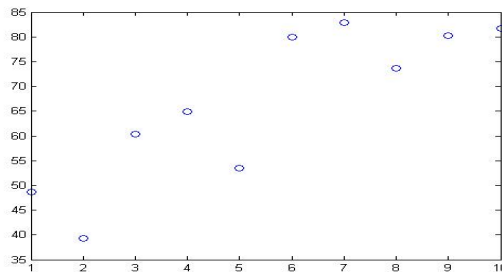


Problema 1

(a) Figura 1: d) fuerte positivo ($r=0.8$); Figura 2: e) fuerte negativa ($r=-0.95$); Figura 3: f) no permite interpretación (hay un punto atípico en la parte inferior derecha; Figura 4: a) muy cercano a cero; Figura 5: f) no permite interpretación (hay dos poblaciones); Figura 6: f) no permite interpretación (la relación es parabólica).

$$(b) \text{cov}(x, y) = \frac{1}{10} \sum x_i y_i - \bar{x} \bar{y} = 36,545$$

$r = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x) \cdot \text{var}(y)}} = 0,86$. Es un coeficiente alto positivo. Para 7:30 horas deberíamos tener un valor relativamente mayor que y_0 .



(c) El coeficiente de correlación no cambia en el primer caso y en el segundo cambia de signo.

Problema 2

(a) Se trata de un test de independencia de dos variables categóricas. Bajo la hipótesis de independencia la tabla de frecuencias es (Tabla 1):

Cuadro 1: Cantidad de vacas por producción de leche y raza. Entre paréntesis las cantidades teóricas bajo la hipótesis de independencia.

Producción Raza	Raza 1	Raza 2	Total
De 10 a 12 litros	27 (19.8)	7 (14.2)	34
De 12 a 15 litros	7 (15.17)	19 (10.83)	26
De 15 a 18 litros	7 (14)	17 (10)	24
De 18 a 20 litros	29 (21)	7 (15)	36
Total	70	50	120

En consecuencia, se tiene que

$$Q = \frac{(27 - 19,8)^2}{19,8} + \frac{(7 - 14,2)^2}{14,2} + \dots + \frac{(7 - 15)^2}{15} = 32,5$$

- (b) El estadístico Q (estadístico de Pearson del contraste de independencia) se interpreta como una medida de diferencia entre las frecuencias empíricas de datos en cada combinación de categorías y las frecuencias teóricas asumiendo independencia.

Bajo la hipótesis nula, de independencia, Q debe seguir una distribución $\chi^2_{(4-1)(2-1)} = \chi^2_3$. Como $Q > 7,81$ y $P(\chi^2_3 > 7,81) = 0,05$ el p-valor es menor a 0,05 y se rechaza la hipótesis de independencia para este nivel de significación.

- (c) Se trata de un test de bondad de ajuste, en que la hipótesis nula es que la cantidad de leche de las vacas de raza 1 (X) presenta la densidad especificada en el enunciado. Los datos vienen dados por la Tabla 2, los cálculos de los valores teóricos se muestran abajo:

Cuadro 2: Datos de cantidad de vacas (Raza 1) por intervalo de producción de leche.

X=Cantidad de leche de vacas de Raza 1	[10;12]	(12;15]	(15;18]	(18;20]
n_i	27	7	7	29
np_i	27.44	7.56	7.56	27.44

$$p_1 = P(10 \leq x \leq 12) = \int_{10}^{12} \frac{3}{250}(x-15)^2 dx = \frac{98}{250} \rightarrow np_1 = 27,44$$

$$p_2 = P(12 < x \leq 15) = \int_{12}^{15} \frac{3}{250}(x-15)^2 dx = \frac{27}{250} \rightarrow np_2 = 7,56$$

$$p_3 = P(15 < x \leq 18) = \int_{15}^{18} \frac{3}{250}(x-15)^2 dx = \frac{27}{250} \rightarrow np_3 = 7,56$$

$$p_4 = 1 - p_1 - p_2 - p_3 = \frac{98}{250} \rightarrow np_4 = 27,44$$

Con ello el valor del estadístico de contraste es $Q = \frac{(27-27,44)^2}{27,44} + \dots + \frac{(29-27,44)^2}{27,44} = 0,178$. Bajo H_0 Q sigue distribución $\chi^2_{4-1} = \chi^2_3$. Como $Q < 7,81$ y $P(\chi^2_3 > 7,81) = 0,05$ el p-valor es mayor a 0,05 y se acepta que la cantidad de leche que producen las vacas de la raza 1 sigue la distribución propuesta, para este nivel de significación.

Problema 3

	Fuente variación	grados libertad	Suma cuadrados	Cuadrado Medio	F	p-valor
(a)	Grupos	2	894.3459	447.1729	52.5946	0.000
	Residuos	272	2312.615	8.5023		
	Total	274	3206.9609			

Se supone que el rendimiento sigue una distribución normal de medias diferentes entre los grupos pero de misma varianza.

El p-valor es nulo: Se concluye que hay diferencia entre los tres grupos.

- (b) Las varianzas insesgados son los cuadrados medios: inter-grupo: 447.17 intra-grupo: 8.50

- (c) La razón de correlación es igual al cociente de la varianza inter-grupo y de la varianza total: $894.3459/3206.9609=0.2789$, lo que no es tan alto.
- (d) El gráfico de cajas muestra las distribuciones de los tres grupos. En la caja hay 50 % de los datos que están alrededor de la mediana; la baja inferior representa el 25 % más bajo y la baja superior el 25 % más alto.
- El p-valor de la tabla ANOVA nos dice que hay algo de diferencias, el gráfico de cajas nos dice cuáles son. Se observa que el fertilizante F2 tiene mejor rendimiento que los otros grupos que son parecidos.
- (e) Si la razón de correlación vale 1, existe una relación funcional del grupo hacia el rendimiento. Si vale 0, no existe; y si vale 0.5 hay una cierta tendencia que relaciona el grupo al rendimiento positivamente.