**PRUEBA SELECCIÓN ANALISTA SOCIOECONÓMICO**

**SUBDEPARTAMENTO DE ESTADÍSTICAS SOCIOECONÓMICAS**

NOMBRE : Valentina Vasquez

RUT :19.124.732-9

**INSTRUCCIONES**

La prueba tiene un **tiempo máximo para su ejecución de 90 minutos**. Este tiempo se hará efectivo una vez que se termine de leer en conjunto las instrucciones y encabezados de los ejercicios, y se hayan aclarado las dudas en relación a los planteamientos y se corrobore que los y las postulantes cuentan con todos los insumos necesarios para la correcta realización de la prueba.

Se encuentra permitido visitar internet pasa resolver dudas. Sin embargo, no está permitido el uso de redes sociales ni cualquier otro medio que permita la comunicación con un tercero para el mismo efecto.

**Pasos a seguir:**

* Corrobore la disponibilidad de las bases de datos **BASE\_PERSONAS** y **BASE\_GASTOS**, disponibles en la carpeta compartida en OneDrive [prueba\_técnica\_sdes](https://inechile-my.sharepoint.com/:f:/g/personal/agustin_arce_ine_cl/EjuBKVfttzhCnwhM2kfLl6QBeBTaaW41ywzw1OTuFQRb8A) (https://inechile-my.sharepoint.com/:f:/g/personal/agustin\_arce\_ine\_cl/EjuBKVfttzhCnwhM2kfLl6QBeBTaaW41ywzw1OTuFQRb8A)
* Las preguntas de desarrollo conceptual, que deban entregarse en Word según las instrucciones, deberán ser contestadas con letra “Calibri” tamaño 11.
* Al finalizar la prueba deberá:
* Guardar este archivo Word con la siguiente nomenclatura “**Apellido\_Nombre.docx”**
* Guardar el *script* o *do-file* con su programación con la nomenclatura “**Apellido\_Nombre.R**” o “**Apellido\_Nombre.do**”
* Enviar a **correo** [**epf@ine.cl**](mailto:epf@ine.cl) **y copia a** [**agustin.arce@ine.cl**](mailto:agustin.arce@ine.cl)los archivos previamente indicados (Word y programación).
* **Se revisará la hora y fecha del envío del correo y de la última modificación** de los archivos enviados, que no pueden exceder el tiempo límite (90 minutos luego de terminar de explicar la evaluación).

**\*\*No cumplir con esta instrucción significará la no revisión de la prueba, entendiéndose no realizada, y por tanto, sin puntaje a asignar\*\***

* Luego de leer las instrucciones en su totalidad, aclare dudas generales referentes a estas.

**PRESENTACIÓN**

El presente instrumento de evaluación está dividido en tres secciones, cada una de ellas con objetivos específicos que procuran medir sus aptitudes, habilidades y conocimiento técnico requerido para el cargo al que postula. Se contempla un puntaje máximo a obtener de 33 puntos, repartidos entre las secciones.

**La primera sección** (abarca el 18,2% del total de puntos a obtener)

Busca medir sus capacidades analíticas estadísticas en la presentación de resultados. Por tanto, se observará:

* Capacidad de análisis estadístico de gráficos
* Capacidad de análisis conjunto
* Capacidad analítica del fenómeno expuesto

**La segunda sección** (contemplael 45,5% del puntaje máximo a obtener)

Tiene por propósito identificar el estado de conocimientos técnicos asociados a uso de software estadístico. Se medirá:

* Exploración de bases de datos
* Herramientas de transformación de datos
* Uso e interpretación de estadísticos

**La tercera sección** (abarca el 24,2% del total de puntos a obtener)

Busca observar su conocimiento en cuanto a modelamiento estadístico e interpretación de resultados. Por tanto, se observará:

* Programación de un modelo que dé cuenta del problema planteado
* Correcta interpretación de resultados del modelo estadístico presentado

**Evaluación Transversal** (comprende el 12,1% del total de puntos a obtener)

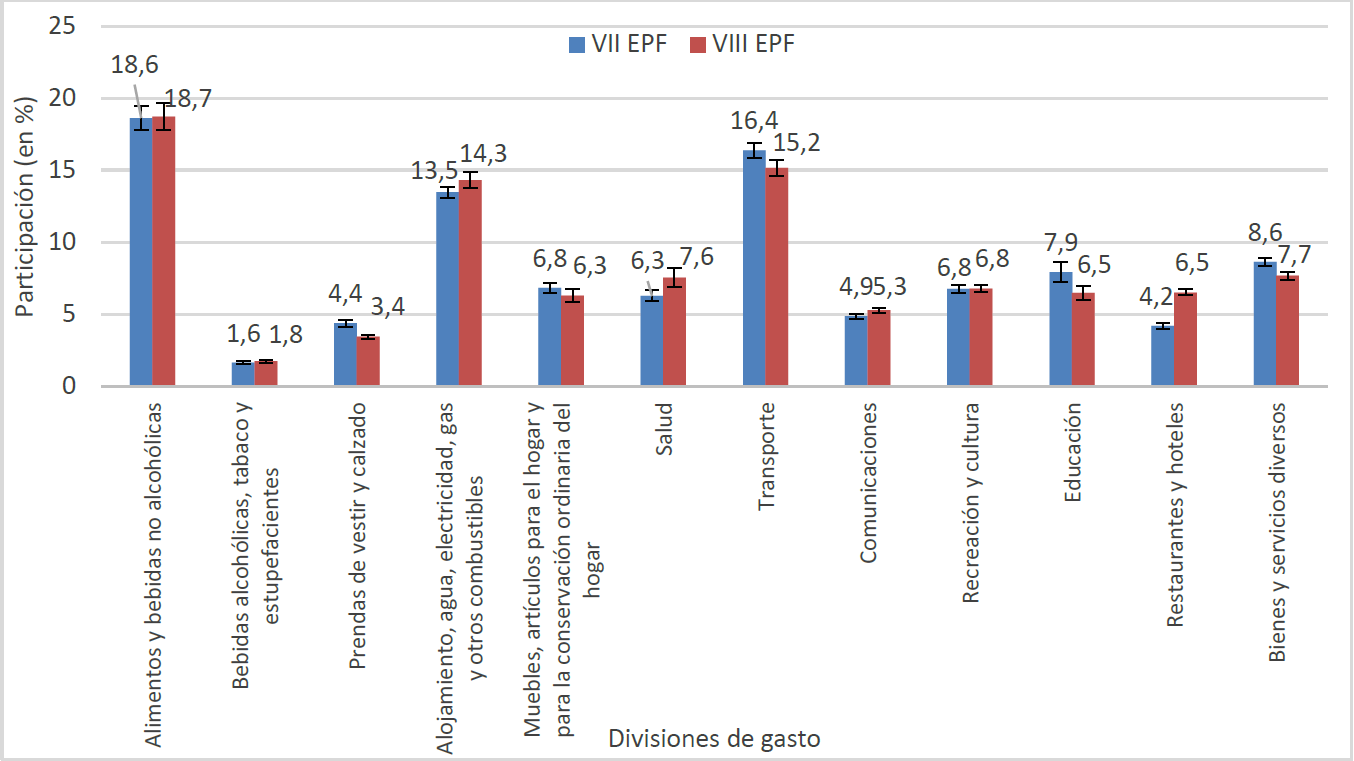
En la interacción del postulante con el instrumento de evaluación que aquí se presenta, se evaluarán las **competencias de redacción** de informes técnicos basados en datos estadísticos. Se busca que los postulantes sean capaces de ordenar ideas, jerarquizándolas adecuadamente; se apoyen en estructuras simples y precisas para la entrega de resultados, valiéndose de elementos adecuados que acompañen una buena redacción (gráficos, tablas, etc.). Esta evaluación se realizará principalmente en base a la Sección I.

**EVALUACIÓN**

**SECCIÓN I**

A continuación se presenta el gráfico 1 que muestra la estructura de gastos promedio mensual de los hogares del total de capitales regionales, basada en la VII EPF (2011-2012) y VIII EPF (2016-2017). Posteriormente se presentan el gráfico 2 y 3 que dan cuenta de información respecto a gasto en Transporte a partir de la VIII EPF.En base a esta información, responda el ítem I.1 en el recuadro indicado más abajo.

***Gráfico 1: Participación de cada una de las divisiones de gastos de la CCIF en el gasto promedio de los hogares VII EPF y VIII EPF. Total capitales regionales***

******

Fuente: Instituto Nacional de Estadísticas (INE) − VIII Encuesta de Presupuestos Familiares (EPF).

|  |  |
| --- | --- |
| **PREGUNTA** | **I.1. Según la participación del gasto presentado en la división de Transporte (gráfico 1) y de acuerdo a lo expuesto en los gráficos 2 y 3 por quintiles de ingreso de los hogares, interprete de forma conjunta.** Recuerde que la respuesta no debe exceder el recuadro disponible para este propósito. |

**Respuesta I.1:**

Tanto en la VII EPF (2011-2012) como en la VIII EPF (2016-2017) los gastos promedio mensuales de los hogares del total de capitales regionales de Chile se concentran en Alimentos y bebidas no alcohólicas (18,6% y 18,7%), Transporte (16,4% y 15,2%) y Alojamiento, electricidad, agua (…) (13,5% y 14,3%). En los dos cortes temporales, estas tres canastas de productos concentran cerca del 50% de los gastos de los hogares. Por otro lado, los menores gastos se concentran en alcohol y cigarrillos.

Tanto en términos relativos como absolutos el quinto quintil de ingresos es el que destina una mayor parte de su presupuesto al pago de transporte. Para este quintil el ítem transporte representa el 17% de su gasto total, lo que representa $379.999.

**SECCIÓN II**

A continuación hará uso de los datos de la VIII Encuesta de Presupuestos Familiares (VIII EPF), que tiene como objetivo la captura de la estructura de gastos e ingresos de los hogares. Utilizando la base de datos “BASE\_PERSONAS” (datos de caracterización a nivel persona) y “BASE\_GASTOS” (gasto total por hogar a nivel de división de gasto), que están disponibles en la carpeta compartida “prueba\_técnica\_sdes”. Genere un *do-file* (Stata) o *script* (R) donde muestre la programación de los tabulados o cálculos que se solicitan.

Si desea indicar algún comentario respecto a la programación, debe hacerlo directamente en la sintaxis (*script* o *do-file*).

|  |  |
| --- | --- |
| **PREGUNTAS** | **II.1** Cuente los hogares presentes en la BASE\_PERSONAS. Considere que la variable FOLIO individualiza a cada hogar.  **II.2** Genere una variable llamada NPERSONAS\_2 que cuente a las personas al interior de cada hogar. Compárela con la variable NPERSONAS (ya existente en la BASE\_PERSONAS), mostrando que no hay diferencias.  **II.3** Genere una tabla que muestre la media, percentil 50, mínimo, máximo y desviación estándar del ingreso disponible por hogar (ING\_DISP\_HOG\_HD) según área de estimación (ZONA), utilizando los valores muestrales. Interprete brevemente los estadísticos de tendencia central y dispersión directamente en el *script* o *do-file*.  **II.4** Genere un tabulado que muestre la cantidad de sustentadores principales (SPRINCIPAL==1) según SEXO, utilizando el factor de expansión (ponderador o peso) de la encuesta (FE).  **II.5** A partir de la variable EDAD ya existente en la BASE\_PERSONAS, genere una variable de edad en tramos, llamándola EDAD\_T. Esta variable debe agrupar las edades en tramos de 10 años (“0 - 9”, “10-19”… hasta “80 o más”). Luego tabule el SEXO según EDAD\_T, utilizando el factor de expansión (FE). En la nueva variable EDAD\_T mantenga los valores -88 y -99 de la variable EDAD.  **II.6** La BASE\_GASTOS que se encuentra disponible en la carpeta compartida “prueba\_técnica\_sdes”, es una modificación de la base de gastos oficial de la VIII EPF. Esta contiene la información de los gastos de los hogares agrupada a nivel de división de gasto (son 12 divisiones en total), por lo tanto, cada observación de esta base corresponde al gasto realizado por un hogar en una división de gasto (cada hogar está repetido como máximo 12 veces, si es que reportó gasto en todas las divisiones).  • A partir de la BASE\_GASTOS, genere una variable llamada GASTOT\_HD\_2 que sume el GASTO total de cada hogar (FOLIO).  • Luego incorpore la variable GASTOT\_HD\_2 recién creada, a la BASE\_PERSONAS, para contar con el gasto de cada hogar y compruebe que la variable GASTOT\_HD\_2 es una réplica de la variable GASTOT\_HD ya existente en esa última base. **Nota:** considere iguales valores con diferencias decimales, estas diferencias se pueden deber a las formas de cálculo o software. |

**SECCIÓN III**

A continuación hará uso de la misma base utilizada durante la sección II, “BASE\_PERSONAS”. Con ella se espera que genere un modelo que permita predecir la probabilidad de que las personas presentes en la VIII EPF se hayan identificado como Administrador o administradora del hogar.

El/la administrador/a es quien habitualmente realiza las compras del hogar, pudiendo haber más de uno/a en cada hogar. La variable que muestra esta característica en la base de datos es ECOMPRAS, donde 1 refiere a que se identifica como administrador del hogar.

|  |  |
| --- | --- |
| **PREGUNTAS** | **III.1** Indique qué tipo de modelo estadístico utilizaría para abordar este problema y justifique brevemente su respuesta. Responda directamente en el *script* o *do-file* (argumente conceptual/teóricamente).  **III.2** Utilizando R o STATA, con la BASE\_PERSONAS, programe un modelo que permita estimar la probabilidad de que una persona se identifique como Administrador/a del hogar (ECOMPRAS), en base a las variables: ingreso disponible del trabajo dependiente (INGDTD\_HD) y EDAD. Solo para personas en edad de trabajar (15 años o más).  **III.3** En base al modelo construido, prediga:   * la probabilidad de que una persona de 20 años de edad, que cuenta con $2.000.000 en ingreso disponible del trabajo dependiente se identifique como Administrador del hogar. * la probabilidad de que una persona de 20 años de edad, que cuenta con $300.000 en ingreso disponible del trabajo dependiente se identifique como Administrador del hogar. * la probabilidad de que una persona de 55 años de edad, que cuenta con $2.000.000 en ingreso disponible del trabajo dependiente se identifique como Administradora del hogar. * la probabilidad de que una persona de 55 años de edad, que cuenta con $300.000 en ingreso disponible del trabajo dependiente se identifique como Administradora del hogar.   **III.4** Interprete conjuntamente los resultados obtenidos en las cuatro predicciones previamente solicitadas (en el punto III.3). Responda directamente en el *script* o *do-file*. |