

## ESTADÍSTICA

### Programa del Curso

#### I. Identificación

<b>Código</b>	
<b>Créditos</b>	5
<b>Duración</b>	Semestral
<b>Ubicación en el plan de estudios</b>	Semestre 3
<b>Requisitos</b>	
<b>Sesiones cátedra semanales</b>	2 cátedras, 1 ayudantía
<b>Profesores</b>	

#### II. Descripción del curso

La asignatura Estadística I, es el primer curso estadístico de la carrera de Ingeniería Comercial y entrega las herramientas necesarias para realizar análisis descriptivos de datos, incluyendo formas simples de modelar relaciones entre variables, con la intención de facilitar e iluminar la toma de decisiones económicas o de negocios. Al mismo tiempo, entrega los fundamentos básicos de la teoría de probabilidades para la modelación de fenómenos naturales con base probabilística. Se espera que los estudiantes sean capaces de sistematizar datos obtenidos de la realidad circundante y elaborar información o hipótesis que expliquen el comportamiento de dicha realidad.

Esta asignatura aspira a enseñar estadística de forma aplicada, haciendo uso de herramientas modernas de programación y situando al estudiante en un rol de analista dentro de una unidad organizacional que necesita comprender la realidad mediante el examen de datos para facilitar la toma de decisiones.

Una característica de este curso es el uso temprano y continuado de diversas bases de datos, el análisis y comunicación de resultados usando herramientas de programación, que ocasionalmente serán contrastados con su contraparte más cercana en planillas electrónicas, para ilustrar y profundizar los conceptos estadísticos. Esto requiere necesariamente de muchas horas de práctica fuera del horario de clases, pues, aunque no es primera vez que usan ese lenguaje de programación ni planillas electrónicas (eso pertenece al curso de Programación, anterior a Estadística I) de todos modos el leer la documentación, corregir errores en los programas y mejorar los análisis, rebasa con creces el tiempo disponible en la cátedra y los ritmos de avance suelen ser dispares entre estudiantes.

En específico, el programa aporta al perfil de egreso del estudiante en los siguientes aspectos:

- i. Desarrollo de competencias analíticas y de resolución de problemas.
- ii. Desarrollar competencias tecnológicas: programación, ciencia de datos, TI.

Aunque este curso sienta las bases de la inferencia al explicar el muestreo simple y las distribuciones muestrales (vía simulación computacional) y dos funciones de probabilidad fundamentales (normal y binomial); las técnicas de estimación, en cambio, junto a las propiedades de los estimadores y la teoría de los tests de hipótesis y varias otras distribuciones importantes, serán vistos en el curso siguiente de esta línea, Estadística II.

En comparación con Estadística II, en Estadística I un menor énfasis en teoremas analíticos respecto de las propiedades de los estimadores o de las distribuciones de probabilidad, descansando casi completamente en el análisis de simulaciones.

### **III. Relación con otros cursos de la línea**

Este curso toma estudiantes que aprendieron en la asignatura de Programación aspectos básicos de un lenguaje de programación y aspectos intermedios de planillas electrónicas (R y Microsoft Excel, respectivamente, en el caso de Programación durante el

2020) y continúa este desarrollo por medio del aprendizaje de flujos de trabajo reproducibles y letrados, pero centrándose en resúmenes estadísticos de variables y relaciones entre variables. Este curso da continuidad a ciertas habilidades introducidas en Programación: tratamiento de datos, visualización, creación de funciones y construcción de informes cuantitativos. Las actividades más importantes donde aplicarán sus habilidades programáticas, son la construcción de análisis reproducibles y la construcción de simulaciones para descubrir una distribución muestral

Sirve además, hacia arriba de la línea, al entregar al estudiante experiencia suficiente para entender la utilidad, pero también las limitaciones de razonar dentro de los límites de la muestra disponible y la necesidad de generalizar más allá de la muestra, introduciendo (vía simulación) la noción de distribución muestral (sampling distribution) de los diversos estimadores vistos en esta asignatura. Este es el puente natural con la asignatura de Estadística II, donde examinará la teoría y práctica de inferencia estadística, incluyendo varios tipos de estimadores y sus propiedades, así como varios tipos de distribución de probabilidad que aparecen en los resultados asintóticos de inferencia estadística. Por ejemplo, en Estadística II volverán a ver las técnicas de regresión lineal simple y múltiple, pero esta vez bajo el lente de inferencia, introduciendo las nociones de test de significancia y especificación, fijando un piso adecuado para que en el curso siguiente, Econometría, discutan la especificación y selección de modelos.

En comparación con Estadística II, en Estadística I un menor énfasis en teoremas analíticos respecto de las propiedades de los estimadores o de las distribuciones de probabilidad, descansando casi completamente en el análisis de simulaciones y tiene más énfasis en la parte exploratoria del análisis estadístico.

#### **IV. Objetivos de aprendizaje**

En esta perspectiva los objetivos generales que se persiguen son:

- Utilizar las técnicas básicas de la estadística descriptiva establecer valores y relaciones empíricas entre variables, que aporten a la toma de decisiones.
- Realizar los análisis en forma reproducible y letrada (*literate*), es decir que sean transparentes para un tercero y que tanto la documentación, el medio elegido (por ejemplo, el software) como la organización del análisis haga factible y fácil reproducir el mismo análisis del autor

## V. Resultados del aprendizaje

Al final de este curso los estudiantes deberán poder

1. Resumir conjuntos de datos usando medidas estadísticas puntuales y de distribución
2. Distinguir y comparar distintas medidas estadísticas descriptivas, eligiendo la o las apropiadas para el tipo de datos y el tipo de pregunta que se quiere responder
3. Realizar estas operaciones por distintos grupos o categorías
4. Elegir y crear gráficos y tablas estadísticas apropiadas para el tipo de variable y la pregunta a la que responden
5. Narrar el procedimiento y los resultados usando rigurosamente el lenguaje de la estadística y de la programación
6. Estructurar y diseñar el análisis de forma tal que un tercero pueda reproducir el análisis de forma fácil y precisa (asegurar la reproducibilidad del trabajo realizado). En particular, utilizar un lenguaje de programación para escribir su análisis de manera reproducible y con la documentación necesaria.
7. Modelar relaciones lineales entre dos o más variables
8. Interpretar y calcular una distribución empírica de un estimador muestral
9. Seleccionar los procedimientos de descripción y modelación que ayuden a eliminar un problema de decisión particular

## VI. Contenidos

**Listado (ver tabla con detalles, más abajo)**

- 1) Estadística Descriptiva y visualización estadística
  - Conceptos Básicos y Datos estadísticos
  - Medidas de localización dispersión y distribución
  - Gráficos estadísticos más usados
- 2) Modelamiento de datos con regresión lineal
  - Simple

- Múltiple
  - Correlación versus causalidad
  - Paradoja de Simpson
- 3) Variables aleatorias y probabilidad
- Experimentos, resultados y variables aleatorias
  - Probabilidad y distribución de probabilidad
  - Distribución de probabilidad binomial
  - Distribución de probabilidad normal
- 4) Estadísticas muestrales como variables aleatorias
- Muestreo simple de un conjunto finito
  - Muestreo simple desde una población infinita: remuestreo
  - Muestreo simple de una población infinita: normal y binomial

Unidad	Contenidos	Referencias	Resultados de aprendizaje	Actividades
<b>I. Estadística descriptiva y visualización estadística</b>	<p>Variables cuantitativas, cualitativas, discretas y continuas, nominales y ordinales</p> <p>Clasificación de Datos</p> <p>Medidas de resumen más usadas</p> <p>Gráficos estadísticos: de dispersión, de barras,</p>	<p><b>Obligatorias:</b></p> <p>Cap 2 y Cap 3 de EAE</p> <p><a href="#">Cap 2 de IMS</a></p> <p><a href="#">Cap 2 de DIVE</a></p> <p><a href="#">“descriptive statistics and graphics”</a></p>	<p>Identifica variables cuantitativas, cualitativas, ordinales, nominales, discretas o continuas</p> <p>Calcula y usa adecuadamente medidas numéricas de localización y variabilidad para caracterizar y comparar variables</p> <p>Calcula y usa adecuadamente medidas de distribución para caracterizar y comparar distribuciones de variables</p>	<p><b>En clases:</b> usando <i>penguins</i> y <i>diamonds</i>:</p> <ul style="list-style-type: none"> <li>- Establecer el tipo estadístico de cada variable (cuanti, cuali, nominal, ordinal, etc)</li> <li>- Producir resúmenes estadísticos numéricos con funciones individuales (como mean) o funciones que invocan funciones (como summary)</li> <li>- Construir una tabla con el resumen</li> <li>- Construir un informe con la tabla y la interpretación de la misma</li> </ul>

	<p>histogramas, de líneas, de caja</p>	<p><b>Complementarias:</b></p> <p><a href="#">Cap 6</a>, <a href="#">7</a>, <a href="#">8</a>, <a href="#">9</a>, <a href="#">10</a> y <a href="#">11</a> de DS</p>	<p>Selecciona el tipo de gráfico adecuado para el tipo de variable</p> <p>Incluye información útil en los gráficos (títulos, leyendas, anotaciones, etiquetas de los ejes) para que un tercero pueda entenderlo fácilmente</p> <p>Crea y usa la información provista en histograma y en un gráfico de caja para caracterizar la distribución de una variable (simetría, dispersión, outliers ...) y comparar distribuciones de distintos grupos o variables</p> <p>Crea y usa diagramas de dispersión para referirse a la presencia y tipo de asociación entre dos variables, incluyendo el uso de una tercera variable para segmentar los casos</p> <p>Crea y usa diagramas de barra para comparar frecuencias en una variable categórica, incluyendo su uso con dos o más grupos</p> <p>Crea y usa diagramas de línea para referirse a la presencia de patrones como ciclos o tendencias</p>	<p><b>En clases:</b> usando <i>penguins</i> y <i>diamonds</i>:</p> <ul style="list-style-type: none"> <li>- Crear gráficos, bien presentados, de dispersión, barras, histogramas, caja, incluyendo comparación entre grupos para una variables o entre distintas variables</li> <li>- Agregar al reporte anterior e interpretar</li> <li>- Usando <i>gapminder</i>, crear y presentar gráficos de líneas. Discutir presencia de patrones temporales como ciclos y tendencias.</li> </ul> <p><b>Fuera de clases</b>, pero para exponer en clases:</p> <ul style="list-style-type: none"> <li>- Crear un reporte reproducible como los anteriores, pero para dos bases de datos de Fundamentos de Administración y Negocios (FAN)</li> <li>- Buscar y explorar paquetes especializados en EDA, como <i>skimr</i>, y producir reportes similares a los anteriores. Discutir pros y contras.</li> </ul> <p><b>Ejercicios relevantes en la bibliografía:</b></p> <ul style="list-style-type: none"> <li>- DIVE: LC2.1 a LC2.7 (scatter plot) , LC2.8 a LC2.13 (de líneas), LC2.14 a LC2.17 histogramas), LC2.18 a LC2.21 (facectas)</li> <li>- IMS: práctica autónoma usando R, los siguientes Labs:</li> </ul>
--	--	---	--	--

			<p>Al explicar las relaciones y cantidades estadísticas, usa el contexto y la información respectiva del caso.</p> <p>Es capaz de producir un reporte <b>reproducible</b> sobre un conjunto de datos, en donde caracterice y compare variables, usando todos los elementos antes mencionados que sean relevantes para ese conjunto de datos.</p>	<ul style="list-style-type: none"> <li>○ <a href="#">Lab 01</a> (intro a exploración, reportes y primer ggplot)</li> <li>○ <a href="#">Lab 02</a> (scatterplot y resumen de datos)</li> <li>○ <a href="#">Tutorial 02</a> (gráficos estadísticos y resúmenes numéricos)</li> <li>- IMS: ejercicios manuales: <ul style="list-style-type: none"> <li>○ 2.1.9 Exercises (en particular 1, 2, 4, 8, 10, 12, 13, 14, 16 y 17)</li> <li>○ 2.2.7 Exercises (en particular 2,3 y 4)</li> </ul> </li> </ul>
<b>II. Modelación de datos con regresión lineal</b>	<p>Una variable explicativa numérica</p> <p>Una variable explicativa categórica</p> <p>Correlación versus causalidad</p> <p>Recta de mejor ajuste</p> <p>una categórica explicativa y una numérica explicativa</p> <p>dos numéricas explicativas</p> <p>selección de modelos</p>	<p><b>Obligatorias:</b></p> <p><a href="#">Cap 5 de DIVE</a></p> <p><a href="#">Cap 6 de DIVE</a></p> <p><a href="#">Cap 3 de IMS</a></p> <p><a href="#">Cap 4 de IMS</a> (sólo secciones 4.1, 4.2 y 4.3)</p> <p><b>Complementarias:</b></p> <p><a href="#">Cap 17 de DS</a> (sólo secciones 17.1 y 17.2)</p>	<p>Puede estimar modelos de regresión lineal con variables explicativas numéricas o categóricas o una combinación de ella</p> <p>Puede interpretar las correlaciones en modelos con una variable explicativa y correlaciones parciales en modelos con más de una variable explicativa</p> <p>Puede graficar datos y regresión en modelos con una variable explicativa numérica</p> <p>Puede graficar datos y regresión en modelos con una variable explicativa numérica y una variable explicativa categórica. Explicar la paradoja de Simpson en este contexto.</p>	<p><b>En clases:</b> usando penguins y un data set económico</p> <ul style="list-style-type: none"> <li>- Estimar la correlación entre dos variables numéricas</li> <li>- Estimar una regresión lineal simple con las mismas variables y comparar coeficiente de regresión con el de correlación</li> <li>- Examinar sus residuos.</li> <li>- Estimar, interpretar y graficar una regresión de dos variables explicativas: una numérica y una categórica, aditiva</li> <li>- Estimar un modelo con la interacción de una numérica y categórica. Interpretar sus coeficientes.</li> </ul>

	<p>coeficiente de correlación</p> <p>paradoja de Simpson</p>		<p>Puede calcular, graficar e interpretar los residuos de una regresión.</p> <p>Puede comparar dos modelos de regresión para la misma variable dependiente.</p>	<ul style="list-style-type: none"> <li>- Estimar un modelos con dos explicativas numéricas. Interpretar sus coeficientes.</li> </ul> <p><b>Fuera de clases:</b></p> <ul style="list-style-type: none"> <li>- Crear un reporte reproducible donde usando un mismo dataset estimen grafiquen e interpreten todos los tipos de regresiones que aparecen en los objetivos de aprendizaje a medida que se expliquen en clases.</li> </ul> <p><b>Ejercicios relevantes en la bibliografía:</b></p> <p><a href="#">Lab 08</a> (regresión lineal simple)</p> <p><a href="#">Tutorial 03</a> (regresión lineal simple)</p> <p><a href="#">Sección 17.5 Exercises de DS</a> (regresión lineal simple)</p> <p>Learning Checks LC5.1 a LC5.8</p> <p>Learning Checks LC6.1 a LC6.3</p>
<b>III. Variables aleatorias y probabilidad</b>	Experimentos, resultados y variables aleatorias	<p>Obligatoria:</p> <p><a href="#">Cap 13 de DS</a></p> <p><a href="#">Cap 14 de DS</a></p>	Puede aplicar las reglas de la probabilidad para calcular probabilidades condicionales, marginales y conjuntas	<p><b>En clases:</b></p> <ul style="list-style-type: none"> <li>- Revisión de las leyes de probabilidad culminando en el teorema de bayes para un numero finito de eventos discretos</li> </ul>



	<p>Probabilidad y distribución de probabilidad</p> <p>Distribución de probabilidad binomial</p> <p>Distribución de probabilidad normal</p>	<p><a href="#">Sección A.2 de DIVE</a></p> <p>Complementaria: Cap 4 de EAE</p> <p>Sección 6.2 de EAE</p> <p>Sección 5.4 de EAE</p> <p><a href="#">Sección 5.3.2 de IMS</a></p> <p><a href="#">Sección 8.8 de DS</a></p>	<p>Grafica y calcula varios conceptos relacionados con estas dos distribuciones: media, cuantiles, probabilidad, densidad. Puede explicar el efecto de variar los parámetros de estas distribuciones.</p> <p>Puede explicar en términos probabilísticos un rango de valores entre dos cuantiles.</p>	<ul style="list-style-type: none"> <li>- Comprar distribuciones conjuntas, marginales y condicionales en un dataset</li> <li>- Presentación de la distribución binomial y sus propiedades</li> <li>- Presentación de la distribución normal y sus propiedades</li> <li>- Comparación del valor esperado y la media muestral a medida que aumenta el tamaño muestral</li> <li>- Cálculo de probabilidades acumuladas, cuantiles, densidades, probabilidades en ambas distribuciones</li> <li>- Explorar efecto del cambio en los parámetros en las probabilidades</li> </ul> <p><b>Fuera de clases:</b></p> <ul style="list-style-type: none"> <li>- Elaborar un informe reproducible en donde: <ul style="list-style-type: none"> <li>o presente gráficamente la distribución binomial usando varios valores de <math>n</math> y de <math>p</math></li> <li>o presente gráficamente la distribución normal usando varios valores de <math>\mu</math> y <math>\sigma</math></li> <li>o Calcula la probabilidad entre dos números para una distribución normal bajo valores distintos de</li> </ul> </li> </ul>
--	--	---	--	---

				<p>mu y sigma y haga el gráfico correspondiente.</p> <ul style="list-style-type: none"> <li>○ Escriba un ejemplo de uso de Bayes con eventos binarios y donde la probabilidad a priori sea un parámetro arbitrario</li> </ul> <p><b>Ejercicios relevantes en la bibliografía:</b></p> <p><a href="#">Sección 13.9 Exercises de DS</a></p> <p><a href="#">Sección 13.14 Exercises de DS</a></p> <p><a href="#">Sección 14.10 Exercises de DS</a></p> <p><a href="#">Sección 14.12 Exercises de DS</a></p> <p><a href="#">Lab 03</a> (probabilidad)</p> <p><a href="#">Lab 04</a> (distribución normal)</p> <p>LCA.1 y LCA.2 de <a href="#">Sección A.2 de DIVE</a></p>
<b>IV. Estadísticas muestrales como variables aleatorias</b>	<p>Muestreo simple de un conjunto finito</p> <p>Muestreo simple desde una población infinita: remuestreo</p>	<p><a href="#">Cap 7 de DIVE</a></p> <p>Secciones <a href="#">8.1</a> y <a href="#">8.2</a> de DIVE</p> <p>Secciones <a href="#">15.1</a>, <a href="#">15.2</a> y <a href="#">15.3</a> de DS</p>	<p>Puede explicar en qué consiste el muestreo aleatorio simple y por qué es importante</p> <p>Puede realizar un m.a.s. desde una población finita o desde una población infinita normal o binomial</p>	<p><b>En clases:</b></p> <ul style="list-style-type: none"> <li>- Muestreo desde una urna virtual simulada en R</li> <li>- Graficar sampling distributions de la proporción de bolas y discutir sus características e introducir el</li> </ul>

	<p>Muestreo simple de una población infinita: normal y binomial</p> <p>Distribución muestral y error estándar mediante simulación de muestras de una urna</p>		<p>Puede generar un número arbitrario de muestras utilizando técnicas de simulación remuestreo (resampling) de conjuntos finitos o muestras generadas a partir de una población infinita</p> <p>Obtiene a partir de muestras así obtenidas un vector de medidas estadísticas de las muestras (medias, medianas, varianzas etc.) y puede graficar la distribución de estos estadísticos.</p>	<p>concepto de error estándar como un tipo de desv estándar</p> <ul style="list-style-type: none"> <li>- Simular una encuesta y analizar los estimadores puntuales y la variabilidad de los mismos</li> </ul> <p><b>Fuera de clases:</b></p> <ul style="list-style-type: none"> <li>- Crear un reporte reproducible donde ilustren y expliquen : <ul style="list-style-type: none"> <li>o la distribución de la proporción muestral de bolas en una urna al tomar repetidas muestras de la urna</li> <li>o lo mismo que el punto anterior pero con remuestreo con reemplazo de una única muestra original</li> </ul> </li> </ul> <p><b>Ejercicios relevantes en la bibliografía:</b></p> <p><a href="#">Lab 5.a de IMS (distribución muestral)</a></p> <p><a href="#">Tutorial 5, Lesson 1, de IMS</a></p> <p>Sección 1.5 Exercises de DS</p> <p>Learning Checks LC7.1 a 7.24 en DIVE</p>
--	---	--	---	---

## **VII. Método de enseñanza**

La asignatura se desarrollará fundamentalmente como talleres guiados, con mayor exposición de conceptos en el caso de las cátedras. El docente introducirá un concepto o un procedimiento y el objetivo a lograr en esa sesión, para luego dar paso al trabajo individual o grupal de los estudiantes. Al final de la sesión provee una solución completa o un esquema de la solución. Salvo excepciones, no hay sesiones puramente expositivas.

En el cronograma del curso, que puede sufrir modificaciones durante el semestre, estarán indicadas las actividades y lecturas correspondientes a cada semana. Las clases asumen que los estudiantes leyeron concentradamente el material una vez antes de la clase respectiva. No asume que dominan el contenido, pero sí que entienden de qué estamos hablando y cual es el contexto, así como la razón por la cual usaremos tal o cual herramienta o función

Semana por medio, tendremos una sesión completamente dedicada a resolver dudas tanto conceptuales como prácticas.

Si un estudiante no puede asistir a una sesión es su responsabilidad revisar los materiales de esa clase, incluyendo la grabación de la misma, si la hubiese.

## **VIII. Descripción general del método de evaluación**

En este curso coexisten las evaluaciones grupales e individuales. Las evaluaciones grupales fomentan el trabajo en equipo, la cooperación y el aprendizaje entre pares, pero las evaluaciones individuales son necesarias para observar el dominio de los contenidos que logró cada individuo, lo que es muy difícil de medir en evaluaciones grupales. Por esta razón hemos incluido ambos tipos, con pesos similares en la nota final del curso. En esta versión, un 55% de la nota final proviene de evaluaciones individuales

Los tipos de evaluaciones obligatorias son tres:

- **Controles online individuales**, con el fin de que obtengan retroalimentación inmediata sobre su comprensión de la materia (definiciones de conceptos, elección de herramientas, empleo de funciones estadísticas)
- **Una prueba solemne individual**, manual, donde demuestren su conocimiento de los métodos y su capacidad de explicar e interpretar resultados.
- **Un examen trabajo semestral grupal/individual**, que consiste en un informe realizado por grupos de dos o tres estudiantes, donde realicen un examen estadístico que permita mejorar la toma de una decisión en contextos económicos o de negocios. Aunque el informe es común, cada estudiante debe realizar una presentación individual para evaluar su dominio de la materia.

Sin perjuicio de lo anterior, durante el semestre les pediremos que expliquen a sus compañeros, durante la cátedra, aspectos de su trabajo grupal (que expongan y reflexiones sobre algún resultado específico) o sobre soluciones a problemas planteados por el profesor durante la clase. También se espera que comenten sobre la exposición que hagan otros compañeros en clase. Esta participación será tomada en cuenta a la hora de corregir su informe final.

Las evaluaciones y ponderaciones son las siguientes:

- **(10%) Promedio de controles.** Son 6 controles en línea (quizzes en Canvas), dos en cada tramo del semestre. Para calcular el promedio se elimina la peor nota. Estos quizzes tendrán lugar durante las cátedras, específicamente en los últimos 20 minutos de ese bloque horario.
  - **Control 1.** Contenido tentativo: tipos de datos y variables. Medidas numéricas de localización y dispersión (Unidad I).
  - **Control 2.** Contenido tentativo: Medidas numéricas de dispersión y visualizaciones de distribución (Unidad I)
  - **Control 3.** Contenido tentativo: Regresiones con una variable explicativa
  - **Control 4.** Contenido tentativo: Regresiones con dos variables explicativas.
  - **Control 5.** Contenido tentativo: Distribuciones normal, binomial y probabilidad
  - **Control 6.** Contenido tentativo: Distribuciones muestrales (sampling distribution)
- **(20%) Prueba Solemne.** Incluye al menos las dos primeras unidades, no les pediremos programar, pero sí demostrar su dominio de los métodos, las herramientas, los tipos de datos y demostrar su capacidad para explicar e interpretar resultados (números, tablas, gráficos, regresiones) estadísticos.

- **(20%) Participación en clases.** Durante el semestre, durante las clases, se le interrogará en tres ocasiones, donde deberán demostrar conocimiento teórico-práctico. La nota final de participación en clases será el promedio de estas tres interrogaciones.
- **(10%) Primera entrega parcial del trabajo:** elección de datos, caracterización de los tipos de variables y estadísticas descriptivas univariadas (globales y por categorías)
- **(15%) Segunda entrega parcial del trabajo:** relaciones entre variables (correlaciones, diagramas de dispersión, regresiones).
- **(25%) Entrega final del trabajo.** Todo lo anterior, integrado de forma orgánica, cuidando la presentación más la incorporación de distribuciones muestrales de indicadores importantes. Salida principal de informe en R, con subproductos para Excel.
  - La mitad de esta nota corresponde a la calificación del informe final y la nota será la misma para todos los integrantes del grupo
  - La otra mitad de la nota corresponde a una presentación individual del informe completo, donde se evaluará el dominio del contenido y los métodos utilizados. Esta nota será individual para cada miembro del grupo

Participación en clases: 20 de la nota final interrogaciones en clases, promedio de tres notas. Las notas posibles son 1, 3, 5 y 7

## **IX. Cronograma**

Calendario (clase-a-clase) a confeccionar, pero una división tentativa podría ser:

- Tramo 1 de 3 del semestre: Unidad I y entrega del primer informe parcial del trabajo
- Tramo 2 de 3 del semestre: Unidad II y entrega del segundo informe parcial del trabajo
- Tramos 3 de 3 del semestre: Unidad III y IV y entrega del informe final

## X. Bibliografía

- **Bibliografía obligatoria**

- (“DIVE”) **Statistical Inference via Data Science: A ModernDive into R and the Tidyverse**. Disponible en línea en <https://moderndive.com/>
- (“IMS”) **Introduction to Modern Statistics**. Disponible en línea en <https://openintro-ims.netlify.app/>
- (“DS”) **Introduction to Data Science, Data Analysis and Prediction Algorithms with R** (Primera edición). Disponible en línea en <https://rafalab.github.io/dsbook/>
- (“EAE”) **Estadística para la Administración y Economía**. Anderson & Sweeney (2008). Décima Edición. Cengage Learning Edition.

- **Bibliografía complementaria**

- (“R4DS”) **R para Ciencia de Datos**: <https://es.r4ds.hadley.nz/>
- (“GUR”) **Getting Used to R, RStudio, and R Markdown** (2019), Chester Ismay and Patrick C. Kennedy, disponible en <https://rbasics.netlify.app/>

## XI. Conocimientos y habilidades previas:

Este curso está pensado para estudiantes con un nivel básico de programación en un lenguaje de alto nivel y un nivel intermedio de planillas de cálculo. Esto quiere decir que el estudiante puede escribir y ejecutar programas sencillos, importar datos desde planillas electrónicas, usar funciones preexistentes, crear sus propias funciones, crear visualizaciones básicas y utilizar la documentación y la ayuda disponible. Para el caso de la planilla electrónica (spreadsheet) puede usar fórmulas condicionales, compuestas, crear visualizaciones y tablas dinámicas.

En esta primera edición hemos elegido como lenguaje de alto nivel a R y como planilla de cálculo a Microsoft Excel (y que corresponden a lo aprendido en el curso Programación I, el prerequisite formal para este curso)

En particular aquí listamos los contenidos que necesitan del curso de programación y dejamos material y actividades para que refresquen sus conocimientos al principio (o durante, si es necesario) del semestre

PRERREQUISITO	Contenidos	Referencias	Resultados de Aprendizaje	Actividades
R básico (del curso de Programación, semestre 1)  Fecha: antes del fin de la primera semana de clases	a) Instalación R, RStudio b) Importar archivos xlsx y csv c) Vectores, dataframes y tibbles d) Crear, guardar, abrir scripts e) Gráficos básicos con función plot f) Comentarios en scripts g) Uso de Help en R h) Funciones básicas de R	a) <a href="#">Cap 1 de DIVE</a> b) <a href="#">Cap 1 de DS</a> c) <a href="#">Cap 2 de DS</a>	1) Puede abrir y ejecutar código de un R script 2) Puede crear un nuevo R script y guardarlo 3) Puede abrir en R una base de datos desde xlsx o csv 4) Identifica filas, columnas, variables y observaciones 5) Puede indexar un vector índices o valores lógicos 6) Puede imprimir resultados en la consola 7) Puede instalar y cargar paquetes 8) Puede hacer gráficos básicos (un vector contra otro o una columna de un dataframe contra otra) 9) Usa la información provista en los archivos de ayuda en R 10) Usa correctamente funciones matemáticas básicas	a) Hacer los Learning Checks LC1.1 (re:instalación de paquetes), LC1.2 (re:carga de paquetes), LC1.3 (re:estructura de un dataframe), LC1.4 (re:tipos de variables), LC1.5 (re:variables en un dataframe), LC1.6 (re:creación de dataframes y tibbles), LC1.7 (re:uso de la ayuda en R) en <a href="#">Cap1 de DIVE</a>  b) Secciones 2.3 Exercises (re:funciones matemáticas básicas), 2.5 Exercises (re:examen de dataframes y vectores), 2. 8 Exercises (re:creación e indexación de vectores), 2.12 Exercises (indexación de vectores) y el primer ejercicio en 2.16 Exercises (re:plot de una columna contra otra) en <a href="#">Cap 2 de DS</a>
Tratamiento de datos (data wrangling)	El operador %>% (pipa)	<a href="#">Cap 3 de DIVE</a> (en especial las	Puede crear nuevas variables, seleccionar filas y columnas basadas en condiciones, agrupar observaciones de acuerdo a los	Hacer los Learning Checks del LC3.1 a LC3.12



	Verbos básicos del paquete dplyr	secciones 3.1 a 3.6)	valores de una variable, puede producir resúmenes numéricos de variables usando funciones	
<p>R intermedio (también del curso de Programación, semestre 1)</p> <p>Fecha: durante la segunda semana de clases</p>	<p>Expresiones condicionales</p> <p>Funciones de usuario</p> <p>Namespaces</p> <p>For loops</p> <p>Vectorización y funcionales</p>	1) <a href="#">Cap 3 de DS</a>	<p>1) Crea vectores y variables usando expresiones condicionales</p> <p>2) Crea funciones y las utiliza (user-defined functions)</p> <p>3) Comprende el concepto de namespace y entiende el uso del operador ::</p> <p>4) Crea y emplea for loops para realizar operaciones iterativas</p> <p>5) Puede usar funciones vectorizadoras como sapply o map (*útil, pero quizás no lo vieron en Programación)</p>	a) Sección 3.6 Exercises (re:loops, condiciones, namespaces, funciones de usuario, vectorizaciones) en <a href="#">Cap 3 de DS</a>

## ¿Qué hacer para refrescar su memoria (respecto de Programación)?

1. Lean los capítulos listados aquí como Referencias
  - a. Tip: si les cuesta leer en inglés, pueden aprovechar que los libros son documentos HTML (“páginas web”) para traducirlas al castellano. Por ejemplo, si están en el navegador Chrome, pueden hacer click-derecho en cualquier parte de la página y elegir la opción Traducir al Español. Por lo que he probado, la traducción es bastante aceptable. Consideren eso sí, la traducción como rueditas chicas de la bicicleta: hay muchos más recursos en inglés que en castellano y no todos son fácilmente traducibles. Eventualmente van a tener que dejar de depender de las traducciones.
  - b. Algunas cosas les parecerán más conocidas y de otras se acordarán menos. Es normal. Tomen nota de cuales son las cosas que entienden mejor y las que menos entienden. Les va ayudar a priorizar su estudio y a sacar el mejor provecho de sus interacciones con ayudantes y profesores.
2. Hagan las Actividades de forma autónoma
  - a. En el caso del libro “DIVE” (ver Bibliografía) los *Learning Check* vienen con respuestas al final del libro.
  - b. En el caso del libro “DS” los Exercises no tienen respuesta, pero están bien integrados con el material que los precede. Si después de leer bien el capítulo 2 no dan con una solución, comenten el ejercicio con sus compañeros (recuerden, en Canvas hay Foros y pueden postear preguntas y respuestas ahí). Es importante que ejercicios de este nivel los puedan resolver solos leyendo el material, solos buscando en Google o hablando con sus compañeros, en ese orden.