Ingeniería Matemática
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE
MA3403-3
Probabilidades y Estadística
Primavera 2012

Guía ejercicios para el examen

1. Se afirma que el 2 % de los conductores olvida su licencia de conducir. Se toma una muestra de 100 conductores, y se observa que todos andan trayendo su licencia.
   a) ¿Cuál es el p-valor del test que contrasta la afirmación con la hipótesis de que menos del 2 % de los conductores olvidan su licencia?
   b) Para $\alpha = 2{,}28\,\%$, ¿cuál es la máxima cantidad de conductores adicionales que traen su licencia tal que la afirmación no se rechaza?
2. Una autopista posee 4 pistas, y se desea investigar si los conductores tienen preferencia por alguna de ellas. Se observó la pista por la que transitaron 1000 automóviles, cuyos resultados se resumen en la siguiente tabla. ¿Hay suficiente evidencia para decir que algunas pistas son preferidas sobre otras? Use $\alpha = 0{,}05$.

Pista	Cantidad observada
1	294
2	276
3	238
4	192

3. La cantidad de accidentes sufridos por maquinistas de una cierta industria se observó durante un periodo de tiempo, cuyos resultados se muestran en la siguiente tabla. Realice un test a nivel 5 % sobre la hipótesis de que los datos provienen de una distribución de Poisson.

Accidentes por maquinista	Cantidad de maquinistas
0	296
1	74
2	26
3	8
4	4
5 ó mas	6

4. a) Considere una m.a.s. $X_1, \dots, X_n$ proveniente de una variable $\text{bin}(m, p)$, con m conocido y p desconocido. Muestre que los estimadores de p de máxima verosimilitud y de los momentos coinciden con $\hat{p} = \bar{X}/m$. Muestre que es insesgado y que converge casi seguramente a p cuando $n \to \infty$.

   En un centro comercial hay 3 tiendas de la misma cadena, y se desea investigar la forma en que los clientes deciden entrar o no entrar en cada una de ellas. Se propone el siguiente modelo: cada cliente decide entrar a una tienda con probabilidad p (desconocida), y decide no entrar con probabilidad $1-p$, independiente de las otras dos tiendas y del resto de los clientes. Se le hace un seguimiento a $n = 64$ clientes, cuyos resultados se resumen en la siguiente tabla:

Cantidad de tiendas visitadas	0	1	2	3
Cantidad de clientes	6	30	18	10

   b) Muestre que el valor del estimador de máxima verosimilitud de p que se obtiene de acuerdo a los datos es $\hat{p} = 1/2$. Plantee las hipótesis del test de bondad de ajuste para el modelo propuesto, especificando el valor de p_i bajo H_0, para cada $i = 0, 1, 2, 3$.
   c) Realice el test y concluya para $\alpha = 10\,\%$.
5. Los empleados en una fábrica deben asistir a una capacitación para aprender a armar los componentes de un determinado producto. Existen dos posibles capacitaciones, la Estándar y la Novedosa. Antes de escoger permanentemente una de ellas, se envía a 9 empleados a capacitarse en la primera y a otros 9 a capacitarse en la segunda, y luego se midió el tiempo (en segundos) que tardó cada empleado en armar el producto. Los datos relevantes de estas mediciones se resumen en la siguiente tabla:

	Estándar	Novedosa
Tiempo promedio	35,22	31,56
$\sum(x_i - \bar{x})^2$	195,56	160,22

   Suponiendo que las muestras provienen de distribuciones normales con la misma varianza, ¿hay suficiente evidencia para afirmar que las capacitaciones generan tiempos de armado promedio distintos? Use $\alpha = 0{,}05$.
6. En una fábrica de motores, se desea comprar una gran cantidad de una cierta pieza específica. Para disminuir la cantidad de motores defectuosos producidos, es muy importante que el grosor de las piezas tenga la menor variabilidad posible. Se dispone de dos proveedores de piezas, A y B. Se adquieren 10 piezas del proveedor A y se mide el grosor de cada una, obteniendo que el valor del estimador insesgado de la varianza es 0,0003. Además, se adquieren 20 piezas del proveedor B y el estimador insesgado de la varianza dio 0,0001. Las muestras se asumen normales e indepenientes. Con $\alpha = 5\,\%$, ¿hay suficiente evidencia para afirmar que el proveedor B produce piezas con varianza estrictamente menor?
7. Se desea investigar la efectividad de una nueva vacuna. Se toma una muestra de 1000 personas, a algunas de las cuales se les da 1 dosis de la vacuna, a otras se les da 2 dosis, y al resto no se les vacuna. Pasadas 2 semanas se observa si las personas enfermaron. Se resume la información en la siguiente tabla:

	Sin vacuna	1 dosis	2 dosis
Enfermo	24	9	13
Sano	289	100	565

   ¿Hay suficiente evidencia para afirmar que la vacuna tiene un efecto observable?
8. En un estudio se clasificó a 80 televidentes en "audiencia de alta violencia" y "audiencia de baja violencia", de acuerdo a los programas de televisión que ellos habituaban ver. Los resultados, segmentados por grupos de edad, se muestran en la siguiente tabla. ¿Hay suficiente evidencia para afirmar, al

nivel 5 %, que el tipo de audiencia es independiente de la edad del televidente? Entregue el p-valor del test (o al menos una cota).

Violencia \ Edad	16-34	35-54	55 ó más
Baja	8	11	21
Alta	18	15	7

9. Consideramos dos indicadores demográficos para 20 países de América Latina: la esperanza de vida, denotada y, y la tasa de natalidad, denotada x. Se estudia la relación que existe entre estas variables, a patir del modelo

$$y \approx \beta_0 + \beta_1 x.$$

Se cuenta con una muestra de tamaño $n = 70$, con $\frac{1}{n}\sum x_i y_i = 24{,}8$, $\frac{1}{n}\sum x_i^2 = 0{,}123$, $\bar{x} = 0{,}305$, $\bar{y} = 78{,}3$. Estime los parámetros β_0 y β_1 en el modelo, de manera que se minimicen los errores al cuadrado. ¿Cuál es la predicción del modelo para $x = 0{,}7$?

10. Se dispone de una variable aleatoria Y, la cual se pretende explicar mediante una combinación lineal de potencias de una variable X, de la forma

$$Y \approx \beta_0 + \beta_1 X + \beta_2 X^2 + \ldots \beta_k X^k.$$

Dada una muestra $(x_1, y_1), \ldots, (x_n, y_n)$, explicite el vector $\mathbb{Y}$ de variables dependientes, el vector β de coeficientes de la regresión, y la matriz $\mathbb{X}$ con los datos de la variable independiente.

11. Se dispone de una muestra aleatoria simple $(X_1, Y_1), \ldots, (X_n, Y_n)$ de datos provenientes del par de variables (X, Y), donde $\mathbb{E}(X)$, $\mathbb{E}(Y)$ y $\text{var}(X) > 0$ son finitas.

   a) Argumente que $\frac{1}{n}\sum_{i=1}^{n} X_i Y_i$ y $\frac{1}{n}\sum_{i=1}^{n} X_i^2$ convergen casi seguramente a $\mathbb{E}(XY)$ y $\mathbb{E}(X^2)$, respectivamente, cuando $n \to \infty$.

   b) Se ajusta un modelo lineal de la forma $Y \approx a + bX$, el cual se resuelve mediante el criterio de minimizar la suma de residuos al cuadrado, con lo cual se obtienen las cantidades $\hat{a}_n$ y $\hat{b}_n$. Muestre que $\hat{b}_n \to \text{cov}(X,Y)/\text{var}(X)$ y $\hat{a}_n \to \mathbb{E}(Y) - \mathbb{E}(X)\text{cov}(X,Y)/\text{var}(X)$ casi seguramente cuando $n \to \infty$.

12. Considere los siguientes datos:

x	-2	-1	0	1	2
y	0	0	1	1	3

   a) Ajuste la recta de mínimos cuadrados a los datos.

   b) ¿Hay suficiente evidencia para afirmar que la pendiente de la recta ajustada es distinta de 0? Utilice $\alpha = 5\,\%$.

   c) Obtenga un intervalo de confianza para la pendiente al nivel 95 %.

13. Existen 4 tamaños predeterminados para un cierto producto. Se desea investigar la posible relación lineal entre la duración del producto y su tamaño, para lo cual se toman muestras de 2 productos por cada tamaño, y se registra su duración y se ajusta la recta de mínimos cuadrados sobre los datos obtenidos.

   a) Se realiza el test para decidir si la constante multiplicativa del modelo es distinta de 0. El valor del estadístico correspondiente, bajo la hipótesis nula, es 1,95. ¿Cuál es el p-valor del test? ¿Cuál es la conclusión si se trabaja con $\alpha = 5\,\%$?

   b) Sabiendo además que el valor estimado de la constante multiplicativa dio 3,9, calcule un intervalo de confianza de dicha constante al nivel 95 %.