Prueba 1 SOL114

Problema I. (10%)

Supongamos que existe una enfermedad rara que afecta al 0.1% de la población. Hay un test médico para detectar esta enfermedad que tiene las siguientes características:

- Probabilidad de que el test sea positivo si una persona tiene la enfermedad:

$$P(\text{Positivo}|\text{Enfermedad}) = 0.99$$

- Probabilidad de que el test sea positivo si una persona no tiene la enfermedad:

$$P(\text{Positivo}|\text{No Enfermedad}) = 0.05$$

Una persona se hace el test y el resultado es positivo. Usa el Teorema de Bayes para determinar cuál es la probabilidad de que esta persona realmente tenga la enfermedad.Es decir: $P(\text{Enfermedad}|\text{Positivo})$.

Solución:

- Probabilidad de tener la enfermedad: $P(\text{Enfermedad}) = 0.001$ (0.1% de la población).
- Probabilidad de que el test sea positivo si tienes la enfermedad: $P(\text{Positivo}|\text{Enfermedad}) = 0.99$.
- Probabilidad de no tener la enfermedad: $P(\text{No Enfermedad}) = 1 - 0.001 = 0.999$.

- Probabilidad de que el test sea positivo si no tienes la enfermedad: $P(\text{Positivo}|\text{No Enfermedad}) = 0.05$.

Luego, la probabilidad de obtener un resultado positivo en general se calcula considerando tanto a las personas que tienen la enfermedad como a las que no la tienen:

$$P(\text{Positivo}) = P(\text{Positivo}|\text{Enfermedad})\times P(\text{Enfermedad})+P(\text{Positivo}|\text{No Enfermedad})\times P(\text{No Enfermedad})$$

Sustituyendo los valores:

$$P(\text{Positivo}) = (0.99 \times 0.001) + (0.05 \times 0.999)$$

$$P(\text{Positivo}) = 0.00099 + 0.04995 = 0.05094$$

Aplicabdo el Teorema de Bayes encontramos la probabilidad de tener la enfermedad dado que el resultado del test es positivo:

$$P(\text{Enfermedad}|\text{Positivo}) = \frac{P(\text{Positivo}|\text{Enfermedad}) \times P(\text{Enfermedad})}{P(\text{Positivo})}$$

Sustituyendo los valores:

$$P(\text{Enfermedad}|\text{Positivo}) = \frac{0.99 \times 0.001}{0.05094} \approx \frac{0.00099}{0.05094} \approx 0.01943$$

Luego, la probabilidad de que la persona realmente tenga la enfermedad dado que el test resultó positivo es aproximadamente **1.94%**. Aunque el test es muy preciso, debido a que la enfermedad es muy rara, la probabilidad de realmente tenerla después de un resultado positivo es bastante baja.

Problema II:

Una empresa está considerando lanzar un nuevo producto al mercado. Antes de hacerlo, realizan una encuesta entre sus consumidores potenciales. Cada encuestado puede dar una respuesta "Positiva" o "Negativa" respecto al producto. Supongamos que la probabilidad de que un encuestado elegido al azar dé una respuesta "Positiva" es r.

Parte i. (5%) Si la respuesta de un encuestado es un evento aleatorio, describe el espacio muestral.

Solución:
El espacio muestral es el conjunto de todos los posibles resultados de la respuesta de un encuestado. En este caso, un encuestado puede dar una respuesta "Positiva" o "Negativa".

El espacio muestral Ω se define como:

$$\Omega = \{\text{"Positiva"}, \text{"Negativa"}\}$$

Esto significa que el espacio muestral consta de dos posibles resultados: que un encuestado dé una respuesta "Positiva" o "Negativa".

Parte ii. (7%) Indica que distribución de probabilidad describe adecuadamente cada respuesta individual. Especifica su función de masa probabilística y explica cada uno de los componentes de la expresión.

Solución:

La distribución de probabilidad que describe adecuadamente cada respuesta individual en este contexto es la **Distribución de Bernoulli**. Esta distribución se utiliza cuando hay un experimento aleatorio con dos posibles resultados: éxito (en este caso, una respuesta "Positiva") y fracaso (una respuesta "Negativa").

La función de masa de probabilidad (PMF) de una variable X que distribuye Bernoulli es:

$$P(X = x) = \begin{cases} r & \text{si } x = 1 \text{ ("Positiva")} \\ 1 - r & \text{si } x = 0 \text{ ("Negativa")} \end{cases}$$

donde

- r: Es la probabilidad de que un encuestado elegido al azar dé una respuesta "Positiva". Esto corresponde a $P(X = 1)$.

- $1 - r$: Es la probabilidad de que un encuestado dé una respuesta "Negativa". Esto corresponde a $P(X = 0)$.
- $X = 1$: Indica que el encuestado ha dado una respuesta "Positiva".
- $X = 0$: Indica que el encuestado ha dado una respuesta "Negativa".

Parte iii. (13%) Dado que r es la probabilidad de que un encuestado seleccionado al azar dé una respuesta "Positiva", expresa la probabilidad de que, al observar las respuestas de 4 encuestados seleccionados al azar (de la misma población y cuyas respuestas son independientes entre si), exactamente 2 den una respuesta "Positiva" y los otros 2 den una respuesta "Negativa". Explica cada uno de los componentes de la expresión.

Solución:

Este problema se puede resolver utilizando la distribución binomial, ya que estamos considerando un número fijo de ensayos (4 encuestados) y cada ensayo tiene dos posibles resultados ("Positiva" o "Negativa"). La probabilidad de que exactamente 2 de los 4 encuestados den una respuesta "Positiva" es:

$$P(X = 2) = \binom{4}{2} r^2 (1 - r)^2$$

Donde: - $\binom{4}{2}$ es el número de combinaciones posibles de 2 respuestas "Positivas" entre 4 encuestados, que es igual a $\frac{4!}{2!(4-2)!} = 6$. - r^2 es la probabilidad de que 2 encuestados den una respuesta "Positiva". - $(1 - r)^2$ es la probabilidad de que los otros 2 encuestados den una respuesta "Negativa".

Por lo tanto, la probabilidad es:

$$P(X = 2) = 6 \times r^2 \times (1 - r)^2$$

Parte iv. (5%) Supongamos que las respuestas "Positivas" y "Negativas" no están influenciadas por el género del encuestado, y que el 60% de los encuestados son mujeres. Determina la probabilidad $P(\text{Mujer} \mid \text{"Positiva"})$.

Solución:

En este caso, se nos dice que la probabilidad de dar una respuesta "Positiva" es independiente del género del encuestado. Esto significa que, dado que un encuestado dio una respuesta "Positiva", la probabilidad de que sea una mujer es simplemente la probabilidad de que cualquier encuestado sea una mujer, sin importar la respuesta.

Podemos usar la definición de independencia para expresar esto:

$$P(\text{Mujer} \mid \text{"Positiva"}) = P(\text{Mujer})$$

Dado que el 60% de los encuestados son mujeres:

$$P(\text{Mujer}) = 0.6$$

La probabilidad de que un encuestado sea una mujer dado que dio una respuesta “Positiva” es **60%**, que es exactamente la proporción de mujeres en la muestra. Esto se debe a que la respuesta “Positiva” no influye en la probabilidad de que un encuestado sea mujer, reflejando la independencia entre la respuesta y el género.

Problema III:

Una socióloga está estudiando la brecha salarial entre hombres y mujeres para el año 2024. Su estudio define la brecha salarial G como la diferencia entre los ingresos de los hombres y los ingresos de las mujeres:

$$G = Y_h - Y_w$$

Donde: - Y_h es la variable aleatoria que representa los ingresos de los hombres, la cual sigue una distribución normal con media μ_h y desviación estándar σ_h.

- $Y_w $es la variable aleatoria que representa los ingresos de las mujeres, la cual sigue una distribución normal con media μ_w y desviación estándar σ_w.

Específicamente, los ingresos de hombres y mujeres tienen las siguientes características:

- $Y_h \sim \text{Normal}(\mu_h = 25, \sigma_h = 2.45)$ (en miles de pesos)
- $Y_w \sim \text{Normal}(\mu_w = 20, \sigma_w = 3.16)$ (en miles de pesos)

Parte i. (10%) Determinar la distribución de la brecha salarial G: tipo de distribución, valor esperado y desviación estándard de G.

Recordar: Si Y_h y Y_w son variables normales independientes, entonces la diferencia $G = Y_h - Y_w$ también sigue una distribución normal.

Solución:

Si Y_h y Y_w son variables normales independientes, entonces la diferencia $G = Y_h - Y_w$ también sigue una distribución normal:

$$G \sim \text{Normal}(\mu_G, \sigma_G^2)$$

Donde:

- El valor esperado de G es:

$$\mu_G = E(G) = E(Y_h - Y_w) = E(Y_h) - E(Y_w) = \mu_h - \mu_w$$

- La varianza de G es:

$$\sigma_G^2 = \text{Var}(G) = \text{Var}(Y_h - Y_w) = \text{Var}(Y_h) + \text{Var}(Y_w) = \sigma_h^2 + \sigma_w^2$$

Y la desviación estándar de G es:

$$\sigma_G = \sqrt{\sigma_h^2 + \sigma_w^2}$$

Luego,

- $$\mu_G = \mu_h - \mu_w = 25 - 20 = 5$$
- $$\sigma_G = \sqrt{(2.45)^2 + (3.16)^2} = \sqrt{6 + 10} = \sqrt{16} = 4$$

Por tanto, la brecha salarial G sigue una distribución normal $G \sim \text{Normal}(\mu_G = 5, \sigma_G = 4)$, donde la media es 5 mil pesos chilenos y una desviación estándard de 4 mil pesos.

Parte ii. (5%) Expresa matemáticamente cómo determinar la probabilidad de la brecha salarial se encuentre entre 3 y 10 mil pesos: $P(3 < G < 10)$ y explica en palabras la intuición detrás de esta expresión.

`*Nota: Hay una integral involucrada. No deben resolver la integral.*`

Solución:

Queremos encontrar la probabilidad de que la brecha salarial G, que es la diferencia entre los ingresos de los hombres y las mujeres, se encuentre entre 3 y 10 miles de pesos. Sabemos que G sigue una distribución normal con media $\mu_G = 5$ miles de pesos chilenos y desviación estándar $\sigma_G = 4$ miles de pesos chilenos. Esto se denota como:

$$G \sim N(\mu_G = 5, \sigma_G^2 = 16)$$

La probabilidad de que G esté entre 3 y 10 se denota matemáticamente como $P(3 < G < 10)$. Esta probabilidad es el área bajo la curva de la distribución normal de G entre los puntos 3 y 10. Para calcular esta probabilidad, necesitamos integrar la función de densidad de probabilidad de G desde 3 hasta 10. La función de densidad de probabilidad para una variable normal se expresa como:

$$f_G(x) = \frac{1}{\sigma_G\sqrt{2\pi}} \exp\left(-\frac{(x-\mu_G)^2}{2\sigma_G^2}\right)$$

Entonces, la probabilidad $P(3 < G < 10)$ se calcula mediante la siguiente integral:

$$P(3 < G < 10) = \int_3^{10} \frac{1}{4\sqrt{2\pi}} \exp\left(-\frac{(x-5)^2}{2(16)}\right) dx$$

Esta integral representa el área bajo la curva de la distribución normal de G entre los valores 3 y 10. Dado que la función de densidad de probabilidad es una función de la distribución normal, calcular esta integralnos da la probabilidad exacta de que la brecha salarial se encuentre dentro de este rango.

Parte iii. (10%) Calcula la probabilidad de que el salario de una mujer sea mayor que el de un hombre.

Nota:Para realizar este cálculo resulta conveniente usar la función Φ después de estandar la variable G

Solución:

Para calcular la probabilidad de que el salario de una mujer sea mayor que el de un hombre, necesitamos encontrar la probabilidad de que la brecha salarial $G = Y_h - Y_w$ sea menor que 0. Esto es:

$$P(Y_w > Y_h) = P(G < 0)$$

Dado que G sigue una distribución normal $G \sim N(\mu_G, \sigma_G^2)$ con $\mu_G = 5$ y $\sigma_G = 4$, queremos encontrar:

$$P(G < 0)$$

Para hacer esto, primero estandarizamos la variable G utilizando la transformación Z, que es una variable normal estándar $Z \sim N(0, 1)$. La transformación se realiza como sigue:

$$Z = \frac{G - \mu_G}{\sigma_G}$$

Sustituyendo los valores de μ_G y σ_G:

$$Z = \frac{G-5}{4}$$

Ahora, queremos calcular $P(G < 0)$, que se convierte en:

$$P\left(\frac{G-5}{4} < \frac{0-5}{4}\right) = P\left(Z < \frac{-5}{4}\right) = P\left(Z < -1.25\right)$$

Esta probabilidad se puede obtener usando la función de distribución acumulativa de la normal estándar, denotada como $\Phi(z)$. Por lo tanto:

$$P(G < 0) = \Phi(-1.25)$$

Usando tablas de la distribución normal estándar o una calculadora, encontramos que:

$$\Phi(-1.25) \approx 0.1056$$

Luego, la probabilidad de que el salario de una mujer sea mayor que el de un hombre es aproximadamente **0.1056**, o **10.56%**.

Parte iv. (10%) Calcula el salario mínimos de un hombre y una mujer que se encuentran en el 10% más rico de las personas de sus respectivo género.

Solución:

Para calcular el salario mínimo de un hombre y una mujer que se encuentran en el 10% más rico de las personas de su respectivo género, necesitamos encontrar el percentil 90 (es decir, el valor por debajo del cual se encuentra el 90% de las observaciones) en la distribución de ingresos de cada género.

Los ingresos de los hombres siguen una distribución normal:

$$Y_h \sim N(\mu_h = 25, \sigma_h = 2.45) \text{ (miles de pesos)}$$

Queremos encontrar el valor y_h^* tal que:

$$P(Y_h > y_h^*) = 0.10$$

Equivalente a encontrar el percentil 90, es decir, el valor y_h^* tal que:

$$P(Y_h \leq y_h^*) = 0.90$$

Esto se traduce en:

$$y_h^* = \mu_h + z_{0.90} \cdot \sigma_h$$

Donde $z_{0.90}$ es el valor correspondiente al percentil 90 de la distribución normal estándar $Z \sim N(0,1)$, que es aproximadamente 1.28. Sustituyendo los valores:

$$y_h^* = 25 + 1.28 \times 2.45 \approx 25 + 3.14 \approx 28.14 \text{ (miles de pesos)}$$

Por su parte, los ingresos de las mujeres siguen una distribución normal:

$$Y_w \sim N(\mu_w = 20, \sigma_w = 3.16) \text{ (miles de pesos)}$$

De manera similar, queremos encontrar el valor y_w^* tal que:

$$P(Y_w > y_w^*) = 0.10$$

O equivalentemente:

$$P(Y_w \leq y_w^*) = 0.90$$

Esto se traduce en:

$$y_w^* = \mu_w + z_{0.90} \cdot \sigma_w$$

Sustituyendo los valores:

$$y_w^* = 20 + 1.28 \times 3.16 \approx 20 + 4.04 \approx 24.04 \text{ (miles de pesos)}$$

Luego,

- El salario mínimo de un hombre que se encuentra en el 10% más rico de su género es **28.14 mil pesos**.
- El salario mínimo de una mujer que se encuentra en el 10% más rico de su género es **24.04 mil pesos**.

Problema VI: Verdadero o Falso (5% cada uno, 25% total)

Justificar brevemente en caso de indicar Falso.

1. La función de densidad de probabilidad $f(x)$ de una variable aleatoria continua X puede ser mayor que 1 para algunos valores de x.

    - **Verdadero**. La función de densidad de probabilidad $f(x)$ puede tomar valores mayores que 1, siempre y cuando la integral total $\int_{-\infty}^{\infty} f(x)\,dx = 1$ se mantenga. Esto puede ocurrir, por ejemplo, en distribuciones con soporte en intervalos muy pequeños.

2. La función de distribución acumulativa $F(x)$ de una variable aleatoria continua es la derivada de la función de densidad de probabilidad $f(x)$.

    - **Falso**. La función de distribución acumulativa $F(x)$ es la integral de la función de densidad $f(x)$ desde (-∞) hasta x. Es decir, $F(x) = \int_{-\infty}^{x} f(t)\,dt$. La derivada de $F(x)$ con respecto a x es $f(x)$.

3. Si dos eventos A y B son independientes, entonces $P(A \cap B) = P(A) \times P(B)$.

    - **Verdadero**. Esta es la definición de independencia entre dos eventos. La probabilidad de que ambos eventos ocurran es el producto de sus probabilidades individuales si los eventos son independientes.

4. El valor de $f(5)$ para una variable aleatoria continua X representa la probabilidad de que X tome exactamente el valor 5.

    - **Falso**. Para una variable aleatoria continua, $f(5)$ no representa la probabilidad de que X tome exactamente el valor 5, ya que esa probabilidad es 0. $f(5)$ representa la densidad de probabilidad en 5, es decir, cuánto contribuye ese punto a la probabilidad en un pequeño intervalo alrededor de 5.

5. El valor de $F(5)$ para una variable aleatoria continua X representa la probabilidad de que X sea menor o igual a 5.

    - **Verdadero**. La función de distribución acumulativa $F(5)$ nos da la probabilidad de que la variable aleatoria continua X tome un valor menor o igual a 5.

Probabilidad acumulada en distribución Normal Standard (mitad positiva)

Z=z	$\Phi(z)$
0.00	0.50000
0.01	0.50399
0.02	0.50798
0.03	0.51197
0.04	0.51595
0.05	0.51994
0.06	0.52392
0.07	0.52790
0.08	0.53188
0.09	0.53586
0.10	0.53983
0.11	0.54380
0.12	0.54776
0.13	0.55172
0.14	0.55567
0.15	0.55962
0.16	0.56356
0.17	0.56749
0.18	0.57142
0.19	0.57535
0.20	0.57926
0.21	0.58317
0.22	0.58706
0.23	0.59095
0.24	0.59483
0.25	0.59871
0.26	0.60257
0.27	0.60642
0.28	0.61026
0.29	0.61409
0.30	0.61791
0.31	0.62172
0.32	0.62552
0.33	0.62930
0.34	0.63307
0.35	0.63683

Z=z	$\Phi(z)$
0.36	0.64058
0.37	0.64431
0.38	0.64803
0.39	0.65173
0.40	0.65542
0.41	0.65910
0.42	0.66276
0.43	0.66640
0.44	0.67003
0.45	0.67364
0.46	0.67724
0.47	0.68082
0.48	0.68439
0.49	0.68793
0.50	0.69146
0.51	0.69497
0.52	0.69847
0.53	0.70194
0.54	0.70540
0.55	0.70884
0.56	0.71226
0.57	0.71566
0.58	0.71904
0.59	0.72240
0.60	0.72575
0.61	0.72907
0.62	0.73237
0.63	0.73565
0.64	0.73891
0.65	0.74215
0.66	0.74537
0.67	0.74857
0.68	0.75175
0.69	0.75490
0.70	0.75804
0.71	0.76115
0.72	0.76424
0.73	0.76730
0.74	0.77035

Z=z	$\Phi(z)$
0.75	0.77337
0.76	0.77637
0.77	0.77935
0.78	0.78230
0.79	0.78524
0.80	0.78814
0.81	0.79103
0.82	0.79389
0.83	0.79673
0.84	0.79955
0.85	0.80234
0.86	0.80511
0.87	0.80785
0.88	0.81057
0.89	0.81327
0.90	0.81594
0.91	0.81859
0.92	0.82121
0.93	0.82381
0.94	0.82639
0.95	0.82894
0.96	0.83147
0.97	0.83398
0.98	0.83646
0.99	0.83891
1.00	0.84134
1.01	0.84375
1.02	0.84614
1.03	0.84849
1.04	0.85083
1.05	0.85314
1.06	0.85543
1.07	0.85769
1.08	0.85993
1.09	0.86214
1.10	0.86433
1.11	0.86650
1.12	0.86864
1.13	0.87076

Z=z	$\Phi(z)$
1.14	0.87286
1.15	0.87493
1.16	0.87698
1.17	0.87900
1.18	0.88100
1.19	0.88298
1.20	0.88493
1.21	0.88686
1.22	0.88877
1.23	0.89065
1.24	0.89251
1.25	0.89435
1.26	0.89617
1.27	0.89796
1.28	0.89973
1.29	0.90147
1.30	0.90320
1.31	0.90490
1.32	0.90658
1.33	0.90824
1.34	0.90988
1.35	0.91149
1.36	0.91309
1.37	0.91466
1.38	0.91621
1.39	0.91774
1.40	0.91924
1.41	0.92073
1.42	0.92220
1.43	0.92364
1.44	0.92507
1.45	0.92647
1.46	0.92785
1.47	0.92922
1.48	0.93056
1.49	0.93189
1.50	0.93319
1.51	0.93448
1.52	0.93574

Z=z	$\Phi(z)$
1.53	0.93699
1.54	0.93822
1.55	0.93943
1.56	0.94062
1.57	0.94179
1.58	0.94295
1.59	0.94408
1.60	0.94520
1.61	0.94630
1.62	0.94738
1.63	0.94845
1.64	0.94950
1.65	0.95053
1.66	0.95154
1.67	0.95254
1.68	0.95352
1.69	0.95449
1.70	0.95543
1.71	0.95637
1.72	0.95728
1.73	0.95818
1.74	0.95907
1.75	0.95994
1.76	0.96080
1.77	0.96164
1.78	0.96246
1.79	0.96327
1.80	0.96407
1.81	0.96485
1.82	0.96562
1.83	0.96638
1.84	0.96712
1.85	0.96784
1.86	0.96856
1.87	0.96926
1.88	0.96995
1.89	0.97062
1.90	0.97128
1.91	0.97193

Z=z	$\Phi(z)$
1.92	0.97257
1.93	0.97320
1.94	0.97381
1.95	0.97441
1.96	0.97500
1.97	0.97558
1.98	0.97615
1.99	0.97670
2.00	0.97725
2.01	0.97778
2.02	0.97831
2.03	0.97882
2.04	0.97932
2.05	0.97982
2.06	0.98030
2.07	0.98077
2.08	0.98124
2.09	0.98169
2.10	0.98214
2.11	0.98257
2.12	0.98300
2.13	0.98341
2.14	0.98382
2.15	0.98422
2.16	0.98461
2.17	0.98500
2.18	0.98537
2.19	0.98574
2.20	0.98610
2.21	0.98645
2.22	0.98679
2.23	0.98713
2.24	0.98745
2.25	0.98778
2.26	0.98809
2.27	0.98840
2.28	0.98870
2.29	0.98899
2.30	0.98928

Z=z	$\Phi(z)$
2.31	0.98956
2.32	0.98983
2.33	0.99010
2.34	0.99036
2.35	0.99061
2.36	0.99086
2.37	0.99111
2.38	0.99134
2.39	0.99158
2.40	0.99180
2.41	0.99202
2.42	0.99224
2.43	0.99245
2.44	0.99266
2.45	0.99286
2.46	0.99305
2.47	0.99324
2.48	0.99343
2.49	0.99361
2.50	0.99379
2.51	0.99396
2.52	0.99413
2.53	0.99430
2.54	0.99446
2.55	0.99461
2.56	0.99477
2.57	0.99492
2.58	0.99506
2.59	0.99520
2.60	0.99534
2.61	0.99547
2.62	0.99560
2.63	0.99573
2.64	0.99585
2.65	0.99598
2.66	0.99609
2.67	0.99621
2.68	0.99632
2.69	0.99643

Z=z	$\Phi(z)$
2.70	0.99653
2.71	0.99664
2.72	0.99674
2.73	0.99683
2.74	0.99693
2.75	0.99702
2.76	0.99711
2.77	0.99720
2.78	0.99728
2.79	0.99736
2.80	0.99744
2.81	0.99752
2.82	0.99760
2.83	0.99767
2.84	0.99774
2.85	0.99781
2.86	0.99788
2.87	0.99795
2.88	0.99801
2.89	0.99807
2.90	0.99813
2.91	0.99819
2.92	0.99825
2.93	0.99831
2.94	0.99836
2.95	0.99841
2.96	0.99846
2.97	0.99851
2.98	0.99856
2.99	0.99861
3.00	0.99865