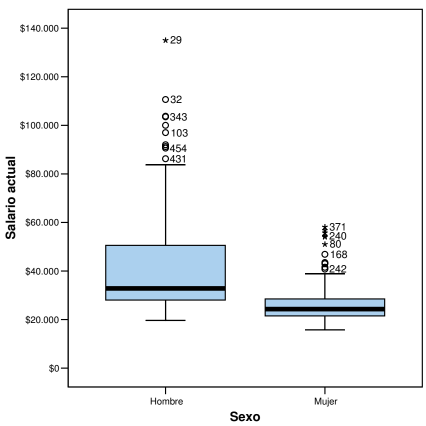
Prueba de Estadística

Apelidos: **Nome:** **DNI:**

Para facilitar la corrección de la prueba, marque en la tabla la letra de la respuesta que considere correcta para cada una de las cuestiones propuestas[1].

Pregunta 1	a	b	c	d
Pregunta 2	a	b	c	d
Pregunta 3	a	b	c	d

1. *(1 punto)* En el siguiente diagrama de cajas se representa el salario mensual en dolares por sexos de una determinada empresa. ¿Qué afirmaciones son correctas?:



   a) Al menos el 50 % del salario de los hombres es superior al 75 % del salario de las mujeres.
   b) El salario medio de los hombres es mayor que el salario medio de las mujeres.
   c) El salario de los hombres es superior al de las mujeres.
   d) Ninguna de las anteriores

   solución correcta: a) y b)

2. *(1 punto)* Un sistema de computación en-línea tiene 4 líneas de entrada. Cada línea cubre un porcentaje de tráfico de entrada y cada línea tiene un % de mensajes con errores. La tabla siguiente describe estos porcentajes.

Línea	% mens. por línea	% mens. sin error
1	40	99.8
2	30	99.9
3	10	99.7
4	20	99.2

   Si un mensaje se recibió erróneamente, ¿cuál es la probabilidad de que entrara por la línea 1?

   a) $\frac{4}{11}$
   b) 0.40
   c) $\frac{1}{4}$
   d) Ninguna de las anteriores. Prob. de bayes = (2*0.4)/(2*0.4+1*0.3+3*.1+8*0.2)=8/30

3. *(1 puntos)* Sea X una variable con distribución $N(5, 0.5)$. El valor de c tal que $P(|X-5| \leq c) = 0.9$ es (redondear a 3 decimales).

   a) 0.641
   b) 0.822
   c) 0.980
   d) Ninguno de los anteriores

   solución correcta b): é o cuantil 0.95 da N(0,.5): round(qnorm(c(0.9,.95,.975), mean=0,sd=0.5),3)

4. *(1 punto)* Los mensajes que llegan a un dispositivo movil lo hacen de acuerdo con una distribución de Poisson con una tasa promedio de 0.1 mensajes por minuto. ¿Determinar el intervalo de tiempo necesario para que la probabilidad de que no llegue ningún mensaje durante ese lapso de tiempo sea 0.8?

   a) 2.231 minutos.
   b) 8 minutos.
   c) 0 minutos.
   d) Ninguna de las anteriores.

   Solución: Sea X_t = número de mensajes en el intervalo $(0, t) \sim Pois(0.1t)$, se pide que:

$$
\begin{aligned}
P(X_t = 0) = \exp(-\lambda t)\frac{(\lambda t)^0}{0!} &= 0.8 \\
\exp(-\lambda t) &= 0.8 \\
\log(\exp(-\lambda t)) &= \log(0.8) \\
\lambda t = 0.2231; t &= 2.231
\end{aligned}
$$

□

[1] 1 respuesta incorrecta penaliza de tal forma que 3 respuestas incorrectas equivale a una respuesta correcta. Las preguntas en blanco no penalizan. Es necesario justificar la opción marcada, usar la parte trasera del enunciado.

1. *(2.5 puntos)* El archivo de datos en formato R *chile.RData*[2] contiene información sobre una encuesta nacional realizada en 1988 en Chile, concretamente las variables contenidas en el data.frame *chile* son:

   - *region*: variabbe que toma valores *C, M, N, S, SA* indicando las regiones Central, area Metropolitana de Santiago, Norte, Sur y Ciudad de Santiago, respectivamente,
   - *poblacion*: tamaño de la población de cada región,
   - *sex*: dos valores *F* - female y *M*-male,
   - *age*: edad,
   - *educacion*: *P* de Primaria, *S* de Secundaria y *PS* Post-Secundaria,
   - *income*: ingresos mensuales en Pesos,
   - *vote*: *A* indicando abstención, *N* votará No (en contra de Pinochet), *U* indeciso, *Y* vota a favor (de Pinochet)

   a) ¿Qué tipo de variables estadísticas son?.

   b) Describe completamente la variable *vote*. Da una representación gráfica adecuada. Interpreta los resultados.

   c) Agrupar la variable *income* en los subintervalos $((0, 10000], (10000, 20000], (20000, 50000], (50000, 250000])$. Da su distribución de frecuencias completa. Dar la representación gráfica más adecuada manteniendo esos intervalos.

   d) Resume numéricamente la variable *income* y la variable *income* agrupada. Interpreta todos los resultados que se indiquen.

2. (*3.5 puntos*) Por razones de seguridad, el 5 % de los mensajes pasados por una red son falsos. Un receptor sabe con la llave apropiada si el mensaje es falso o no. De 400 mensajes enviados: *a)* ¿cuál es el número esperado de mensajes falsos? suponiendo que ellos se intercalan aleatoria e independientemente entre los válidos, *b)* ¿cuál es la probabilidad de que el número de mensajes falsos exceda de 15 pero sea menor que 25?. Dar la probabilidad exacta y usando la corrección por continuidad, *c)* ¿cuantos mensajes hay que enviar para que la probabilidad de que el número de mensajes falsos menor que 25 sea menor que 6 décimas?. Razonar detalladamente todos los pasos.

   Solución: Sea la variable aleatoria *X= número de mensajes falsos*, entonces X sigue una distribución binomial porque cuanta cuantos mensajes falsos se envían de forma independiente en n pruebas y p.

   $X \sim Bi(400, 0.05)$

   a) El número medio de mensajes falsos enviados en 400 es $E(X) = n * p = 400 * 0.05 = 20$.

   b) **Probabilidad exacta**:

   $$P(15 < X < 25) = \sum_{16}^{24} \binom{n}{i} p^i (1-p)^{n-i} =$$

   pbinom(24,size=400,prob=0.05) - pbinom(15,size=400,prob=0.05)=0.6990499

   o equivalentemente:

   sum(dbinom(16:24, size=400,prob=0.05))=0.6990499

   Probabilidad aproximada:

   $$P(15 < X < 25) \approx P\left(15 + 0.5 < \tilde{X} < 25 - 0.5\right)$$

   con $\tilde{X} \sim N(np, \sqrt{npq})$. La media y sd de la variable $\tilde{X}$ es por lo tanto $\mu = 20$ y $\sigma = \sqrt{20 * 0.05 * 0.95}$. Si aplicamos la corrección de continuidad:

   $$P\left(15 + 0.5 < N(20, \sqrt{19}) < 25 - 0.5\right) = P\left(N(20, \sqrt{19}) < 24.5\right) - P\left(N(20, \sqrt{19}) < 15.5\right) =$$

   pnorm(24.5,mean=20, sd= sqrt(19)) - pnorm(15.5, mean=20,sd=sqrt(19))=0.6981016

[2] El archivo de datos *chile.RData* se puede descargar desde la url *http://dl.dropboxusercontent.com/u/29008031/chile.RData*

c Ahora la variable es $X \sim Bi(n, 0.05)$ y la condición indica que n debe verificar que

$$P(X < 25) \leq 0.6$$

Usando la aproximación por la normal obtenemos que puedo aproximar X por $\tilde{X} \sim N(0.05 * n, \sqrt{0.0475 * n})$. Tipificando la ecuación anterior obtenemos que:

$$P\left(\tilde{X} < 25\right) = P\left(\frac{\tilde{X} - 0.05n}{\sqrt{0.0475 * n}} < \frac{25 - 0.05n}{\sqrt{0.0475 * n}}\right) \leq 0.4$$

donde ahora la varible tipificada es N(0,1) y por lo tanto el valor pedido es su cuantil 0.4, es decir, *qnorm(0.6, mean=0,sd=1)=-0.2533471*. Entonces

$$\frac{25 - 0.05n}{\sqrt{0.0475 * n}} \leq -0.2533471$$
$$25 - 0.05n \leq -0.2533471 * \sqrt{0.0475 * n}$$

Sólo resta resolver la inecuación:

$$\begin{aligned} 25 - 0.05n &= -0.2533471 * \sqrt{0.0475 * n} \\ (25 - 0.05n)^2 &= 0.2533471^2 * 0.0475 * n \\ 625 - 2 * 25 * 0.05n + 0.0025n^2 &= 0.2533471^2 * 0.0475 * n \\ 0.0025n^2 - \left(2.5 + 0.2533471^2 * 0.0475\right) * n + 625 &= 0 \\ 0.0025n^2 - 2.503049n + 625 &= 0 \end{aligned}$$

Soluciones de la ecuación de 2º grado dadas por: n= 475.909 (raíz negativa) y n = 525.3105 (raíz positiva). Realizando las comprobaciones se obtiene que para que se verifique la probabilidad de inferior a 0.4 debe ser $n \geq 526$.

Algoritmo computacional alternativo de búqueda:

```
#
k <- 24
p <- 0.05
n <- 401
# Busqueda Exacta
  while( pbinom( k, size= n, prob=p) >= 0.4) n = n+1
  print (n)
# aproximación á normal para <= 24
  n <- 401
  while( pnorm( k, mean= n*p, sd=sqrt(n*p*(1-p))) >= 0.4) n = n+1
  print (n)
# aproximación á normal para <= 24+0.5
  n <- 401
  while( pnorm( k+0.5, mean= n*p, sd=sqrt(n*p*(1-p))) >= 0.4) n = n+1
  print (n)
# aproximación á normal para < 25
  n <- 401
  while( pnorm( k+1, mean= n*p, sd=sqrt(n*p*(1-p))) >= 0.4) n = n+1
  print (n)
```

□