



Universidad Simón Bolívar
Departamento de Cómputo Científico y Estadística
Septiembre - Diciembre 2019
Estadística para Ingenieros (CO3321)
Estadística para Matemáticos (CO3322)

Proyecto Final

Datos

Supóngase que el departamento de ventas de una empresa quiere estudiar la influencia que tiene la publicidad a través de distintos canales sobre el número de ventas de un producto. Se dispone de un conjunto de datos que contiene los ingresos (en millones) conseguido por ventas en 4 regiones, así como la cantidad de presupuesto, también en millones, destinado a anuncios en facebook, periódico, instagram, tv y radio. Los datos se presentan con el nombre **datosproy.txt**.

Trabajo asignado

Suponiendo que *ventas* es la variable dependiente:

1. Realice un análisis descriptivo y exploratorio de los datos. Incluya en este análisis la matriz de correlación.
2. Calcular el intervalo de confianza del 95% para las medias de ventas por región. Discuta los resultados.
3. Encuentre el modelo de regresión simple que mejor se ajuste a los datos; realice las pruebas estadísticas que considere conveniente para justificar su respuesta, incluyendo un análisis de residuales.
4. Consiga el **modelo múltiple** más apropiado. Realice, como en el inciso 3, todas las pruebas estadísticas que considere conveniente para justificar su respuesta, incluyendo un análisis de residuos. Considere un nivel del 5%.
5. Estudios previos indican que las ventas en la región 1 muestran un precio de 150 (millones), aunque estudios suponen que dicha cantidad es superior a la mostrada por este análisis. Con un nivel de confianza que usted considere necesario, realice un código en el software estadístico **R** que muestre el resultado de dicho análisis. Analice los resultados y concluya.
6. Para el modelo de regresión lineal simple obtenido en el inciso 3, realice la predicción correspondiente a 5 ventas que se anexan a la muestra, los datos se presentan en el Cuadro 1. Grafique los intervalos de predicción y de confianza respectivamente. Realice el análisis respectivo.
7. ¿Existe suficiente evidencia que permita concluir que las ventas media de las variables de estudio difieren con respecto a las regiones? Use el procedimiento de análisis de varianza para un diseño de un factor. ¿Qué concluiría usted con un nivel de significancia $\alpha = 0.03$?
Sugerencia: Incluya un boxplot en el análisis.
8. ¿Cómo compararía los resultados del intervalo de confianza del inciso 3 con el ANOVA?

NOTA: Realice transformaciones en las variables si lo considera necesario.

Variable	ventas 1	ventas 2	ventas 3	ventas 4	ventas 5
ventas					
facebook	300	320	338	350	400
periodico	116	130	180	195	210
instagram	52	58	63	79	81
tv	920	970	988	52 1020	1100
ebay	2300	2400	2850	3000	3200

Table 1: Datos que se anexan a los datos originales

Criterios de corrección para el proyecto

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.

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.

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

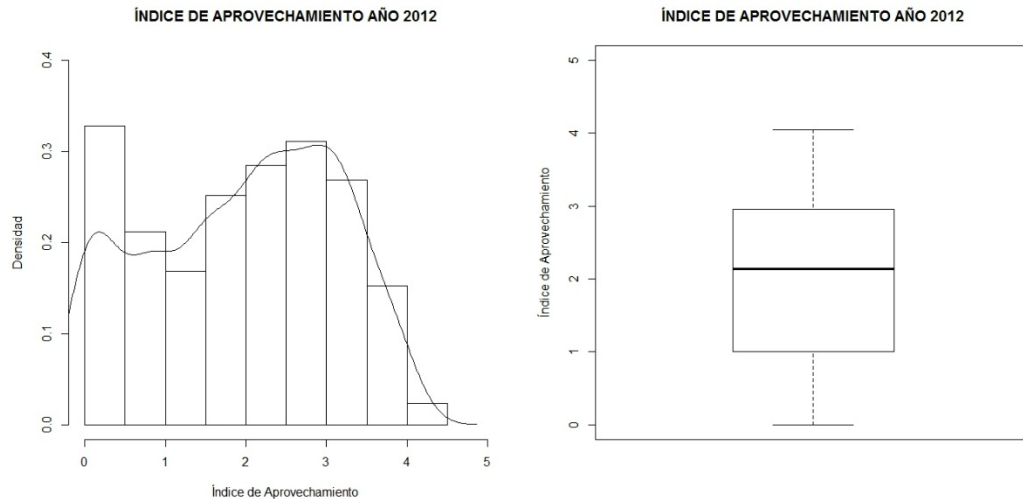


Gráfico 1. Histograma y gráfico de caja para la variable Índice de aprovechamiento.

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>.

Condiciones de entrega

- a El proyecto debe ser realizado en grupos de 3 personas.
- b El informe debe ser entregado en forma electrónica y en formato “.pdf”.
- c La entrega se realizará al correo electrónico dvillalta@usb.ve (con copia a povallesgarcia@usb.ve) a más tardar el viernes 29 de noviembre de 2019 a las 11:59 p.m. El asunto del correo DEBE ser: “*Proyecto. CO3321*”.
- d No se corregirán informes entregados fuera del tiempo establecido para la entrega.