Tarea 08 - Análisis de datos
Giovanni Gamaliel López Padilla

Problema 2

Para investigar si una moneda es justa o no, alguien decid que va a lanzar la moneda 100 veces. Si el número de veces de obtener sol es entre 42 % y 58 %, va a apoyar la hipótesis de que la moneda es justa. Calcula el nivel de significancia correspondiente.

Problema 3

Por experiencia se sabe que el número de accesos ,X, durante una hora a una base de datos sigue una distribución Poisson:

$$P(X = x) = exp(-\lambda)\frac{\lambda^x}{x!} \qquad \text{para } x = 0, 1, 2, \ldots \text{ y } \lambda > 0$$

Calcula el estimador de máxima verosimilitud para λ en una muestra.

La función de máximo verosimilitud de la distribución es Poisson es:

$$\mathcal{L} = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^x}{x!}$$

Entonces, la función log-verosimilitud es:

$$\begin{aligned}
log(\mathcal{L}) &= log\left(\prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^x}{x!}\right) \\
&= \sum_{i=1}^{n} log\left(e^{-\lambda}\frac{\lambda^x}{x!}\right) \\
&= \sum_{i=1}^{n}\left[log(e^{-\lambda}) + log(\lambda^{x_i}) - log(x!)\right] \\
&= \sum_{i=1}^{n}\left[-\lambda + x_i log(\lambda) - log(x_i)\right] \\
log(\mathcal{L}) &= -\lambda n + log(\lambda)\sum x_i - \sum log(x_i!)
\end{aligned}$$

Por lo que calculando el valor crítico de $log(\mathcal{L})$ se obtiene que:

$$\frac{\partial \log(\mathcal{L})}{\partial \lambda} = 0$$
$$-n + \frac{\sum x_i}{\lambda} = 0$$
$$\hat{\lambda} = \frac{1}{n} \sum x_i$$

donde $\hat{\lambda}$ corresponde a la media aritmética.

Calculando la segunda derivada de $log(\mathcal{L})$ para comprobar que el valor crítico $(\hat{\lambda})$ se trata de un máximo se obtiene lo siguiente:

$$\frac{\partial^2 \log(\mathcal{L})}{\partial \lambda^2} \geq 0$$
$$-\frac{1}{\lambda^2} \sum x_i \geq 0$$

como $\lambda^2 > 0$ y $x_i \in \{0, 1, 2, \dots\}$, entonces, la segunda derivada de $log(\mathcal{L})$ será siempre positiva, por lo tanto $log(\mathcal{L})$ es el estimador de máxima verosimilitud.

Problema 4

Considera los siguientes datos de un estudio en Bélgica sobre la intención de voto entre 1000 parejas. Las variables X_1 , X_2 indican si la mujer, respectivamente el hombre, votará para un partido de la coalición (0) o de la oposición (1) en caso de que hubieran elecciones en ese momento.

	$X_1 = 0$	$X_1 = 1$
$X_2 = 0$	245	170
$X_2 = 1$	218	367

a) Calcula el oddsratio $\hat{R}$

b) Se puede mostrar que si el tamaño de la muestra va a ∞, la distribución de log(R) converge a una normal con promedio $\log(\hat{R})$, el verdadero log-oddsratio de la distribución subyacente, y con varianza

$$\frac{1}{n_{0,0}} + \frac{1}{n_{0,1}} + \frac{1}{n_{1,0}} + \frac{1}{n_{1,1}}$$

donde $n_{i,j}$ es el número de observaciones con $X_1 = i$ y $X_2 = j$. ¿Apoyas la hipótesis que la pareja vota de manera independiente $(\alpha = 0.05)$?

Problema 5

Sea c una cierta cadena binaria de longitud 100. Se quiere verificar si proviene de una muestra $Bern(0.5)(= H_0)$. Para eso se calcula el número de cambios. Un cambio es un 1 seguido por un 0 o un 0 seguido por un 1 en la cadena.

a) Calcula T el número de cambios en una cadena usando una sola linea de código en R.

b) Usando muchas simulaciones de cadenas bajo H_0 , estima y visualiza la distribución de T.

c) Calcula el valor de p para H_0 si T=42.

Problema 6

¿Hay alguna semejanza entre una taza y su dueño? Para eso se decide hacer un pequeño experimento. Se muestran a n voluntarios 5 fotos de personas y 5 tazas en orden al azar. Se pide a cada persona asociar cada taza con una persona (una a una). ¿Cómo formular una prueba de hipótesis para este problema? Propon una estadistica de prueba. Estima su distribución con simulaciones de respuestas bajo H_0.

Problema 7

Durante los juegos olimpicos de Salt Lake City surgió en un periódico la discusión si en las pruebas de 1500m de patinaje, la persona en el carril exterior no tendria ventaja sobre el carril interior. Se organizaron 24 pruebas (una se canceló por una caida). Abajo los tiempos. Aplica una(s) pruebas de estadistica relevante para contestar esta pregunta.