

TERCER PARCIAL

15 de Mayo de 2020

Indicaciones generales

- Este es un examen **individual** con una duración de **110 minutos: de 9:00 a 10:50**. Los diez minutos entre las 10:50 y 11:00 se destinarán a subir la información de forma correcta.
- Debe entrar al aula virtual de CASUR 308 (por zoom) y encender la cámara web (puede ser mediante el celular).
- Sólo se contestarán preguntas sobre los enunciados del parcial, durante los primeros 10 minutos.
- Únicamente puede utilizar R como calculadora para resolver los problemas, exceptuando el cálculo de los cuantiles y el punto 4.
- Las respuestas deben estar totalmente justificadas. Puede determinar los cuantiles con R, pero debe explicar cómo lo hizo.
- En todos los puntos que use R, debe justificar el procedimiento tanto en papel (no el código) y debe colocar el código usado en un archivo Rmarkdown (o archivo .R) con nombre **NombreApellido.Rmd**, y posteriormente adjuntarlo en **e-aulas**.
- Puede enviar fotos rápidas de los procesos de cada punto al acabar el parcial. Posteriormente, puede enviar las imágenes escaneadas (o foto) con buena calidad de los procedimientos detallados de forma clara y ordenada. El archivo .Rmd (o .R) deben adjuntarlo al acabar el parcial.
- Cualquier incumplimiento de lo anterior conlleva la anulación del examen.

- [25 ptos.] Cinco monedas de 1000 pesos fueron lanzadas simultáneamente 1000 veces y en cada lanzamiento se observó el número de caras. Se recopilaron los números de lanzamientos durante los cuales se obtuvieron 0, 1, 2, 3, 4 y 5 caras (frecuencia observada). Con estos datos se realizó un ajuste para una distribución binomial resultando en un modelo cuya distribución binomial ajustada está dada por $f(x) = (0.494)^x(0.506)^{5-x}$. La siguiente tabla muestra el número de caras, las probabilidades (del número de caras), las frecuencias esperadas (por el ajuste) y las frecuencias observadas.

Número de caras (x)	$P(x \text{ caras})$	Frecuencia esperada	Frecuencia observada
0	0.0332	33.2	38
1	0.1619	161.9	144
2	0.3162	316.2	342
3	0.3087	308.7	287
4	0.1507	150.7	164
5	0.0294	29.4	25

Se desea realizar una prueba de bondad de ajuste (prueba Chi-cuadrada) del procedimiento realizado (¿los datos siguen una distribución binomial?).

- [5 ptos.] Calcule el estadístico de prueba.
- [10 ptos.] Determine la región de rechazo para un nivel de significancia del 5 % y concluya.
- [10 ptos.] Determine el *valor p* y concluya.



2. [25 ptos.] Con los siguientes datos:

x_i	4.0	4.2	4.5	4.7	5.1	5.5	5.9	6.3	6.8	7.1
y_i	102.56	113.18	130.11	142.05	167.53	195.14	224.87	256.73	299.50	326.72

- a) [5 ptos.] Determine la aproximación de una recta de mínimos cuadrados.
- b) [10 ptos.] Determine la aproximación de mínimos cuadrados de la forma bx^a .
- c) [10 ptos.] ¿Cuál de las dos aproximaciones realiza un mejor ajuste sobre los datos? Justifique cuantitativamente su respuesta.
3. [25 ptos.] ¿Cuál es el mínimo tamaño de muestra necesario para poder concluir que un coeficiente de correlación es mayor que cero, a un nivel de significancia de 0.05? **Sugerencia:** inicie con el cuantil correspondiente a infinitos grados de libertad.
4. [25 ptos.] Cargue el dataset **Boston** de la librería **MASS**. Este dataset contiene **medv** (el valor medio de casas) para 506 barrios en Boston. Para predecir **medv** se pueden usar hasta 13 descriptores (predictores, o variables independientes) tales como: **rm** (número promedio de habitaciones por casa), **age** (promedio de edad de las casas), **lstat** (porcentaje de hogares con bajo nivel socioeconómico), entre otros. Utilice la función **lm()** para realizar una regresión lineal simple con **medv** como variable dependiente y **age** como variable independiente. Use la función **summary()** para mostrar los resultados. Con los resultados mostrados, responda con justificación estadística las siguientes preguntas:
- a) [10 ptos.] ¿Hay una relación entre la variable independiente y la variable dependiente?
- b) [10 ptos.] ¿Qué tan fuerte es la relación entre la variable dependiente y la variable independiente?
- c) [5 ptos.] ¿Existe una relación positiva o negativa entre la variable dependiente y la variable independiente?