

**PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ**  
**FACULTAD DE CIENCIAS E INGENIERÍA**

**ESTADÍSTICA PARA INGENIERÍA**

**Segunda práctica (tipo a)**  
**(Primer semestre 2023)**

**Indicaciones generales:**

- Duración: 1h 50 min.
- Materiales o equipos a utilizar: Con apuntes de clase y calculadora.
- No está permitido el uso de ningún material o equipo electrónico adicional al indicado.
- Defina todos los eventos y justifique su procedimiento usando propiedades de probabilidad. Interprete los resultados numéricos teniendo en cuenta el contexto de cada ejercicio.
- **La presentación, la ortografía y la gramática de los trabajos influirán en la calificación.**

Puntaje total: 20 puntos (12 puntos del cuestionario y 8 puntos de una tarea)

**Cuestionario:**

**Pregunta 1** (4.0 puntos)

Se han recolectado datos de 500 localidades, en las cuáles se ha medido el porcentaje de habitantes que son usuarios de bicicleta (biking) y el porcentaje de habitantes que presentan una enfermedad cardíaca (heart.disease).

Inicialmente se consideró un modelo de regresión lineal que explique el porcentaje de habitantes que presentan una enfermedad cardíaca en función del porcentaje de habitantes que son usuarios de bicicleta.

```
reg=lm(heart.disease~biking,data = d)
summary(reg)

##
## Call:
## lm(formula = heart.disease ~ biking, data = d)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 17.697884   0.146780  120.57  <2e-16 ***
## biking      -0.199091   0.003378  -58.94  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 1.618 on 496 degrees of freedom
## Multiple R-squared:  0.8751, Adjusted R-squared:  0.8748
## F-statistic: 3474 on 1 and 496 DF, p-value: < 2.2e-16
```



- a) (1.0 punto) Interprete los coeficientes estimados.
- b) (1.0 punto) Evalúe la verdad o falsedad de la siguiente afirmación: "Considerando los resultados del modelo, se puede decir que aproximadamente para el 87.51% de las localidades evaluadas, el porcentaje observado de habitantes que presentan enfermedad cardíaca se predice sin error". Justifique su respuesta.
- c) (1.0 punto) En base a las siguientes estadísticas estime un modelo de regresión que permita estimar el porcentaje de habitantes que presentan una enfermedad cardíaca en base al porcentaje de habitantes que son fumadores. Indique cuál de los dos modelos explica mejor el porcentaje de habitantes que presentan una enfermedad cardíaca. Justifique su respuesta.

Variable	Media	Varianza	Covarianza
Porcentaje de habitantes que fuman	15.4	68.7	11.7
Porcentaje de habitantes que presentan una enfermedad cardíaca	10.2	20.9	

- d) (1.0 punto) Estime usando el modelo elegido en la parte c), el porcentaje de habitantes que presentan una enfermedad cardíaca en una localidad donde el 15% de habitantes fuma y el 30% son usuarios de bicicleta.

### **Pregunta 2** (4.0 puntos)

De acuerdo a la encuesta de ENSUSALUD 2015 realizada a la población de médicos del país se conoce que aproximadamente

- El 74.7% de los médicos encuestados es de sexo masculino.
- El 66.5% de los médicos encuestados tiene especialidad.
- El 51.7% de los médicos encuestados tiene un ingreso mayor a los 5000 soles.
- El 8% de los médicos encuestados es del sexo femenino y tiene un ingreso mayor a los 5000 soles.
- El 89.2% de los médicos encuestados es del sexo masculino o tiene especialidad.
- El 14.6% de los médicos encuestados es del sexo masculino, no tienen especialidad y tienen un ingreso menor o igual a los 5000 soles.

Se selecciona al azar uno de los médicos encuestados. Basado en la información disponible, calcule:

- a) (1.0 punto) La probabilidad de que el médico seleccionado sea del sexo femenino y no tenga especialidad.



- b) (1.0 punto) La probabilidad de que el médico seleccionado sea del sexo masculino y tenga un ingreso mayor a los 5000 soles.
- c) (2.0 puntos) La probabilidad de que el médico seleccionado sea del sexo masculino, no tenga especialidad y tenga un ingreso mayor a los 5000 soles.

**Pregunta 3** (4.0 puntos)

Una empresa tiene dos sedes, en Lima y en Panamá. Los trabajadores de la gerencia de planeamiento de inventario de la empresa pueden pertenecer a una de tres categorías: analista senior, analista junior y asistente. En Lima, trabajan 10 personas en esta gerencia, 5 analistas senior, 3 analistas junior y 2 asistentes. En Panamá trabajan 8 personas en la gerencia, 1 analista senior, 3 analistas junior y 4 asistentes.

- a) (2.0 punto) Suponga que se va a seleccionar del conjunto de trabajadores de las dos sedes, cinco trabajadores al azar y sin reemplazo. Calcule la probabilidad de seleccionar más de tres asistentes.
- b) (2.0 puntos) Suponga que de cada sede se seleccionan al azar y sin reemplazo tres trabajadores. Halle la probabilidad de seleccionar dos analistas senior en Lima o un analista senior en Panamá. Sugerencia: calcule previamente el número de elementos del espacio muestral.

Profesores del curso:

Cristian Bayes / Ana Valdivia / Enver Tarazona / Miluska Osorio / Stefany Necioşup / Jim Silvestre

San Miguel, 25 de abril de 2023



## Solución PC2

1) Coeficientes :  $a = 17,697884$   
 $b = -0,199091$

### a) Interpretación:

a : Cuando el porcentaje de habitantes que son usuarios de bicicleta es 0%, entonces el porcentaje de habitantes que presentan una enfermedad cardíaca es 17,70% . aprox. ✓

b : Cuando " " bicicleta aumenta en 1%, entonces ... cardíaca disminuye en 0,20% . aprox. ✓

### b) Falso. ✓

Si bien el coeficiente de determinación ( $R^2$ ) es 0,8751, habla de variabilidad (no lo predice sin error).

c) X : % de habitantes que fuman  
Y : % de habitantes que presentan una enfermedad cardíaca

Datos:  $\bar{X} = 15,4$        $S_y^2 = 20,9$   
 $\bar{Y} = 10,2$        $S_{xy} = 11,7$   
 $S_x^2 = 68,7$

Se cumple :

↑ coeficiente de correlación de Pearson

$$r = \frac{S_{xy}}{S_x S_y} = \frac{11,7}{(\sqrt{68,7}) \cdot \sqrt{20,9}} = 0,3088$$

La ecuación de regresión estimada es :

$$\hat{Y} = a + bX$$

Dónde:  $b = r \cdot \frac{S_y}{S_x} = 0,1703$

$$a = \bar{Y} - b\bar{X} = 7,5774$$

Sys



Modelo de regresión estimado:

$$\hat{y} = 7,5774 + 0,1703 \cdot x$$

Se cumple que  $R^2 = r^2$

$$R^2 = 0,0953$$

Como el coeficiente de determinación ( $R^2$ ) es mayor en el modelo anterior (0,8751), el modelo que explica mejor el % de habitantes que presentan una enfermedad cardíaca es el de los usuarios de bicicleta.

d) Modelo elegido en c)

$$\hat{y} = 17,6979 + 0,1991 \cdot x$$

$$x = 30$$

$$\Rightarrow \hat{y} = 11,72$$

∴ El porcentaje de habitantes que presentan una enfermedad cardíaca en una localidad donde el 30% son usuarios de bicicleta es 11,72%.

2)

M: el médico elegido es de sexo masculino

E: " " Tiene especialidad

I: " " Tiene un ingreso mayor a 5000

$$* P(M) = 0,7470$$

$$P(E) = 0,6650$$

$$P(I) = 0,5170$$

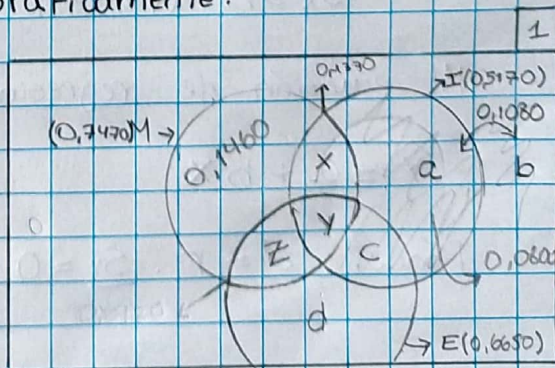
Gráficamente:

$$* P(M \cap I) = 0,0800$$

$$P(M \cup E) = 0,8920$$

$$P(M \cap E \cap I) = 0,1460$$

Evento: elegir uno al azar



Sys



a) Piden:  $P(M^c \cap E^c) = ?$

Se cumple que  $M^c \cap E^c = (M \cup E)^c$

Así,  $P((M \cup E)^c) = 1 - P(M \cup E)$

$= 1 - 0,8920$

$= 0,1080 = 10,80\%$

b) Piden:  $P(M \cap I) = ?$

Se cumple:

$P(I) = P(M \cap I) + P(M^c \cap I)$

$0,5170 = x + 0,0800$

$P(M \cap I) = 0,4370$

c) Piden:  $P(M \cap E^c \cap I) = ?$

$x : P(M \cap E^c \cap I)$

$y : P(M \cap E \cap I)$

Se cumple:

$z : P(M \cap E \cap I^c)$

$x + y = 0,4370 \rightarrow P(M \cap I)$

$x + y + z = 0,7470 - 0,1460$

$z = 0,1640$

a :  $P(M^c \cap E^c \cap I)$  d :  $P(M^c \cap E \cap I^c)$

b :  $P(M^c \cap E^c \cap I^c)$

c :  $P(M^c \cap E \cap I)$

Se cumple:  $0,1080 = x + y + z$

$a + b = 0,1080$

$a + c = 0,0800$

$0,7470 + c + d = 0,8920$

$c + d = 0,1450$

$z + y + c + d = 0,6650$

$y = 0,6650 - 0,1450 - 0,1640$

$y = 0,3560$

Sys



$$\therefore X + Y = 0,4370$$

$$X = 0,437 - 0,356$$

$$X = 0,081$$

$$\Rightarrow P(M \cap E' \cap I) = \underline{0,0810} //$$

3)

	S	J	A	S : analista senior
				J : " junior
				A : asistente
Lima	5	3	2	
Panamá	1	3	4	

a) Evento : seleccionar 5 y sin reemplazo

X : # asistentes seleccionados

Piden  $X > 3$  :

$$(X = 4) \text{ o } (X = 5)$$

4 A y 1 restante + 5 A  $\rightarrow$  casos favorables

$$\frac{{}^6C_4 \cdot {}^{12}C_1 + {}^6C_5}{{}^8C_5} = \frac{31}{1428} = \underline{0,0217} //$$

b) Evento : seleccionar 3 de cada sede y sin reemplazo

X : # seniors en Lima

Usamos:  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Y : # seniors en Panamá

La intersección:

$$X = 2 \text{ y } Y = 1$$

Piden:  $X = 2$  o  $Y = 1$

$\hookrightarrow$  2 S y 1 restante

$$\text{En Lima: } \frac{{}^5C_2 \cdot {}^5C_1}{{}^{10}C_3} = \frac{5}{12}$$

$$\text{En Panamá: } \frac{{}^1C_1 \cdot {}^7C_2}{{}^8C_3} = \frac{3}{8}$$

$$\therefore \frac{5}{12} + \frac{3}{8} - \left( \frac{5}{12} \cdot \frac{3}{8} \right) = \frac{61}{96} = \underline{0,6354} //$$

Sys