Fundamentos de Ciencias de Datos

Segundo semestre 2022

Proyecto 1: El Impacto de la Pandemia en el Empleo en Chile

# Introducción

La Encuesta Nacional de Empleo (ENE) es un instrumento que permite conocer la situación de trabajo que tienen las personas que viven en Chile. Esta permite responder preguntas cómo: ¿cuántas personas están trabajando actualmente?, ¿cuántas están desocupadas?, etc. El resultado principal de la encuesta es la tasa de desocupación del país.

Esta encuesta la aplica el Instituto Nacional de Estadísticas (INE) y se levanta durante todo el año, abarcando todas las regiones del país. Los encuestadores del INE visitan una cantidad determinada de hogares, en donde entrevistan a cada uno de los miembros del hogar que tienen 15 años y más. Se entrevistan aproximadamente 12.000 viviendas mensuales y se capturan datos de alrededor de 35.000 personas en todo el país[[1]](#footnote-2).

# Objetivo

El objetivo de este proyecto es, a través del uso de los datos resultantes de la ENE, estudiar el mercado del trabajo en Chile y cómo ha sido afectado por la pandemia de COVID-19.

# Descripción

Para la realización de este trabajo se cuenta con los siguientes archivos: nene1.csv y nene2.csv.

Actividades a realizar

**Primera parte**

1. Realice la unión de los archivos nene1.csv y nene2.csv utilizando la variable **id**.
2. Realice la correspondiente limpieza de datos: considere valores nulos, duplicados, transformación de variables a valores numéricos de ser necesario, etc. Cada paso que desarrolle debe quedar correctamente documentado: lo que se obtiene y los criterios utilizados. ¿Qué diferencias tiene la nueva base obtenida con la anterior?
3. Descripción de los datos. Esta descripción debe ser por variable ya sea desde las métricas que se obtienen del dataframe y gráficos. Además. indique en cada uno de los casos el tipo de variable, esto es cualitativa: nominal u ordinal, cuantitativa: discreta o continua.

**Segunda parte**

Realice un análisis que permita determinar si existen diferencias en la fuerza laboral antes de la pandemia y después de la pandemia por año, trimestre y mes, género (sexo), edad, tipo de trabajo, ocupación formal versus informal. Para ello responda las siguientes preguntas.

Sea explícito en su desarrollo y análisis de los resultados.

1. ¿Qué género (hombre y mujer) fue el más afectado por la pandemia?
2. ¿Ha existido mayor precariedad en los empleos después de la pandemia?, considere precariedad con CAE\_ESPECIFICO 3 y 4.
3. ¿Existe algún nivel de variación después de la pandemia para quienes tienen estudios secundarios? ¿Qué se podría decir para cada nivel de estudio y la variación del trabajo formal?
4. ¿Hay alguna actividad económica que haya sufrido más con la pandemia en relación al empleo? ¿Se puede asociar dicha actividad con elementos de presencialidad?

Sea explícito en su desarrollo y análisis de los resultados.

Finalmente, en no más de una plana, haga un resumen de los principales hallazgos.

# Entregables

* Una presentación **grupal** en formato PowerPoint o Google Slides.
* Un archivo ipynb **grupal**.

La semana del 12 de septiembre en clase se realizarán sesiones de coaching en torno al proyecto, esto es, los grupos trabajarán en el desarrollo de este pudiendo resolver dudas con el profesor/a.

La semana del 3 de octubre los grupos deberán presentar sus hallazgos en no más de 10 minutos. Las presentaciones se harán vía zoom y se escogerá al azar a cualquier integrante del equipo para estos efectos.

# Evaluación de Pares

Nota entre 1 y 100 que cada alumno asigna a sus compañeros de grupos de acuerdo al aporte realizado por el mismo en el trabajo. El promedio de estas calificaciones será el factor por el cual se multiplica la nota obtenida por el grupo. Si un alumno no realiza esta evaluación será calificado en el proyecto con la nota mínima.

# Anexo. Descripción de variables

Encuesta a nivel nacional (todas las regiones), los datos contienen 808.947 observaciones o filas y 12 variables, sin contar con el ID de cada fila o registro.

La data corresponde a mediciones trimestrales desde al 2019 al 2021, donde los trimestres corresponden a: Enero-Febrero-Marzo; Abril-Mayo-Junio; Julio-Agosto-Septiembre; y Octubre-Noviembre- Diciembre. La variable mes\_central corresponde al número o nombre del mes medio de cada trimestre, así para el primer trimestre Enero-Febrero-Marzo el valor es 2 o Febrero.

|  |  |
| --- | --- |
| **Variable** | **Descripción** |
| ID | Índice de los datos, no representa nada más |
| ano\_trimestre | Año del trimestre, en este caso 2019, 2020 y 2021 |
| ano\_encuesta | Año de encuesta, en este caso 2019, 2020 y 2021 |
| mes\_central | Mes central del trimestre, por ejemplo primer trimestre Enero, Febrero y Marzo, el mes central es Febrero o el mes 2 |
| mes\_encuesta | Mes de la entrevista |
| Región | Región del encuestado (1 a 16 o por el nombre) |
| Sexo | Sexo, hombre o mujer, o 1 y 2, donde 1 representa hombre y 2 mujer |
| Edad | Edad del encuestado al momento de aplicar la encuesta |
| cine | Categoría laboral (CISE) – ver notas |
| cae\_especifico | Código sumario de empleo (específico) |
| ocup\_form | 0, 1 y 2. 0 representa menor de 15 años, 1 representa ocupación formal, 2 ocupación informal |
| b14\_rev4cl\_caenes | B14. Rama de Actividad Económica de la Empresa donde Trabaja, CIIU Revisión 4.cl – 1 Dígito, según CAENES – ver notas |
| fact\_cal | Factor de expansión |

Las variables consideradas en la BBDD PODRIAN ser clasificadas de acuerdo a su contenido.

**Período de Tiempo**

1. **ano\_trimestre:** Año del trimestre móvil al que pertenece la base de datos. Desde 2010 en adelante
2. **ano\_encuesta**: Año que se realizó la encuesta Desde 2010 en adelante
3. **mes\_central:** Número que identifica el mes central del trimestre

| **Número** | **mes\_central** |
| --- | --- |
| 1 | Diciembre – Febrero |
| 2 | Enero – Marzo |
| 3 | Febrero – Abril |
| 4 | Marzo – Mayo |
| 5 | Abril - Junio |
| 6 | Mayo - Julio |
| 7 | Junio - Agosto |
| 8 | Julio - Septiembre |
| 9 | Agosto – Octubre |
| 10 | Septiembre - Noviembre |
| 11 | Octubre - Diciembre |
| 12 | Noviembre - Enero |

1. **Mes\_encuesta:** Mes que se realizó la encuesta

| **Número** | **mes\_encuesta** |
| --- | --- |
| 1 | Enero |
| 2 | Febrero |
| 3 | Marzo |
| 4 | Abril |
| 5 | Mayo |
| 6 | Junio |
| 7 | Julio |
| 8 | Agosto |
| 9 | Septiembre |
| 10 | Octubre |
| 11 | Noviembre |
| 12 | Diciembre |

**Características de la fuerza laboral**

1. **Región**: Identifica la región donde vive el encuestado, con 16 categorías.

| **Número** | **Regiones** | **Capital** |
| --- | --- | --- |
| 15 | Región de Arica y Parinacota | Arica |
| 1 | Región de Tarapacá | Iquique |
| 2 | Región de Antofagasta | Antofagasta |
| 3 | Región de Atacama | Copiapó |
| 4 | Región de Coquimbo | La Serena |
| 5 | Región de Valparaíso | Valparaíso |
| 13 | Región Metropolitana de Santiago | Santiago |
| 6 | Región del Libertador General Bernardo O'Higgins | Rancagua |
| 7 | Región del Maule | Talca |
| 8 | Región del Biobío | Concepción |
| 9 | Región de La Araucanía | Temuco |
| 14 | Región de Los Ríos | Valdivia |
| 10 | Región de Los Lagos | Puerto Montt |
| 11 | Región de Aysén del General Carlos Ibáñez del Campo | Coyhaique |
| 12 | Región de Magallanes y de la Antártica Chilena | Punta Arenas |
| 16 | Región de Ñuble | Chillán |

1. **Sexo**: Señala el sexo de cada una de las personas encuestadas

| **Valor** | **Descripción** |
| --- | --- |
| 1 | Hombre |
| 2 | Mujer |

1. **Edad:** Señala la edad de la persona, al momento de aplicar la encuesta

**Características de la fuerza laboral**

1. **Cine:** Clasificación Internacional de Nivel Educacional (CINE) 1997

| **Valor** | **Descripción** |
| --- | --- |
| 1 | Nunca estudió |
| 2 | Educación preescolar |
| 3 | Educación primaria (nivel 1) |
| 4 | Educación primaria (nivel 2) |
| 5 | Educación secundaria |
| 6 | Educación Técnica (Educación Superior no Universitaria) |
| 7 | Educación universitaria |
| 8 | Post títulos y maestría |
| 9 | Doctorado |
| 999 | Nivel ignorado |

1. **Cae\_especifico:** Condición de Actividad Económica (CAE) Específica. Corresponde a la clasificación de todas las personas de 15 años y más, según su situación laboral, basado en un código sumario empleo a 29 categorías.

Fuerza de trabajo (FT) Aquellos encuetados cuya variable cae especifico toma valores entre 1 y 9.

| **Valor** | **Descripción** |
| --- | --- |
| 1 | Ocupado tradicional |
| 2 | Ocupado sin remuneración tradicional |
| 3 | Ocupado no tradicional |
| 4 | Ocupado sin remuneración no tradicional |
| 5 | Ocupado ausente con vínculo efectivo |
| 6 | Ocupado ausente con pronto retorno |
| 7 | Ocupado ausente con sueldo o ganancia |
| 8 | Cesante |
| 9 | Busca trabajo por primera vez |

1. **Ocup\_form:** corresponde a una variable que clasifica a la población ocupada según informalidad en su ocupación
2. **B14\_rev4cl\_caenes:** Rama de actividad económica de la empresa donde trabaja la persona

| **Código** | **Sección** |
| --- | --- |
| 1 | A Agricultura, ganadería, silvicultura y pesca |
| 2 | B Explotación de minas y canteras |
| 3 | C Industrias manufactureras |
| 4 | D Suministro de electricidad, gas, vapor y aire acondicionado |
| 5 | E Suministro de agua; evacuación de aguas residuales, gestión de desechos y descontaminación |
| 6 | F Construcción |
| 7 | G Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas |
| 8 | H Transporte y almacenamiento |
| 9 | I Actividades de alojamiento y de servicio de comidas |
| 10 | J Información y comunicaciones |
| 11 | K Actividades financieras y de seguros |
| 12 | L Actividades inmobiliarias |
| 13 | M Actividades profesionales, científicas y técnicas |
| 14 | N Actividades de servicios administrativos y de apoyo |
| 15 | O Administración pública y defensa; planes de seguridad social de afiliación obligatoria |
| 16 | P Enseñanza |
| 17 | Q Actividades de atención de la salud humana y de asistencia social |
| 18 | R Actividades artísticas, de entretenimiento y recreativas |
| 19 | S Otras actividades de servicios |
| 20 | T Actividades de los hogres como empleadores; actividades no diferenciadas de los hogares como productores de bienes y servicios para uso propio |
| 21 | U Actividades de organizaciones y órganos extraterritoriales |

1. **fat\_cal**: Factor de expansión trimestral calibrado.

El factor de expansión se interpreta como la cantidad de personas en la población, que representa una persona en la muestra.

La estimación de un total dado para una variable se obtiene, primero, ponderando el valor de la variable en cada persona por su factor de expansión y luego, sumando todas las personas de la muestra.

De acuerdo con el diseño de la muestra, corresponde aplicar este factor de expansión a cada hogar y persona seleccionada, lo que depende del número de viviendas que tiene la sección geográfica y el número de secciones que tiene el estrato. Los factores de expansión incluyen un ajuste de población, según las proyecciones del INE a la fecha de la encuesta, para de esta manera aumentar la precisión de las estimaciones

1. Fuente: INE (https://www.ine.cl/ine-ciudadano/definiciones-estadisticas/economia/encuesta-nacional-del-empleo) [↑](#footnote-ref-2)