Examen 4

1. Se desea explicar el tiempo de arranque de un equipo informático (en segundos) a partir del número de iconos presentes en el Escritorio, para lo cual se han tomado datos en 40 equipos con siguientes resultados:

	Media	Coeficiente de Variación
Y: Tiempo de arranque	45	0.3
X: Número de iconos	30	0.7

La covarianza es $s_{XY} = 250$.

a) Calcula la desviación típica de ambas variables.
b) Calcula el coeficiente de correlación lineal de Pearson e interprétalo.
c) Ajusta el modelo de regresión lineal para explicar el tiempo de arranque de un equipo informático a partir del número de iconos presentes en el escritorio.
d) En base a ese modelo, ¿cuál es el tiempo de arranque predicho para un equipo con 50 iconos en el escritorio?

SOLUCIONES:

a) Como tenemos el coeficiente de variación y la media, podemos despejar la desviación típica de la fórmula:

$$CV_X = \frac{s_X}{|\bar{x}|}$$

$$0.7 = CV_X = \frac{s_X}{30}$$

$$s_X = 30 * 0.7 = 21$$

Y para la otra variable:

$$CV_Y = \frac{s_Y}{|\bar{y}|}$$

$$0.3 = CV_Y = \frac{s_Y}{45}$$

$$s_Y = 45 * 0.3 = 13.5$$

b) El coeficiente de correlación lineal de Pearson se puede calcular como:

$$r_{XY} = \frac{s_{XY}}{s_X\, s_Y} = \frac{250}{21 * 13.5} = 0.88$$

Teniendo en cuenta que siempre está entre -1 y 1, el valor es positivo y bastante cercano a 1, por lo cual, parece ser que hay una relación lineal positiva (o directa) y fuerte entre las dos variables.

c) Nos piden el modelo de regresión que explique la variable Y en función de la X:

$$\hat{y} = a + bx$$

Donde los parámetros del modelo se calculan:

$$b = \frac{s_{XY}}{s_X^2} = \frac{250}{21^2} = 0.57$$

$$a = \bar{y} - b\bar{x} = 45 - 0.57 * 30 = 27.9$$

Entonces la ecuación queda:

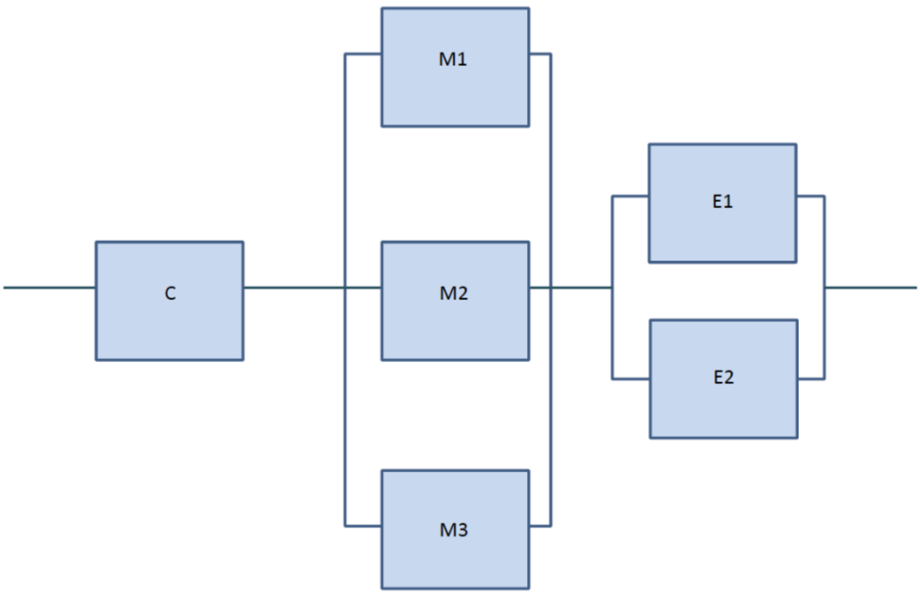
$$\hat{y} = 27.9 + 0.57x$$

d) Nos dicen que un equipo tiene 50 iconos, y nos piden calcular el tiempo de arranque del equipo usando el modelo de regresión. Sustituimos $x = 50$ y hacemos los cálculos:

$$\hat{y} = 27.9 + 0.57 * 50 = 56.4$$

Entonces, el equipo tarda 56.4 segundos en arrancar.

2. Supongamos que en un proceso industrial se sigue el esquema:



Según dicho esquema, para que una pieza se fabrique correctamente debe pasar por tres subprocesos diferentes: Chequeo previo (C), Montaje, que se podrá realizar a través de tres tipos diferentes de maquinaria (M1, M2 o bien M3), y Embalaje, que se podrá realizar en cajas de cartón o de metal (E1 o E2). Sabiendo que cada uno de estos subprocesos se pasa correctamente (o está operativo) con una probabilidad del 95% y que los subprocesos ocurren de forma independiente unos de otros.

a) Determine la probabilidad de que una pieza elegida al azar se fabrique correctamente.
b) Si una caja contiene 10 piezas, ¿cuál es la probabilidad de que en la caja no halla ninguna pieza fabricada correctamente?
c) ¿Y la probabilidad de que halla al menos 1 pieza fabricada correctamente?
d) Sabiendo que en la caja hay al menos 1 pieza fabricada correctamente, ¿cuál es la probabilidad de que haya más de 9 piezas fabricadas correctamente?

SOLUCIONES:

a) Sea C el suceso “la fase o el subproceso de chequeo previo se pasa correctamente”.

   De modo análogo, denotemos por M_i, si $i = 1,2,3$ y por E_j si $j = 1,2$ los sucesos que indican que los demás subprocesos se pasan correctamente, es decir, Montaje y Embalaje, respectivamente.

   Sabemos que $P(C) = P(M_i) = P\left(E_j\right) = 0.95$, si $i = 1,2,3$ y $j = 1,2$.
   Sea $A = C \cap (M_1 \cup M_2 \cup M_3) \cap (E_1 \cup E_2)$. Nos piden $P(A)$:

$$P(A) = P\big(C \cap (M_1 \cup M_2 \cup M_3) \cap (E_1 \cup E_2)\big)$$

   Por independencia, podemos transformar las intersecciones en productos de probabilidades:

$$P(A) = P(C) \cdot P(M_1 \cup M_2 \cup M_3) \cdot P(E_1 \cup E_2)$$

   Vamos a calcular por separado las dos últimas probabilidades y luego las sustituimos.

   En la primera vamos a tener una probabilidad de la unión de 3 sucesos. Un truco es usar el complemento, y la propiedad de que el complemento de la unión es la intersección de los complementos:

   $P(M_1 \cup M_2 \cup M_3) = 1 - P((M_1 \cup M_2 \cup M_3)^C) = 1 - P(M_1^C \cap M_2^C \cap M_3^C) =$
   $1 - P\left(M_i^C\right)^3 = 1 - 0.05^3 = 1 - 0.000125 = 0.9999$

   Luego tenemos que calcular también:
   $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2) = 0.95 * 2 - 0.95^2 = 0.9975$

   Y sustituyendo:
   $P(A) = P(C) \cdot P(M_1 \cup M_2 \cup M_3) \cdot P(E_1 \cup E_2) = 0.95 * 0.9999 * 0.9975 =$
   0.9475

b) Del inciso anterior sabemos que la probabilidad de que una pieza elegida al azar se fabrique correctamente es $p = 0.9475$.

Ahora tenemos una caja de 10 piezas. Denotemos por X la variable aleatoria que cuenta el número de piezas fabricadas correctamente en la caja de 10 piezas. Esta nueva v.a. tiene distribución Binomial con parámetros $n = 10$ y $p = 0.9475$

$$X \sim Binomial(n = 10, p = 0.9475)$$

Nos piden la probabilidad de que en la caja no halla ninguna pieza fabricada correctamente. Esto es equivalente a decir que $X = 0$, y para calcular la probabilidad de ese suceso, usaremos la función de probabilidad de la Binomial:

$P(X = 0) = \binom{10}{0} 0.9475^0 (1 - 0.9475)^{10-0} = \frac{10!}{0!(10-0)!} 1 * 0.0525^{10} = 1.6 * 10^{-13}$

c) Para la probabilidad de que halla al menos 1 pieza fabricada correctamente, podemos usar la propiedad del complemento, y el hecho de que los valores de esta Binomial son 0,1,2,...,10:

$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - 1.6 * 10^{-13} \approx 1$

d) Nos dicen que sabemos que en la caja hay al menos 1 pieza fabricada correctamente, y nos piden calcular la probabilidad de que haya más de 3 piezas fabricadas correctamente, sabiendo la anterior condición. Entonces nos piden una probabilidad condicional:

$$P(X > 9 | X \geq 1) = \frac{P((X > 9) \cap (X \geq 1))}{P(X \geq 1)}$$

El conjunto $(X > 9) \cap (X \geq 1)$ es igual a $(X > 9)$ porque $(X \geq 1)$ está incluido y se trata de una intersección entre los dos.
Entonces:

$$P(X > 9 | X \geq 1) = \frac{P(X > 9)}{P(X \geq 1)}$$

El denominador lo tenemos calculado en el inciso anterior, es aproximadamente igual a 1. Falta calcular el numerador:

$P(X > 9) = P(X = 10) = \binom{10}{10} 0.9475^{10} (1 - 0.9475)^{10-10} = \frac{10!}{10!(10-10)!} 0.9475^{10} * 0.0525^0 = 0.9475^{10} = 0.5832$

Entonces

$$P(X > 9 | X \geq 1) = \frac{P(X > 9)}{P(X \geq 1)} \approx 0.5832$$

3. En un campo de concentración alemán durante la Segunda Guerra Mundial, el coronel jefe del comité de fugas aliado, tiene que decidir qué plan autoriza a llevar a cabo de entre dos que le han propuesto. Necesita calcular cuál tiene más probabilidad de éxito. Para ello han estado observando el movimiento de los vigilantes del patio del campo durante las últimas 48 horas. La frecuencia con la que pasan por el punto por donde es más fácil escapar es la siguiente:

Número de vigilantes	Número de horas
0	7
1	9
2	6
3	10
4	8
5	5
6	3

Los planes propuestos son:
Plan A: Para tener éxito han de pasar durante 1 hora menos de tres vigilantes.
Plan B: Para tener éxito, durante un mínimo de 30 minutos no puede pasar ningún vigilante.

a) Calcule la probabilidad de éxito de cada uno de los planes y diga cuál es el elegido, suponiendo que los vigilantes aparecen manteniendo una media estable y de forma independiente.
b) Ante la sospecha de un plan de fuga los oficiales alemanes deciden incrementar la vigilancia con un nuevo grupo de soldados. Estos actúan de forma independiente a los que ya había en el campo y su aparición sigue un proceso de Poisson de media 2 vigilantes a la hora. ¿Cómo queda la probabilidad de fuga del mejor plan tras la incorporación de este nuevo grupo de vigilantes?

SOLUCIONES:

a) Estamos ante un proceso de Poisson. Para conocer la media:

$$\lambda = \frac{0*7+1*9+2*6+3*10+4*8+5*5+6*3}{48} = 2.625$$

Para conocer la probabilidad del plan A, analizaremos la variable aleatoria X = "Número de vigilantes por hora", que sigue una distribución de Poisson:

$$X \sim Poisson(\lambda = 2.625)$$

La probabilidad de que funcione el plan A (que en una hora pasen menos de tres vigilantes), es la siguiente:

$$P(Plan\ A) = P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$$

Cada una de las tres probabilidades se puede calcular usando la función de probabilidad de la Poisson:

$P(X = 0) = \frac{2.625^0}{0!} e^{-2.625} = 0.0724$

$P(X = 1) = \frac{2.625^1}{1!} e^{-2.625} = 0.1902$

$P(X = 2) = \frac{2.625^2}{2!} e^{-2.625} = 0.2496$

Entonces:

$P(Plan\ A) = P(X < 3) = 0.0724\ + 0.1902 + 0.2496 = 0.5122$

Para conocer la probabilidad del Plan B, analizaremos la variable aleatoria exponencial T = "Tiempo entre el paso de dos vigilantes". Concretamente tendríamos que hallar la probabilidad de $T > 0.5$ horas (el tiempo entre 2 vigilantes tiene que ser como mínimo de 30 minutos = 0.5 horas). Como es una variable continua da igual si ponemos $>$ o $\geq$. Además esta variable tiene distribución:

$$T \sim Exp(\frac{1}{\lambda} = \frac{1}{2.625})$$

Por lo cual, la probabilidad del Plan B es:

$$P(Plan\ B) = P(T > 0.5) = e^{(-2.625*0.5)} = 0.2691$$

Como tenemos que $P(Plan\ A) = 0.5122 > 0.2691 = P(Plan\ B)$, el plan elegido será el Plan A.

b) La distribución de los vigilantes originales X_1 y la de los nuevos X_2 son sendas distribuciones de Poisson. Como además son independientes, se puede afirmar que la suma de las dos S es una Poisson con media igual a la suma de las respectivas medias de X_1 y X_2:

$X_1 \sim Poisson(\lambda_1 = 2.625)$
$X_2 \sim Poisson(\lambda_2 = 2)$
$S \sim Poisson(\lambda_1 + \lambda_2 = 4.625)$

Para recalcular la probabilidad del Plan A:

$P(S = 0) = \frac{4.625^0}{0!} e^{-4.625} = 0.0098$

$P(S = 1) = \frac{4.625^1}{1!} e^{-4.625} = 0.0453$

$P(S=2)=\frac{4.625^2}{2!}e^{-4.625}=0.1048$

Entonces:

$$P(Plan\ A\ nuevo)=P(S<3)=0.0098\ +0.0453+0.1048=0.1599$$

4. Aplicando el Teorema Central del Límite a una sucesión de 100 variables aleatorias independientes X_i de tipo Bernoulli(0.3) se obtiene que $Y=\sum_{i=1}^{100} X_i$ puede ser aproximada con una Normal de parámetros:
   a) $\mu=0.3, \sigma=0.003$
   b) $\mu=30, \sigma^2=21$
   c) $\mu=30, \sigma=0.03$
   d) $\mu=0.003, \sigma=0.021$

SOLUCIONES:

Denotemos como S la suma de Bernoullis. Por el TCL, S se puede aproximar por una Normal con los siguientes parámetros:

$$S=X_1+\cdots+X_{100}\sim Normal(n\mu, n\sigma^2)$$

Donde μ es la media de las variables originales, en nuestro caso, la media de las Bernoulli

$$\mu=p=0.3$$

Y σ^2 es la varianza de las variables originales, en nuestro caso, la varianza de las Bernoulli

$$\sigma^2=p(1-p)=0.3*0.7=0.21$$

Entonces la distribución para S quedaría:

$$S\sim Normal\big(np, np(1-p)\big)=Normal(\mu=30, \sigma^2=21)$$

La opción (b) es la correcta.

Notemos que esto es la demostración de cómo por el TCL se puede aproximar a la Binomial por una Normal, ya que la suma de n Bernoullis es Binomial.