FORMULARIO

ESTADÍSTICA DESCRIPTIVA

Datos: $\{x_1, x_2, ..., x_N\}$

Media: $\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$

Mediana: $\{x_1, x_2, ..., x_N\}$ ordenados de menor a mayor

$$Me = \begin{cases} x_{\frac{N+1}{2}} & \text{si } N \text{ es impar} \\ \\ \frac{x_{\frac{N}{2}} + x_{\frac{N}{2}+1}}{2} & \text{si } N \text{ es par} \end{cases}$$

Moda: Valor muestral con la frecuencia más alta

Percentiles: k-ésimo percentil

1. Ordenar las N observaciones de menor a mayor
2. Calcular $\frac{N \cdot k}{100}$
   (a) Si $\frac{N \cdot k}{100}$ no es un entero: considerar el entero inmediato posterior y determinar el valor ordenado correspondiente
   (b) Si $\frac{N \cdot k}{100}$ es un entero, digamos j: calcular la media de las observaciones ordenadas j-ésima y $(j+1)$-ésima

Rango: $\max_{i=1,...,N}\{x_i\}$ - $\min_{i=1,...,N}\{x_i\}$

Rango intercuartílico: Diferencia entre el tercer y primer cuartil

Varianza: $s^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1} = \frac{\sum_{i=1}^{N} x_i^2 - N\bar{x}^2}{N-1}$

Desviación típica: s $= \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}} = \sqrt{\frac{\sum_{i=1}^{N} x_i^2 - N\bar{x}^2}{N-1}}$

Coeficiente de variación: $\frac{s}{\bar{x}}$

Coeficiente de asimetría: $CA = \frac{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^3}{N-1}}{s^3}$

Coeficiente de curtosis: $CC = \frac{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^4}{N-1}}{s^4}$

Datos: $\{(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)\}$

Covarianza: $s_{xy} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{N-1} = \frac{\sum_{i=1}^{N} x_i y_i - N\bar{x}\bar{y}}{N-1}$

Coeficiente de correlación: $r_{xy} = \frac{s_{xy}}{s_x s_y}$

Recta de regresión de Y sobre X:

$$y \text{ - } \bar{y} = r_{xy}\frac{s_y}{s_x}(x - \bar{x})$$

$$Y = a + b \cdot X :$$

$$b = \frac{N\sum_{i=1}^{N} x_i \cdot y_i - (\sum_{i=1}^{N} x_i) \cdot (\sum_{i=1}^{N} y_i)}{N(\sum_{i=1}^{N} x_i^2) - (\sum_{i=1}^{N} x_i)^2}; \quad a = \frac{\sum_{i=1}^{N} y_i - b(\sum_{i=1}^{N} x_i)}{N}$$

Coeficiente de determinación: $R^2 = r_{xy}^2$

X \ Y	y_1	...	y_j	...	y_c	Total
x_1	o_{11}	...	o_{1j}	...	o_{1c}	$T_{1.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
x_i	o_{i1}	...	o_{ij}	...	o_{ic}	$T_{i.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
x_r	o_{r1}	...	o_{rj}	...	o_{rc}	$T_{1.}$
Total	$T_{.1}$	...	$T_{.j}$	...	$T_{.c}$	T

$T_{i.}$ es el total de observaciones de la fila i-ésima, $T_{.j}$ es el total de observaciones de la columna j-ésima y T es el total de observaciones.

El coeficiente de contingencia es:

$$C = \sqrt{\frac{\chi^2}{T + \chi^2}}$$

donde χ^2 se calcula de la siguiente forma:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(o_{ij} - e_{ij})^2}{e_{ij}},$$

siendo $e_{ij} = T_{i.} \cdot T_{.j}$ / T

PROBABILIDAD

Leyes de DeMorgan: $(\bigcap_i A_i)^c = \bigcup_i A_i^c$ y $(\bigcup_i A_i)^c = \bigcap_i A_i^c$

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

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

$P(A_1 \cap A_2 \cap ... \cap A_k) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \ldots \cdot P(A_k|A_1 \cap A_2 \cap ... \cap A_{k-1})$

A_1, A_2,..., A_k independientes: $P(A_1 \cap A_2 \cap ... \cap A_k) = P(A_1) \cdot P(A_2) \cdot P(A_3) \cdot \ldots \cdot P(A_k)$

Teorema de la probabilidad total: $P(A) = \sum_{i=1}^{N} P(B_i)P(A|B_i)$

Teorema de Bayes: $P(B_k|A) = \frac{P(B_k)P(A|B_k)}{P(A)}$

DISTRIBUCIONES DISCRETAS:

Uniforme discreta:

$$P(X = x_i) = \frac{1}{n} \qquad \text{i = 1, ..., n}$$

Binomial(n, p):

$$P(X = x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, ..., n, \quad q = 1-p \quad \mu = n \cdot p, \text{ y } \sigma^2 = n \cdot p \cdot q$$

$$\binom{n}{x} = \frac{n!}{x! \cdot (n-x)!} \qquad \text{siendo} \qquad n! = n \cdot (n-1) \cdot (n-2) \cdot ... \cdot 2 \cdot 1$$

Poisson(λ):

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0,\ 1,\ 2,\ 3,\ ... \qquad (x \in \mathbb{N}) \quad \mu = \lambda \text{ y } \sigma^2 = \lambda$$

Sea $X \sim Bi(n, p)$. Si n es grande, p pequeña y $n \cdot p$ moderada, podremos aproximar X por una Poisson $(n \cdot p)$

ESTIMACIÓN

Estimador puntual de p: $\frac{X}{N}$, donde X es el número de éxitos en los N experimentos

Estimador puntual de μ: $\overline{X}$

Estimador puntual de σ^2: $S^2 = \frac{\sum_{i=1}^{N}(X_i - \overline{X})^2}{N-1}$

INTERVALOS DE CONFIANZA: tamaño muestral $=N$, nivel de significación α

- A) Intervalo de confianza para μ, con σ^2 conocida: $(\overline{x}$ - $z_{\alpha/2}\frac{\sigma}{\sqrt{N})}$, $\overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}})$ con P(Z $\geq z_{\alpha/2}) = \alpha/2$, $Z \sim$ N(0,1)
- B) Intervalo de confianza para μ, con σ^2 desconocida, para Normales: $(\overline{x}$ - $t_{\alpha/2}\frac{s}{\sqrt{N})}$, $\overline{x} + t_{\alpha/2}\frac{s}{\sqrt{N}})$ con P(T $\geq t_{\alpha/2}) = \alpha/2$, T es t- Student con $N-1$ grados de libertad
- C) Intervalo de confianza para μ, con σ^2 desconocida y N grande ($N \geq 30$): $(\overline{x}$ - $z_{\alpha/2}\frac{s}{\sqrt{N})}$, $\overline{x} + z_{\alpha/2}\frac{s}{\sqrt{N}})$ con P(Z $\geq z_{\alpha/2}) = \alpha/2$, $Z \sim$ N(0,1)

- Selección del tamaño de la muestra (media): $N = (\frac{z_{\alpha/2} \cdot \sigma}{Error})^2$
- D) Intervalo de confianza para la diferencia de medias μ_1 - μ_2, con σ_1^2 y σ_2^2 conocidas, para muestras aleatorias independientes (N_1 = tamaño muestral de la muestra de la población 1, N_2 = tamaño muestral de la muestra de la población 2):

  $(\overline{x}_1 - \overline{x}_2 \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}})$ con P($Z \geq z_{\alpha/2}) = \alpha/2$, $Z \sim$ N(0,1)
- E) Intervalo de confianza para la diferencia de medias μ_1 - μ_2, con σ_1^2 y σ_2^2 desconocidas, para muestras aleatorias independientes y tamaños muestrales grandes (N_1 = tamaño muestral de la muestra de la población 1, N_2 = tamaño muestral de la muestra de la población 2):

  $(\overline{x}_1 - \overline{x}_2 \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}})$ con P($Z \geq z_{\alpha/2}) = \alpha/2$, $Z \sim$ N(0,1)
- F) Intervalo de confianza para la diferencia de medias μ_1 - μ_2 de poblaciones normales independientes, con varianzas desconocidas pero iguales ($\sigma_1^2 = \sigma_2^2$) (N_1 = tamaño muestral de la muestra de la población 1, N_2 = tamaño muestral de la muestra de la población 2):

  $(\overline{x}_1 - \overline{x}_2 \pm t_{\alpha/2}\sqrt{\frac{(N_1-1)s_1^2+(N_2-1)s_2^2}{N_1+N_2-2}}\sqrt{\frac{N_1+N_2}{N_1N_2}})$ con P($T \geq t_{\alpha/2}) = \alpha/2$, T es t-Student con $N_1 + N_2 - 2$ grados de libertad
- G) Intervalo de confianza para la diferencia de medias μ_1 - μ_2 de poblaciones normales independientes, con varianzas σ_1^2, σ_2^2 desconocidas y desiguales (N_1 = tamaño muestral de la muestra de la población 1, N_2 = tamaño muestral de la muestra de la población 2):

  $(\overline{x}_1 - \overline{x}_2 \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}})$ con P($T \geq t_{\alpha/2}) = \alpha/2$, T es t-student con $\frac{(\frac{s_1^2}{N_1}+\frac{s_2^2}{N_2})^2}{\frac{(s_1^2/N_1)^2}{N_1-1}+\frac{(s_2^2/N_2)^2}{N_2-1}}$ grados de libertad
- H) Intervalo de confianza para la diferencia de medias μ_1 - μ_2 para muestras apareadas:

  $(\overline{d} \pm t_{\alpha/2}\ \frac{s_d}{\sqrt{N}})$ donde $\overline{d}$ es la media de las diferencias y s_d es la desviación típica de las diferencias. Además, P($T \geq t_{\alpha/2}) = \alpha/2$, T es t-Student con N - 1 grados de libertad, N es el número de objetos (parejas) de que disponemos
- I) Intervalo de confianza para σ^2 en una población normal:

  $(\frac{(N-1)s^2}{\chi^2_{\alpha/2}}, \frac{(N-1)s^2}{\chi^2_{1-\alpha/2}})$ con P($\chi^2 > \chi^2_{\alpha/2}) = \alpha/2$, χ^2 es chi- cuadrado con $N-1$ grados de libertad
- J) Intervalo de confianza para el cociente σ_1^2/σ_2^2 de varianzas de dos poblaciones normales independientes:

  $(\frac{s_1^2}{s_2^2}\frac{1}{F_{\alpha/2}}, \frac{s_1^2}{s_2^2}\frac{1}{F_{1-\alpha/2}})$ donde P(F $> F_{\alpha/2}) = \alpha/2$ y F es F de Sndecor con (N_1-1, N_2-1) grados de libertad
- K) Intervalo de confianza para una proporción p (de una Binomial) cuando N es grande y la proporción no es cercana a cero:

$(\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{N}})$, donde P($Z > z_{\alpha/2}) = \alpha/2$ $Z \sim$ N(0,1) y $\hat{p}$ = X /N, $\hat{q}$ = 1 - $\hat{p}$, X = número de éxitos

- Selección del tamaño de la muestra (proporción):

  $N = p(1-p) \cdot (\frac{z_{\alpha/2}}{E})^2 \leq \frac{1}{4}\ (\frac{z_{\alpha/2}}{E})^2$

- L) Intervalo de confianza para una proporción p, si ésta es muy cercana a cero:

  $(0,\ \frac{1}{2N}\chi^2_\alpha)$ con P$(\chi^2 > \chi^2_\alpha) = \alpha$, χ^2 es chi- cuadrado con $2(X+1)$ grados de libertad, X = número de éxitos

- M) Intervalo de confianza para la diferencia de dos proporciones, con N_1 y N_2 grandes (N_1 = tamaño muestral de la muestra de la población 1, N_2 = tamaño muestral de la muestra de la población 2):

  $(\hat{p_1}$ - $\hat{p_2} \pm z_{\alpha/2}\sqrt{\frac{\hat{p_1}\hat{q_1}}{N_1} + \frac{\hat{p_2}\hat{q_2}}{N_2}})$, donde P($Z > z_{\alpha/2}) = \alpha/2$ $Z \sim$ N(0,1), $\hat{p_1} = X_1\ /N_1$, $\hat{q_1}$ = 1 - $\hat{p_1}$, X_1 = número de éxitos en las N_1 pruebas y $\hat{p_2} = X_2\ /N_2$, $\hat{q_2}$ = 1 - $\hat{p_2}$, X_2 = número de éxitos en las N_2 pruebas

CONTRASTE DE HIPÓTESIS: tamaño muestral $=N$, nivel de significación α

- Contraste de hipótesis para μ, con N grande

  $Z = \frac{\bar{X}-\mu_0}{S/\sqrt{N}} \sim$ N(0,1) $H_0 : \mu = \mu_0$

H_1	Región crítica
$\mu < \mu_0$	$(-\infty, -z_\alpha)$
$\mu \neq \mu_0$	$(-\infty, -z_{\alpha/2}) \cup (z_{\alpha/2}, \infty)$
$\mu > \mu_0$	(z_α, ∞)

- Contraste de hipótesis para μ, con σ^2 desconocida para una población Normal

  $T = \frac{\bar{X}-\mu_0}{S/\sqrt{N}} \sim t_{N-1}$ $H_0 : \mu = \mu_0$

H_1	Región crítica
$\mu < \mu_0$	$(-\infty, -t_\alpha)$
$\mu \neq \mu_0$	$(-\infty, -t_{\alpha/2}) \cup (t_{\alpha/2}, \infty)$
$\mu > \mu_0$	(t_α, ∞)

- Pruebas con tablas de contingencia:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(o_{ij}-e_{ij})^2}{e_{ij}},$$

  Bajo H_0, sigue aproximadamente una distribución χ^2 con $(r-1)\cdot(c-1)$ grados de libertad. La región crítica (a nivel α) es: $(\chi^2_\alpha,\ \infty)$.

CONSEJOS:

1. ESTUDIA y haz problemas.
2. Lleva calculadora científica, el formulario y las tablas de distribuciones (y si has de entregar algún trabajo) y el DNI.
3. No copies (sobre todo no copies de alguien que no haya estudiado).
4. Lee detenidamente cada pregunta y asegúrate de que respondes lo que te preguntan (no te precipites).
5. Lee todo el examen y comienza por las preguntas que domines más y de entre ellas por las que más puntuen, te dará confianza. Si te lías con alguna, déjala para el final, no te "obsesiones en resolverla"
6. Escribe todos los pasos y cálculos que realices, no olvides que la profesora no puede leer el pensamiento y únicamente corregirá lo que quede escrito en el papel. Además, cuanto más explicado y claro (que se pueda leer) mejor.
7. Cuando obtengas un resultado piensa si es ilógico o imposible (una varianza no puede ser negativa, r_{xy} estará entre -1 y 1, una probabilidad estará entre 0 y 1, ¿concuerda con los datos?, etc.), en tales casos repasa los cálculos o el razonamiento, incluso coge un papel en blanco y comiénzalo de cero. Si no logras encontrar el error, señala tus dudas.
8. Si algún problema no sabes resolverlo, busca semejanzas con otros problemas que conozcas, simplifica tus datos, hazte gráficos, ... y al menos, inténtalo, no se pierde nada por intentarlo.
9. Repasa los cálculos y el examen antes de entregarlo.
10. Descansa la noche antes del examen, cuanto más despejad@ y relajad@ estés mejor
11. ¡Suerte! y ¡sé positiv@!