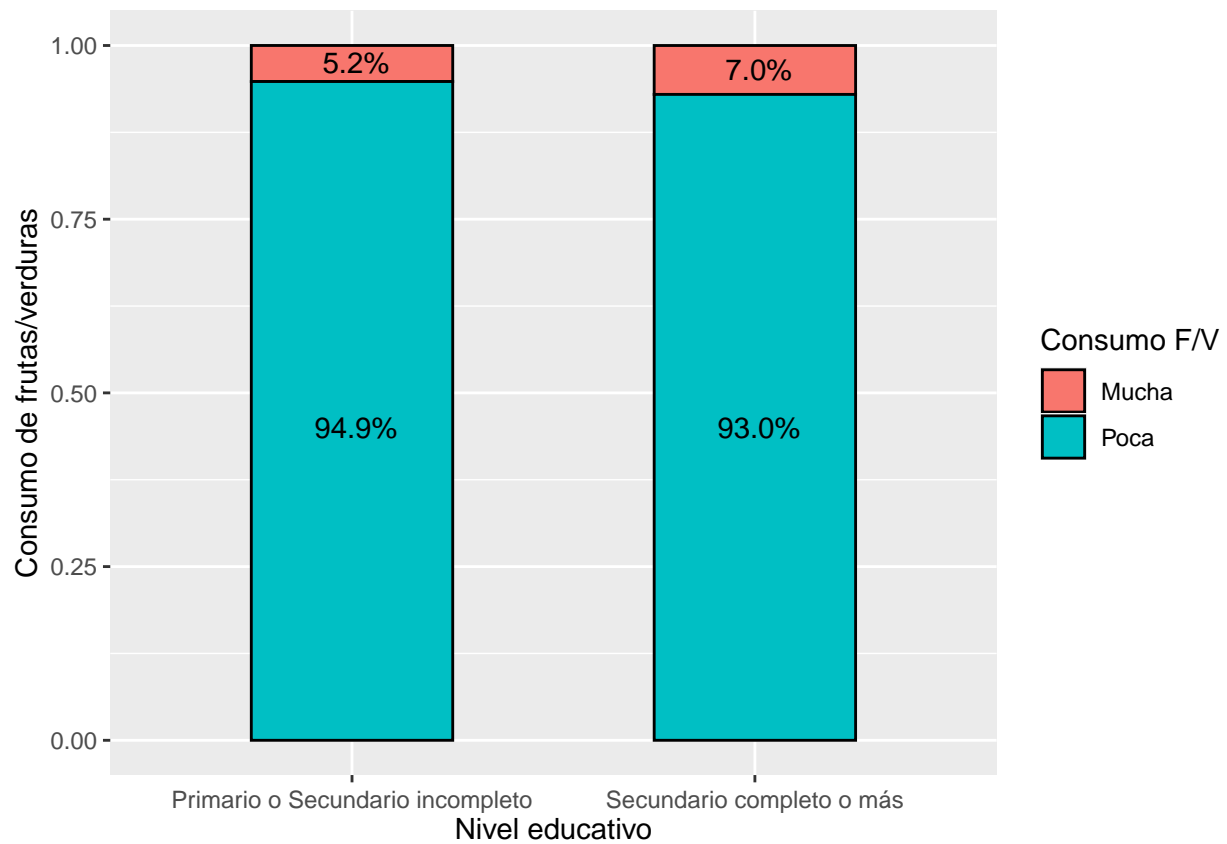
*Tabla 1:* Media de edad de los individuos encuestados según el consumo de muchas o pocas porciones diarias frutas/verduras.

En la *Tabla 1* vemos que el promedio de edad de la gente que consume 5 o más porciones de fruta y verdura diarias es un poco mayor a los que consumen menos.

## *Variables explicatorias*

### A nivel del individuo

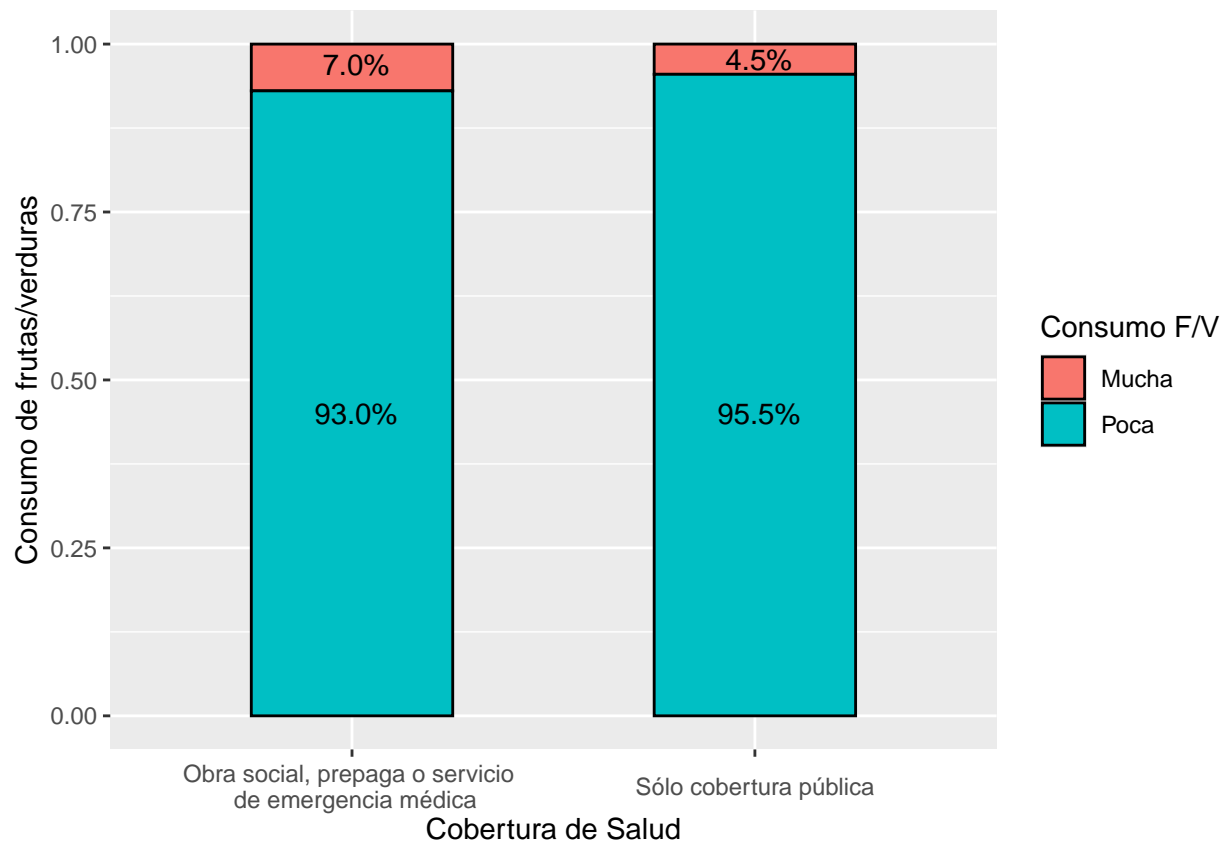
#### Consumo según nivel de instrucción



*Figura 5:* Gráfico de barras de proporción de consumo de muchas o pocas porciones diarias de frutas/verduras según el nivel de instrucción (“primario o secundario incompleto” o “secundario completo o más”).

El consumo de frutas o verduras parece ser mayor en personas con secundario completo o más.

## Consumo según cobertura de salud



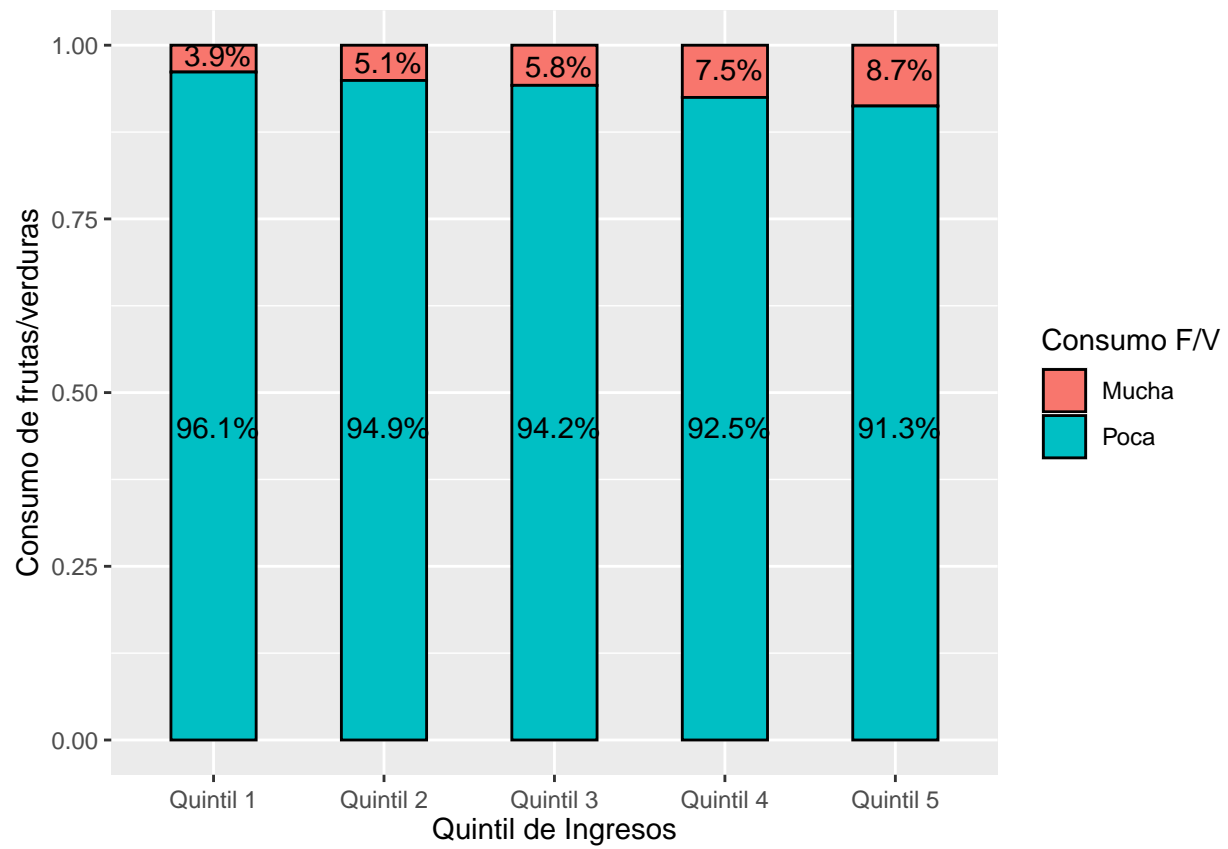
*Figura 6:* Gráfico de barras de proporción de consumo de muchas o pocas porciones diarias de frutas/verduras según el tipo de cobertura de salud (“obra social, prepaga o servicio de emergencia médica” o “sólo cobertura de salud”).

Se observa una mayor proporción de personas que consumen muchas frutas/verduras en la primera categoría (obra social, prepaga o servicio de emergencia médica) en relación a la segunda (cobertura pública).



## A nivel del hogar

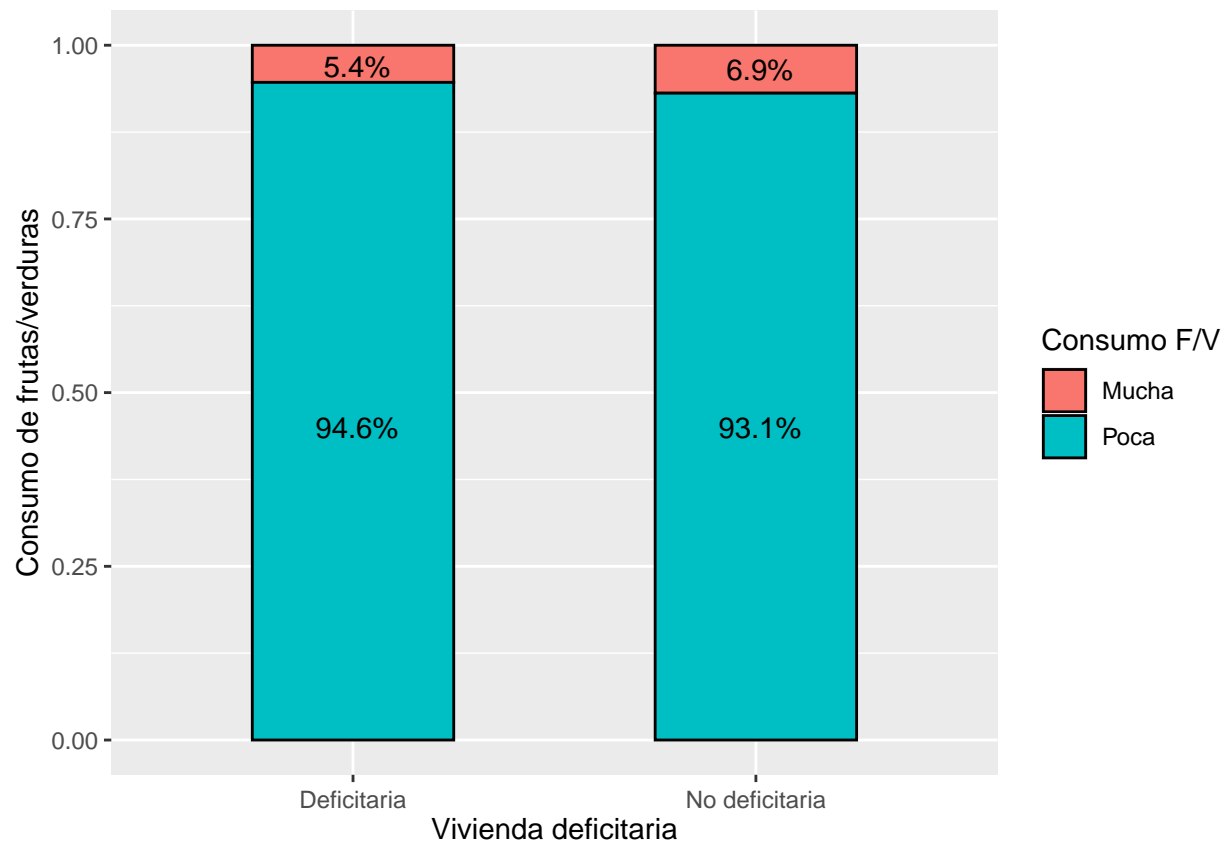
Consumo según quintil de ingresos del jefe del hogar



*Figura 7:* Gráfico de barras de proporción de consumo de muchas o pocas porciones diarias de frutas/verduras según el quintil de ingresos del jefe del hogar (del 1 al 5).

Se puede observar que el consumo de frutas y verduras es mayor a medida que los individuos pertenecen a quintiles de mayores ingresos, lo que pareciera indicar un efecto dosis-respuesta.

### Consumo según vivienda deficitaria sí o no



*Figura 8:* Gráfico de barras de proporción de consumo de muchas o pocas porciones diarias de frutas/verduras según si la vivienda se clasifica como deficitaria o no.

La proporción de gente que come muchas frutas/verduras es levemente mayor en el caso de las viviendas no deficitarias.

## A nivel de la provincia

Gráfico de consumo según NBI agrupado

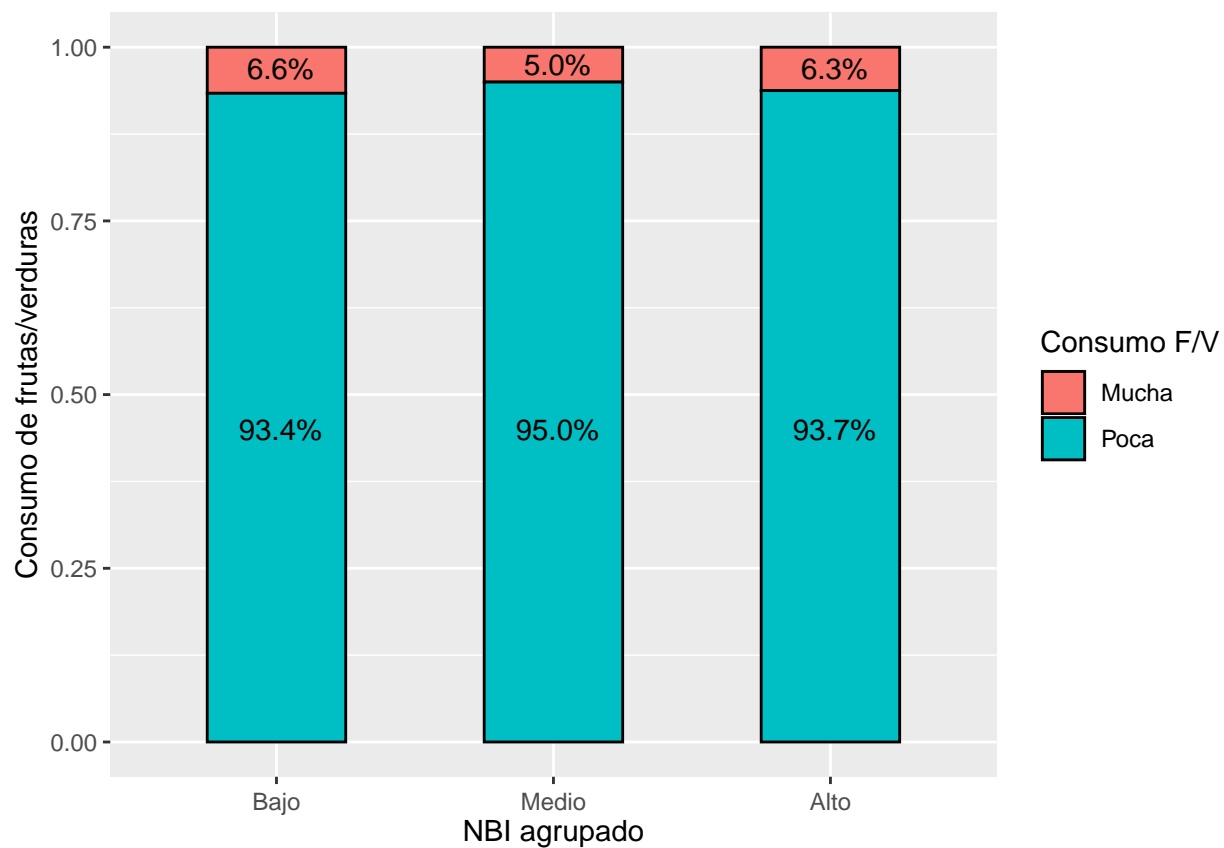


Figura 9: Gráfico de barras de proporción de consumo de muchas o pocas porciones diarias de frutas/verduras según el NBI de la provincia de residencia del encuestado (“bajo”, “medio”, “alto”).

No parece haber asociación entre el NBI y el consumo de frutas/verduras, o al menos no la asociación esperada (el consumo de 5 o más porciones de F/V es mayor en las provincias con NBI bajo y alto).

## Tabla resumen de la estadística descriptiva

<b>Variables explicatorias</b>	<b>Total n=28471</b>	<b>% Total</b>	<b>% Mucha</b>	<b>% Poca</b>
<b><i>Género</i></b>				
Mujer	16257	57.10	6.91	93.09
Varón	12214	42.90	5.31	94.69
<b><i>Nivel Educativo</i></b>				
Primario o Secundario incompleta	12548	44.07	5.19	94.81
Secundario completo o mas	15923	55.93	7.04	92.96
<b><i>Cobertura Médica</i></b>				
Solo cobertura pública	8555	30.05	4.50	95.50
Obra social, prepaga	19916	69.95	6.96	93.04
<b><i>Quintil de ingresos</i></b>				
1	5513	19.36	3.86	96.14
2	5778	20.29	5.07	94.93
3	5555	19.51	5.78	94.22
4	5867	20.61	7.52	92.48
5	5758	20.22	8.75	91.25
<b><i>Vivienda deficitaria</i></b>				
Si	12401	43.56	5.35	94.65
No	16070	56.44	6.90	93.10

*Tabla 2:* Tabla resumen de estadística descriptiva. Se muestra el total de individuos encuestados correspondientes a cada nivel de las variables explicatorias (Total n), el porcentaje de individuos en cada nivel (% Total), el porcentaje de individuos dentro de ese nivel que consumen 5 o más porciones diarias de frutas y verduras (% Mucha), y el porcentaje de individuos dentro de ese nivel que consumen menos de 5 porciones diarias de frutas y verduras (% Poca).

## Modelo teórico

### El modelo tiene tres componentes

#### Predictor lineal

$$\eta_i = \beta_0 + \beta_1 * Edad + \beta_2 * Mujer + \beta_3 * Cobertura\ salud + \beta_4 * Secundario\ Completo\ o\ más + \\ \beta_5 * Vivienda\ deficitaria + \beta_6 * Quintil\ 2 + \beta_7 * Quintil\ 3 + \\ \beta_8 * Quintil\ 4 + \beta_9 * Quintil\ 5 + \beta_{10} * NBI\ Medio + \beta_{11} * NBI\ Alto + \\ \beta_{12} * Provincia\ 1 + \dots + \beta_{36} * Provincia\ 24$$

#### Componente aleatorio

$$Y_i : \text{consumo de 5 o más frutas o verduras diarias o no consumo de 5 o más frutas o verduras diarias} \\ Y_i \sim \text{Bernoulli}(\pi_i)$$

#### Función de enlace

$$\eta_i = \text{logit } \pi_i = \ln \frac{\pi_i}{1 - \pi_i} = \ln (\text{odds } \pi_i)$$

## Planteamos los supuestos y construimos el modelo

### Supuestos

- 1) Muestra aleatoria, observaciones independientes.
- 2) No-colinealidad. Como las VE que probablemente son colineales son cualitativas, no podemos calcular VIF ni hacer gráficos de correlación. Sin embargo, es probable que haya colinealidad entre las variables, porque 5 de ellas hablan sobre el nivel socioeconómico, y es razonable, por ejemplo, que las personas con bajos ingresos posean solamente cobertura de salud pública. Para analizar si hay colinealidad vamos a plantear distintos modelos, agregando variables de a una, y analizar cambios en las significancias.
- 3) Linealidad (entre las VE cuantitativas y el logaritmo natural de los odds estimados a partir de la muestra para cada nivel) (ver ANEXO II).

### Aclaración sobre los resultados presentados

Presentamos las salidas de los “**drop1**” para cada modelo, en las cuales evaluamos la significancia de todas las variables incluidas en el modelo. Los summaries, en los que evaluamos que los estimadores para cada variable dieran números razonables, se encuentran en el ANEXO III.

#### Modelo 1

Primero, incluimos las 3 variables de control.

```
M1 <- glm(Consumo_FV ~ Edad + Genero + Provincia, data = datos1, family = binomial)
```

## single term deletions

```
Model:
Consumo_FV ~ Edad + Genero + Provincia
              Df Deviance   AIC    LRT Pr(>Chi)
<none>                12702 12754
Edad             1      12765 12815   63.17 1.901e-15 ***
Genero           1      12730 12780   27.98 1.224e-07 ***
Provincia       23      13164 13170  463.01 < 2.2e-16 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Los coeficientes de “edad” y de “género” dan significativos. Si bien la variable “provincia” también lo es, al tener más de 2 niveles no estamos controlando el error global en el summary, por lo que los p-valores observados no son confiables. Además, los coeficientes son diferencias de medias respecto al nivel de referencia (CABA), lo que no es de interés en este caso.

**Para evaluar potencial colinealidad entre las VEs**, vamos a ir agregando las VEs de a una, y evaluando significancias y errores estándar de los coeficientes de regresión parciales a medida que agregamos VEs.

**Probamos agregando las variables explicatorias de a una al modelo con las variables control:**

### Modelo 2

“Cobertura de salud”.

```
M2 <- glm(Consumo_FV ~ Cobertura_salud + Edad + Genero + Provincia, data = datos1, family = binomial)
```

## single term deletions

```
Model:
Consumo_FV ~ Cobertura_salud + Edad + Genero + Provincia
              Df Deviance   AIC    LRT Pr(>Chi)
<none>                12671 12725
Cobertura_salud  1      12702 12754   30.38 3.557e-08 ***
Edad             1      12704 12756   33.19 8.379e-09 ***
Genero           1      12697 12749   25.87 3.649e-07 ***
Provincia       23      13134 13142  462.83 < 2.2e-16 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Significativa “cobertura de salud”.

### Modelo 3

“Nivel educativo”

```
M3 <- glm(Consumo_FV ~ Nivel_Educ_agrup + Edad + Genero + Provincia, data = datos1, family = binomial)
```

## single term deletions

Model:

Consumo\_FV ~ Nivel\_Educ\_agrup + Edad + Genero + Provincia

	Df	Deviance	AIC	LRT	Pr(>Chi)
<none>		12630	12684		
Nivel_Educ_agrup	1	12702	12754	71.74	< 2.2e-16 ***
Edad	1	12729	12781	99.39	< 2.2e-16 ***
Genero	1	12652	12704	22.41	2.206e-06 ***
Provincia	23	13090	13098	460.44	< 2.2e-16 ***

---

signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Significativa “nivel educativo”.

## Modelo 4

“Quintil de ingresos”

## single term deletions

Model:

Consumo\_FV ~ Quintil\_ingresos + Edad + Genero + Provincia

	Df	Deviance	AIC	LRT	Pr(>Chi)
<none>		12585	12645		
Quintil_ingresos	4	12702	12754	116.12	< 2.2e-16 ***
Edad	1	12633	12691	47.32	6.027e-12 ***
Genero	1	12620	12678	34.34	4.631e-09 ***
Provincia	23	13027	13041	441.33	< 2.2e-16 ***

---

signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Significativa “quintil de ingresos”. Siguen siendo significativas las variables de control.

## Modelo 5

“Vivienda deficitaria”.

## single term deletions

Model:

Consumo\_FV ~ Vivienda\_deficitaria + Edad + Genero + Provincia

	Df	Deviance	AIC	LRT	Pr(>Chi)	
<none>		12662	12716			
Vivienda_deficitaria	1	12702	12754	39.41	3.442e-10	***
Edad	1	12714	12766	51.75	6.303e-13	***
Genero	1	12689	12741	26.96	2.082e-07	***
Provincia	23	13145	13153	483.12	< 2.2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Significativa “vivienda deficitaria”.

## Modelo 6

“NBI agrupado”.

En este modelo, excluimos a “provincia”, porque “NBI\_agrupado” y “provincia” están perfectamente anidados (a cada provincia le corresponde un valor de NBI), por lo que el modelo no se puede ajustar incluyendo las dos variables.

## single term deletions

Model:

Consumo\_FV ~ NBI\_agrupado + Edad + Genero

	Df	Deviance	AIC	LRT	Pr(>Chi)	
<none>		13148	13158			
NBI_agrupado	2	13164	13170	16.445	0.0002686	***
Edad	1	13215	13223	67.164	2.498e-16	***
Genero	1	13175	13183	27.278	1.762e-07	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Significativo “NBI”.

Ahora vamos complejizando el modelo, sumando de a una VE.

## Modelo 7

```
M7 <- glm(Consumo_FV ~ Cobertura_salud + Nivel_Educ_agrup
          + Edad + Genero + Provincia, data = datos1, family = binomial)
```



## single term deletions

Model:

Consumo\_FV ~ Cobertura\_salud + Nivel\_Educ\_agrup + Edad + Genero +  
Provincia

	Df	Deviance	AIC	LRT	Pr(>Chi)	
<none>		12617	12673			
Cobertura_salud	1	12630	12684	12.58	0.0003903	***
Nivel_Educ_agrup	1	12671	12725	53.95	2.062e-13	***
Edad	1	12681	12735	63.50	1.600e-15	***
Genero	1	12639	12693	21.73	3.139e-06	***
Provincia	23	13078	13088	461.13	< 2.2e-16	***

---

signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Ambas dan significativas.

## Modelo 8

Agregamos “quintil de ingresos”.

```
M8 <- glm(Consumo_FV ~ Cobertura_salud + Nivel_Educ_agrup + Quintil_ingresos  
+ Edad + Genero + Provincia, data = datos1, family = binomial)
```

## single term deletions

Model:

Consumo\_FV ~ Cobertura\_salud + Nivel\_Educ\_agrup + Quintil\_ingresos +  
Edad + Genero + Provincia

	Df	Deviance	AIC	LRT	Pr(>Chi)	
<none>		12560	12624			
Cobertura_salud	1	12562	12624	1.34	0.247	
Nivel_Educ_agrup	1	12582	12644	21.89	2.891e-06	***
Quintil_ingresos	4	12617	12673	57.01	1.230e-11	***
Edad	1	12614	12676	53.33	2.823e-13	***
Genero	1	12588	12650	28.37	1.000e-07	***
Provincia	23	13006	13024	445.40	< 2.2e-16	***

---

signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Al agregar “quintil de ingresos” como VE, deja de dar significativa la variable “cobertura de salud”. Esto indica que **probablemente haya colinealidad entre estas dos variables**, y por lo tanto el aporte independiente de cada una de ellas es pequeño, por lo cual uno de los coeficientes se vuelve no significativo.

## Modelo 9

Agregamos, ahora, “vivienda deficitaria”, dejando “cobertura de salud” (pero eliminando “quintil de ingresos”).

```
M9 <- glm(Consumo_FV ~ Cobertura_salud + Nivel_Educ_agrup + Vivienda_deficitaria +  
Edad + Genero + Provincia, data = datos1, family = binomial)
```

single term deletions

Model:

Consumo\_FV ~ Cobertura\_salud + Nivel\_Educ\_agrup + Vivienda\_deficitaria +  
Edad + Genero + Provincia

	Df	Deviance	AIC	LRT	Pr(>Chi)	
<none>		12600	12658			
Cobertura_salud	1	12609	12665	8.94	0.002784	**
Nivel_Educ_agrup	1	12642	12698	41.88	9.727e-11	***
vivienda_deficitaria	1	12617	12673	17.07	3.598e-05	***
Edad	1	12656	12712	55.87	7.754e-14	***
Genero	1	12622	12678	21.76	3.083e-06	***
Provincia	23	13074	13086	473.97	< 2.2e-16	***

---

signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Todos los coeficientes de las variables explicatorias relacionadas con nivel socioeconómico dan significativos por ahora.

## Modelo 10

Agregamos la última VE, que es NBI agrupado, pero sacando provincia (porque estas dos están perfectamente anidadas).

```
M10 <- glm(Consumo_FV ~ Cobertura_salud + Nivel_Educ_agrup + Vivienda_deficitaria + NBI_agrupado + Edad + Gen
```

single term deletions

Model:

Consumo\_FV ~ Cobertura\_salud + Nivel\_Educ\_agrup + Vivienda\_deficitaria +  
NBI\_agrupado + Edad + Genero

	Df	Deviance	AIC	LRT	Pr(>Chi)	
<none>		13054	13070			
Cobertura_salud	1	13064	13078	9.960	0.001600	**
Nivel_Educ_agrup	1	13102	13116	47.355	5.923e-12	***
vivienda_deficitaria	1	13061	13075	7.321	0.006815	**
NBI_agrupado	2	13074	13086	19.955	4.642e-05	***
Edad	1	13116	13130	62.416	2.781e-15	***
Genero	1	13076	13090	21.499	3.540e-06	***

---

signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Todas dan significativas.

## Modelo 11

Entre “quintil de ingresos” y “cobertura de salud” detectamos problemas de colinealidad. Ahora probamos incluyendo “quintil de ingresos”, sin “cobertura de salud”, y agregando “vivienda deficitaria”.

```
M11 <- glm(Consumo_FV ~ Quintil_ingresos + Nivel_Educ_agrup + Vivienda_deficitaria  
+ Edad + Genero + Provincia, data = datos1, family = binomial)
```

## single term deletions

```
Model:
Consumo_FV ~ Quintil_ingresos + Nivel_Educ_agrup + vivienda_deficitaria +
  Edad + Genero + Provincia
      Df Deviance   AIC    LRT Pr(>Chi)
<none>          12550 12614
Quintil_ingresos    4    12609 12665   58.75 5.302e-12 ***
Nivel_Educ_agrup    1    12569 12631   18.85 1.417e-05 ***
vivienda_deficitaria 1    12562 12624   11.21 0.000813 ***
Edad                1    12608 12670   57.31 3.731e-14 ***
Genero              1    12579 12641   28.43 9.699e-08 ***
Provincia           23    13006 13024  455.68 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Todas son significativas.

## Modelo 12

Agregamos, al modelo con “quintil de ingresos”, “NBI agrupado”, y sacamos la provincia.

```
M12 <- glm(Consumo_FV ~ Quintil_ingresos + Nivel_Educ_agrup + Vivienda_deficitaria +
  NBI_agrupado + Edad + Genero , data = datos1, family = binomial)
```

## single term deletions

```
Model:
Consumo_FV ~ Quintil_ingresos + Nivel_Educ_agrup + vivienda_deficitaria +
  NBI_agrupado + Edad + Genero
      Df Deviance   AIC    LRT Pr(>Chi)
<none>          12981 13003
Quintil_ingresos    4    13064 13078  83.134 < 2.2e-16 ***
Nivel_Educ_agrup    1    12999 13019  17.955 2.262e-05 ***
vivienda_deficitaria 1    12983 13003   1.956 0.1619
NBI_agrupado        2    13006 13024  25.020 3.689e-06 ***
Edad                1    13043 13063  62.253 3.021e-15 ***
Genero              1    13011 13031  30.235 3.827e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Todas son significativas.

## Selección de modelos

Rankeamos los modelos por su devianza residual que nos da una medida de la falta de ajuste y por Akaike (AIC) que tiene en cuenta la verosimilitud y número de parámetros a estimar por los modelos.

```
## Analysis of Deviance Table
##
## Model 1: Consumo_FV ~ Cobertura_salud + Nivel_Educ_agrup + Edad + Genero +
##   Provincia
## Model 2: Consumo_FV ~ Cobertura_salud + Nivel_Educ_agrup + Quintil_ingresos +
```

```
##      Edad + Genero + Provincia
## Model 3: Consumo_FV ~ Cobertura_salud + Nivel_Educ_agrup + Vivienda_deficitaria +
##      Edad + Genero + Provincia
## Model 4: Consumo_FV ~ Cobertura_salud + Nivel_Educ_agrup + Vivienda_deficitaria +
##      NBI_agrupado + Edad + Genero
## Model 5: Consumo_FV ~ Quintil_ingresos + Nivel_Educ_agrup + Vivienda_deficitaria +
##      Edad + Genero + Provincia
## Model 6: Consumo_FV ~ Quintil_ingresos + Nivel_Educ_agrup + Vivienda_deficitaria +
##      NBI_agrupado + Edad + Genero
##      Resid. Df Resid. Dev   Df Deviance  Pr(>Chi)
## 1      28391      12617
## 2      28387      12560    4      57.01 1.230e-11 ***
## 3      28390      12600   -3     -39.94 1.098e-08 ***
## 4      28411      13054  -21    -454.02 < 2.2e-16 ***
## 5      28387      12550   24     503.83 < 2.2e-16 ***
## 6      28408      12981  -21    -430.65 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##      df      AIC
## M7  28 12673.17
## M8  32 12624.16
## M9  29 12658.09
## M10 8 13070.11
## M11 32 12614.28
## M12 11 13002.94
```

Teniendo en cuenta los dos criterios, el modelo 11 (que incluye quintil de ingresos, nivel educativo, vivienda deficitaria, edad, género y provincia) es el ganador.

## Modelo seleccionado en términos teóricos

### Predictor lineal

$$\eta_i = \beta_0 + \beta_1 * \text{Quintil 2} + \beta_2 * \text{Quintil 3} + \beta_4 * \text{Quintil 4} + \beta_5 * \text{Quintil 5} + \\ \beta_6 * \text{Secundario Completo o más} + \beta_7 \text{Vivienda deficitaria} + \beta_8 * \text{Edad} + \beta_9 * \text{Género} + \\ \beta_{10} * \text{Provincia 1} + \dots + \beta_{34} * \text{Provincia 24}$$

### Componente aleatorio

$Y_i$  : consumo de 5 o más frutas o verduras diarias o no consumo de 5 o más frutas o verduras diarias

$$Y_i \sim \text{Bernoulli}(\pi_i)$$

### Función de enlace

$$\eta_i = \text{logit } \pi_i = \ln \frac{\pi_i}{1 - \pi_i} = \ln (\text{odds } \pi_i)$$

## Validación del modelo

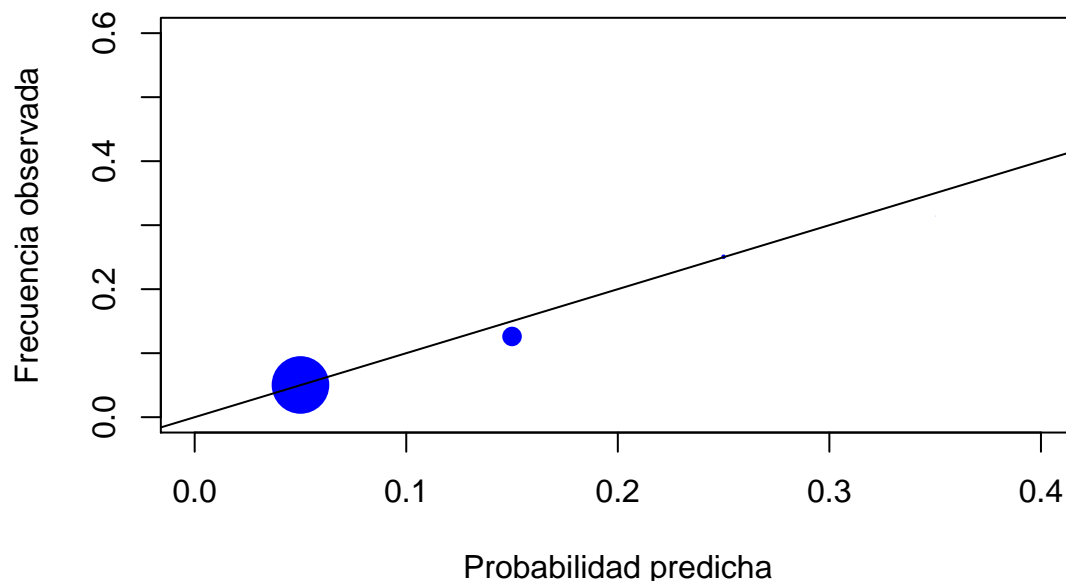
### Prueba de Hosmer-Lemeshow

```
##  
## Hosmer and Lemeshow goodness of fit (GOF) test  
##  
## data:  datos1$Consumo_FV, fitted(M11)  
## X-squared = 3.0417, df = 8, p-value = 0.9317
```

Con un p-valor = 0,957 para la prueba de Hosmer-Lemeshow y una confianza del 95%, no hay evidencias para rechazar la hipótesis de un buen ajuste al modelo.

### Gráfico de frecuencia observada vs. frecuencia esperada para verificar el ajuste

El tamaño de los puntos es proporcional al número de observaciones.



```
## integer(0)
```

*Figura 10:* Gráfico frecuencia observada vs. frecuencia predicha (según el modelo seleccionado) para consumo de muchas o pocas porciones diarias de frutas/verduras donde el tamaño de cada punto es proporcional al número de casos.

### Prueba de Kolmogorov-Smirnov

Con un p-valor = 0,8007 para la prueba de Kolmogorov-Smirnov y un NC95% tampoco se encontraron evidencias para rechazar un buen ajuste al modelo (ver ANEXO IV).

### Curva ROC

La utilizamos para determinar el punto de corte que maximiza la sensibilidad y la especificidad de las predicciones del modelo. **El punto de corte ideal es 0.06.** Es decir, si el modelo estima una probabilidad de consumo mayor a 0.06

para un individuo, asumimos que la predicción es que este individuo consume muchas porciones diarias de frutas/verduras (ver ANEXO IV).

Utilizamos este punto de corte para crear la matriz de confusión

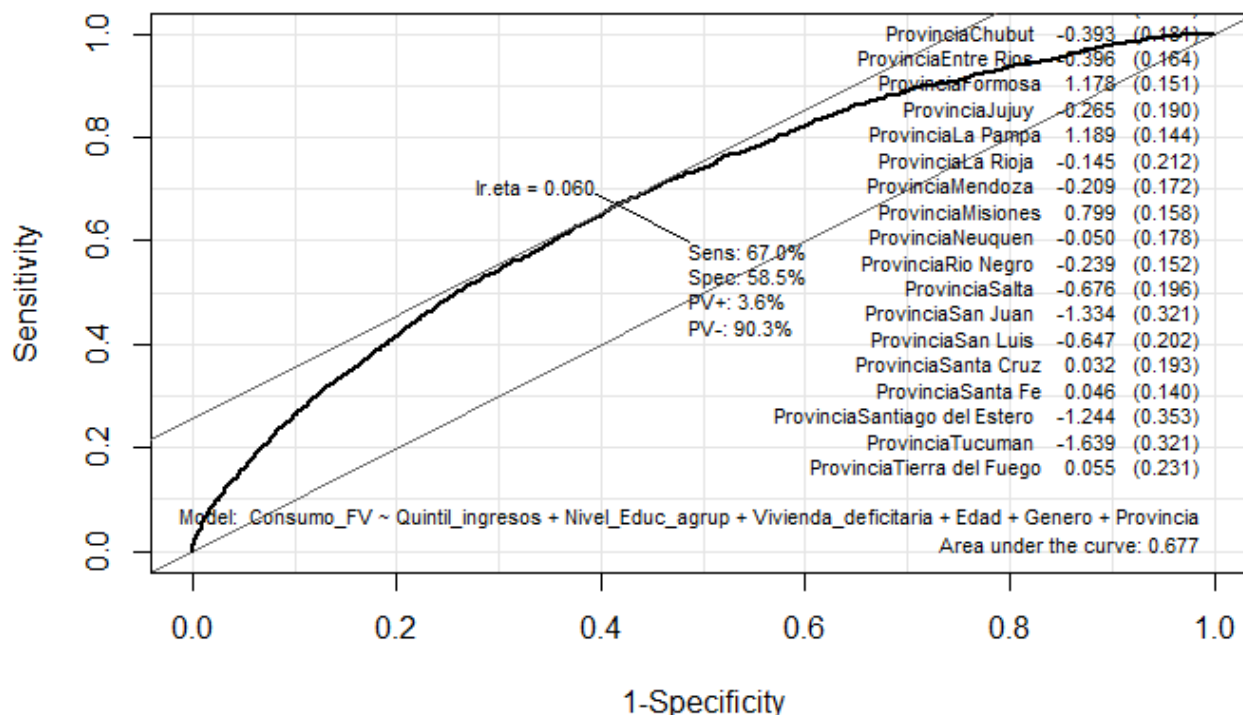


Figura 11: curva ROC, gráfico donde se ilustra la sensibilidad (tasa de verdaderos positivos) y 1-especificidad (tasa de falsos positivos) de cada uno de los posibles puntos de corte a partir de los cuales se supone que el modelo estima que el individuo consume muchas frutas/verduras. Como se observa en el gráfico, el punto de corte ideal es 0.06.

El área bajo la curva ROC (AUC) vale 0.68 lo que le confiere una capacidad predictiva “moderada”.

```
## Confusion Matrix and Statistics
##
##
##      0      1
## 0 15482   578
## 1 11165  1194
##
##               Accuracy : 0.5868
##               95% CI : (0.581, 0.5925)
##      No Information Rate : 0.9376
##      P-Value [Acc > NIR] : 1
##
##               Kappa : 0.0673
##
##  Mcnemar's Test P-Value : <2e-16
##
##      Sensitivity : 0.58100
##      Specificity : 0.67381
##      Pos Pred Value : 0.96401
##      Neg Pred Value : 0.09661
##      Prevalence : 0.93765
##      Detection Rate : 0.54478
##      Detection Prevalence : 0.56511
```

```
##          Balanced Accuracy : 0.62741
##
##          'Positive' Class : 0
##
```

*Tabla 3:* Matriz de confusión y estadísticas. En las columnas de la tabla están los valores observados de consumo de frutas/verduras (0: poco, 1: mucho), en las filas está lo predicho por el modelo respecto del consumo (considerando un punto de corte 0.06).

El modelo tiene una **validez (*accuracy*) del 58%** (la **tasa de error global es del 42%**).

Además, tiene un **58% de sensibilidad** (probabilidad de detectar una persona que consume muchas frutas/verduras cuando efectivamente consume muchas), y un **67% de especificidad** (probabilidad de predecir que una persona no consume suficientes frutas/verduras cuando no lo hace) y por lo tanto, una tasa del 33% de falsos positivos.

## Inferencias a partir del modelo seleccionado

### Comparaciones múltiples para “Quintil de ingresos”

```
## contrast          odds.ratio      SE  df asymp.LCL asymp.UCL
## Quintil 5 / Quintil 4      1.13 0.0797 Inf      0.931      1.37
## Quintil 5 / Quintil 3      1.40 0.1099 Inf      1.127      1.73
## Quintil 5 / Quintil 2      1.54 0.1289 Inf      1.227      1.94
## Quintil 5 / Quintil 1      1.88 0.1777 Inf      1.456      2.44
## Quintil 4 / Quintil 3      1.24 0.0959 Inf      1.002      1.53
## Quintil 4 / Quintil 2      1.37 0.1107 Inf      1.094      1.70
## Quintil 4 / Quintil 1      1.67 0.1527 Inf      1.300      2.14
## Quintil 3 / Quintil 2      1.10 0.0934 Inf      0.876      1.39
## Quintil 3 / Quintil 1      1.35 0.1265 Inf      1.044      1.74
## Quintil 2 / Quintil 1      1.22 0.1150 Inf      0.945      1.58
##
## Results are averaged over the levels of: Nivel_Educ_agrup, Vivienda_deficitaria, Genero, Provincia
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 5 estimates
## Intervals are back-transformed from the log odds ratio scale
```

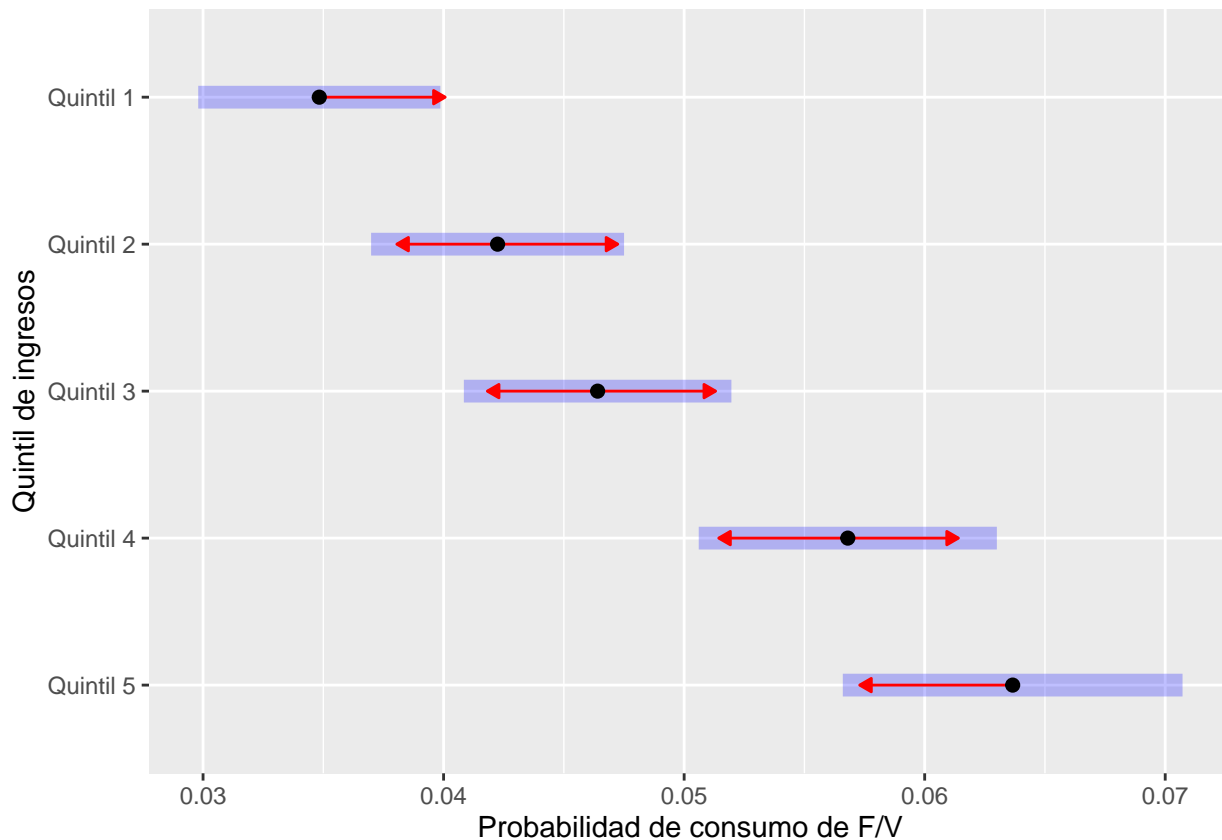


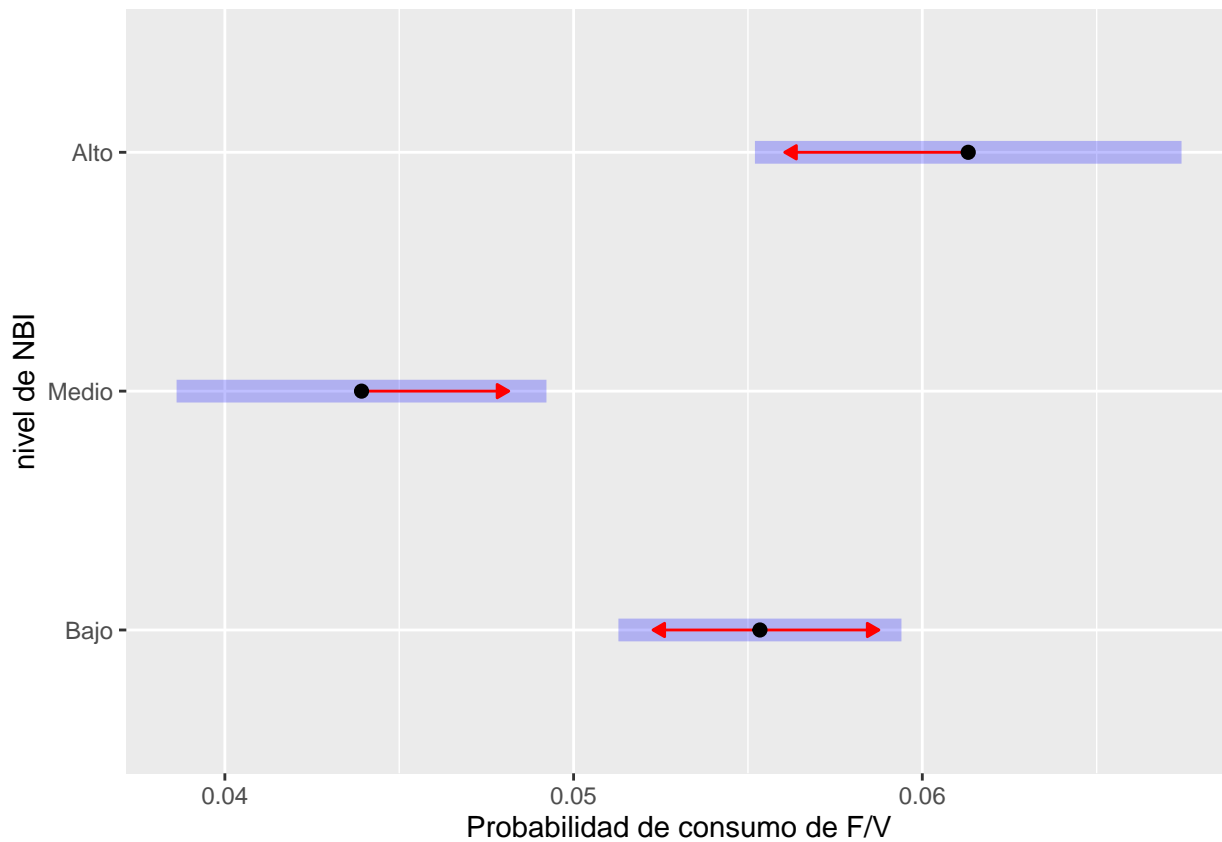
Figura 12: Probabilidad de consumo de 5 o más porciones diarias de F/V para los 5 niveles de quintil de ingresos, a escala de la VR. Las barras grises son los intervalos de confianza (con un nivel de confianza del 95%) para las medias, las flechas rojas son para las comparaciones entre grupos. Si una flecha se superpone a otra de otro grupo, la diferencia no es significativa.

Observamos que entre quintiles lejanos, hay diferencias significativas en cuanto a la probabilidad en el consumo de 5 o más porciones diarias de frutas/verduras (hay mayor consumo en los quintiles superiores); pero entre quintiles cercanos no hay diferencias significativas.

### Comparaciones múltiples para “NBI” (en el modelo 10, que no es el ganador, pero que incluye NBI).

```
## contrast      odds.ratio      SE  df asymp.LCL asymp.UCL
## Bajo / Medio    1.275 0.0887 Inf    1.084    1.501
## Bajo / Alto     0.897 0.0595 Inf    0.768    1.048
## Medio / Alto    0.703 0.0583 Inf    0.579    0.854
##
## Results are averaged over the levels of: Cobertura_salud, Nivel_Educ_agrup, Vivienda_deficitaria, Genero
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 3 estimates
## Intervals are back-transformed from the log odds ratio scale
```





*Figura 13:* Probabilidad de consumo de 5 o más porciones diarias de F/V para los 3 niveles de NBI, a escala de la VR. Las barras grises son los intervalos de confianza para las medias, las flechas rojas son para las comparaciones entre grupos. Si una flecha se superpone a otra de otro grupo, la diferencia no es significativa.

Observamos que tanto las provincias con NBI bajo como alto presentan un consumo significativamente mayor de frutas y verduras respecto a las provincias con NBI medio. A su vez, no hay diferencias significativas en el consumo entre provincias con NBI alto y bajo.

Tabla con los OR ajustados (no incluimos a la provincia)

Variables explicatorias	OR (95% CI)	aCtrlOR (95% CI)	aOR (95% CI)
Edad	1.011 (1.009-1.014)	-	1.011 (1.008-1.014)
Genero (mujer)	1.32 (1.20-1.46)	-	1.32 (1.19-1.46)
Vivienda deficitaria (si)	0.76 (0.69-0.84)	0.64 (0.55-0.74)	0.78 (0.67-0.90)
Nivel educativo (Sec Completo o mas)	1.38 (1.25-1.53)	1.58 (1.42-1.75)	1.29 (1.15-1.45)
Quintiles	1	1	1
Segundo vs primero	1.33 (1.04-1.31)	1.26 (0.98-1.63)	1.22 (0.95-1.58)
Tercero vs primero	1.53 (1.19-1.95)	1.46 (1.14-1.88)	1.35 (1.04-1.74)
Cuarto vs primero	2.03 (1.61-2.56)	1.90 (1.49-2.42)	1.67 (1.30-2.14)
Quinto vs primero	2.39 (1.90-3.00)	2.27 (1.78-2.89)	1.88 (1.46-2.44)

Tabla 4: Tabla resumen de OR obtenidos para las variables explicatorias. En la columna "OR (95%CI)" figuran los valores de cada variable por separado, sin ajustar por las demás, en la columna "aCtrlOR (95% CI)" ajustando por las variables control y en aOR (95% CI)" ajustando por todas las demás variables.

# Gráfico con los Odds Ratios (OR)

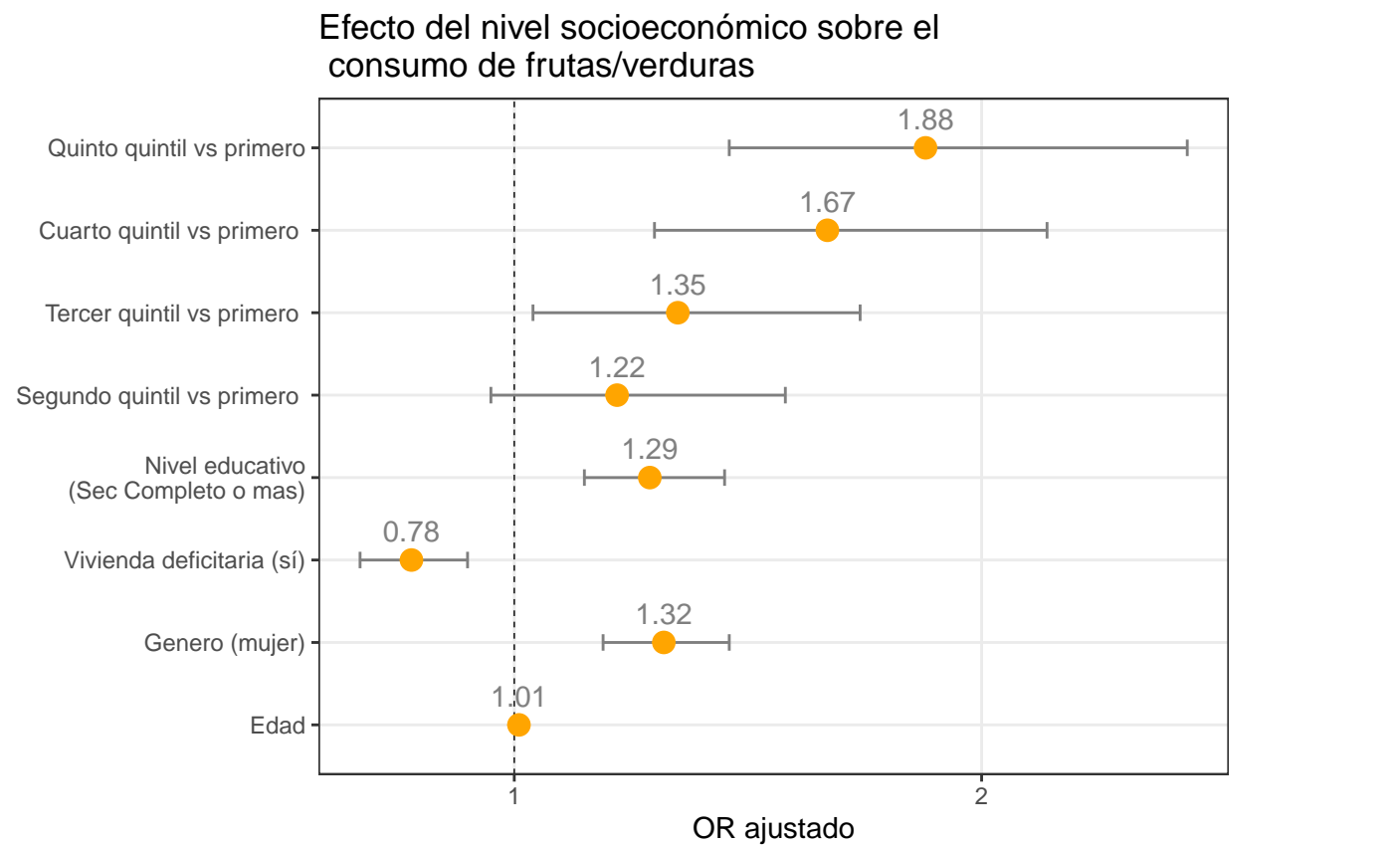


Figura 14: Gráfico de comparaciones, a escala de odds, del consumo de 5 o más porciones de frutas/verduras y sus intervalos de confianza entre niveles de las variables explicatorias del modelo elegido. El punto naranja indica el estimador puntual, y las barras negras, el IC. El valor del OR para cada variable se ajusta por el resto de las variables.

## Conclusiones

Por cada año que aumenta la edad de una persona, el odds estimado de consumo de al menos 5 porciones diarias de frutas o verduras aumenta 1%, controlando por las otras variables (género, provincia de residencia, vivienda deficitaria/no deficitaria, nivel educativo, quintil de ingresos).\*\*

El género mujer está asociado con un aumento estimado en el odds de consumo de al menos 5 porciones diarias de frutas o verduras de entre un 19% y un 46% respecto de ser hombre, controlando por las otras variables, con un 95% de confianza.

Habitar en una vivienda deficitaria está asociado con una disminución estimada en el odds de consumo de al menos 5 porciones diarias de frutas o verduras de entre un 10% y un 33% respecto de habitar en una vivienda no deficitaria, controlando por las otras variables, con un 95% de confianza.

Poseer un nivel educativo que incluya el secundario completo o más está asociado con un aumento estimado en el odds de consumo de al menos 5 porciones diarias de frutas o verduras de entre un 15% y un 45% respecto de poseer un nivel educativo inferior, controlando por las otras variables, con un 95% de confianza.

Un ingreso del jefe del hogar perteneciente al segundo quintil no está asociado con un aumento estimado en el odds de consumo de al menos 5 porciones diarias de frutas o verduras respecto del primer quintil, controlando por las otras variables, con un 95% de confianza. Si el jefe del hogar pertenece al tercer quintil, se produce un aumento en el odds de entre 4% y 74%. Si pertenece al cuarto quintil, aún mayor: de entre 30% y 114%. Y si pertenece al quinto quintil, de entre 46% y 144%.

Según este análisis, el hábito saludable de consumir 5 o más porciones diarias de frutas y verduras está asociado positivamente con niveles socioeconómicos en general más altos. Por otro lado se observa una mayor proporción de mujeres con esta conducta y, que tanto para mujeres como hombres, con los años el consumo saludable va en aumento.

## Gráfico de probabilidades a nivel de la VR

Según edad y género

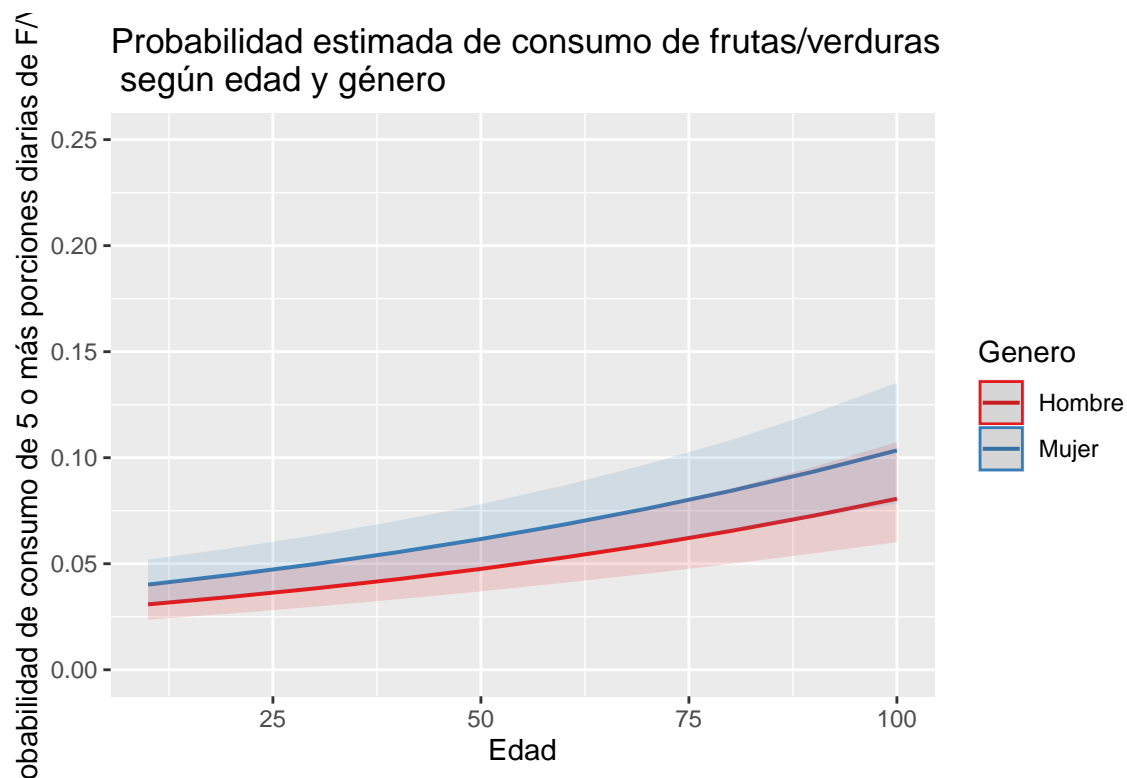


Figura 15: Gráfico de probabilidades estimadas de consumo de 5 o más porciones de frutas/verduras diarias según edad y género. Se grafican los intervalos de confianza, con un nivel de confianza del 95%.

**En mujeres** la probabilidad de consumir 5 porciones diarias de frutas y verduras es mayor con respecto a hombres.

Ademas, **por cada año que aumenta la edad de una persona** la probabilidad de consumir 5 porciones de frutas y verduras diarias aumenta.

## Según nivel de instrucción

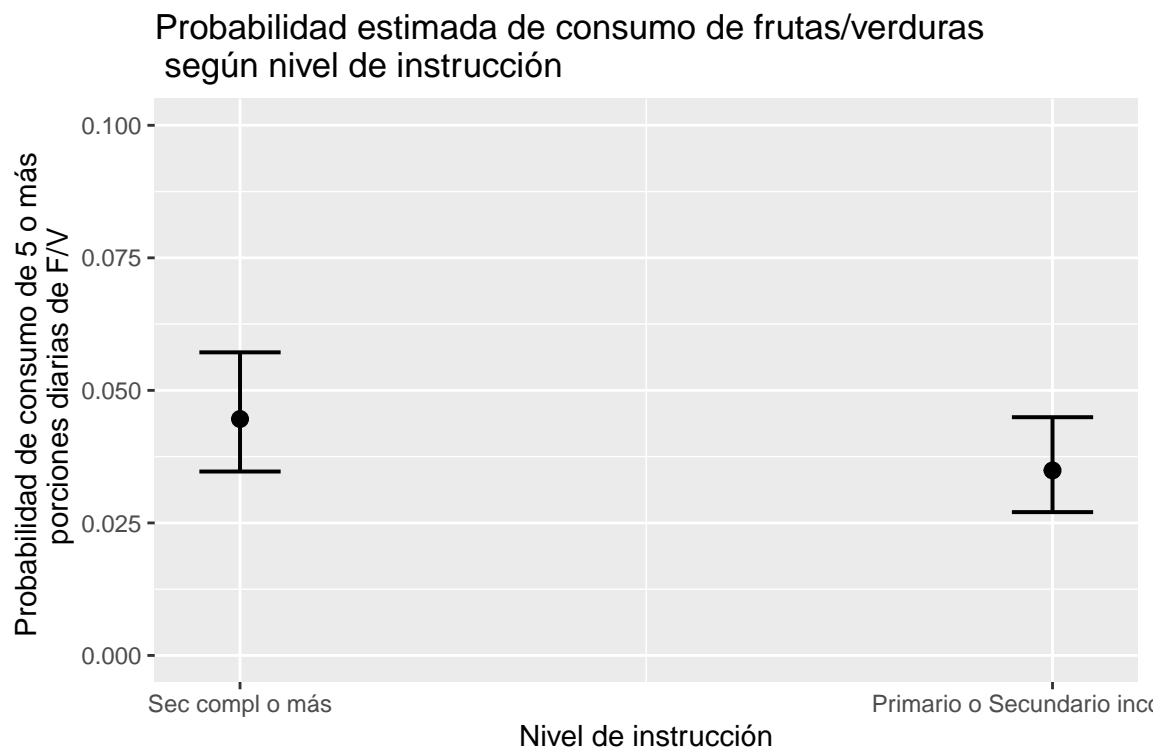
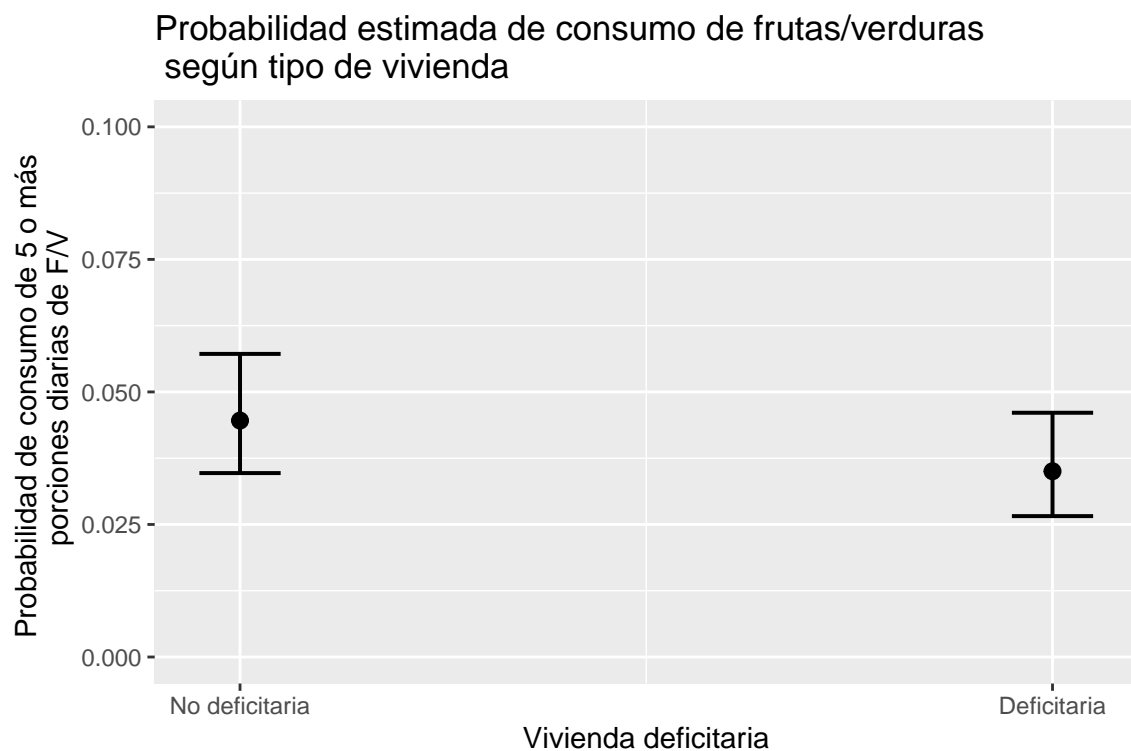


Figura 16: Gráfico de probabilidades estimadas de consumo de 5 o más porciones de frutas/verduras diarias según el nivel de instrucción. Se grafican los intervalos de confianza, con un nivel de confianza del 95%.

La probabilidad de consumir 5 porciones o más de frutas y verduras es menor si el individuo **tiene el secundario incompleto o menos**.

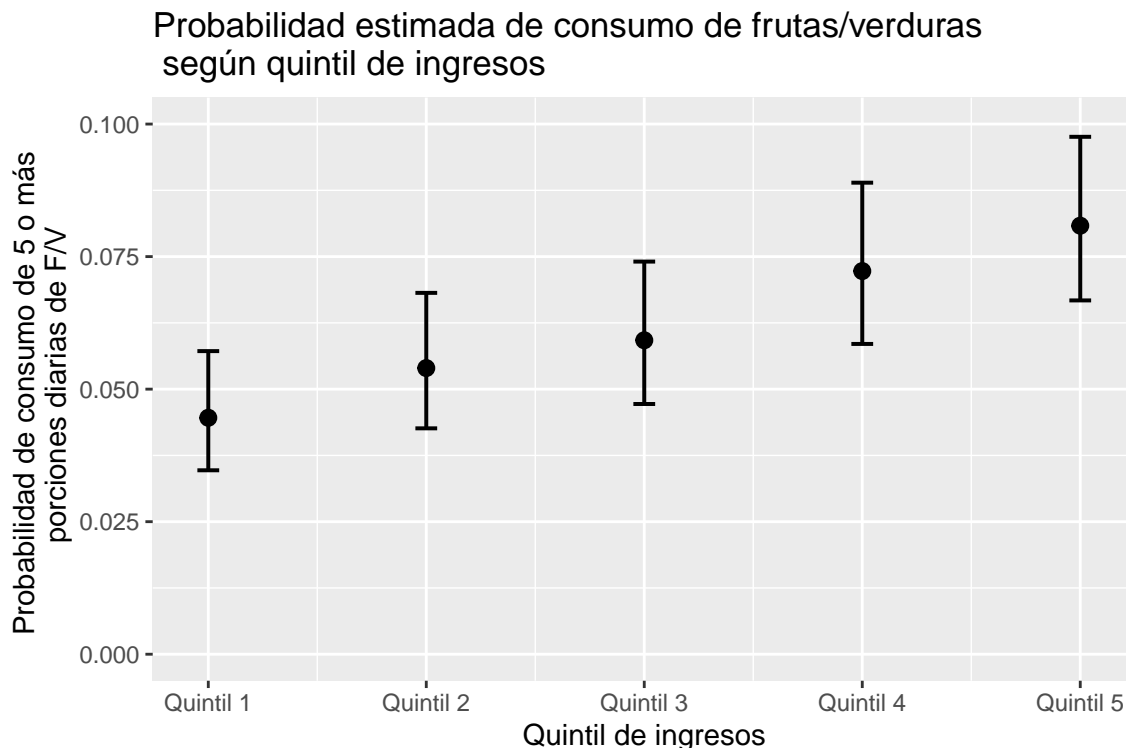
## Según vivienda deficitaria o no deficitaria



*Figura 17:* Gráfico de probabilidades estimadas de consumo de 5 o más porciones de frutas/verduras diarias según si la vivienda es o no deficitaria. Se grafican los intervalos de confianza, con un nivel de confianza del 95%.

La probabilidad de consumir 5 porciones diarias de frutas y verduras es menor si **el individuo habita una vivienda deficitaria**.

### Según vivienda quintiles de ingresos



*Figura 18:* Gráfico de probabilidades estimadas de consumo de 5 o más porciones de frutas/verduras diarias según el quintil de ingresos del jefe de hogar. Se grafican los intervalos de confianza, con un nivel de confianza del 95%.

**A medida que el ingreso del jefe del hogar sea de un quintil superior** la probabilidad de consumir 5 porciones diarias de frutas y verduras también ira aumentando.

## Conclusiones y discusión

Este estudio permitió identificar que el nivel de instrucción, el hecho de habitar en una vivienda adecuada (casa o departamento, con acceso a gas de red, baño y agua dentro de la vivienda) y el nivel de ingresos del jefe del hogar están asociados positivamente con el consumo de frutas y verduras en la población argentina mayor a 18 años residente en regiones urbanizadas (de 5000 o más habitantes), controlando por la edad, el género y la provincia de residencia, con un 95% de confianza.

A pesar de que no era parte de nuestra hipótesis, también hallamos que las mujeres consumen más frutas/verduras que los hombres, y que con el aumento en la edad aumenta la probabilidad de consumir frutas/verduras.

Resultó muy interesante el efecto observado de dosis-respuesta en cuanto a los quintiles de ingresos: sistemáticamente, quintiles más altos poseían un odds ratio más alto de consumo de frutas/verduras respecto a quintiles más bajos, estimándose un aumento del consumo entre el primer y el quinto quintil de entre 46% y 144%. Por lo tanto, la dieta baja en consumo de frutas y verduras impactaría especialmente sobre las familias de menores ingresos.

Si bien esperábamos que a mayor NBI, el consumo diario de frutas y verduras fuera menor, no observamos esta tendencia en los resultados. Se verificó que el consumo de provincias con nivel “bajo” no difirió significativamente con las de “alto”, mientras que el nivel “medio” fue significativamente menor que los dos anteriores. Sin embargo, la mayoría de las provincias que fueron agrupadas en NBI alto (Corrientes, Chaco, Santiago del Estero, Misiones, Formosa, Salta y Jujuy)

forman parte de las principales productoras de hortalizas en el país (Giacobone, Castronuovo, Tiscornia y Allemandi, 2018). Esto podría brindar a la población de estas provincias un acceso directo a las verduras, posiblemente a un precio menor que en el resto del país. Otra posibilidad, es el surgimiento de ferias de agricultura familiar gestadas a la par de la desindustrialización y un escenario de precarización, pobreza y concentración de la tierra rural, de la que participan organismos gubernamentales y no (INTA y Ministerio de Desarrollo Social de la Nación) fomentando diferentes estrategias como posibilidad de generar ingresos a productores familiares principalmente en el noreste, noroeste (donde se concentran las provincias de mayor NBI) y zona pampeana, acercando las producciones fruto-hortícolas directamente del productor al consumidor. A modo de ejemplo, la provincia de Formosa, que posee el mayor NBI del país, de 19.7%, es una de las provincias que más frutas/verduras consume. Es una forma posible de explicar lo observado.

La inclusión de “provincia” como variable de control resulta explicativa. Cabe destacar que esta variable controla más factores que el nivel socioeconómico, y agrega información que no podemos indentificar, como la accesibilidad a las frutas/verduras, la disponibilidad de los alimentos, los precios, las distancias, las costumbres, las políticas dirigidas a la promoción del consumo, etcétera.

Con los resultados expuestos, podemos concluir que las variables nivel de instrucción, deficiencia en la vivienda y nivel de ingresos medidos en quintiles, constituyen factores de riesgo para el consumo bajo de alimentos saludables, actuando a escalas individuo y hogar, para una población mayor a 18 años argentina habitante en áreas de 5000 habitantes o más. Mientras que a escala provincial, no podemos concluir que el NBI sea un factor con una asociación directa al bajo consumo, sino que estaría atravesada por otras variables no contempladas en este estudio.

## Bibliografía

- 1) OMS (2014). Informe sobre la situación mundial de las enfermedades no transmisibles. Recuperado de [https://apps.who.int/iris/bitstream/handle/10665/149296/WHO\\_NMH\\_NVI\\_15.1\\_spa.pdf?sequence=1](https://apps.who.int/iris/bitstream/handle/10665/149296/WHO_NMH_NVI_15.1_spa.pdf?sequence=1).
- 2) OMS (2010). Informe sobre la situación mundial de las enfermedades no transmisibles. Recuperado de: [https://www.who.int/nmh/publications/ncd\\_report2010/es/](https://www.who.int/nmh/publications/ncd_report2010/es/).
- 3) Phelan JC, Link BG, Tehranifar P. Social Conditions as Fundamental Causes of Health Inequalities: Theory, Evidence, and Policy Implications. *Journal of Health and Social Behavior*. 2010; 51(1\_suppl): S28-S40. <https://doi.org/10.1177/0022146510383498>
- 4) Marmot M, Allen JJ. Social determinants of health equity. *Am J Public Health*. 2014 Sep; 104 Suppl 4 (Suppl 4): S517-9. <https://doi.org/10.2105/AJPH.2014.302200>
- 5) Di Cesare M, Khang, YH, Asaria P, Blakely T, Cowan MJ, Farzadfar F et al. Inequalities in Non-Communicable Diseases and Effective Responses. *The Lancet*. 2014; 381, 585-597. [https://doi.org/10.1016/S0140-6736\(12\)61851-0](https://doi.org/10.1016/S0140-6736(12)61851-0)
- 6) Bulletin of the World Health Organization 2009; 87:84-84. doi: 10.2471/BLT.08.062695
- 7) Resultados definitivos de la 4° Encuesta Nacional de Factores de Riesgo, INDEC 2018. [https://www.indec.gov.ar/ftp/cuadros/publicaciones/enfr\\_2018\\_resultados\\_definitivos.pdf](https://www.indec.gov.ar/ftp/cuadros/publicaciones/enfr_2018_resultados_definitivos.pdf)
- 8) Informe Censal 2010, datos de NBI. <http://www2.mecon.gov.ar/hacienda/dinrep/Informes/archivos/NBIAmpliado.pdf>
- 9) Giacobone G, Castronuovo L, Tiscornia V y Allemandi L. Análisis de la cadena de suministro de frutas y verduras en Argentina. Fundación Interamericana del Corazón, 2018.