

**Estadística****PROYECTO (10 + 5 puntos)***Profesor: Pedro Ovalles**Enero - Marzo 2025***1. Datos**

Cada grupo recibirá un conjunto de datos que contiene 100 observaciones de 9 variables. Las observaciones o unidades muestrales corresponden a empresas, y las variables están relacionadas a variables económicas, éstas son:

“y” = ventas de la empresa durante el periodo de estudio, en miles de dólares.

“x1” = índice económico, que es una variable que estudia la salud de la economía en el periodo de estudio.

“x2” = número de canales de venta que utiliza la empresa.

“x3” = presupuesto para publicidad en televisión durante el periodo de estudio, en miles de dólares.

“x4” = número de promociones hechas por la empresa durante el periodo de estudio.

“x5” = presupuesto para publicidad en radio durante el periodo de estudio, en miles de dólares.

“x6” = estrato de la población a la que está dirigida predominantemente los productos de la empresa.

“x7” = presupuesto para publicidad en redes sociales durante el periodo de estudio, en miles de dólares.

“x8” = presupuesto para publicidad en medios impresos durante el periodo de estudio, en miles de dólares.

**2. Trabajo asignado**

1. Realizar un análisis descriptivo de los datos.
2. Utilice intervalos de confianza y/o pruebas de hipótesis para sacar conclusiones acerca de los presupuestos que asignan las empresas para televisión, radio, redes sociales y medios impresos.
3. Realizar un gráfico de dispersión y una matriz de correlación de las variables.
4. Realizar una prueba de bondad de ajuste para determinar si la variable “y” se distribuye en forma normal.
5. Encuentre el modelo de regresión simple que mejor se ajuste a los datos, donde “y” es la variable dependiente.
6. Halle un modelo lineal que explique mejor la variable “y”. Incluya todas las pruebas necesarias para llegar a este modelo, así como un análisis de residuos del modelo final.
7. Con los datos *datos\_prediccion.txt* haga una predicción de la variable “y” (con el mejor modelo) y haga un histograma, diagrama de cajas y resumen estadístico de los residuos de predicción (valor observado vs. predicción del modelo para concluir con relación al poder predictivo del modelo).
8. Realice un análisis de varianza para estudiar si existen diferencias entre las ventas media de las empresas de acuerdo al estrato.
9. En el caso de conseguir diferencias significativas, realice pruebas de medias para determinar cuales son las medias que difieren.

### 3. Asignación de Grupos

Grupo 1		Grupo 2	
1611299	Jorge Guzmán	2010265	Mauricio Fragachan
1510779	Mario Leyton	2010332	Jesús Gutiérrez
1611342	Andrea Zambrano	2010527	Rosa Ramírez
Grupo 3		Grupo 4	
2010349	Víctor Hernández	2010300	Aarón Girón
2110447	Brian Orta	1910216	Pedro Moreno
2010479	Ángel Pacheco	1710506	Samuel Quintana
Grupo 5		Grupo 6	
1910181	Leonardo Dolande	1910072	Jesús Bovea
1910211	Jesús Prieto	1810292	Gabriel Orejarena
1900036	Gabriel Seijas	1811093	Samuel Sánchez
Grupo 7		Grupo 8	
1410749	Carlos Ochoa	1810451	Carlo Herrera
1610882	José Pérez	1506637	Eribert Márquez
1611238	Blanyer Vielma		

### 4. Criterios de corrección para el informe

La estructura que debe tener el informe es:

- Portada con resumen (en la misma hoja).
- Planteamiento del problema (incluyendo los objetivos del trabajo), descripción de la base de datos y la metodología a emplear.
- Desarrollo (donde se realizan las asignaciones).
- Conclusiones y recomendaciones.
- Bibliografía.
- Anexos (+ códigos en R).

En la portada se debe encontrar el título del proyecto, el resumen y la identificación de los autores.

Una de las partes más importantes del informe es el resumen; en este se deben plantear los objetivos del proyecto y una breve descripción de la base de datos y de la metodología empleada. También se deben encontrar los resultados del proyecto (o por lo menos, los más substanciales), y se debe aclarar las implicaciones de estos resultados, las conclusiones y recomendaciones (simplificadas) que hace el analista.

El cuerpo principal del informe, debe comenzar con el planteamiento del problema, y luego describir la base de datos y la metodología que se empleará durante la resolución del mismo. Se deben usar tablas y gráficos para facilitar la lectura del informe y obtener la atención del cliente; las tablas y gráficos deben estar comentadas, no se permiten tablas o gráficos a las que no se hacen referencia. Debido a que el informe no debe tener más de diez (10) páginas (desde la portada a la bibliografía), se debe resumir la información en tablas o diagramas y se deben seleccionar los gráficos más relevantes.

En las conclusiones se presentan los resultados obtenidos conjuntamente con las implicaciones que tienen esos resultados (sin profundizar en terrenos del área en el que se desenvuelve el cliente, a menos de que se esté seguro del impacto de las implicaciones). Recuerde que este es un trabajo parecido al de asesoría y que el cliente es el que toma las decisiones, el analista sólo plantea alternativas y puede sugerir alguna de las soluciones al problema.

## 4.1. Presentación de resultados

- Presente sus resultados en tablas ordenadas e interprete.
- Identifique en los diagramas de caja si hay datos atípicos, cómo es la distribución de los datos, si es sesgada a la derecha, etc.
- Los gráficos tiene que tener su título y los nombres de los ejes (todo en español).

Es INACEPTABLE:

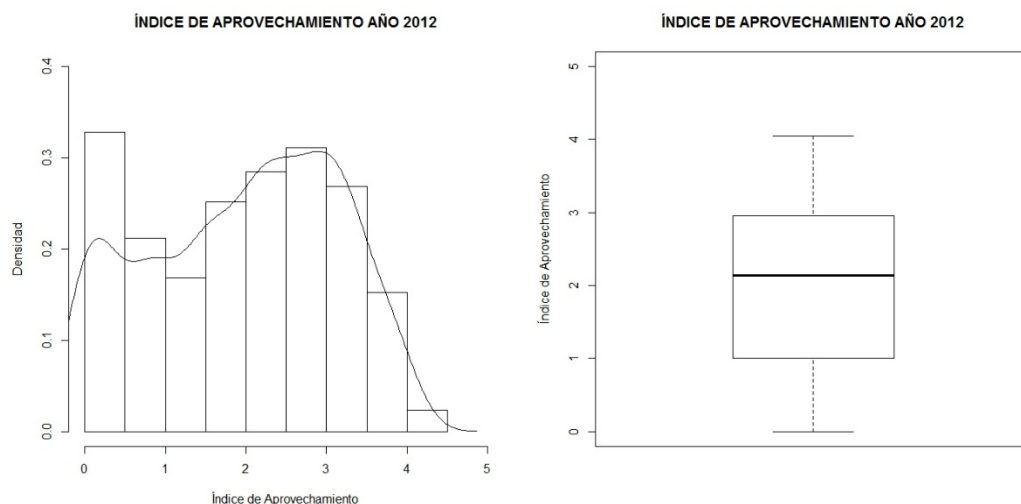
- No se aceptará presentación de resultados con manuscritos escaneados.
- Se anulará la evaluación de aquellos que compartan fotografías tomadas desde la pantalla de la computadora.
- No se aceptará un copy y paste de los resultados.
- No se aceptarán títulos de las gráficas generados por defecto en el programa.

## 4.2. Ejemplos

Por último se exponen unos ejemplos para la presentación de los resultados (Gráficas y Tablas), para mayor información se puede consultar las normas de la Universidad Simón Bolívar para la elaboración de trabajos.

**Tabla 1.** Resumen estadístico para la variable Índice de Aprovechamiento

Resumen Estadístico							
Variable	Mínimo	Primer	Mediana	Media	Tercer	Máximo	Desviación
IAP	0	1	2.14	1.97	2.95	4.05	1.16



**Gráfico 1.** Histograma y gráfico de caja para la variable Índice de aprovechamiento.

**NOTA:** recuerde que existen normas para la elaboración de trabajos propias de la USB, es recomendable revisar las mismas para la escritura del proyecto. Por ejemplo, es muy común cometer errores en la bibliografía. Recuerde que el autor debe ser mencionado en el texto, y posteriormente señalar la referencia en la bibliografía.

Ejemplo:

“Para Gelman y otros (2014), el muestreador de Gibbs es un método de gran utilidad en problemas donde el espacio de parámetros es multidimensional.”

“En este trabajo se aplicó el programa R Development Core Team (2015).”

“Según Gil, J. (s/f), los métodos...”

En la bibliografía

Gelman, A., Carlin, J., Stern, H. y Rubin, D. (2004). Bayesian data analysis. Second Edition. Chapman & Hall/ CRC.

Gil, J. (s/f). Modelos de medición: desarrollos actuales, supuestos, ventajas e inconvenientes. Universidad de Sevilla. [Revista en Línea]. Disponible: <http://innoevalua.us.es/files/irt.pdf> [Consulta: 2015, Diciembre, 09].

R Development Core Team (2015). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0, Disponible: <http://www.R-project.org>.

## 5. Condiciones de entrega

- a El informe final debe ser entregado en forma electrónica y en formato “.pdf”.
- b La entrega del informe final se realizará al correo electrónico **povallesgarcia@usb.ve** a más tardar el lunes 31 de marzo de 2025 a las 8:00 p.m. El asunto del correo DEBE ser: “*Proyecto. CO3321*”.
- c No se corregirán informes entregados fuera del tiempo establecido para la entrega.