



Estadística para Ingenieros

PROYECTO (20 puntos)

Profesor: Pedro Ovalles

Abril -Julio 2021

1. Datos

En el archivo *broadway-shows.txt* hay un resumen de lo ocurrido en las temporadas de Broadway entre 1984 y 2016. Para cada uno se registran:

Season = El año de la temporada.

Gross(\$M) = Lo recaudado en millones de dólares.

Attendance = La concurrencia total en miles de personas.

Playing weeks = La suma de semanas en las que estuvieron activas todas obra.

New Productions = Número de obras nuevas en la temporada.

Mean ticket = Precio promedio del ticket.

Pct.sold = Porcentaje de tickets que no se venden.

LogGross = El logaritmo de lo recaudado.

2. Trabajo asignado

Para este grupo de variables se solicita el siguiente trabajo:

1. Realizar un análisis descriptivo de los datos.
2. Realice un intervalo de confianza del 97 % para la media de cada variable en estudio. Analice lo obtenido.
3. Pruebe, a un nivel de 0.05, que el promedio obras nuevas inferior a 40.
4. Realizar una prueba de bondad de ajuste para determinar si la variable **Gross(\$M)** tiene distribución normal.
5. Realizar un gráfico de dispersión y una matriz de correlación de las variables.
6. Haga un muestreo para dividir los datos en dos subconjuntos, uno con 80 % y 20 % de los datos.
7. Con el subconjunto del 80 % de los datos, halle un modelo lineal que explique mejor la variable **Gross(\$M)**. Incluya todas las pruebas necesarias para llegar a este modelo, así como un análisis de residuos del modelo final. (No utilice las variables **Season** y **LogGross**)
8. Repita la parte anterior, pero ahora explicar **LogGross**. Decida si es preferible estudiar lo recaudado o su logaritmo.
9. Con los datos del 20 % restante, haga una predicción de la variable **Gross(\$M)** o **LogGross** (con el mejor modelo) y haga un resumen estadístico de los residuos de predicción (valor observado vs. predicción del modelo) para concluir con relación al poder predictivo del modelo.
10. Haga un análisis de varianza para estudiar si los promedios de recaudo, concurrencia y nuevas producciones son iguales por década.

3. 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.

Presentación de resultados

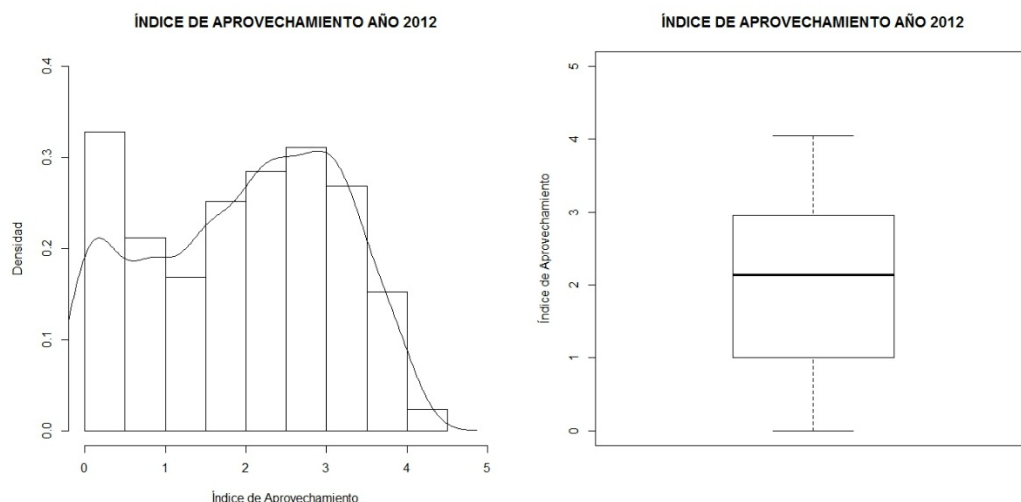
- Presente sus resultados en tablas ordenadas e interprete.
- Identifique en los diagramas de caja si hay datos atípicos, cómo es la distribución de los datos, si es sesgada a la derecha, etc.
- Los gráficos tiene que tener su título y los nombres de los ejes (todo en español).

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.

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



3.1. 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.

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

Gleman, A., Carlin, J., Stern, H. y Rubien, 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>.

4. 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 domingo 25 de julio de 2021 a las 2:00 p.m. El asunto del correo DEBE ser: “*Proyecto. CO3321*”.
- c No se corregirán informes entregados fuera del tiempo establecido para la entrega.