

EJERCICIO 1

Estadística

Tema 1. Fenómenos aleatorios. Regularidad estadística. Concepto de probabilidad: axiomas. Sucesos independientes y teorema de bayes.

*Esta publicación no tiene carácter oficial, se trata de material didáctico cuyo objeto es servir de apoyo en la preparación de la oposición al Cuerpo Superior de Actuarios, Estadísticos y Economistas de la Administración de la Seguridad Social. Esta documentación es **orientativa y no es exclusiva ni única** para el correcto desarrollo de los temas. Tampoco vincula al órgano convocante ni al Tribunal actuante. Los errores o desactualizaciones que pudieran contenerse en la documentación no serán imputables a la SESSP, debiéndose atender en todo caso a la legislación vigente publicada en el Boletín Oficial del Estado.*

Aviso: La SESSP se reserva las acciones legales que pudieran corresponder por el uso lucrativo de esta información. Queda prohibido expresamente la comercialización o venta del presente material.

Edición Enero 2025

1.1 Fenómenos aleatorios

El campo de aplicación y estudio de la Estadística y su fundamento matemático (el cálculo de probabilidades) lo constituyen los llamados fenómenos aleatorios.

Se define **fenómeno aleatorio** como aquel fenómeno aquel aspecto de la realidad que puede ser observado y del que conocemos todas sus posibles respuestas, pero no la respuesta que efectivamente va a tener lugar o realizarse, tras un experimento aleatorio.

Esto es, la concurrencia de circunstancias fijas no permite ver cuál será el efecto producido y por tanto los resultados son imprevisibles, pero pertenecen a un conjunto de resultados que puede ser descrito de forma completa previamente → conocemos todos los posibles resultados, pero no el resultado exacto que resultará en la i -ésima realización del experimento.

Se entenderá por experimento la acción que puede dar lugar a fenómenos cuantificables.

En un experimento determinista se conoce exactamente el resultado que va a darse, mientras que en un experimento aleatorio los resultados del experimento bajo las mismas condiciones iniciales, puede dar lugar a distintos resultados que sólo serán conocidos al final del experimento y se podrán caracterizar de forma probabilística.

Tipos de fenómenos aleatorios en función del ambiente:

- Ambiente de riesgo o estocástico: Podemos probabilizar la aparición de los diferentes resultados
- Ambiente de incertidumbre: No es posible probabilizar los posibles resultados.

En ambos casos se conocen los resultados posibles, pero sólo en un ambiente de riesgo o estocástico podremos probabilizar la ocurrencia de los mismos.

Propiedades:

- 1- En las mismas condiciones iniciales dan lugar a resultados diferentes.
- 2- Se conocen de antemano los posibles resultados.
- 3- No se puede predecir el resultado exacto de cada experimento en particular.
- 4- Si un experimento de este tipo se repite un número indeterminado de veces, n , denotando por n_i las veces que se da un determinado resultado, el cociente n/n_i (al que se denominará frecuencia relativa) tiende a estabilizarse en un valor fijo p_i cuando el número de repeticiones aumenta (Regularidad estadística – Teorema de convergencia en probabilidad – Caso particular Teorema de Bernoulli).

1.2 Regularidad Estadística

Si un experimento de este tipo se repite un número indeterminado de veces, n , denotando por n_i las veces que se da un determinado resultado, el cociente n/n_i (al que se denominará frecuencia relativa) tiende a estabilizarse en un valor fijo p_i cuando el número de repeticiones aumenta.

Teorema de Bernoulli (Convergencia en probabilidad):

Sea un suceso A con probabilidad p y sea n el número de pruebas de un experimento aleatorio, se denomina f a la frecuencia relativa del suceso A que se define:

$$f_i = \frac{n_i}{n}$$

Sea ε un número real positivo, infinitamente pequeño, tal que:

$$\varepsilon > 0 (\varepsilon \in \mathbb{R}^+)$$

Entonces:

$$\lim_{n \rightarrow \infty} P(|f_i - p_i| \geq \varepsilon) = 0 \quad \forall_i$$

El límite en probabilidad, cuando n tiende a infinito de que la diferencia en valor absoluto entre la frecuencia relativa y la probabilidad para el resultado del experimento i , sea mayor o igual que un número ε infinitamente pequeño, es cero, para todo i resultado probable del experimento aleatorio.

De esta manera queda formalizada la Ley del Azar o **Ley de estabilidad de las frecuencias relativas** que es donde descansa el concepto de **regularidad estadística**, base de la Teoría Frecuentista.

Este Teorema de Bernoulli se usa para determinar el número de pruebas necesarias para que, con una probabilidad dada, se encuentre la probabilidad p a la que tiende la frecuencia relativa.

1.3 Concepto de probabilidad: axiomas.

“El planteamiento axiomático de Kolmogorov pone en relación la teoría de la probabilidad con la teoría de conjuntos y la teoría de medida y establece una conexión entre el mundo real y el matemático recurriendo a la base empírica de la teoría frecuentista” (Martín-Pliego)

Se consigue formalizar la regularidad estadística sobre las regularidades en las frecuencias relativas de acaecimientos de un **suceso aleatorio**.

Se denomina **suceso elemental** a cada uno de los resultados de un experimento aleatorio y **suceso compuesto** a los que se generan a partir de los primeros a través de operaciones. Los sucesos forman parte de un espacio muestral que contiene todos los posibles resultados del experimento aleatorio.

Las **operaciones con sucesos** que pueden establecerse son:

- Unión de sucesos: El suceso C contiene todos los sucesos elementales de A y de B.

$$A \cup B = C$$

- Intersección de sucesos: El suceso C contiene únicamente los sucesos que coincidan en los subconjuntos A y B.

$$A \cap B = C$$

Propiedades de las operaciones de unión e intersección:

- Conmutativa:
 - Para la unión: Si $A \cup B \rightarrow B \cup A$
 - Para la intersección: Si $A \cap B \rightarrow B \cap A$
- Asociativa:
 - Para la unión: $A \cup (B \cup C) = (A \cup B) \cup C$
 - Para la intersección: $A \cap (B \cap C) = (A \cap B) \cap C$
- Distributiva:
 - $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
 - $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Sucesos complementarios: Sea A (Sucede A) se denomina suceso complementario \bar{A} al suceso en el que A no sucede. La unión de un suceso y su complementario ocupan todo el espacio muestral E.

Sucesos disjuntos: Su intersección es el espacio vacío. Sean S_i y S_j dos sucesos disjuntos, cuando ocurre uno no ocurre el otro.

- $S_i \cap S_j = \emptyset \forall i \neq j$ Disjuntos dos a dos
- $\bigcap_{i=1}^n S_i = \emptyset$ Los nb sucesos son completamente disjuntos
- Si $S_i \cap S_j = \emptyset \forall i \neq j$ y $\bigcap_{i=1}^n S_i = E$ (E Espacio muestral) $\rightarrow S_i$ forman una partición del espacio muestral. Esta definición se amplía para cualquier conjunto infinito numerable de disjuntos dos a dos.

Establecidas las definiciones básicas de la teoría de conjuntos de lo que es un espacio muestral y los sucesos que lo componen, se procede a establecer lo que es un **espacio probabilizable**, bajo el **álgebra de Boole o de sucesos**.

Sea A una colección de subconjuntos de E (espacio muestral), tal que:

- $E \in A$
- Si un suceso $S \in A \Rightarrow \bar{S} \in A$ y, por tanto
 - $S \cup \bar{S} = E$
 - $E \cup \bar{E} = E \cup \emptyset$
 - $E \in A \Rightarrow \emptyset \in A$
- Si $S_1, S_2 \in A \Rightarrow S_1 \cup S_2 \in A$ y por tanto $S_1 \cap S_2 \in A$

A es un **álgebra de Boole (Álgebra de sucesos)** que se denota por Ω

Definido el álgebra de sucesos se establecen las condiciones que deben ser dadas para definir **la Probabilidad**.

Axiomática de Kolmogorov

- 1) Sea S un suceso, $S \in \Omega$ (siendo Ω el álgebra de sucesos) $\Rightarrow P(S) \geq 0$
- 2) $P(E) = 1$
- 3) $P(\cup S_i) = \sum P(S_i)$ si $S_i \cap S_j = \emptyset \forall i \neq j$

Se define espacio probabilístico al conjunto (E, Ω, P) donde:

- E es el espacio muestral
- Ω es el álgebra de sucesos y
- P es la medida de probabilidad

Consecuencias:

1.- Si $B \subset A \Rightarrow P(B) \leq P(A)$

Demostración:

$$P(A) = P(B \cap A) + P(\bar{B} \cap A), P(A \cap B) = P(B) \text{ porque } B \text{ está incluido en } A$$

$$\Rightarrow P(A) = P(B) + P(\bar{B} \cap A) \Rightarrow P(B) \leq P(A) \text{ ya que } P(\bar{B} \cap A) \geq 0 \text{ por el axioma 1}$$

2.- $P(S_i) \leq 1$

Demostración:

$S \in E \Rightarrow$ Dado que:

- $P(S) \leq P(E)$
- $P(E) = 1$

3.- $P(\bar{S}) = 1 - P(S)$

Demostración:

$$\bar{S} \cup S = E, P(\bar{S} \cup S) = P(E),$$

$$S \cap \bar{S} = \emptyset,$$

$$P(\bar{S} \cup S) = P(S) + P(\bar{S}) = P(E) = 1 \Rightarrow P(S) = 1 - P(\bar{S})$$

1.4 Sucesos independientes y teorema de bayes.

Dos sucesos son independientes si sucedido el primero, la probabilidad de que ocurra el segundo no se ve modificada.

Se define la **probabilidad condicionada** de un suceso A sobre otro suceso B como

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Dos **sucesos son independientes** si:

$$P(A \cap B) = P(A)P(B)$$

De modo que:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

El suceso B no modifica la probabilidad de A.

Teorema de la probabilidad total

- ➔ Sean A_i ,, $i = 1, 2, \dots, n$, n particiones del espacio muestral de forma que $A_i \cap A_j = \emptyset \forall i \neq j$
- ➔ Sea A un suceso perteneciente al espacio muestral

Se puede escribir la probabilidad del suceso A en función de las intersecciones de éste con las diferentes particiones B_i del espacio muestral ,,

$$P(B) = \sum_{i=1}^n P(B \cap A_i)$$

Regla de multiplicación de sucesos:

$$P(\cap S_i) = P(S_1)P(S_2|S_1) * P(S_3|S_1 \cap S_2) * \dots * P(S_n|S_1 \cap S_2 \cap \dots \cap S_{n-1})$$

Para el caso particular de dos sucesos:

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

Teorema de Bayes:

- ➔ Sea $A_i \forall i = 1, 2, \dots, n$ un conjunto de n sucesos mutuamente excluyentes ($A_i \cap A_j = \emptyset \forall i \neq j$) con probabilidad distinta de cero y sea B un suceso cualquiera del que se conocen las probabilidades condicionales $P(B|A_i)$, el teorema de Bayes permite conocer la probabilidad de cualquiera de los i sucesos A una vez ocurrido B.

Aplicando el teorema de la probabilidad total al denominador de la probabilidad condicionada y la regla de multiplicación al numerador, el teorema de Bayes se define como:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^n P(B \cap A_i)}$$

Donde $P(A_i)$ son las probabilidades a priori y $P(A_i|B)$ son las probabilidades a posteriori.

EJERCICIO 1

Estadística

Tema 2. Variable aleatoria unidimensional.
Distribución de probabilidad. Función de densidad.
Características de tendencia central y de dispersión.
Desigualdad de Chebychev.

*Esta publicación no tiene carácter oficial, se trata de material didáctico cuyo objeto es servir de apoyo en la preparación de la oposición al Cuerpo Superior de Actuarios, Estadísticos y Economistas de la Administración de la Seguridad Social. Esta documentación es **orientativa y no es exclusiva ni única** para el correcto desarrollo de los temas. Tampoco vincula al órgano convocante ni al Tribunal actuante. Los errores o desactualizaciones que pudieran contenerse en la documentación no serán imputables a la SESSP, debiéndose atender en todo caso a la legislación vigente publicada en el Boletín Oficial del Estado.*

Aviso: La SESSP se reserva las acciones legales que pudieran corresponder por el uso lucrativo de esta información. Queda prohibido expresamente la comercialización o venta del presente material.

Edición Enero 2025

2.1 Variable aleatoria unidimensional

Una **variable aleatoria** es una realización funcional que nos hace pasar del espacio muestral E a la recta \mathbb{R} , esto es, cada fenómeno aleatorio es modelizado a través de una variable aleatoria.

Dado (E, Ω, P) un espacio probabilístico, llamamos variable aleatoria a la aplicación

$$\zeta: (E, \Omega) \rightarrow (\mathbb{R}, \mathcal{B}_R),,$$

- E es el espacio muestral
- Ω es el álgebra de sucesos
- \mathbb{R} la recta de los números reales
- \mathcal{B}_R es el σ -álgebra de Borel (esto es, contenido en el álgebra de sucesos, es la mínima álgebra)

Entonces

$$\zeta^{-1}((-\infty, x]) = \{\omega \in E / \zeta(\omega) \leq x\} = \{\zeta \leq x\} \in \Omega$$

$(-\infty, x]$ definido en el σ -álgebra de Borel

ω sucesos del espacio muestral

Por convención se escribe con letras griegas las variables aleatorias y con letras latinas los valores de la variable (x en este caso).

2.2 Distribución de probabilidad.

Dado un espacio probabilístico (E, Ω, P) y una variable aleatoria ζ se denomina **función de distribución de probabilidad** de la variable aleatoria a aquella

$$F: \mathbb{R} \rightarrow \mathbb{R}$$

Que asigna a cada x la probabilidad de que la variable aleatoria tome cierto valor de la variable o cualquiera de los valores a su izquierda.

$$F(x) = P_{\zeta}((-\infty, x]) = P(\zeta^{-1}((-\infty, x])) = P(\{\omega \in E / \zeta(\omega) \leq x\}) = P(\zeta \leq x) \quad \forall x \in \mathbb{R}$$

También se le denomina **función de distribución acumulativa** y proporciona la cantidad de masa que hay en el punto x y a su izquierda hasta el extremo inferior.

Esta definición es la misma tanto para variables aleatorias de tipo discreto como de tipo continuo.

Propiedades

1.- $0 \leq F(x) \leq 1 \quad \forall x \in \mathbb{R}$ por ser $F(x)$ una probabilidad

2.- $F(-\infty) = 0$

3.- $F(\infty) = 1$

→ Nota: se indica $-\infty$ e ∞ para el dominio general. Se refiere a los extremos inferior y superior de los valores ordenados de ζ (variable aleatoria) $F(x) = P(\zeta \leq x) \quad \forall x \in I$ siendo I el intervalo de la variable aleatoria.

4.- $F(x)$ es monótona creciente $F(x_1) \leq F(x_2)$ si $x_1 \leq x_2$

5.- Continua por la derecha

6.- Si es discontinua por la izquierda es una variable aleatoria discreta.

2.3 Función de densidad.

Cuando una variable aleatoria viene definida en un conjunto finito o infinito numerable de respuestas (**variables discretas**), se puede asignar a cada una de estas respuestas una probabilidad. En este caso el espacio probabilístico queda definido de tal forma que la probabilidad inducida P_ζ queda definida como partes del espacio muestral → se puede asignar a cada suceso ω de E (espacio muestral) una probabilidad $P(\omega)$ y en ese caso se está ante una **función de probabilidad o función de cuantía**.

Cuando por el contrario la variable aleatoria queda definida en un espacio infinito no numerable no es posible asignar una probabilidad a cada infinito valor de la variable aleatoria pues la masa de un suceso es infinitesimal y su probabilidad en el punto (probabilidad puntual) es cero. Se debe pues trabajar con densidades y no con masas puntuales. La probabilidad así medida será calculada mediante intervalos y la función que mide estas probabilidades se denomina **función de densidad**.

De esta forma, las funciones de distribución de probabilidad (la función acumulada) serán:

- Caso de variable discreta : $F(x) = P(\zeta \leq x) = \sum_{-\infty}^x P(\zeta = x_i)$ donde $P(\zeta = x_i)$ es la **función de cuantía**
- Caso de variable continua : $F(x) = P(\zeta \leq x) = P(\zeta < x) = \int_{-\infty}^x f(x) dx$ donde $f(x)$ es la **función de densidad**

Propiedades

- 1) $f(x) \geq 0 \quad \forall x \in \mathbb{R}$
- 2) $\int_{-\infty}^{\infty} f(x) dx = 1$ el área bajo la función de densidad en todo el campo de variación es 1 por ser 1 el valor máximo que puede tomar una probabilidad. Cuando esto no se cumple se debe introducir una constante K tal que:

$$k \int_{-\infty}^{\infty} g(x) dx = 1 \Rightarrow K = \frac{1}{\int_{-\infty}^{\infty} g(x) dx}$$

De modo que $f(x) = K g(x)$

- 3) La función de densidad es no negativa. Toma valores positivos sólo en el soporte de la variable aleatoria

$$f: \mathbb{R} \rightarrow \mathbb{R}$$

$$f(x) < 0 \Leftrightarrow x \in \text{Soporte}$$

Y verifica que $P(a < x < b) = \int_a^b f(x)dx$ y que la probabilidad en todo el intervalo es la unidad $\int_{INT} f(x)dx = 1$

2.4 Características de tendencia central y de dispersión.

Las medidas de tendencia central son medidas estadísticas con las que se pretende resumir un conjunto de valores de una variable a un único valor, representando un centro (de ahí su nombre) en torno al cual se sitúan los valores de la variable.

Las **medidas de tendencia central** más comunes son:

- **Media o esperanza matemática:** Generalmente se usan estos dos conceptos de forma indistinta sin embargo lo correcto es utilizar el término media cuando se establece esta medida en datos de una muestra a partir de las frecuencias de aparición de los valores de la variable y esperanza matemática cuando se desea obtener el valor en una población de la que se desconocen las frecuencias, pero se conocen las probabilidades o densidades.

- Media muestral:

$$\bar{x} = \frac{\sum x_i n_i}{n}$$

Donde n_i es el número de apariciones (frecuencia absoluta) del i -ésimo valor y x_i el valor de la variable de ese i -ésimo suceso. También puede verse $n_i/n=f_i$ como la frecuencia relativa del i -ésimo valor.

- Esperanza matemática variable discreta:

$$E(x) = \sum x_i p_i$$

Donde p_i es la probabilidad (función de cuantía) $p_i = P(\zeta = x_i)$

- Esperanza matemática variable continua:

$$\int_{INT} x f(x) dx$$

Donde $f(x)$ es la función de densidad (INT=en todo el intervalo)

La media o esperanza matemática es uno de los estadísticos más conocidos y utilizados, sin embargo, se debe tener en cuenta que no es estadísticamente robusta, dado que se ve muy afectado por los valores extremos. La media debería ir siempre acompañada – al menos - de una medida de dispersión sobre la misma con la que se informe de su representatividad. Otras cantidades que pueden acompañar a este estadístico para informar de su fuerza informativa son el valor mínimo, el máximo y el rango.

Momentos respecto del origen:

A partir de la esperanza se definen los momentos respecto del origen como:

$$E(\zeta^r) = \alpha_r$$

De modo que la esperanza matemática es el momento donde $r=1$, primer momento respecto del origen.

- **Mediana:** Es aquel valor que deja el mismo número de datos a la derecha y a la izquierda; Por tanto, se puede definir también como el valor del percentil 50, el cuartil 2 o el decil 5.

La mediana no se ve tan afectada por los valores extremos por lo que es más robusta que la media o esperanza. Ante variables con fuerte asimetría es bueno acompañar el valor de la media con la mediana dado que la diferencia de valores entre una y otra dará una idea de la asimetría de la distribución. En una distribución perfectamente simétrica ambos valores coinciden.

- En una muestra

$$Me = L_{i-1} + \frac{\frac{N}{2} - F_{i-1}}{f_i} a$$

- L_{i-1} es el límite inferior
- a es la amplitud
- F_{i-1} es la frecuencia acumulada anterior al intervalo mediano
- f_i frecuencia relativa del intervalo mediano

En el caso de población se sustituye la frecuencia acumulada por la función de distribución y la frecuencia relativa por la función de cuantía o densidad.

- Variable discreta: Es aquel valor tal que $P[\zeta \leq Me] \geq \frac{1}{2}$ y $P[\zeta \geq Me] \geq \frac{1}{2}$
- Variable continua: es el valor de x que hace que $\int_{-\infty}^{Me} f(x)dx = 1/2$

Moda: es el valor o valores que más se repiten en los datos (esto hace que realmente no deba ser incluido de forma estricta en la definición de tendencia central, aunque muchos autores lo hagan).

- Variable discreta: $P(\zeta = x_i) \geq P(\zeta = x_j) \quad \forall i \neq j$
- Variable continua: Es el valor que anula la primera derivada siendo el valor de la segunda derivada negativa (esto es, se produce un máximo)

Propiedades de la Esperanza

- 1.- $E(\zeta)$ no siempre existe. Existe sólo si $\int_{INT} |x|f(x)dx < \infty$ (converge absolutamente) \rightarrow si ζ es acotada $E(\zeta)$ existe.
- 2.- La esperanza de una constante es el valor de la constante $E(C)=C$
- 3.- La esperanza de la suma es la suma de las esperanzas $E(\zeta_1 + \zeta_2 + \zeta_3 + \dots + \zeta_n) = E(\zeta_1) + E(\zeta_2) + E(\zeta_3) + \dots + E(\zeta_n)$
- 4.- si $a \leq \zeta \leq b \implies E(a) \leq E(\zeta) \leq E(b)$
- 5.- Si $\zeta \geq 0 \implies E(\zeta) \geq 0$
- 6.- Si $\zeta \geq a$ y $E(\zeta) = a$ entonces $\zeta = a$ es una constante

7.- La esperanza se ve afectada por los cambios de origen y de escala $E(a\zeta + b) = a E(\zeta) + b$

8.- $E(\zeta - E(\zeta)) = 0$ y además $\sum E(\zeta_i - E(\zeta)) = 0$ la suma de las desviaciones de los valores de la variable respecto de la media es cero.

9.- Si $\zeta_1, \zeta_2, \zeta_3, \dots, \zeta_n$ son variables aleatorias independientes e igualmente distribuidas (vaid) $E(\zeta_1 * \zeta_2 * \zeta_3 * \dots * \zeta_n) = E(\zeta_1) * E(\zeta_2) * E(\zeta_3) * \dots * E(\zeta_n)$ la esperanza del producto es el producto de las esperanzas.

10.- **Momentos respecto a la media:** Los momentos respecto a la media se pueden obtener respecto de los momentos respecto del origen. Sea $E(\zeta) = \mu$ y μ_r el r-ésimo momento respecto de la media, se define como:

$$\mu_r = E((\zeta - \mu)^r) = \sum_{k=0}^r (-1)^k \binom{r}{k} \mu^k E(\zeta^{r-k})$$

$$E(\zeta^{r-k}) = \alpha_{r-k}$$

Siendo α_r el momento r-ésimo respecto del origen

Medidas de dispersión

Las medidas de dispersión valoran lo centrales que son las medidas anteriores, concretamente sobre la media y la mediana. De esta manera se puede observar lo representativo de las medidas de tendencia central.

Dispersión sobre la media (esperanza matemática)

Dado que $E(\zeta - E(\zeta)) = 0$ para calcular la dispersión respecto del valor de la esperanza existen varias soluciones.

- **Desviación media respecto de la media ($D_{\bar{x}}$):** Las desviaciones se toman en valor absoluto para evitar que se anulen $D_{\bar{x}} = E(|\zeta - E(\zeta)|)$.
 - También existe la medida de dispersión análoga respecto de la mediana
- **Varianza (σ^2):** Se elevan al cuadrado las desviaciones para evitar que se anulen. $E(\zeta - E(\zeta))^2 = \sigma^2$, es el momento 2 respecto de la media. En función de los momentos respecto del origen quedaría como.

$$E(\zeta - E(\zeta))^2 = \sigma^2 = \alpha_2 - \alpha_1^2 = \mu_2$$

La varianza es la medida de dispersión más extendida pero dado que el resultado que ofrece no es en las mismas unidades que la variable de estudio sino en unidades al cuadrado, se obtiene su raíz cuadrada y a dicho resultado se le denomina desviación típica o desviación estándar.

$$\sigma = \sqrt{\sigma^2}$$

La desviación típica tiene la ventaja de estar en las mismas unidades, sin embargo, se debe tener muy en cuenta que, ante distribuciones con mucha asimetría, no es una medida de dispersión representativa. En casos de distribuciones con fuerte asimetría es mejor usar medidas más sofisticadas como la entropía.

Propiedades de la varianza:

- 1.- La varianza de una constante es cero. $V(c)=0$
- 2.- Los cambios de escala afectan a la varianza $V(c \zeta) = c^2 V(\zeta)$
- 3.- $V(\zeta \pm \eta) = V(\zeta) + V(\eta) \pm 2Cov(\zeta, \eta)$
* Caso particular: Si ζ y η son independientes, $Cov(\zeta, \eta) = 0$
- 4.- $V(-\zeta) = V((-1) \zeta) = (-1^2)V(\zeta) = V(\zeta)$ La varianza SIEMPRE es positiva
- 5.- La varianza no se ve afectada por las traslaciones $V(a + b\zeta) = b^2V(\zeta)$

Coefficiente de variación de Pearson

Cuando se desea comparar dos poblaciones con datos distintos de media y varianza o con unidades distintas o ambas, lo mejor es usar el coeficiente de variación de Pearson que por un lado es adimensional (no depende de las unidades) y por otro lado permite comparar la representatividad de la media en dos poblaciones con datos de media y desviación típica distintas.

$$CVP = \frac{\sigma}{\mu}$$

Un valor cercano a cero indica que la media es representativa, al contrario, un valor cercano a 1 indica que la media no es representativa, esto es, la población no es homogénea, la dispersión es mayor. Si CVP es mayor que 1 indica alta dispersión.

2.5 Desigualdad de Chebychev.

La desigualdad de Chebychev es un caso particular de la desigualdad de Markov.

Desigualdad de Markov

$$P[g(\zeta) \geq k] \leq \frac{E(g(\zeta))}{k}$$
$$k > 0$$

A partir de aquí, sea ζ una variable aleatoria tal que:

$$\mu = E(\zeta)$$

$$\sigma^2 = V(\zeta)$$

Con $g(\zeta) = (\zeta - \mu)^2$ y $K = k^2$

$$P[(\zeta - \mu)^2 \geq k^2] \leq \frac{E((\zeta - \mu)^2)}{k^2} = \frac{\sigma^2}{k^2}$$

$$P[|\zeta - \mu| \geq k] \leq \frac{\sigma^2}{k^2}$$

$$P[|\zeta - \mu| < k] \geq 1 - \frac{\sigma^2}{k^2}$$

Si hacemos que $k = k\sigma$

$$P[|\zeta - \mu| < k\sigma] \geq 1 - \frac{\sigma^2}{k\sigma^2} = 1 - \frac{1}{k^2} = 1 - \alpha$$

Se usa cuando se desconoce la distribución de probabilidad de la variable aleatoria ζ y se conocen μ y σ de tal manera que se pueden establecer cotas de probabilidad. Cuando k crece existe poca probabilidad de que se tomen valores fuera del intervalo.

Se usa también para construir intervalos de confianza (se denomina $1 - \alpha$ al nivel de confianza, y α al nivel de significación)

EJERCICIO 1

Estadística

Tema 3. Distribución binomial. Características. Teoremas de Bernoulli. Ajuste de una distribución binomial. Aplicaciones. Distribución binomial negativa. Distribución geométrica. Distribución hipergeométrica. Propiedades.

*Esta publicación no tiene carácter oficial, se trata de material didáctico cuyo objeto es servir de apoyo en la preparación de la oposición al Cuerpo Superior de Actuarios, Estadísticos y Economistas de la Administración de la Seguridad Social. Esta documentación es **orientativa y no es exclusiva ni única** para el correcto desarrollo de los temas. Tampoco vincula al órgano convocante ni al Tribunal actuante. Los errores o desactualizaciones que pudieran contenerse en la documentación no serán imputables a la SESSP, debiéndose atender en todo caso a la legislación vigente publicada en el Boletín Oficial del Estado.*

Aviso: *La SESSP se reserva las acciones legales que pudieran corresponder por el uso lucrativo de esta información. Queda prohibido expresamente la comercialización o venta del presente material.*

Edición Enero 2025

3.1 Distribución binomial. Características.

Para estudiar un fenómeno aleatorio se asigna un modelo de distribución que describa razonablemente bien su comportamiento, para a continuación, decidir los valores de los parámetros que mejor se ajusten al experimento. De esta manera las distribuciones siempre van a presentarse en función de uno o varios parámetros que deben conocerse o estimarse.

Sea ζ una variable aleatoria “Nº de éxitos (o fracasos) tras n pruebas” de un fenómeno aleatorio dicotómico (éxito/fracaso). Esta variable aleatoria con una distribución binomial se define como un proceso de Bernoulli en n repeticiones independientes, teniendo todas ellas la misma probabilidad de éxito “ p ”.

Se trata por tanto de un tipo de distribución contadora (cuenta eventos bajo una condición) y por ello la variable aleatoria es de tipo discreto.

Su **espacio muestral** son todos los valores entre 0 y n . Su **campo de variación** es:

$$\zeta: x_i \quad \forall i = 0, 1, 2, \dots, n \in \mathbb{N}^+ \text{ y } 0$$

Probabilidad de éxito: p

Probabilidad de no éxito (o fracaso): $1-p=q$

$$\zeta \rightarrow B(n, p)$$

Si $n=1$ la distribución es la de Bernoulli.

Parámetros:

n : número de pruebas (*no es realmente un parámetro que haya que estimar, sino el número de pruebas o tamaño*)

p : probabilidad de éxito

Función de cuantía:

Probabilidad de que en n pruebas se den x éxitos

$$P(\zeta = x) = \binom{n}{x} p^x q^{n-x}$$

$$\binom{n}{x} = \frac{n!}{(n-x)! x!}$$

Función de distribución:

Probabilidad de que en n pruebas haya x o hasta x éxitos

$$F(x) = \sum_{i=1}^x P(\zeta = i)$$

Esperanza matemática:

$$E(x) = np$$

Varianza:

$$V(x) = npq$$

Propiedades.

La distribución de Bernoulli es una binomial de parámetros $n=1$, p , por tanto, todo lo que aquí se indique respecto de la Binomial se traslada a la distribución de Bernoulli haciendo $n=1$.

Distribución de Bernoulli

La distribución de Bernoulli es un caso particular de la distribución binomial con $n=1$

Probabilidad de no éxito (o fracaso): $1-p=q$

$$\zeta \rightarrow B(1, p)$$

Parámetros:

n: número de pruebas

p: probabilidad de éxito

Función de cuantía:

Probabilidad de que en n pruebas se den x éxitos

$$P(\zeta = x) = p^x q^{1-x}$$

Función de distribución:

Probabilidad de que en n pruebas haya x o hasta x éxitos

$$F(x) = \sum_{i=1}^x P(\zeta = x)$$

Esperanza matemática:

$$E(x) = p$$

Varianza:

$$V(x) = pq$$

Partiendo de lo anterior, se puede definir la Binomial como una suma de n distribuciones $B(1, p)$, tal que,

$$\zeta \rightarrow B(n, p), \zeta = \sum_{i=1}^n \zeta_i \text{ con } \zeta_i \rightarrow B(1, p)$$

Probabilidad máxima o modal: es el valor de la distribución que verifique que

$$np - q \leq Mo \leq np + q$$

Propiedad aditiva o reproductiva: La suma de variables independientes binomiales también es una binomial.

La asimetría de la distribución aumenta con la diferencia (q-p) y la distribución es simétrica para p=q=0,5

Proposición:

Sea $\zeta \rightarrow B(n, p)$ y $\eta \rightarrow B(n, 1 - p) \implies P(\zeta = k) = P(\eta = n - k) \quad \forall k = 0, 1, \dots, n$

3.2 Teorema de Bernoulli.

Teorema de Bernoulli

Sea un suceso A con probabilidad p y sea n el número de pruebas de un experimento aleatorio, se denomina f a la frecuencia relativa del suceso A que se define:

$$f_i = \frac{n_i}{n}$$

Sea ε un número real positivo, infinitamente pequeño, tal que:

$$\varepsilon > 0 (\varepsilon \in \mathbb{R}^+)$$

Entonces:

$$\lim_{n \rightarrow \infty} P(|f_i - p_i| \geq \varepsilon) = 0 \quad \forall i$$

El límite en probabilidad, cuando n tiende a infinito de que la diferencia en valor absoluto entre la frecuencia relativa y la probabilidad para el resultado del experimento i, sea mayor o igual que un número épsilon infinitamente pequeño, es cero, para todo i resultado probable del experimento aleatorio.

De esta manera queda formalizada la Ley del Azar o **Ley de estabilidad de las frecuencias relativas** que es donde descansa el concepto de **regularidad estadística**, base de la Teoría Frecuentista.

Este Teorema de Bernoulli se usa para determinar el número de pruebas necesarias para que, con una probabilidad dada, se encuentre la probabilidad p a la que tiende la frecuencia relativa.

Vamos a desarrollar este teorema a partir del uso de la distribución binomial:

Sea $\zeta \rightarrow B(n, p)$, se considera una nueva variable aleatoria función de la anterior tal que $\gamma = \frac{\zeta}{n}$ que cumple que:

$$E(\gamma) = \frac{E(\zeta)}{n} = \frac{np}{n} = p$$

$$V(\gamma) = \frac{V(\zeta)}{n^2} = \frac{npq}{n^2} = \frac{pq}{n}$$

Considerando la desigualdad de Markov:

$$P[g(\gamma) \geq k] \leq \frac{E[g(\gamma)]}{k} \quad g(\gamma) \text{ y } k > 0$$

Haciendo

$$g(\gamma) = (\gamma - p)^2$$

$$k = \varepsilon^2$$

$$P[(\gamma - p)^2 \geq \varepsilon^2] \leq \frac{E[(\gamma - p)^2]}{\varepsilon^2} = \frac{V(\gamma)}{\varepsilon^2} = \frac{\frac{pq}{n}}{\varepsilon^2} = \frac{pq}{n\varepsilon^2} \Rightarrow P[|\gamma - p| \geq \varepsilon] \leq \frac{pq}{n\varepsilon^2}$$

Cuando $n \rightarrow \infty$ tenemos que $P[|\gamma - p| \geq \varepsilon] \rightarrow 0 \Rightarrow \lim_{n \rightarrow \infty} P\left(\frac{\gamma}{n}\right) = p$

El teorema de Bernoulli nos dice que se puede emplear $\frac{\zeta}{n}$ para estimar la probabilidad de p cuando $n \rightarrow \infty$. Realmente n va a ser siempre finito, pero se puede determinar un número para n de modo que $\frac{\zeta}{n}$ esté comprendida entre $(p - \varepsilon)$ y $(p + \varepsilon)$ con una probabilidad grande.

Si se desconoce p , que la necesitamos para calcular $\frac{pq}{n\varepsilon^2} = \frac{p-p^2}{n\varepsilon^2}$ el máximo va a aparecer haciendo la primera y segunda derivada parcial respecto de p

$$\frac{\partial}{\partial p} \left(\frac{p - p^2}{n\varepsilon^2} \right) = \frac{1 - 2p}{n\varepsilon^2} = 0 \Rightarrow p = 1/2$$

$$\frac{\partial^2}{\partial p^2} \left(\frac{p - p^2}{n\varepsilon^2} \right) = \frac{-2}{n\varepsilon^2} < 0 \Rightarrow \text{máx}$$

$$\frac{pq}{n\varepsilon^2} < \frac{1}{2} \frac{1}{2} = \frac{1}{4n\varepsilon^2}$$

Por lo que el teorema de Bernoulli se puede expresar como

$$P[|\gamma - p| \geq \varepsilon] \leq \frac{pq}{n\varepsilon^2} < \frac{1}{4n\varepsilon^2}$$

3.3 Ajuste de una distribución binomial.

Para estudiar un fenómeno aleatorio le asignamos un modelo de distribución que describa de forma razonable su comportamiento.

La propuesta de asignación de un modelo teórico que ajuste los datos observados se hace tras una tarea de exploración y descripción de las observaciones.

El modelo propuesto es en realidad una familia de distribuciones con una misma función cuya forma, posición y dispersión que define a unas y otras depende del valor del parámetro o parámetros que se asignen al modelo.

Los parámetros son valores que definen el modelo sobre la familia de distribuciones de ese modelo y están estrechamente relacionados con sus momentos (un método de estimación de los mismos es precisamente la estimación por momentos).

Cuando se tiene una muestra se puede querer saber si la distribución de frecuencias de la misma se ajusta a una binomial. Es importante notar que, aunque la binomial se define como $B(n,p)$ n no es un parámetro sino el número de pruebas y no hay que estimarlo en sentido paramétrico (se puede estudiar el valor necesario para una probabilidad dada, pero no se estima).

Los parámetros de la binomial son $B(n,p)$ por lo que se debe estimar el valor de estos.

Teniendo en cuenta que el valor esperado de una binomial es $E(\zeta) = \mu$, vamos a estimar el valor de μ

$$\hat{\mu} = \bar{x} = \frac{\sum x_i n_i}{N}$$

Si queremos ajustar a una $B(m,p)$ donde se verifica que

$$\mu = mp \Rightarrow mp = \bar{x} \Rightarrow p = \frac{\bar{x}}{m}$$

Otro método para estimar el valor de p es el de estimación por máxima verosimilitud:

Sea $B(m,p)$ la distribución de una variable aleatoria ζ donde se toma una muestra aleatoria simple (m.a.s) de tamaño n :

$$P(\zeta = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$L(x; p) = \prod_{i=1}^n \binom{n}{x_i} p^{\sum x_i} (1-p)^{nm - \sum x_i}$$

$$\text{Log}L(x; p) = \sum_{i=1}^n \text{Log} \binom{n}{x_i} + \sum x_i \text{Log} p + (nm - \sum x_i) \text{Log} (1-p)$$

$$\frac{\partial \text{Log}L(x; p)}{\partial p} = \frac{\sum x_i}{p} - \frac{(nm - \sum x_i)}{1-p} = 0 \implies \hat{p} = \frac{\sum x_i}{nm} = \frac{\bar{x}}{m}$$

Llegando al mismo resultado que con el método de los momentos. Es un estimado óptimo (sin sesgo y de varianza mínima) ya que iguala la cota de Cramer-Rao podemos decir que es eficiente.

Se calcularán las probabilidades de B(m,p) con los parámetros estimados tal que:

$$P(\zeta = x) = \binom{m}{x} p^x (1-p)^{m-x}$$

Obteniendo las frecuencias teóricas n_{it} , se compararán los resultados de éstas con los que ofrecía la muestra (frecuencias muestrales) de modo que si las diferencias no son grandes se podrá afirmar que la muestra se distribuye binomialmente.

Se puede hacer una prueba de bondad de ajuste como el Test Chi-cuadrado χ_{k-r-1} donde k son las clases y r el número de parámetros (en este caso únicamente es uno, p).

$$\chi_{k-r-1} = \sum \frac{(\text{frecuencias observadas} - \text{valores teóricos})^2}{\text{valores teóricos}} = Id$$

Ese valor Id se compara en una tabla χ_{k-r-1} donde k-r-1 son los grados de libertad; de tal forma que si el valor de lo anterior supera el valor de la Chi-cuadrado de la tabla se rechaza el ajuste.

3.4 Aplicaciones.

Ejemplos: Hombre/mujer, Trabajador/no trabajador, Siniestro/no siniestro, fraude/no fraude

- 1) La distribución binomial se utiliza cuando:
 - Fenómeno aleatorio es dicotómico (éxito/fracaso)
 - La varianza es inferior a la media (distribución infradispersa)
 - Se trata de una variable aleatoria contadora (pertenece al conjunto de números naturales positivos más el cero) → cuenta el número de éxitos/fracasos en n pruebas
- 2) Esta distribución pertenece a la familia paramétrica exponencial:

$$f(x; \theta) = h(x)B(\theta)e^{Q(x)R(\theta)}$$

Esto permite su uso en Modelos Lineales Generalizados donde su función link canónica es la logarítmica (Regresión logística: modelos logit, modelos probit)

- 3) Pertenece a la familia de la clase (a,b) de distribuciones, una forma funcional que permite generar secuencias de forma iterativa de probabilidad correspondientes a una distribución particular (Algoritmo de Panjer)
- 4) En mixturas con la distribución de Poisson (también contadora) genera la distribución de Hermite.

3.5 Distribución binomial negativa.

La distribución binomial negativa define la variable aleatoria $\zeta \rightarrow \mathbb{N}^0$ de éxitos tras el r-ésimo fracaso. Es por tanto una función contadora.

$$\zeta \rightarrow BN(r, p)$$

La probabilidad de una secuencia con k fracasos y r-1 éxitos antes del éxito r-ésimo será:

$$p^{r-1}q^k p = p^r q^k$$

Pero dado que los r-1 éxitos previos y k fracasos previos que se pueden producir pueden aparecer en diferente disposición.,

$$\frac{(r-1+k)!}{k!(r-1)!} = \binom{k+r-1}{k} = (-1)^k \binom{-r}{k}$$

Quedando pues definida la función de cuantía como:

$$\binom{k+r-1}{k} p^r q^k$$

O también como:

$$(-1)^k \binom{-r}{k} p^r q^k = (-q)^k p^r \binom{-r}{k}$$

Esta última forma es por la que recibe el nombre de binomial negativa.

Para k=0,1,...,n

Esperanza:

$$E(\zeta) = \frac{rq}{p}$$

Varianza:

$$V(\zeta) = \frac{rq}{p^2}$$

- 1) La distribución binomial se utiliza cuando:
 - Se quiere contar el número de ocurrencias de sucesos raros tras un número determinado de no sucesos.
 - Se trata de una variable aleatoria contadora (pertenece al conjunto de números naturales positivos más el cero)
- 2) Esta distribución pertenece a la familia paramétrica exponencial:

$$f(x; \theta) = h(x)B(\theta)e^{Q(x)R(\theta)}$$

Esto permite su uso en Modelos Lineales Generalizados donde su función enlace canónica es la logarítmica.

- 3) Pertenece a la familia de la clase (a,b) de distribuciones, una forma funcional que permite generar secuencias de forma iterativa de probabilidad correspondientes a una distribución particular (algoritmo de Panjer)
- 4) Se usa en modelos Zero-truncados ZTBN (truncados por $1 - P(\zeta = 0)$)
- 5) La BN puede tener distintas parametrizaciones, una de ellas es la que se puede obtener como una composición de la función Poisson con una Gamma

$$\zeta \rightarrow P(\lambda)$$

$$\lambda \rightarrow \gamma(\alpha, r)$$

$$BN(r, p = \frac{\alpha}{1 + \alpha})$$

En ocasiones, sobre todo en el mundo asegurador, se usa esta composición que suele parametrizarse de la siguiente manera buscando el número de siniestros (fracasos) siendo α el tiempo que transcurre entre uno y el siguiente (Gamma como generalización de la distribución exponencial).

$$B(\alpha, \beta)$$

En este caso $p = \frac{1}{1 + \beta}$

Y la función de cuantía queda determinada como:

$$P(\zeta = x) = \binom{x - \alpha - 1}{x} \left(\frac{1}{1 + \beta}\right)^\alpha \left(\frac{\beta}{1 + \beta}\right)^x$$

La esperanza y varianza se obtienen igual que en líneas anteriores teniendo en cuenta $p = \frac{1}{1 + \beta}$

3.6 Distribución geométrica. Distribución hipergeométrica. Propiedades.

Geométrica

La distribución geométrica es un caso particular de la distribución binomial negativa cuando ésta tiene el parámetro $r=1$, de tal forma que

$$\zeta = N^{\circ} \text{ de fracasos antes del primer éxito} \rightarrow G(p)$$

Se realizan experimentos de Bernoulli sucesivos de probabilidad p de tal forma que se busca la probabilidad de que aparezca por primera vez el suceso “éxito” tras r fracasos, esto es, que aparezca un éxito tras r fracasos.

$$P(\zeta = r) = q^r p$$

Para $r=0,1,2,\dots,K$

Esperanza:

$$E(\zeta) = \frac{q}{p}$$

Varianza

$$V(\zeta) = \frac{q}{p^2}$$

Al contrario que la binomial, la distribución geométrica no verifica la propiedad aditiva ya que la suma de muchas variables aleatorias distribuidas como una geométrica no siguen una distribución geométrica, en su caso la distribución resultante será una binomial negativa.

Tiene la característica de la falta de memoria, de tal forma que:

- Si se aplica en experimentos de repetición de un fenómeno, ninguna observación condiciona la posterior
- Si una variable aleatoria toma valores enteros no negativos y cumple con la probabilidad de falta de memoria, esta variable se distribuye como una distribución geométrica.

Hipergeométrica

Una distribución hipergeométrica

$$\zeta \rightarrow H(N, n; p)$$

Donde $\zeta \rightarrow N^\circ$ de individuos con la característica x de una muestra aleatoria simple de n en una población de tamaño N

$$p = \frac{N_1}{N}$$

$$q = \frac{N_2}{N}$$

Función de cuantía:

$$P(\zeta = x) = \frac{\binom{Np}{x} \binom{Nq}{n-x}}{\binom{N}{n}}$$

Esperanza:

$$E(\zeta) = np$$

Varianza:

$$V(\zeta) = \frac{N-n}{N-1} npq$$

Cuando $n \rightarrow \infty$, $H(N, n; p) \rightarrow B(n, p)$ siempre que $p = \frac{N_1}{N}$ permanezca estable.

La distribución Hipergeométrica es un caso particular de la distribución de Polyá (o distribución de contagio) con un parámetro de contagio $c = -1$

**** Nota final: no se han incluido gráficas que pertenecen a cada una de estas distribuciones por poder localizarse fácilmente en cualquier manual de estadística básica, pero se recomienda su inclusión si en alguna de las preguntas del ejercicio inicial se pide la descripción de alguna de estas distribuciones.*

EJERCICIO 1

Estadística

Tema 4. Distribución de Poisson: Características.
Distribución de Poisson como límite de la binomial.
Ajuste de una distribución de Poisson. Aplicaciones.
Distribución exponencial.

*Esta publicación no tiene carácter oficial, se trata de material didáctico cuyo objeto es servir de apoyo en la preparación de la oposición al Cuerpo Superior de Actuarios, Estadísticos y Economistas de la Administración de la Seguridad Social. Esta documentación es **orientativa y no es exclusiva ni única** para el correcto desarrollo de los temas. Tampoco vincula al órgano convocante ni al Tribunal actuante. Los errores o desactualizaciones que pudieran contenerse en la documentación no serán imputables a la SESSP, debiéndose atender en todo caso a la legislación vigente publicada en el Boletín Oficial del Estado.*

Aviso: La SESSP se reserva las acciones legales que pudieran corresponder por el uso lucrativo de esta información. Queda prohibido expresamente la comercialización o venta del presente material.

Edición Enero 2025

Distribución de Poisson: Características.

Se dice que una variable aleatoria describe una distribución de Poisson de parámetro λ cuando su probabilidad queda definida por la siguiente distribución de cuantía

$$P(\zeta = x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

Tomando todos los valores enteros positivos

$$x \in \mathbb{N}^+ \cup 0$$

Es una distribución de probabilidad, por tanto, de tipo discreto y contadora. Se usa para estudiar la probabilidad de ocurrencia de sucesos raros, pero requiere de muestras homogéneas para su aplicación, ya que una de sus características es que su media es igual a su varianza, por lo que no admite sobredispersión.

Esperanza:

$$E(\zeta) = \lambda$$

Varianza:

$$V(\zeta) = \lambda$$

Valor modal

$$1 - \lambda \leq Mo \leq \lambda$$

Cumple con la **propiedad aditiva o reproductiva:**

Sean $\zeta_1, \zeta_2, \dots, \zeta_n$, variables aleatorias independientes igualmente distribuidas $\zeta_i \rightarrow P(\lambda_i)$ entonces

$$\xi = \zeta_1 + \zeta_2 + \dots + \zeta_n \rightarrow P(\lambda = \sum_i^n \lambda_i)$$

Distribuciones condicionadas:

Sean ζ_1 y ζ_2 variables aleatorias independientes igualmente distribuidas $\zeta_i \rightarrow P(\lambda_i)$ ($i=1,2$) entonces la distribución de la variable aleatoria ζ_1 condicionada por la variable suma $\zeta_1 + \zeta_2$ es una distribución binomial.

$$\frac{\zeta_1}{\zeta_1 + \zeta_2 = \gamma} = B\left(\gamma, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)$$

Aproximación a la Normal:

Como la suma de variables aleatorias independientes igualmente distribuidas que siguen una distribución de Poisson es, como se ha indicado, una Poisson con parámetro la suma de los parámetros. Cuando $\lambda > 5$ se puede considerar que la $P(\lambda) \sim N(\lambda, \sqrt{\lambda})$

Distribución de Poisson como límite de la binomial.

Sea ζ una variable aleatoria $B(n,p)$ con un valor de p muy bajo y un valor de n muy alto, en el que además $np = \lambda$ siendo λ constante, en ese caso se puede aproximar la distribución binomial $B(n,p)$ a una $Poisson(\lambda)$.

Si cada intervalo es una Bernoulli (ocurre/no ocurre), la suma de los intervalos es una Binomial $p = c \frac{t}{n}$, cuando $n \rightarrow \infty$

$$\lim_{n \rightarrow \infty} \binom{n}{x} p^x q^{n-x} = \frac{e^{-ct} (ct)^x}{x!}$$
$$ct = \lambda$$

¿Cuándo se puede entender que p es pequeña? Algunos autores indican que cuando $p \leq 0,01$ pero también existen autorías que indican que la aproximación es buena si se cumple que $p \leq 0,1$ y $np < 5$.

El valor de λ puede estimarse a partir de n y p conocidos estimando p a partir de la media de una muestra de valores observados.

También, como cualquier parámetro de distribución, puede estimarse mediante por máxima verosimilitud.

Ajuste de una distribución de Poisson.

Para estudiar un fenómeno aleatorio le asignamos un modelo de distribución que describa de forma razonable su comportamiento.

La propuesta de asignación de un modelo teórico que ajuste los datos observados se hace tras una tarea de exploración y descripción de las observaciones.

El modelo propuesto es en realidad una familia de distribuciones con una misma función cuya forma, posición y dispersión que define a unas y otras depende del valor del parámetro o parámetros que se asignen al modelo.

Los parámetros son valores que definen el modelo sobre la familia de distribuciones de ese modelo y están estrechamente relacionados con sus momentos (un método de estimación de los mismos es precisamente la estimación por momentos).

Para ajustar a una Poisson en primer lugar se debe estar ante una variable de tipo contador y se requieren datos homogéneos. Si se cumple lo anterior y además en la muestra se observa que la media y la varianzas muestrales son prácticamente iguales, se puede suponer que estamos ante una distribución de Poisson.

Dibujado el histograma se aprecia una cadencia.

Si n es muy grande el histograma irá tomando la forma de campana y en ese caso el ajuste más correcto será a una $N(\lambda, \sqrt{\lambda})$

Se estimará el valor de λ mediante verosimilitud o bien a través de $\lambda = np \Rightarrow \hat{p} = \frac{\bar{x}}{n}$ donde es el tamaño de la muestra y \bar{x} la media muestral. El resultado a partir del método de máxima verosimilitud es el mismo.

$$L(X; \lambda) = e^{-n\lambda} \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!}$$

$$\ln(L(X; \lambda)) = -n\lambda + \ln \lambda \sum_{i=1}^n x_i - \sum_{i=1}^n \ln x_i!$$

$$\frac{\partial \ln(L(X; \lambda))}{\partial \lambda} = -n + \frac{\sum_{i=1}^n x_i}{\lambda} = 0$$

$$\frac{\partial^2 \ln(L(X; \lambda))}{\partial^2 \lambda} = -\frac{n}{\bar{x}} < 0$$

Verifica la condición de máximo. Y como se quería demostrar, el EMV es precisamente la media muestral:

$$\frac{\sum_{i=1}^n x_i}{n} = \hat{\lambda} = \bar{x}$$

Una vez estimado el valor del parámetro se calculan los valores teóricos y se comparan con los valores que se observan en la distribución de frecuencias de la muestra. Cuanto menor sea la diferencia, mejor será el ajuste.

Se puede hacer un test de bondad de ajuste como es el Test Chi-cuadrado χ_{k-r-1} donde k son las clases y r el número de parámetros (en este caso únicamente es uno).

$$\chi_{k-r-1} = \sum \frac{(\text{frecuencias observadas} - \text{valores teóricos})^2}{\text{valores teóricos}} = Id$$

Ese valor Id se compara en una tabla χ_{k-r-1} donde k-r-1 son los grados de libertad; de tal forma que si el valor de lo anterior supera el valor de la Chi-cuadrado de la tabla se rechaza el ajuste.

Aplicaciones.

Se aplica cuando se trata de sucesos raros, esto es con una probabilidad de ocurrencia muy pequeña en cada observación, se trata de obtener probabilidades de que ocurra el suceso tras un número muy grande de observaciones, por este motivo se ha usado tradicionalmente en la estadística actuarial no vida.

Se usa en la Teoría de Colas para modelizar la distribución de las llegadas a un determinado lugar donde se irán formando colas.

La distribución de Poisson es contadora y por ello puedo tomar cualquier valor positivo y el cero, sin embargo, la probabilidad de que tome valores cada vez mayores decrece de forma acelerada y la mayoría de la probabilidad se acumula en los primeros valores (depende del valor del parámetro, obviamente). Su distribución es escalonada.

Cuando en la forma de su distribución el cero tiene la mayor masa, estamos ante distribuciones de tipo Poisson zero-truncadas.

Esta distribución pertenece a la familia paramétrica exponencial:

$$f(x; \theta) = h(x)B(\theta)e^{Q(x)R(\theta)}$$

Esto permite su uso en Modelos Lineales Generalizados donde su función link canónica es la logarítmica, permitiendo modelizar λ a partir de variables independientes y ajustar dicho valor al riesgo. En ese caso $\lambda = E(N)$ valor esperado del número de siniestros, el modelo GLM se modela con función link logarítmica como:

$$\ln(\lambda) = \beta_0 + \sum_{i=1}^n \beta_i x_i$$

La distribución de Poisson requiere de homogeneidad en los datos de la población.

Compuesta con una Gamma ($\lambda \rightarrow G(\alpha, \beta)$) resulta una Binomial negativa, que admite más dispersión.

Igualmente se puede considerar que λ se distribuye como una Inversa Gaussiana ($r=-0,5$) obteniendo una Binomial Negativa Extendida y Truncada (BNET)

Distribución exponencial.

La distribución exponencial, un caso particular de la distribución Gamma, es una distribución continua que representa el tiempo transcurrido entre un suceso y el siguiente. Su parámetro es λ . Para valores $t > 0$

$$f(t) = \lambda e^{-\lambda t}$$

Su **generalización a la Gamma** (λ, k) será el tiempo transcurrido hasta la presencia de k acontecimientos

$$f_{T_k} = \frac{\lambda^k}{(k-1)!} e^{-\lambda t} t^{k-1}$$

Su función de distribución es:

$$F(x) = 1 - e^{-\frac{x}{\lambda}}$$

Esperanza:

$$E(x) = \lambda$$

Varianza:

$$V(x) = \lambda^2$$

Por su parte, el número de acontecimientos que pueden ocurrir en el intervalo de tiempo sigue una distribución de Poisson $P(\lambda t)$. La exponencial como generalización de la Gamma cuenta el tiempo transcurrido o espacio necesario para la presentación de 1 suceso (tiempo entre dos sucesos).

La generalización a n sucesos es otro caso particular de la Gamma, la distribución de Erlang (muy usada en teoría de Colas), $G(n, n\lambda)$.

La Gamma (y por tanto la exponencial y la Erlang) pertenecen a la familia exponencial

$$f(x; \theta) = h(x)B(\theta)e^{Q(x)R(\theta)}$$

***Véase el apartado de la binomial negativa en el que se compone una Poisson con una Gamma.*

**** Nota final: no se han incluido gráficas que pertenecen a cada una de estas distribuciones por poder localizarse fácilmente en cualquier manual de estadística básica, pero se recomienda su inclusión si en alguna de las preguntas del ejercicio inicial se pide la descripción de alguna de estas distribuciones.*

EJERCICIO 1

Estadística

Tema 5. Distribución normal y log-normal.
Características. Ajuste de una distribución normal y log-normal. Importancia de la distribución normal en la estadística

*Esta publicación no tiene carácter oficial, se trata de material didáctico cuyo objeto es servir de apoyo en la preparación de la oposición al Cuerpo Superior de Actuarios, Estadísticos y Economistas de la Administración de la Seguridad Social. Esta documentación es **orientativa y no es exclusiva ni única** para el correcto desarrollo de los temas. Tampoco vincula al órgano convocante ni al Tribunal actuante. Los errores o desactualizaciones que pudieran contenerse en la documentación no serán imputables a la SESSP, debiéndose atender en todo caso a la legislación vigente publicada en el Boletín Oficial del Estado.*

Aviso: La SESSP se reserva las acciones legales que pudieran corresponder por el uso lucrativo de esta información. Queda prohibido expresamente la comercialización o venta del presente material.

Edición Enero 2025

Distribución normal y log-normal. Características.

Normal

Se dice que una variable aleatoria describe una distribución Normal de parámetros $[\mu, \sigma$ ($\mu, \sigma \in \mathbb{R}$ $\sigma > 0$)] si su función de densidad es:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Es una variable continua.

Esta distribución es la que mayor aplicación tiene en el cálculo de probabilidades y en inferencia. Es la que se usa generalmente para establecer intervalos de confianza y su importancia la pone de manifiesto en el teorema central del límite:

“Si en un experimento aleatorio el resultado es el efecto de un gran número de causas aleatorias elementales e independientes y cada una cuantitativamente insignificante, el resultado global es aproximadamente Normal”.

Se denomina también campana de gauss por la forma de su gráfica, aunque fue De Moivre quien estableció la expresión matemática, Gauss junto a Laplace la obtuvieron de forma empírica estudiando la distribución de los errores de medición.

$$f(\varepsilon) = \frac{1}{b\sqrt{2\pi}} e^{-\frac{(\varepsilon)^2}{2b^2}}$$

Su función de distribución es:

$$F(x) = P(\zeta < x) = \int_0^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{u-\mu}{\sigma}\right)^2} du$$

Muchas distribuciones tienden a ésta cuando crece el tamaño de la muestra (o población).

Propiedades

1. Es continua en toda la recta real
2. Es simétrica respecto del parámetro μ , esto es $f(\mu + t) = f(\mu - t)$
3. Nunca toma el valor cero, pero cuando $x \rightarrow \pm\infty$, tiende al eje OX (asíntotas horizontales)

$$\lim_{x \rightarrow -\infty} f(x) = \lim_{x \rightarrow \infty} f(x) = 0$$

4. Es estrictamente creciente si $x < \mu$
5. Es estrictamente decreciente si $x > \mu$
6. Máximo en $x = \mu$, $y = \frac{1}{\sigma\sqrt{2\pi}}$
7. Puntos de inflexión en $\mu \pm \sigma$, donde toman el valor $y = \frac{1}{\sigma\sqrt{2\pi}e}$
8. Convexa si $|x| < \mu - \sigma$ y cóncava si $|x| > \mu - \sigma$
9. Esperanza:

$$E(\zeta) = \mu$$

10. Varianza

$$V(\zeta) = \sigma^2$$

11. Asimetría: Nula en la distribución normal (perfectamente simétrica) (μ_3 es el tercer momento respecto la media)

$$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

12. Curtosis: Nula en la distribución normal (μ_4 es el cuarto momento respecto a la media)

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$$

13. En una distribución Normal las distancias interválicas son las siguientes:

- $(\mu \pm \sigma)$ se encuentra aproximadamente el 68,26% de la densidad de la distribución
- $(\mu \pm 2\sigma)$ se encuentra aproximadamente el 95,45% de la densidad de la distribución
- $(\mu \pm 3\sigma)$ se encuentra aproximadamente el 99,73% de la densidad de la distribución

Si en una muestra se pueden ajustar los datos a una Normal esto va a permitir dar intervalos más precisos que los que se obtienen con la desigualdad de Chebichev.

14. Propiedad aditiva o reproductiva:

Sea $\zeta_i \rightarrow N(\mu_i, \sigma_i)$ si tenemos una distribución $\gamma = a + \sum b_i \zeta_i \rightarrow N(a + \sum b_j \mu_j, \sqrt{\sum b_i^2 \sigma_i^2})$

$N(0,1)$

Si se tipifica la variable (se le resta la media μ y se divide entre su desviación típica σ , se obtiene la $N(0,1)$ que suele llamarse Z. Esta distribución es sumamente importante para la estadística, especialmente cuando no se disponía de tecnología suficiente para trabajar con la $N(\mu, \sigma)$ para hallar $P(X < x)$.

LogNormal

Si $\eta = e^\zeta$ y $\zeta \rightarrow N(\mu, \sigma)$, entonces $\eta \rightarrow \text{Log}N(\mu, \sigma)$

Es el resultado de un número elevado de causas independientes y con efectos multiplicativos y teniendo cada una de las causas efectos despreciables respecto del global.

El campo de variación de $\eta \in \mathbb{R}^+ \cup \{0\}$

La distribución tiene forma de campana con asimetría positiva, cola hacia la derecha, siendo más leptocúrtica cuanto más pequeña sea σ y más platicúrtica cuanto más grande sea σ .

Aplicaciones: Predicción de seísmos, precio de acciones (Blacks-Scholes), patrones de abundancia de especies.

Se conoce como distribución de McAlister o ley de efecto proporcional por considerar que es el resultado de un número elevado de causas independientes con efectos positivos y que se componen de manera multiplicativa y cada una de esas causas tiene un efecto despreciable respecto al global.

Función de distribución:

$$G(y) = P(\eta \leq y) = P(e^\zeta \leq y) = P(\zeta \leq \ln(y)) = F(\ln(y))$$

Función de densidad:

$$g(y) = \frac{1}{y} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}}$$

Esperanza:

$$e^{-\mu + \frac{1}{2}\sigma^2}$$

Varianza:

$$(e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$$

Ajuste de una distribución normal y log-normal.

Para estudiar un fenómeno aleatorio le asignamos un modelo de distribución que describa de forma razonable su comportamiento.

La propuesta de asignación de un modelo teórico que ajuste los datos observados se hace tras una tarea de exploración y descripción de las observaciones.

El modelo propuesto es en realidad una familia de distribuciones con una misma función cuya forma, posición y dispersión que define a unas y otras depende del valor del parámetro o parámetros que se asignen al modelo.

Los parámetros son valores que definen el modelo sobre la familia de distribuciones de ese modelo y están estrechamente relacionados con sus momentos (un método de estimación de los mismos es precisamente la estimación por momentos).

Tanto en la Normal como en la LogNormal sus parámetros son μ, σ .

Se estimarán los parámetros por máxima verosimilitud o por el método de los momentos. En el caso de la Normal el estimador máximo verosímil y por momentos del parámetro μ es la media muestral sin embargo en el caso del segundo parámetro, la varianza poblacional su estimador máximo verosímil, si se conoce μ , no es la varianza muestral ya que el mismo depende de μ y no de la media muestral. En el caso de que se desconozcan ambos, el estimador es la varianza muestral. No obstante, es mejor usar como estimador de la varianza poblacional la cuasivarianza muestral si n no es muy grande, dado que la varianza muestral es un estimador sesgado, cosa que no pasa con la cuasivarianza muestral.

Dados los datos de una muestra, si el histograma presenta una forma acampanada aproximadamente se puede tratar de ajustar la misma a una distribución normal. Además, se tendrán en cuenta las siguientes características para ajustar a una distribución normal:

- Se buscará que la media muestral, la mediana muestral y la moda muestral sean aproximadamente iguales.

- Se buscará que el coeficiente de asimetría y curtosis en la muestra sean aproximadamente nulos.
- En un QQ-plot se buscará que la nube de puntos esté cercana a la bisectriz.

En el caso de la lognormal, el ajuste se intentará cuando la forma acampanada tenga asimetría positiva y cola a la derecha, por tanto. Además, la LogNormal se tratará de ajustar cuando la variable de la muestra tenga un campo de variación positivo (y cero).

En cualquiera de los dos ajustes el primer paso es estimar el valor de los parámetros. De forma puntual se pueden estimar por el método de la máxima verosimilitud o el de los momentos. En estas distribuciones cualquiera de los dos métodos hacen que los mejores estimadores de μ, σ sean la media muestral para el primer parámetro y la cuasidesviación típica para el segundo.

Estimados los parámetros se calcularán los valores teóricos bajo la distribución candidata y sus resultados se compararán con los valores muestrales que ofrece la distribución de frecuencias.

Se puede hacer un test de bondad de ajuste como el Test Chi-cuadrado χ_{k-r-1} donde k son las clases y r el número de parámetros (en este caso únicamente es uno, p).

$$\chi_{k-r-1} = \sum \frac{(\text{frecuencias observadas} - \text{valores teóricos})^2}{\text{valores teóricos}} = Id$$

Ese valor Id se compara en una tabla χ_{k-r-1} donde k-r-1 son los grados de libertad; de tal forma que si el valor de lo anterior supera el valor de la Chi-cuadrado de la tabla se rechaza el ajuste.

Test de Kolmogorov-Smirnov, muy habitual para test de normalidad, basado en la discrepancia máxima en valor absoluto entre la distribución teórica y la observada.

Test de Shapiro-Wilks: Relacionado con la PP-plot, mide el ajuste a la recta.

Importancia de la distribución normal en la estadística.

Es la distribución límite de muchísimas sucesiones de variables aleatorias, tanto discretas (binomial, Poisson) como continuas. Esto queda demostrado a través del Teorema Central del límite "Si en un experimento aleatorio el resultado es el efecto de un gran número de efectos aleatorios elementales independientes y cada uno cuantitativamente insignificante, el resultado global sigue una distribución aproximadamente Normal.

Esta distribución pertenece a la familia paramétrica exponencial:

$$f(x; \theta) = h(x)B(\theta)e^{Q(x)R(\theta)}$$

Esto permite su uso en Modelos Lineales Generalizados donde su función link canónica es la función identidad y de ésta se obtiene el caso particular de Regresión Lineal Ordinaria, muy usada en la estadística, en series temporales, en modelos económicos aditivos en los que a partir de una serie de variables se trata de explicar otra.

La desventaja de la Normal es que al ser simétrica no es buena candidata a la modelización de variables como salarios, cuantía de siniestros, etc.... donde existen muchos casos (alta densidad) en valores bajos y casos con poca probabilidad, pero de valores muy altos que no deben ser desdeñados (y que en un ajuste simétrico se pierden).

Su propiedad reproductiva (aditiva) y las distribuciones que derivan de la misma son aplicadas al muestreo, a la inferencia estadística (intervalos de confianza y contrastes de hipótesis)

Teorema central del límite

Sea una sucesión de variables aleatorias $\{\zeta_n\} = \zeta_1, \zeta_2, \dots, \zeta_n$ converge en probabilidad a una variable aleatoria ζ si

$$\lim_{n \rightarrow \infty} P(\zeta_n \leq x) = P(\zeta \leq x)$$

En todos los puntos donde $F_\zeta(x) = P(\zeta \leq x)$ es continua. Esta característica se denomina convergencia en distribución y se denota como

$$\zeta_n \xrightarrow{D} \zeta$$

El Teorema Central del límite nos informa del comportamiento de las sumas aleatorias independientes e igualmente distribuidas, no necesariamente como una Normal, de modo que sea $\{\zeta_n\}$ una sucesión de variables aleatorias igualmente distribuidas con media μ y desviación típica σ la suma $S_n = \zeta_1 + \zeta_2 + \dots + \zeta_n$ verifica que

$$E\left(\sum_{i=1}^n \zeta_i\right) = \sum_{i=1}^n E(\zeta_i) = n\mu$$

$$V\left(\sum_{i=1}^n \zeta_i\right) = \sum_{i=1}^n V(\zeta_i) = n\sigma^2$$

El TCL dice que cuando n es grande, la variable sigue aproximadamente una $N(0,1)$:

$$\frac{\sum_{i=1}^n \zeta_i - E(\sum_{i=1}^n \zeta_i)}{\sqrt{V(\sum_{i=1}^n \zeta_i)}} = \frac{\sum_{i=1}^n \zeta_i - n\mu}{\sigma\sqrt{n}} = \frac{\bar{\zeta} - \mu}{\sigma/\sqrt{n}}$$

Teorema de Moivre-Laplace (aproximación de probabilidad de binomial y Poisson con una Normal)

Este teorema indica que si $\zeta \rightarrow B(n, p)$ cuando $n \rightarrow \infty$

$$\frac{\zeta - np}{\sqrt{npq}} \rightarrow N(0,1)$$

Como extensión a este teorema se generaliza a una Poisson si n es muy grande, de tal forma que:

$$P(\lambda) \rightarrow N(\lambda, \sqrt{\lambda})$$

$$\frac{\zeta - \lambda}{\sqrt{\lambda}} \rightarrow N(0,1)$$

**** Nota final: no se han incluido gráficas que pertenecen a cada una de estas distribuciones por poder localizarse fácilmente en cualquier manual de estadística básica, pero se recomienda su inclusión si en alguna de las preguntas del ejercicio inicial se pide la descripción de alguna de estas distribuciones.*

EJERCICIO 1

Estadística

Tema 6. Distribuciones derivadas de la normal. Distribución χ^2 («ji cuadrado»). Distribución t de Student. Distribución F de Snedecor. Características e importancia de estas distribuciones en la práctica estadística.

*Esta publicación no tiene carácter oficial, se trata de material didáctico cuyo objeto es servir de apoyo en la preparación de la oposición al Cuerpo Superior de Actuarios, Estadísticos y Economistas de la Administración de la Seguridad Social. Esta documentación es **orientativa y no es exclusiva ni única** para el correcto desarrollo de los temas. Tampoco vincula al órgano convocante ni al Tribunal actuante. Los errores o desactualizaciones que pudieran contenerse en la documentación no serán imputables a la SESSP, debiéndose atender en todo caso a la legislación vigente publicada en el Boletín Oficial del Estado.*

Aviso: La SESSP se reserva las acciones legales que pudieran corresponder por el uso lucrativo de esta información. Queda prohibido expresamente la comercialización o venta del presente material.

Edición Enero 2025

Distribuciones derivadas de la normal.

Las distribuciones derivadas de la normal están diseñadas para asignar medidas probabilísticas a la variable aleatoria que aparecen en la estimación de parámetros de distribución tipo. En estas distribuciones no importan tanto sus momentos u otras características como su relación con la Normal.

Distribución χ^2 («ji cuadrado»).

Sea

$$\zeta_i \rightarrow N(0,1)$$

Sea

$$\eta_n = \zeta_1^2 + \zeta_2^2 + \dots + \zeta_n^2 = \sum_{i=1}^n \zeta_i^2$$

Entonces se define la distribución chi-cuadrado (o ji-cuadrado) como:

$$\eta_n \rightarrow \chi_n^2$$

Una Chi-cuadrado con n grados de libertad (DF). Los grados de libertad se refieren al número de variables que intervienen.

Su esperanza es igual al número de variables, esto es, a sus grados de libertad (n) y su varianza a dos veces la esperanza (2n).

Propiedad aditiva:

La χ_n^2 tiene la propiedad aditiva, de modo que si tenemos m variables aleatorias independientes igualmente distribuidas como χ_n^2

$$\begin{aligned} & \chi_{n_1}^2, \chi_{n_2}^2, \dots, \chi_{n_m}^2 \\ & \chi_{n_1}^2 + \chi_{n_2}^2 + \dots + \chi_{n_m}^2 = \sum_{i=1}^m \chi_{n_i}^2 \rightarrow \chi_N^2 \\ & N = \sum_{i=1}^m n_i \end{aligned}$$

Convergencia en distribución (aproximadamente $n > 30$)

$$\begin{aligned} & \chi_n^2 - n \rightarrow \infty \rightarrow N(n, 2n) \\ & \sqrt{2\chi_n^2 - n} - n \rightarrow \infty \rightarrow N(\sqrt{2n-1}, 1) \end{aligned}$$

En este segundo caso la varianza no depende de n.

La forma de su distribución es asimétrica positiva.

Aplicaciones

Para inferencia, se usa esta distribución:

- Para contrastes de bondad de ajuste
- Pivotes en la estimación de la varianza de una población normal (o la desviación típica)
- Contrastes no paramétricos como estudios de independencia o estudios de homogeneidad en tablas de contingencia
- Así como para cualquier inferencia sobre desviaciones al cuadrado.

Para inferencia se usa

$$P(\chi_n^2 > k) = \alpha$$

Siendo α el nivel de significación.

Relación con otras distribuciones:

Esta distribución se particulariza en una Gamma cuando.

$$\chi_n^2 \sim G\left(\frac{n}{2}, 2\right)$$

Distribución t de Student.

La distribución t de Student es una distribución con forma acampanada como la Normal pero más platicúrtica y con colas más largas por lo que presenta mayor dispersión y esto es muy útil cuando se quieren hacer inferencias cuando se desconoce el valor de la varianza poblacional y usamos como estimador la cuasivarianza muestral.

La t de Student se define como

$$t_n = \frac{\zeta_{0i}}{\sqrt{\frac{\sum_{i=1}^n \zeta_{0i}^2}{n}}}$$

Donde $\zeta_{0i} \rightarrow N(0, \sigma)$ esto es, normales centradas.

Si hacemos que sean $\xi \rightarrow N(0,1)$ y $\zeta \rightarrow \chi_n^2$

$$t_n = \frac{\xi}{\sqrt{\frac{\sum_{i=1}^n \zeta_{0i}^2}{n}}} = \frac{\xi}{\sqrt{\frac{\chi_n^2}{n}}} = \frac{\xi}{\sqrt{\zeta/n}}$$

Su función de densidad es

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}$$

Para $-\infty < t < \infty$

Características.

Es simétrica respecto al origen $P(t_n > x) = P(t_n < -x) \quad \forall x$

Sus extremos son $(-\infty, -2)$ y $(2, +\infty)$ aproximadamente.

Convergencia

La t de Student converge a una $N(0,1)$ cuando la n crece (aproximadamente $n < 30$)

Esperanza y varianza

Su esperanza es 0 cuando $n \geq 2$. Para $n=1$ la media no existe

En el caso de la varianza es

$$V(t_n) = \frac{n}{n-2}$$

Cuando $n \geq 3$. Si n es menor, la varianza no existe.

Aplicaciones

La aplicación más importante es en inferencia sobre la media poblacional cuando la varianza es desconocida y en muestras de tamaño pequeño.

En regresión se usa para el contraste de variables una a una

$$H_0 = \beta_i$$

así como en análisis ANOVA para cada una de las medias tomadas individualmente

$$H_0 = \mu_i$$

Distribución F de Snedecor.

Sean dos variables aleatorias independientes igualmente distribuidas como Chi-cuadrado, con n y m grados de libertad respectivamente, tales que:

$$X \rightarrow \chi_n^2$$

$$Y \rightarrow \chi_m^2$$

Entonces se define la distribución F de Snedecor como:

$$\zeta \rightarrow F_{n,m} = \frac{X/n}{Y/m} = \frac{\chi_n^2/n}{\chi_m^2/m}$$

Esperanza y varianza

$$E(\zeta) = \frac{m}{m-2} \quad m > 2$$

$$V(\zeta) = \frac{2m^2(n+m-2)}{n(m-4)(m-2)^2} \quad \text{si } m > 4$$

Se cumple que

$$P(F_{n,m} < x) = 1 - P(F_{m,n} > x) = 1 - \alpha$$

Siendo α el nivel de significación y $1 - \alpha$ el nivel de confianza.

Se puede reescribir como:

$$F_{n,m} = \frac{1}{F_{m,n}}$$

Además, cuando $n=1$, se reescribe como una t de Student

$$F_{1,m} = t_m^2$$

Aplicaciones

- Para el análisis ANOVA bajo la hipótesis $H_0 = \mu_1 = \mu_2 = \dots = \mu_n$
- Para la significación global en un modelo de regresión $H_0 = \beta_1 = \beta_2 = \dots = \beta_n$
- Inferencia en razón de varianzas de una población Normal

Características e importancia de estas distribuciones en la práctica estadística.

En los apartados de cada una de las distribuciones descritas se han ido indicando las características de estas distribuciones y las aplicaciones más relevantes en la práctica estadística.

**** Nota final: no se han incluido gráficas que pertenecen a cada una de estas distribuciones por poder localizarse fácilmente en cualquier manual de estadística básica, pero se recomienda su inclusión si en alguna de las preguntas del ejercicio inicial se pide la descripción de alguna de estas distribuciones.*

EJERCICIO 1

Estadística

Tema 7. Distribuciones multidimensionales.
Distribuciones marginales y condicionadas.
Momentos. Coeficientes de correlación.
Independencia.

*Esta publicación no tiene carácter oficial, se trata de material didáctico cuyo objeto es servir de apoyo en la preparación de la oposición al Cuerpo Superior de Actuarios, Estadísticos y Economistas de la Administración de la Seguridad Social. Esta documentación es **orientativa y no es exclusiva ni única** para el correcto desarrollo de los temas. Tampoco vincula al órgano convocante ni al Tribunal actuante. Los errores o desactualizaciones que pudieran contenerse en la documentación no serán imputables a la SESSP, debiéndose atender en todo caso a la legislación vigente publicada en el Boletín Oficial del Estado.*

Aviso: La SESSP se reserva las acciones legales que pudieran corresponder por el uso lucrativo de esta información. Queda prohibido expresamente la comercialización o venta del presente material.

Edición Enero 2025

C.S Actuarios, Estadísticos y Economistas de la Administración de la Seguridad Social

Distribuciones multidimensionales.

Si el resultado de un experimento se materializa en la ocurrencia de dos o más observaciones conjuntas, simultáneas, estamos ante una variable aleatoria multidimensional; ahora se miden n características.

Consideramos el espacio de probabilidad (E, Ω, P) y en él n variables aleatorias $\xi_1, \xi_2, \dots, \xi_n$ el vector $\mathbb{Z} = (\xi_1, \xi_2, \dots, \xi_n)$ es una variable aleatoria n -dimensional.

La probabilidad inducida de una variable aleatoria n -dimensional se define como sigue,

Sea $\mathbb{Z}: (E, \Omega, P) \rightarrow (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ una variable n -dimensional se define la probabilidad inducida, se define la probabilidad inducida en $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ como:

$$P_{\zeta}(\mathcal{B}) = P(\zeta^{-1}(\mathcal{B})) \quad \forall \mathcal{B} \in \mathcal{B}_{\mathbb{R}^n}$$

El espacio de probabilidad inducida queda $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, P_{\zeta})$

Distribución conjunta de una variable aleatoria n -dimensional

Sea $\mathbb{Z} = (\xi_1, \xi_2, \dots, \xi_n)$ una variable aleatoria n -dimensional cuyo campo de variación es $-\infty < x_i < \infty \quad \forall i = 1, 2, \dots, n$

Función de distribución (generalización)

$$F(x_1, x_2, \dots, x_n) = P(\xi_1 \leq x_1, \xi_2 \leq x_2, \dots, \xi_n \leq x_n)$$

Propiedades:

- $F(-\infty, -\infty, \dots, -\infty) = 0$ y $F(x_1, -\infty, \dots, -\infty) = 0$ Esto es., con una única que sea igual a $-\infty$, la función de distribución conjunta es 0.
- $F(\infty, \infty, \dots, \infty) = 1$
- $F(\mathbb{Z})$ es monótona no decreciente en cada variable
- $F(\mathbb{Z})$ es continua por la derecha en cada variable (o argumento)
- $F(\mathbb{Z}) \geq 0$ (positiva) por ser probabilidades

Caso particular bidimensional: Función de distribución

$$F(x, y) = P(-\infty < \zeta \leq x; -\infty < \eta \leq y) = P(\zeta \leq x; \eta \leq y)$$

Variables discretas

Una variable aleatoria n -dimensional es discreta si el conjunto de valores que toma es finito o infinito numerable.

Se denominará **función de probabilidad o cuantía (discreta) conjunta** a la función

$$P(\zeta_1 = x_1; \zeta_2 = x_2; \dots; \zeta_n = x_n)$$

Su **función de distribución conjunta** será

$$F(x_1, x_2, \dots, x_n) = \sum P(\zeta_1 = x_1; \zeta_2 = x_2; \dots; \zeta_n = x_n)$$

En todo su dominio (existiría un sumatorio para cada una de las n variables).

Variables continuas

Es aquella variable aleatoria cuyos componentes son variables unidimensionales de tipo continuo.

Su **función de distribución conjunta** será

$$F(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f(\zeta_1, \zeta_2, \dots, \zeta_n) d\zeta_1 d\zeta_2 \dots d\zeta_n$$

Como en el caso unidimensional no existe la probabilidad en un punto por lo que de querer conocer la probabilidad en un intervalo será dicho intervalo el que se usará en los extremos de las n-integrales.

Distribuciones marginales y condicionadas.

Sea $\mathbb{Z} = (\zeta_1, \zeta_2, \dots, \zeta_n)$ una variable aleatoria n-dimensional, se presentan dos tipos de distribuciones derivadas:

- La distribución marginal: Surge de ignorar n-1 variables de las que integran \mathbb{Z}
- La distribución condicionada: Supedita el comportamiento de un conjunto de unas sobre otras

Distribución marginal n-dimensional

Generalización

Al tomar el intervalo completo de cada una de las n-1 variables que se quieren ignorar, el valor de su función de distribución es 1.

$$F_1(x_1) = P(\zeta_1 \leq x_1, \zeta_2 < \infty, \dots, \zeta_n < \infty) = F(x_1, \infty, \dots, \infty)$$

$$F_2(x_2) = P(\zeta_1 < \infty, \zeta_2 \leq x_2, \dots, \zeta_n < \infty) = F(\infty, x_2, \dots, \infty)$$

.

.

.

$$F_n(x_n) = P(\zeta_1 < \infty, \zeta_2 < \infty, \dots, \zeta_n \leq x_n) = F(\infty, \infty, \dots, x_n)$$

En el caso particular bivalente

$$F_1(x) = P(\zeta \leq x; \eta < \infty) = F(x, \infty)$$

$$F_2(y) = P(\zeta < \infty; \eta \leq y) = F(\infty, y)$$

Variables discretas

$$F(x_1, \infty, \dots, \infty) = \sum_{i_1=1}^{x_1} \sum_{i_2=1}^{\infty} \dots \sum_{i_n=1}^{\infty} P(\zeta_{i_j} = x_{i_j}) = \sum_{i_1=1}^{x_1} p_{i_1 \dots i_n} = \sum_{i_1=1}^{x_1} p_{i_1}$$

$$F_n(x_n) = F(\infty, \infty, \dots, x_n) = \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{\infty} \dots \sum_{i_n=1}^{x_n} P(\zeta_{i_j} = x_{i_j}) = \sum_{i_n=1}^{x_n} p_{\dots i_n} = \sum_{i_n=1}^{x_n} p_{i_n}$$

Caso particular bivalente

$$F_1(x) = F(x, \infty) = \sum_{i=1}^x p_{i\blacksquare}$$

$$F_2(y) = F(\infty, y) = \sum_{j=1}^y p_{\blacksquare j}$$

Variables continuas (generalización)

$$F(x_1, \infty, \dots, \infty) = \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n = \int_{-\infty}^{x_1} f(x_1) dx_1$$

.....

$$F(\infty, \infty, \dots, x_n) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{x_n} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n = \int_{-\infty}^{x_n} f(x_n) dx_n$$

Distribuciones condicionadas

Surgen cuando en una distribución conjunta se establece una condición sobre una de las variables (o un conjunto de ellas sobre otro conjunto).

La distribución incondicional expresa el comportamiento probabilístico de una variable cuando las otras están sujetas a ciertas condiciones, por ejemplo, variando el intervalo dado.

Se va a desarrollar primero el caso bivalente para posteriormente generalizar. Se parte de la definición de la probabilidad condicionada:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

En el caso discreto y atendiendo a lo que se ha indicado en líneas anteriores

$$P(\zeta = x | \eta = y) = \frac{P(\zeta = x, \eta = y)}{P(\eta = y)} = \frac{p_{ij}}{p_{\blacksquare j}}$$

$$P(\eta = y | \zeta = x) = \frac{P(\zeta = x, \eta = y)}{P(\zeta = x)} = \frac{p_{ij}}{p_{i\blacksquare}}$$

Funciones de distribución bivariantes, discretas

$$F(x_i | y) = P(\zeta \leq x_i | \eta = y) = \sum_{i=1}^x \frac{p_{ij}}{p_{\blacksquare j}}$$

$$F(y_j | x) = P(\eta \leq y_j | \zeta = x) = \sum_{j=1}^y \frac{p_{ij}}{p_{i\blacksquare}}$$

Generalización n-dimensional y de variable continua

$$F(x_1, x_2, \dots, x_q | x_{q+1}, x_{q+2}, \dots, x_n) = \frac{\int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \int_{-\infty}^{x_n} f(x_1, x_2, \dots, x_n) dx_1 dx_2, \dots, dx_n}{\int_{-\infty}^{x_{q+1}} \int_{-\infty}^{x_{q+2}} \int_{-\infty}^{x_n} f(x_{q+1}, x_{q+2}, \dots, x_n) dx_{q+1} dx_{q+2}, \dots, dx_n}$$

Momentos.

Momentos respecto del origen- conjuntos y marginales

Se explican a partir de los casos particulares bidimensionales pudiendo extenderse hasta n dimensiones.

Esperanza

Si se escribe $\mathbb{Z} = (\zeta_1, \zeta_2, \dots, \zeta_n)$ traspuesta (como matriz columna)

$$\mathbb{Z}^t = \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \dots \\ \zeta_n \end{pmatrix}$$

$$E(\mathbb{Z}^t) = E \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \dots \\ \zeta_n \end{pmatrix} = \begin{pmatrix} E(\zeta_1) \\ E(\zeta_2) \\ E(\zeta_3) \\ \dots \\ E(\zeta_n) \end{pmatrix}$$

La esperanza de una matriz es la matriz de las esperanzas.

Caso particular bidimensional

Por definición el r-ésimo momento respecto al origen es el valor de la Esperanza de la variable elevado a r. Dado lo que se ha indicado anteriormente y particularizando a dos variables, se escribiría:

$$\alpha_{r,m} = E(\zeta^r, \eta^m)$$

Se observa rápidamente que:

$$\alpha_{1,0} = E(\zeta^1, \eta^0) = E(\zeta)$$

$$\alpha_{0,1} = E(\zeta^0, \eta^1) = E(\eta)$$

Esto es, las esperanzas de las distribuciones marginales y en el caso siguiente

$$\alpha_{1,1} = E(\zeta^1, \eta^1) = E(\zeta\eta)$$

el valor esperado conjunto

Teorema de la existencia de $\alpha_{1,1}$

Si existen los momentos marginales de segundo orden respecto del origen

$$\alpha_{2,0} = E(\zeta^2) \text{ y } \alpha_{0,2} = E(\eta^2)$$

existe $\alpha_{1,1} = E(\zeta\eta)$

Demostración

$$(\zeta + \eta)^2 = \zeta^2 + \eta^2 + 2\zeta\eta \geq 0$$

$$\zeta^2 + \eta^2 \geq |2\zeta\eta|$$

$$\frac{\zeta^2 + \eta^2}{2} \geq |\zeta\eta|$$

Tomando esperanzas

$$E(|\zeta\eta|) \leq \frac{1}{2}(E(\zeta^2) + E(\eta^2)) < \infty$$

Converge absolutamente, luego $E(\zeta\eta)$ existe porque está acotada superiormente.

Lo anterior se generaliza a n variables teniendo en cuenta la matriz de esperanzas inicialmente expuesta.

Esperanza conjunta: Caso continuo

Sea $f_{\zeta,\eta}(x, y)$ la función de densidad conjunta de una variable bidimensional

$$E(\zeta\eta) = \iint_{-\infty}^{\infty} xy f_{\zeta,\eta}(x, y) d\zeta d\eta$$

Esperanza conjunta: Caso discreto

$$E(\zeta\eta) = \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} xy P(\zeta = x; \eta = y)$$

Por otro lado, si ζ y η son variables aleatorias independientes, entonces

$$E(\zeta\eta) = E(\zeta)E(\eta)$$

Momentos respecto de la media- conjuntos y marginales

Caso continuo

$$\mu_{r_1 r_2 \dots r_n} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (x_1 - \mu_1)^{r_1} (x_2 - \mu_2)^{r_2} \dots (x_n - \mu_n)^{r_n} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$

Caso discreto

$$\mu_{r_1 r_2 \dots r_n} = \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} \dots \sum_{-\infty}^{\infty} (x_1 - \mu_1)^{r_1} (x_2 - \mu_2)^{r_2} \dots (x_n - \mu_n)^{r_n} (\zeta_1 = x_1; \zeta_2 = x_2; \dots; \zeta_n = x_n)$$

Caso particular bidimensional

Sea $f_{\zeta, \eta}(x, y)$ la función de densidad conjunta de una variable bidimensional

Caso continuo

$$\mu_{r,s} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_1)^r (y - \mu_2)^s f(x,y) dx dy$$

Caso discreto

$$\mu_{r,s} = \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} (x - \mu_1)^r (y - \mu_2)^s P(\zeta = x; \eta = y)$$

Casos particulares:

Varianzas marginales

$$\mu_{2,0} = V(\zeta)$$

$$\mu_{0,2} = V(\eta)$$

Varianza conjunta

$$\mu_{2,2} = V(\zeta\eta) = E\left((\zeta - \alpha_{1,0})^2 (\eta - \alpha_{0,1})^2\right) = \sigma_{\zeta, \eta}^2$$

Covarianza

$$\mu_{1,1} = E\left((\zeta - \alpha_{1,0})(\eta - \alpha_{0,1})\right) = Cov(\zeta, \eta) = \sigma_{\zeta, \eta} = \alpha_{1,1} - \alpha_{1,0}\alpha_{0,1}$$

Su signo indica el sentido de la relación que puede existir entre las variables

$Cov(\zeta, \eta) > 0 \rightarrow$ relación directa

$Cov(\zeta, \eta) < 0 \rightarrow$ relación indirecta

Teorema de la existencia

Si existen $E(\zeta^2)$ y $E(\eta^2)$ existen los momentos anteriores y por tanto las varianzas marginales también, luego existe $\alpha_{1,1} = E(\zeta\eta)$

Matriz de covarianzas

$$\Sigma = \begin{bmatrix} \sigma_{1,1}^2 & \sigma_{1,2} & \dots & \sigma_{1,n} \\ \sigma_{2,1} & \sigma_{2,2}^2 & \dots & \sigma_{2,n} \\ \sigma_{n,1} & \sigma_{n,2} & \dots & \sigma_{n,n}^2 \end{bmatrix}$$

- La matriz de covarianzas es simétrica $\sigma_{i,j} = \sigma_{j,i}$
- Su diagonal es el vector de varianzas conjuntas
- $\Sigma \geq 0$ es no negativa

Sea $C \neq 0$ un vector de constantes de orden $(n,1)$

$$C^t \Sigma C = C^t E((Z - \mu)(Z - \mu)^t) C = E(C^t (Z - \mu)(Z - \mu)^t C) = E(C^t (Z_s - \mu)^2) \geq 0$$

$(Z - \mu)(Z - \mu)^t$ Escalares

Positiva por estar al cuadrado

- Si existe dependencia lineal entre las variables aleatorias ζ_i del vector Z entonces la matriz Σ es estrictamente positiva.
- Si $\sigma_{i,j} = \sigma_{j,i} = 0 \forall i \neq j$ entonces Σ es diagonal siendo su traza

$$\text{tr}(\Sigma) = \sum_{i=1}^n \sigma_i^2 = \sigma_{\zeta_1 + \zeta_2 + \dots + \zeta_n}^2$$

- Si $\zeta_1, \zeta_2, \dots, \zeta_n$ son estadísticamente independientes Σ será diagonal y su traza es la misma que en el punto anterior.

Esperanza condicionada

Vamos a estudiar el caso particular bidimensional. Dada una variable aleatoria bidimensional (ζ, η) , se define la esperanza condicionada de ζ dado $\eta = y$ y de η dado $\zeta = x$ como:

$$E(\zeta|\eta = y) = \sum_{i=1}^{\infty} x_i P(\zeta = x_i | \eta = y)$$

$$E(\eta|\zeta = x) = \sum_{i=1}^{\infty} y_i P(\eta = y_i | \zeta = x)$$

Si son discretas y como:

$$E(\zeta|\eta = y) = \int_{-\infty}^{\infty} x f_{\zeta|\eta}(x|y) dx$$

$$E(\eta|\zeta = x) = \int_{-\infty}^{\infty} y f_{\eta|\zeta}(y|x) dy$$

Si es continua.

Se puede generalizar a cualquier transformación de x :

$$E(\zeta|\eta = y) = \sum_{i=1}^{\infty} g(x_i) P(\zeta = x_i | \eta = y)$$

$$E(\zeta|\eta = y) = \int_{-\infty}^{\infty} g(x) f_{\zeta|\eta}(x|y) dx$$

Y análogamente para la otra variable.

La esperanza condicionada tiene una interpretación y es que la esperanza condicionada a un valor fijo, como una constante, supone corresponder con el valor marginal de la esperanza de la variable siempre que las variables sean independientes.

$$E(\zeta|\eta = y) = E(x)$$

$$E(\eta|\zeta = x) = E(y)$$

Además, esto se cumple de forma iterativa

$$E(E(\zeta|\eta = y)) = E(x)$$

$$E(E(\eta|\zeta = x)) = E(y)$$

Coeficientes de correlación.

El coeficiente de correlación lineal entre dos variables aleatorias ζ, η se define como

$$\rho = \frac{\sigma_{\zeta,\eta}}{\sigma_{\zeta}\sigma_{\eta}}$$

La covarianza de las dos variables entre el producto de sus desviaciones típicas marginales.

El signo de ρ viene del signo de la covarianza y tiene el mismo significado.

Cuando ρ es negativo las variables están incorreladas linealmente. Si es ± 1 la correlación lineal es perfecta (directa o indirecta según el signo)

La correlación lineal no implica independencia estadística. Pero si hay independencia, la correlación lineal será cero.

Independencia.

Diremos que dos variables son independientes a través de las funciones de distribución condicionadas, si:

$$F(x/y)=F(x) \text{ y } F(y/x)=F(y)$$

Se podrá escribir que:

$$f(x,y)=f(x)f(y)$$

$$F(X,y)=F(x)F(y)$$

Consecuentemente:

$$E(\zeta|\eta = y) = E(x)$$

$$E(\eta|\zeta = x) = E(y)$$

Así mismo como consecuencia de lo anterior $\sigma_{\zeta,\eta} = 0$ lo recíproco no se puede afirmar.

EJERCICIO 1

Estadística

Tema 8. Regresión lineal simple. Hipótesis: Propiedades. Estimación. Contraste de hipótesis. Predicción.

*Esta publicación no tiene carácter oficial, se trata de material didáctico cuyo objeto es servir de apoyo en la preparación de la oposición al Cuerpo Superior de Actuarios, Estadísticos y Economistas de la Administración de la Seguridad Social. Esta documentación es **orientativa y no es exclusiva** ni única para el correcto desarrollo de los temas. Tampoco vincula al órgano convocante ni al Tribunal actuante. Los errores o desactualizaciones que pudieran contenerse en la documentación no serán imputables a la SESSP, debiéndose atender en todo caso a la legislación vigente publicada en el Boletín Oficial del Estado.*

Aviso: La SESSP se reserva las acciones legales que pudieran corresponder por el uso lucrativo de esta información. Queda prohibido expresamente la comercialización o venta del presente material.

Edición Enero 2025

Regresión lineal simple.

La regresión lineal es un caso particular de los modelos lineales generalizados (GLM) cuando la función link es la identidad y la distribución del modelo es la Normal, que pertenece a la familia exponencial.

$$f(x; \theta) = h(x)B(\theta)e^{Q(x)R(\theta)}$$

A través de un modelo se trata de explicar el comportamiento de una variable aleatoria en función de otra (regresión lineal ordinaria simple-RLOS) u otras (regresión lineal ordinaria múltiple- RLOM) en el caso en el que la relación entre una y otra(s) sea de tipo lineal.

En el caso de RLOS el modelo se describe

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

y_i es la variable dependiente, variable objetivo o variable respuesta

β_0 es el término independiente, no depende de ninguna variable y contiene toda la información que no es explicada por el resto de variables independientes o los errores

β_1 es el parámetro que acompaña a la variable independiente o explicativa. Su signo nos dirá la relación entre la variable respuesta y la explicativa.

β_0 y β_1 son los coeficientes del modelo

x_i es la variable independiente o explicativa

ϵ_i es el término que recoge los errores

Hipótesis: Propiedades.

- Ausencia de multicolinealidad:

La multicolinealidad se da cuando existe alta correlación entre variables independientes. En el caso de la regresión lineal simple, al hacerse un modelo sobre una única variable independiente, no existe problema de multicolinealidad.

Cuando se trata de regresión lineal múltiple la alta correlación entre variables explicativas puede causar que el modelo estime de forma errónea, por lo que es necesario hacer análisis de correlación dos a dos antes de proponer variables al modelo (se verá en el tema de regresión lineal múltiple).

- Normalidad de los errores con desviación típica constante (homocedasticidad; lo contrario es heterocedasticidad) $\epsilon_i \rightarrow N(0, \sigma)$
- Independencia de los errores $E(\epsilon_i \epsilon_j) = 0 \forall i \neq j$

Si no se da esta independencia puede hacer perder eficiencia de los estimadores del modelo. Una forma de verificarlo es a través del test de Durbin-Watson (estudio de la incorrelación de los estimadores)

$$d = \frac{\sum (e_i - e_{i-1})^2}{\sum e_i^2}$$
$$0 < d < 4$$

Si $d=2$ o aproximadamente 2 indica que no hay correlación entre los residuos.

Si es mucho menor que 2 indica una correlación positiva, mientras que si es mayor que 2 indica que la correlación es negativa. Valores por encima de 3 o debajo de 1 pueden dar lugar a subestimar el nivel de significación del modelo.

En muchas ocasiones se aplica un modelo RLO sin verificar previamente el cumplimiento de las hipótesis, lo cual puede llevar a conclusiones desacertadas.

Otras cuestiones a tener en cuenta:

- En la muestra posibles datos ausentes: Si una variable tiene muchos datos ausentes no debería entrar en el modelo. En ocasiones las variables que recogen, por ejemplo, en una encuesta, datos económicos, vienen sin dato o con un dato poco fiable. Si el número de ausentes es bajo se pueden imputar valores para que podamos usar la variable en el modelo. La imputación de valores tiene que ser teniendo en cuenta que el valor imputado, si es puntual, debería ser robusto (no una media, por ejemplo, mejor una mediana) e idealmente se pueden imputar valores haciendo uso de otro modelo de regresión teniendo como respuesta la variable sobre la que queremos imputar valores (siempre que no sean demasiados o generará una evidente multicolinealidad en la regresión en la que ésta sea explicativa)
- En la muestra posibles outliers que sean realmente errores y no outliers (una edad de 903 años, por ejemplo)

Por ello antes de iniciar cualquier modelización se debe trabajar previamente con el conjunto de datos y hacer un estudio exploratorio y descriptivo de cada una de las variables de forma univariante y posteriormente bivariante para cada par de explicativas y entre cada explicativa y respuesta. También puede ser recomendable hacer un análisis multivariante y valorar si recoger el efecto de un conjunto de variables explicativas en un único factor si se ha detectado que existe correlación entre ellas o crear clases (clúster) y hacer distintos modelos para cada clase que será más homogénea que el total de las clases.

Variables categóricas (alfanuméricas): Crear dummies

Además, es necesario que aquellas variables que sean categóricas se pasen a n-1 dummies de sus categorías esto es, si una variable tiene como respuestas posibles {A, B, C, D}, se deben crear 3 variables con respuesta [0,1] que informen de la presencia o ausencia de A, la primera, de B la segunda, etc. De una de ellas no se crea dummy ya que si las n-1 categorías customizadas (dummy) tienen como respuesta 0, significa que la última tiene como respuesta 1 y dicha información aparecerá recogida en el término independiente.

Estimación.

Nuestro objetivo es estimar el valor de los coeficientes de tal manera que el resultado de \hat{y}_i se aproxime o se ajuste lo más cerca posible a y_i de estructura desconocida.

Los coeficientes son funciones de los valores muestrales.

Estimación por MCO (mínimos Cuadrados ordinarios)

$$\widehat{\beta}_0 = \bar{y} - \beta_1 \bar{x} = \bar{y} - \frac{S_{xy}}{S_x^2} \bar{x}$$

$$\widehat{\beta}_1 = \frac{S_{xy}}{S_x^2}$$

Distribución de los estimadores

Los coeficientes siguen una distribución normal cuando se conoce la desviación típica poblacional.

$$\beta_0 \rightarrow N\left(\beta_0, \frac{\sigma}{\sqrt{n}} \sqrt{1 + \frac{\bar{x}^2}{S_x^2}}\right)$$

$$\beta_1 \rightarrow N\left(\beta_1, \frac{\sigma}{S_x \sqrt{n}}\right)$$

Estos estimadores son insesgados, consistentes y eficientes.

La covarianza entre los coeficientes nos da la relación entre los estimadores.

$$\text{Cov}(\beta_0, \beta_1) = -\bar{x} \frac{1}{\chi_n^2}$$

Si se conoce σ (desviación típica)

$$\sigma^2 = \frac{\sum e_i^2}{n} = \frac{\sum (y_i - \hat{y}_i)^2}{n} = \frac{SCE}{n}$$

Los intervalos de confianza serán $P\left(Z > Z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$

$$\beta_0 \in [\widehat{\beta}_0 \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \sqrt{1 + \frac{\bar{x}^2}{S_x^2}}]$$

$$\beta_1 \in [\widehat{\beta}_1 \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n} S_x}]$$

- S_x es la cuasidesviación típica
- n el tamaño de la muestra
- $Z_{\frac{\alpha}{2}}$ es el valor de la variable tipificada $N(0,1)$ para una significación de $\alpha/2$
- \bar{x} es la media muestral

En el caso de que se desconozca la desviación típica (que será lo habitual y que es el supuesto bajo el que trabajan los software estadísticos), ésta se estimará por la **varianza residual** (es la varianza de los residuos)

$$\widehat{\sigma}^2 = S_R^2 = \frac{\sum e_i^2}{n-2} = \frac{\sum (y_i - \hat{y}_i)^2}{n-2} = \frac{n}{n-2} \left(S_y^2 - \frac{S_{xy}^2}{S_x^2} \right)$$

Y se verifica que

$$S_R \rightarrow t_{n-2}$$

$$\frac{\sum e_i^2}{\sigma^2} = \frac{(n-2)S_R^2}{\sigma^2} \rightarrow \chi_{n-2}^2$$

Los **intervalos de confianza para los coeficientes** quedan como $(P(t_{n-2} > t_{\frac{\alpha}{2}}) = \frac{\alpha}{2})$

$$\beta_0 \in \left[\widehat{\beta}_0 \pm t_{\frac{\alpha}{2}} \frac{S_R}{\sqrt{n}} \sqrt{1 + \frac{\bar{x}^2}{S_x^2}} \right]$$

$$\beta_1 \in \left[\widehat{\beta}_1 \pm t_{\frac{\alpha}{2}} \frac{S_R}{\sqrt{n} S_x} \right]$$

Y será con la t de Student con la distribución que se verificarán las hipótesis del contraste para cada uno de los coeficientes (uno en el caso de RLOS)

Descomposición de la variabilidad y análisis de correlación

Denominando:

$$SCT \text{ Suma de cuadrados total } SCT = \sum (y_i - \bar{y})^2$$

$$SCE \text{ Suma de cuadrados de los errores } SCE = \sum (y_i - \hat{y}_i)^2$$

$$SCR \text{ Suma de cuadrados de la regresión } SCR = \sum (\hat{y}_i - \bar{y})^2$$

$$SCR = SCT - SCE$$

Cuando $\hat{y}_i \rightarrow y_i$ SCR \rightarrow SCT

Se denominará **razón de correlación** al cociente entre la variabilidad explicada por la regresión entre la variabilidad total.

$$R^2 = \frac{SCR}{SCT} = \frac{SCT - SCE}{SCT} = 1 - \frac{SCE}{SCT} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

Dividiendo numerador y denominador por n

$$R^2 = \frac{\sigma_e^2}{\sigma_y^2}$$

Otra forma de expresarlo es:

$$R^2 = \frac{SCR}{SCT} = \frac{SCR/n}{SCT/n} = \frac{\sum(\hat{y}_i - \bar{y})^2/n}{\sum(y_i - \bar{y})^2/n} = \frac{\sigma_R^2}{\sigma_y^2}$$

En el caso de que la regresión sea lineal, esta razón coincide con el **coeficiente de determinación** (Sustituyendo por las estimaciones de β_0 y β_1)

$$R^2 = \frac{S_{xy}^2}{S_x^2 S_y^2}$$

Siendo su raíz cuadrada el **coeficiente de correlación lineal**:

$$\rho = \sqrt{R^2} = \frac{S_{xy}}{S_x S_y}$$

No tiene dimensión, toma los valores siempre entre -1 y 1 (siendo -1 correlación perfecta inversa y 1 correlación perfecta directa). Si existe independencia su valor es cero (el recíproco no tiene por qué ser cierto)

Contraste de hipótesis.

En el contraste de la regresión se trata de valorar la siguiente hipótesis:

$$H_0 \Rightarrow \beta_1 = 0$$

$$H_1 \Rightarrow \beta_1 \neq 0$$

Siendo el estadístico de contraste

$$t = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{S_R^2}{n S_x^2}}} \rightarrow t_{n-2}$$

Siendo el intervalo de confianza, como se ha indicado antes:

$$\beta_1 \in \left[\hat{\beta}_1 \pm t_{\alpha/2} \frac{S_R}{\sqrt{n} S_x} \right]$$

Así mismo teniendo en cuenta:

$$P\left(t_{n-2} > t_{\alpha/2}\right) = \frac{\alpha}{2}$$

Siendo el estadístico de contraste

$$t = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\frac{e^t e}{n - k - 1} \sqrt{u_{ii}}}} \rightarrow t_{n-k-1}$$

Se rechaza si $t \geq t_{n-k-1}$ a un nivel de significación α . La variable va a ser significativa en cuanto $t \geq \pm 2$

Este valor t tendrá asociada una probabilidad que se denomina p-valor. Cuanto mayor sea t menor será el p-valor. Cuando el p-valor sea menor que α es porque el valor $t \geq \pm 2$. Un coeficiente con un p-valor cercano a cero hará que se rechace la hipótesis nula y que por tanto se admita que la variable de dicho coeficiente es significativa en el modelo.

Otro contraste para valorar el modelo es sobre la relación de dependencia entre x e y; si el coeficiente es cero eso significa que la pendiente de la recta es cero y por tanto no existe relación de dependencia entre x e y, por tanto, x no puede explicar y por ello también se puede establecer que:

$$H_0 \Rightarrow \rho = 0$$

$$H_1 \Rightarrow \rho \neq 0$$

Si se rechaza la H_0 los datos son consistentes con una relación lineal, aunque podrían tener otro tipo de relación (esto no implica que no puedan tener relaciones de otro tipo). Si se acepta entonces la relación entre las variables o no existe o si existe, no es lineal.

En este caso, el estadístico de contraste es:

$$\frac{R\sqrt{n-2}}{\sqrt{1-R^2}} \rightarrow t_{n-2}$$

O También

$$\frac{R^2(n-2)}{1-R^2} \rightarrow F_{1,n-2}$$

En el caso de que la hipótesis se haga sobre un valor concreto del coeficiente,

$$H_0 \Rightarrow \rho = \rho_0$$

Se usa la transformación de Fisher:

$$Z = \frac{1}{2} \ln \left(\frac{1+R}{1-R} \right) \rightarrow N \left(\frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right), \sqrt{\frac{1}{n-3}} \right)$$

Predicción.

Estimados los coeficientes la predicción consiste en ver qué respuesta se obtiene y_i al ir dando distintos valores a x_i que se multiplicarán por el valor del coeficiente estimado (más el término independiente estimado)

En un análisis de datos y proposición de modelo se suele usar un porcentaje (ej. 80%) de la muestra para hacer el modelo y el 20% restante para validar el mismo y se comprueba si al meter nuevos datos el modelo sigue funcionando. Se realizará una pila de comprobaciones

como contrastes de bondad de ajuste, estudio del coeficiente de determinación, comprobar si existe sobreajuste, etc.

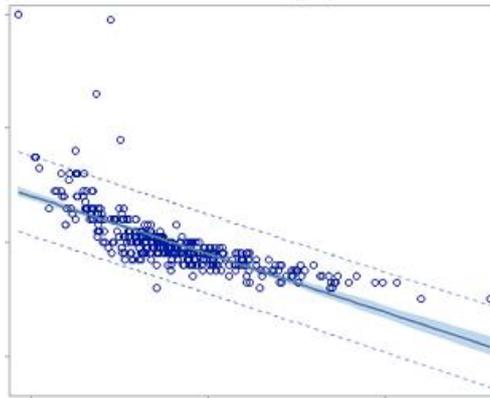
Para observar el ajuste que tiene la predicción sobre el modelo se estudian los residuos.

Estudio de residuos

$$e_i = y_i - \beta_0 - \beta_1 x_i = y_i - \hat{y}_i$$

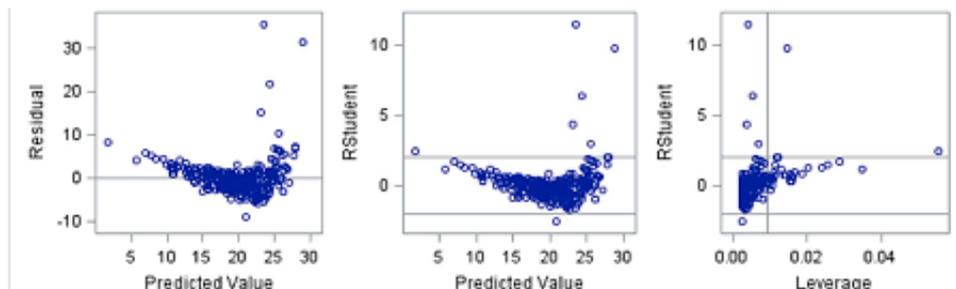
Se pueden estandarizar dividiendo e_i por una estimación de s desviación estándar.

Sirven para detectar outliers, puntos palanca así como la influencia de otras variables explicativas que no se hayan tenido en cuenta. También ayudan a observar si existe heterocedasticidad, patrones temporales, etc.



Los residuos que se observan fuera de las líneas de puntos (que indicarían los límites del intervalo de confianza) son outliers y pueden ser además puntos palanca (leverage).

Otra gráfica que puede observarse es la de valores predichos frente a residuos estandarizados (tipificados con la cuasidesviación típica). Estas gráficas nos ayudan a visualizar y diagnosticar el modelo bajo las medidas de la t de Student, además de ver si existen puntos leverage, cuanto más alejados estén los valores resultantes del eje de coordenadas (cero) mayor el valor del residuo y por tanto más probable que sean valores leverage. Se puede pedir, con un software estadístico, se listen aquellas observaciones de la regresión donde en valor absoluto, sus residuos estandarizados sean mayores que 2.

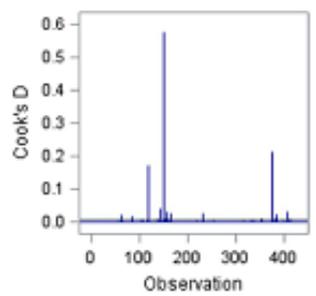


Otro gráfico que nos ayuda a ver si el modelo es bueno es el de la D de Cook, con él se mide la distancia que provoca cada observación respecto del modelo planteado, esto es, puntos palanca.

El valor más bajo de la D de Cook es el cero y cuanto más se aleje de éste, más influencia tiene esa observación en el modelo (más efecto palanca). El umbral se suele establecer en $4/n$ siendo n el número de observaciones utilizadas. Con un buen software se puede pedir que se listen todas las observaciones que superan ese punto.

Generalmente una palanca con un valor mayor de $\frac{2k+2}{n}$ debe estudiarse y examinarse pormenorizadamente (k es el número de predictores, en regresión lineal simple 1, y n el número de observaciones).

En la siguiente gráfica se observan 3 puntos palanca con fuerte influencia.



Una medida muy relacionada con la D de Cook son los DFFITs. Estas medidas combinan la información sobre los residuos y los puntos palanca el total de los coeficientes (las betas). El umbral para DFFITs se establece en $2\sqrt{k/n}$ (k número de predictores, 1 en regresión lineal simple, y n número de observaciones). Cuanto más cercano a cero sea el valor, menor influencia tiene dicha observación sobre el modelo.

Para considerar de forma más específica cómo influyen las observaciones en cada coeficiente, se estudia la medida DFBETA y se crea para cada uno de los coeficientes (1 en el caso de regresión lineal simple). El umbral ahora será $2\sqrt{n}$

EJERCICIO 1

Estadística

Tema 9. Regresión lineal múltiple: Hipótesis. Estimación. Propiedades. Contraste de hipótesis. Predicción.

*Esta publicación no tiene carácter oficial, se trata de material didáctico cuyo objeto es servir de apoyo en la preparación de la oposición al Cuerpo Superior de Actuarios, Estadísticos y Economistas de la Administración de la Seguridad Social. Esta documentación es **orientativa y no es exclusiva** ni única para el correcto desarrollo de los temas. Tampoco vincula al órgano convocante ni al Tribunal actuante. Los errores o desactualizaciones que pudieran contenerse en la documentación no serán imputables a la SESSP, debiéndose atender en todo caso a la legislación vigente publicada en el Boletín Oficial del Estado.*

Aviso: La SESSP se reserva las acciones legales que pudieran corresponder por el uso lucrativo de esta información. Queda prohibido expresamente la comercialización o venta del presente material.

Edición Enero 2025

Regresión lineal múltiple

La regresión lineal es un caso particular de los modelos lineales generalizados (GLM) cuando la función link es la identidad y la distribución del modelo es la Normal, que pertenece a la familia exponencial.

$$f(x; \theta) = h(x)B(\theta)e^{Q(x)R(\theta)}$$

A través de un modelo se trata de explicar el comportamiento de una variable aleatoria en función de otra (regresión lineal ordinaria simple-RLOS) u otras (regresión lineal ordinaria múltiple- RLOM) en el caso en el que la relación entre una y otra(s) sea de tipo lineal.

En el caso de RLOM el modelo se describe

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \epsilon_i$$

y_i es la variable dependiente, variable objetivo o variable respuesta

β_0 es el término independiente, no depende de ninguna variable y contiene toda la información que no es explicada por el resto de variables independientes o los errores

β_i es el parámetro que acompaña a cada variable independiente o explicativa. Su signo nos dirá la relación entre la variable respuesta y la explicativa.

β_0 y las β_i son los coeficientes del modelo

x_{ij} es la variable independiente o explicativa

ϵ_i es el término que recoge los errores

Matricialmente se expresa como

$$Y = X\beta + \epsilon$$

Hipótesis

- $n \geq k+1$ y $\text{rango}(X) = k+1$ (siendo k el número de predictores, betas, y n el número de observaciones)
- Ausencia de multicolinealidad:
La colinearidad indica que hay variables explicativas que son combinación lineal de otras explicativas que están en el modelo.

Cuando se trata de regresión lineal múltiple la alta correlación entre variables explicativas puede causar que el modelo estime de forma errónea, por lo que es necesario hacer análisis de correlación dos a dos antes de proponer variables al modelo (se verá en el tema de regresión lineal múltiple). La colinearidad perfecta de las variables impedirá calcular los estimadores de las betas dado que $(X^t X)$ resulta ser singular.

Una alta colinearidad vuelve muy imprecisas las estimaciones y puede hacer que variables que son importantes y significativas, no lo parezcan.

Es importante por tanto revisar, antes de aplicar un modelo RLOM, un análisis de correlación bivariante entre cada una de las variables candidatas a ser explicativas. Si dos variables candidatas a ser explicativas están muy correlacionadas se debe elegir una u otra o bien buscar una nueva variable, transformada de las anteriores, para el modelo.

Se puede conocer si el modelo tiene multicolinealidad estudiando la inflación de la varianza del factor (VIF), de modo que si el valor de ésta es mayor a 10 se indica que existe multicolinealidad. La medida de tolerancia es $1/VIF$ y en ese caso se revisará cuando $TOL > 0,1$. Estos valores se deben pedir al software con el que se esté estudiando la posibilidad de aplicar un modelo RLOM. Además, se podrá pedir, según lo potente que sea el software, que se estudie lo anterior incluyendo y no incluyendo el término independiente.

$$TOL = I_c = \sqrt{\frac{\text{Autovalor máx}}{\text{Autovalor Mín}}}$$

- $E(\epsilon_i) = 0 \forall i = 1, 2, \dots, n$
- $V(\epsilon_i) = \sigma^2 \rightarrow$ homocedasticidad (varianza constante de los errores) $\forall i = 1, 2, \dots, n$

Si la heterocedasticidad es baja no hay problema, si es alta puede ser debido a la omisión de alguna variable que contiene información importante.

En caso de heterocedasticidad la prueba t de Student pierde potencia. Gráficamente se puede apreciar homocedasticidad mediante la ausencia de patrones en las gráficas de residuos (si hay patrón pensaremos que existe heterocedasticidad). Además de la prueba gráfica es necesario hacer contrastes de homocedasticidad (Test de Goldfeld y Quandt, de Breusch y Pagan, Test de White, prueba de Levene ..).

A través del coeficiente de correlación de Pearson para cada par de variables explicativas se observan sus relaciones. La matriz de correlación debe tener valores ± 1 en la diagonal y lo más próximo a cero en el resto de la matriz.

Gráficamente se observa con los llamados Scatter-plot o gráfico de dispersión, que debe presentar ausencia de patrones lineales entre las dos variables explicativas que se observan. Esto significa que un par de variables puede tener una relación pero que esta no sea lineal.

Por el contrario, el scatter-plot debe presentar algún tipo de patrón o relación lineal entre cada explicativa y la que se pretende explicar.

- $Cov(\epsilon_i, \epsilon_j) = 0$ porque los errores son independientes

Si no se da esta independencia puede hacer perder eficiencia de los estimadores del modelo. Una forma de verificarlo es a través del test de Durbin-Watson (estudio de la incorrelación de los estimadores)

$$d = \frac{\sum(e_i - e_{i-1})^2}{\sum e_i^2}$$

$$0 < d < 4$$

Si $d=2$ o aproximadamente 2 indica que no hay correlación entre los residuos.

Si es mucho menor que 2 indica una correlación positiva, mientras que si es mayor que 2 indica que la correlación es negativa. Valores por encima de 3 o debajo de 1 pueden dar lugar a subestimar el nivel de significación del modelo.

Como $\epsilon_i \rightarrow N(0, \sigma)$ Normalmente distribuidos, la matriz de varianzas-covarianzas es, por tanto, diagonal

$$\Sigma_{\epsilon} = Cov(\epsilon_i, \epsilon_j) = \sigma^2 I_n$$

Siendo I_n la matriz identidad de orden n .

En muchas ocasiones se aplica un modelo RLO sin verificar previamente el cumplimiento de las hipótesis, lo cual puede llevar a conclusiones desacertadas.

Otras cuestiones a tener en cuenta:

- En la muestra posibles datos ausentes: Si una variable tiene muchos datos ausentes no debería entrar en el modelo. En ocasiones las variables que recogen, por ejemplo, en una encuesta, datos económicos, vienen sin dato o con un dato poco fiable. Si el número de ausentes es bajo se pueden imputar valores para que podamos usar la variable en el modelo. La imputación de valores tiene que ser teniendo en cuenta que el valor imputado, si es puntual, debería ser robusto (no una media, por ejemplo, mejor una mediana) e idealmente se pueden imputar valores haciendo uso de otro modelo de regresión teniendo como respuesta la variable sobre la que queremos imputar valores (siempre que no sean demasiados o generará una evidente multicolinealidad en la regresión en la que ésta sea explicativa)
- En la muestra posibles outliers que sean realmente errores y no outliers (una edad de 903 años, por ejemplo)

Por ello antes de iniciar cualquier modelización se debe trabajar previamente con el conjunto de datos y hacer un estudio exploratorio y descriptivo de cada una de las variables de forma univariante y posteriormente bivariante para cada par de explicativas y entre cada explicativa y respuesta. También puede ser recomendable hacer un análisis multivariante y valorar si recoger el efecto de un conjunto de variables explicativas en un único factor si se ha detectado que existe correlación entre ellas o crear clases (clúster) y hacer distintos modelos para cada clase que será más homogénea que el total de las clases.

Variables categóricas (alfanuméricas): Crear dummies

Además, es necesario que aquellas variables que sean categóricas se pasen a $n-1$ dummies de sus categorías esto es, si una variable tiene como respuestas posibles {A, B, C, D}, se deben crear 3 variables con respuesta [0,1] que informen de la presencia o ausencia de A,

la primera, de B la segunda, etc. De una de ellas no se crea dummy ya que si las n-1 categorías customizadas (dummy) tienen como respuesta 0, significa que la última tiene como respuesta 1 y dicha información aparecerá recogida en el término independiente.

Estimación. Propiedades.

Matricialmente, los **predictores** se estiman:

$$\hat{\beta} = (X^t X)^{-1} (X^t Y)$$

$\hat{\beta}$ es combinación lineal de las componentes del vector Y, por lo que se distribuye según una variable aleatoria normal

Llamando u_{ii} a los elementos (i, i) de la matriz (la diagonal)

$$(X^t X)^{-1} = \begin{pmatrix} u_{00} & u_{01} & u_{0k} \\ u_{10} & u_{11} & u_{1k} \\ \dots & \dots & \dots \\ u_{k0} & u_{k1} & u_{kk} \end{pmatrix}$$

$$\hat{\beta}_i \rightarrow N(\beta_i, \sigma \sqrt{u_{ii}}) \quad \forall i, j = 1, \dots, k$$

Luego la estimación del modelo se escribe matricialmente como:

$$\hat{Y} = X \hat{\beta}$$

Teniendo en cuenta que:

$$E(\hat{\beta}) = \beta$$

Los **coeficientes** representan el efecto de la variable explicativa sobre la variable explicada, su signo informa sobre la relación directa o inversa y miden el cambio esperado en la respuesta ante pequeñas variaciones en las variables explicativas:

$$\beta_i = \frac{\partial}{\partial x_i} E(y)$$

Y la **matriz de varianzas-covarianzas** de los predictores es

$$\Sigma_{\hat{\beta}} = Cov(\hat{\beta}_i \hat{\beta}_j) = \sigma^2 (X^t X)^{-1}$$

Estimación de la **varianza de los residuos**

Sea

$$e = \hat{Y} - Y$$

$$\sigma^2 = \frac{E(e^t e)}{n - k - 1}$$

Por lo que el estimador es:

$$S_R^2 = \hat{\sigma}^2 = \frac{e^t e}{n - k - 1} = \frac{Y^t Y - \hat{\beta}^t X^t Y}{n - k - 1}$$

Teniendo en cuenta que

$$\frac{1}{\sigma^2} e^t e = \frac{1}{\sigma^2} \sum_{i=1}^n e_i^2 \rightarrow \chi_{n-k-1}^2$$

Coefficiente de determinación: R cuadrado

Es el cociente entre la variabilidad explicada por la regresión (SCR) y la variabilidad total (SCT)

$$R^2 = \frac{SCR}{SCT} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

$$\sum(y_i - \bar{y})^2 = \sum(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n e_i^2 = \sum(\hat{y}_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2$$

$$R^2 = \frac{SCT - SCE}{SCT} = 1 - \frac{SCE}{SCT} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

Matricialmente:

$$R^2 = \frac{\hat{\beta}^t X^t Y - n\bar{y}^2}{Y^t Y - n\bar{y}^2} = \frac{Y^t X \hat{\beta} - n\bar{y}^2}{Y^t Y - n\bar{y}^2}$$

R cuadrado ajustado

Al aumentar el número de regresores (predictores) este coeficiente aumenta ya que disminuye la variabilidad no explicada. Este incremento nos puede confundir. Debe existir una relación entre el número de observaciones y el de predictores ya que es muy sensible a los valores de n y de k. Se usa entonces el **“R cuadrado ajustado”** que penaliza el número de variables regresoras que se incluyen en el modelo, ajustándolo a los grados de libertad (observaciones).

$$\overline{R^2} = \frac{\frac{1}{n - k - 1} \sum_{i=1}^n e_i^2}{\frac{1}{n - 1} \sum(y_i - \bar{y})^2} = 1 - \frac{n - 1}{n - k - 1} (1 - R^2)$$

Coefficiente de correlación parcial

Se hace entre pares de variables independientes una vez se han eliminado en ambas variables los efectos debido al resto.

En un conjunto con $X_1, X_2, X_3, \dots, X_k$ el coeficiente de correlación parcial entre X_1, X_2 se denota $r_{12 \cdot 3 \dots k}$

- En primer lugar, se calcula una regresión lineal entre X_1 respecto del resto de variables distintas de X_2 , donde los residuos serán $e_{1 \cdot 3 \dots k}$
- Se calcula la recta de regresión de X_2 respecto del resto de variables distintas de X_1 donde los residuos serán $e_{2 \cdot 3 \dots k}$
- El coeficiente de correlación parcial entre X_1, X_2 es el coeficiente de correlación lineal simple entre las variables $e_{1 \cdot 3 \dots k}$ y $e_{2 \cdot 3 \dots k}$

Por su parte para calcular el coeficiente de correlación parcial entre la variable respuesta Y y cualquiera de las regresoras X_i se utiliza el estadístico de contraste individual de la t respecto a la variable X_i

$$t = \frac{\hat{\beta}_i}{S_R \sqrt{c_{ii}}} \quad \forall i = 1, 2, \dots, k$$

Obteniendo:

$$r_{Y_i \cdot C}^2 = \frac{t_i^2}{t_i^2 + n - k - 1}$$

Donde $C = \{1, 2, \dots, i - 1, i + 1, \dots, k\}$ el conjunto de índices de todas las variables regresoras excepto el que ocupa precisamente el i -ésimo.

Contraste de hipótesis.

Contraste global

$$H_0 \Rightarrow \beta_i = 0 \quad \forall i = 1, \dots, k$$

La hipótesis alternativa es que al menos uno de los predictores es distinto de cero, en cuyo caso el modelo se será globalmente por bueno.

El contraste se hace a través de la tabla ANOVA, con el estadístico F de Snedecor de modo que cuando el p -valor sea inferior al nivel de significación se rechazará la hipótesis nula de que todos los predictores son iguales entre sí e iguales a cero.

ANOVA

Variación	Suma de cuadrados	Grados de libertad	Media cuadrática	F valor
Explicada por la regresión	SCE	K	SCE/k	$\frac{SCE/k}{SCR/(n - k - 1)}$
Residual	SCR	n-k-1	SCR/n-k-1	
Total	SCT	n-1		

$$F = \frac{\frac{1}{k}R^2}{\frac{1}{n-k-1}(1-R^2)} = \frac{SCE/k}{SCR/(n-k-1)} \rightarrow F_{k,n-k-1}$$

Se rechaza si $F \geq F_{k,n-k-1}$ a un nivel de significación α .

Este valor F tendrá asociada una probabilidad que se denomina p-valor. Cuanto mayor sea F menor será el p-valor. Un coeficiente con un p-valor cercano a cero hará que se rechace la hipótesis nula y que por tanto se admita que la variable de dicho coeficiente es significativa en el modelo.

El Test t

Sirve para contrastar la hipótesis nula para cada uno de los coeficientes de forma individual de modo que observamos la significatividad de la variable i-ésima

$$H_0 \Rightarrow \beta_i = 0$$

$$H_1 \Rightarrow \beta_i \neq 0$$

Siendo el estadístico de contraste

$$t = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\frac{e^t e}{n-k-1} \sqrt{u_{ii}}}} \rightarrow t_{n-k-1}$$

Se rechaza si $t \geq t_{n-k-1}$ a un nivel de significación α . La variable va a ser significativa en cuanto $t \geq \pm 2$

Este valor t tendrá asociada una probabilidad que se denomina p-valor. Cuanto mayor sea t menor será el p-valor. Cuando el p-valor sea menor que α es porque el valor $t \geq \pm 2$. Un coeficiente con un p-valor cercano a cero hará que se rechace la hipótesis nula y que por tanto se admita que la variable de dicho coeficiente es significativa en el modelo.

Predicción.

Para observar el ajuste que tiene la predicción sobre el modelo se estudian los residuos.

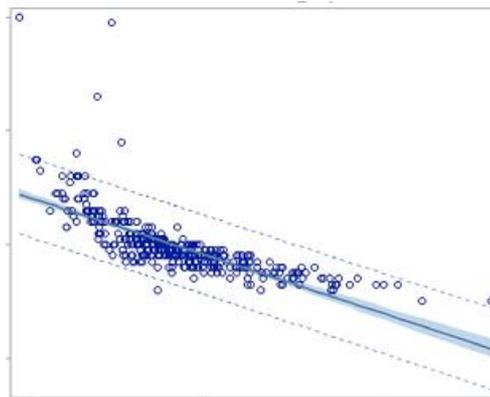
Estimados los coeficientes la predicción consiste en ver qué respuesta se obtiene y_i al ir dando distintos valores a las x_i , (que se multiplican por el valor de los coeficientes estimados)

En un análisis de datos y proposición de modelo se suele usar un porcentaje (ej. 80%) de la muestra para hacer el modelo y el 20% restante para validar el mismo y se comprueba si al meter nuevos datos el modelo sigue funcionando. Se realizará una pila de comprobaciones como contrastes de bondad de ajuste, estudio del coeficiente de determinación, comprobar si existe sobreajuste, etc.

Estudio de residuos

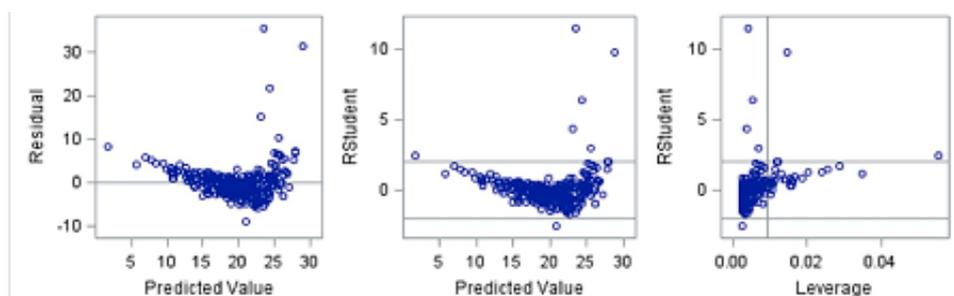
Se pueden estandarizar dividiendo e_i por una estimación de s desviación estándar.

Sirven para detectar outliers, puntos palanca, así como la influencia de otras variables explicativas que no se hayan tenido en cuenta. También ayudan a observar si existe heterocedasticidad, patrones temporales, etc.



Los residuos que se observan fuera de las líneas de puntos (que indicarían los límites del intervalo de confianza) son outliers y pueden ser además puntos palanca (leverage).

Otra gráfica que puede observarse es la de valores predichos frente a residuos estudentizados (tipificados con la cuasidesviación típica). Estas gráficas nos ayudan a visualizar y diagnosticar el modelo bajo las medidas de la t de Student, además de ver si existen puntos leverage, cuanto más alejados estén los valores resultantes del eje de coordenadas (cero) mayor el valor del residuo y por tanto más probable que sean valores leverage. Se puede pedir, con un software estadístico, se listen aquellas observaciones de la regresión donde en valor absoluto, sus residuos estudentizados sean mayores que 2.

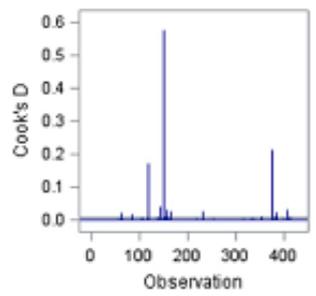


Otro gráfico que nos ayuda a ver si el modelo es bueno es el de la D de Cook, con él se mide la distancia que provoca cada observación respecto del modelo planteado, esto es, puntos palanca.

El valor más bajo de la D de Cook es el cero y cuanto más se aleje de éste, más influencia tiene esa observación en el modelo (más efecto palanca). El umbral se suele establecer en $4/n$ siendo n el número de observaciones utilizadas. Con un buen software se puede pedir que se listen todas las observaciones que superan ese punto.

Generalmente una palanca con un valor mayor de $\frac{2k+2}{n}$ debe estudiarse y examinarse pormenorizadamente.

En la siguiente gráfica se observan 3 puntos palanca con fuerte influencia.



Una medida muy relacionada con la D de Cook son los DFFITs. Estas medidas combinan la información sobre los residuos y los puntos palanca el total de los coeficientes (las betas). El umbral para DFFITs se establece en $2\sqrt{k/n}$. Cuanto más cercano a cero sea el valor, menor influencia tiene dicha observación sobre el modelo.

Para considerar de forma más específica cómo influyen las observaciones en cada coeficiente, se estudia la medida DFBETA y se crea para cada uno de los coeficientes. El umbral ahora será $2\sqrt{n}$

EJERCICIO 1

Estadística

Tema 10. Análisis de series temporales.
Componentes fundamentales. Tendencia general.
Desestacionalización de series históricas.
Componente cíclica.

*Esta publicación no tiene carácter oficial, se trata de material didáctico cuyo objeto es servir de apoyo en la preparación de la oposición al Cuerpo Superior de Actuarios, Estadísticos y Economistas de la Administración de la Seguridad Social. Esta documentación es **orientativa y no es exclusiva ni única** para el correcto desarrollo de los temas. Tampoco vincula al órgano convocante ni al Tribunal actuante. Los errores o desactualizaciones que pudieran contenerse en la documentación no serán imputables a la SESSP, debiéndose atender en todo caso a la legislación vigente publicada en el Boletín Oficial del Estado.*

Aviso: La SESSP se reserva las acciones legales que pudieran corresponder por el uso lucrativo de esta información. Queda prohibido expresamente la comercialización o venta del presente material.

Edición Enero 2025

Análisis de series temporales.

Una serie temporal es una secuencia ordenada de observaciones cada una de las cuales se asocia a un momento de tiempo. Por tanto, estamos ante procesos estocásticos, asociados a distintos momentos de tiempo, sobre los que se establecen las siguientes condiciones:

- El conjunto de datos está ordenado de forma cronológica.
- No existe independencia entre las observaciones y esta dependencia es objeto del estudio y parte de la naturaleza de estas series.

Debido a estas particularidades las series temporales deben ser desarrolladas bajo modelos específicos que recogen esta dependencia y la aprovechan en su estudio entre las observaciones ordenadas.

El estudio de estas series de observaciones se denomina Análisis de Series Temporales y el instrumento con el que se analizan es a través de un modelo que va a tratar de reproducir el comportamiento de la variable objeto de estudio para poder, a partir del mismo, hacer predicciones a futuro.

Estos modelos pueden ser:

- Univariantes: Cuando la serie se analiza únicamente en función de su pasado.
- Multivariantes: Se analizan varias series temporales a la vez, existiendo cierta dependencia o relación entre los pasados de las distintas series (un ejemplo en Seguridad Social la serie temporal de accidentes de trabajo con la de afiliación y la de presupuesto destinado a prevención de riesgos)

Una serie temporal univariante se denotaría por

$$Y_t$$

El subíndice t indica el tiempo en el que se observa el dato. La serie temporal puede ser discreta si $t=1,2,\dots,T$ (días, meses, años..) esto es, finito o infinito numerable; o puede ser continua si se considera el tiempo como lo que es, una variable continua ($\in \mathbb{R}$) con un número infinito de observaciones.

Las variables con las que se trabaja en la serie temporal pueden ser denominadas variables flujo o variables stock.

Variables flujo: Aquellas que se miden respecto a un intervalo de tiempo (consumo mensual de un determinado producto)

Variables Stock: aquellas que se miden en un determinado momento del tiempo (afiliación a la seguridad social a final del año)

Objetivos a la hora de analizar una serie temporal.

- **Describir:** La descripción permite conocer el comportamiento que ha tenido la variable a lo largo del tiempo de estudio. Esta descripción se puede hacer mediante algunos estadísticos resumen y representaciones gráficas. Dada la particularidad de la serie respecto a la dependencia entre sus observaciones, suele añadirse

algún tipo de función que describa el valor medio de la serie, la variación media entre dos periodos, etc.

- **Predecir:** Las predicciones que se realicen a partir de una serie y, por tanto, a partir de datos pasados, no son más que extrapolaciones del comportamiento del pasado hacia el futuro, si bien nos pueden dar un punto de referencia para luego utilizar modelos más sofisticados de predicción. Si la predicción puede darse de forma exacta entonces la serie se dice que es determinista, pero generalmente son estocásticas, esto es, los valores futuros tendrán una distribución de probabilidad que se condiciona al conocimiento de las observaciones pasadas.

Generalmente los modelos de series temporales en el ámbito de la economía suelen responder a una estructura en el que se da una **parte sistemática** o regular de la variable y una parte aleatoria que se denomina **innovación**.

$$Y_t = PS_t + \alpha_t$$

PS_t parte sistemática

α_t innovación

Ambas dependen de t. Concretamente S_t de determinará únicamente en función de la información de las observaciones pasadas

$$PS_t = f(Y_t, Y_{t-1}, Y_{t-2}, \dots)$$

Componentes fundamentales.

Se basa en la idea de que una serie temporal puede ser considerada como la superposición de varios componentes fundamentales no observables:

- **Tendencia:** Movimiento a largo plazo que señala la evolución general o de conjunto. T_t
- **Estacionalidad:** Oscilaciones que se dan dentro de un año, generalmente, por cambios que se producen por costumbres sociales como las vacaciones de semana santa o navidad o cambios climáticos en primavera, verano, otoño e invierno. S_t
- **Ciclo:** Esta componente capta – en el caso de que existan – oscilaciones que son periódicas y se dan dentro de la tendencia a medio plazo. C_t
- **Error:** Esta componente recoge todo aquello que no se puede explicar por las otras tres y se representa por una variable aleatoria de valor esperado nulo y varianza constante. r_t

Cualquier método de descomposición en componentes fundamentales de una serie temporal va a tratar de aislar estos componentes y estudiarlos de forma separada para luego observar la forma en la que actúan conjuntamente, dando lugar a distintos modelos de descomposición: **aditivo, multiplicativo o mixtos**.

Aditivo

Es adecuado cuando no se producen interacciones en el modelo, esto es, cuando un componente no depende de los otros.

$$Y_t = T_t + C_t + S_t + r_t$$

Multiplicativo

Al contrario que en el esquema aditivo, si los componentes varían con los otros, se debe usar un modelo multiplicativo.

$$Y_t = T_t \cdot C_t \cdot S_t \cdot r_t$$

Al contrario que otros modelos paramétricos como los ARIMA que tratan de obtener la representación de la serie en términos de la interrelación temporal de sus elementos y que caracterizan las series como sumas o diferencias de las variables aleatorias o de las series resultantes (estas sumas o diferencias pueden a su vez estar ponderadas según ciertos pesos).

Análisis gráfico

El primer estudio que debe hacerse sobre una serie temporal es el estudio gráfico, donde se va a observar si existe tendencia (al alza o a la baja), y se podrá observar si hay valores outliers, la amplitud que pueda darse entre las oscilaciones estacionales, posibilidad de ciclos, etc.

Series estacionarias/no estacionarias

Una serie será estacionaria cuando su comportamiento medio y la variabilidad son constantes en el tiempo (homocedasticidad).

Será no estacionaria cuando la media y su varianza varían con el tiempo.

Tendencia general.

En el análisis gráfico previo se habrá podido detectar si existe una tendencia lineal, en este caso se tratará de aplicar un **modelo de tendencia lineal** aplicando el criterio de mínimos cuadrados ordinarios (MCO) para la estimación de los parámetros de la recta (visto en el tema de regresión lineal).

$$Y_t = \beta_0 + \beta_1 T_t + \varepsilon_t$$

En el caso de que el esquema sea multiplicativo, se tomarán logaritmos y se modelizará, obteniendo los parámetros de la recta también a través de MCO (la tendencia observada sería **exponencial**)

$$\ln(Y_t) = \beta_0 + \beta_1 T_t + \varepsilon_t$$

Si se observa un modelo de **tendencia cuadrático** se ajustará a:

$$Y_t = \beta_0 + \beta_1 T_t + \beta_2 T_t^2 + \varepsilon_t$$

Otra forma de observar la tendencia es con el método de **medias móviles (MM)**, en el que se realiza un cálculo de medias aritméticas consecutivas con el mismo número de datos en cada una de ellas, pero retirando el primer dato de media anterior e incorporando el dato siguiente. La longitud, tamaño y orden de las MM es el número de datos utilizado.

Con este método se suaviza la tendencia, se eliminan las variaciones a corto y medio plazo y algunas irregularidades no controlables, de modo que la gráfica de MM nos va a permitir observar la tendencia.

Desestacionalización de series históricas.

En primer lugar, cualquiera de los métodos explicados en el apartado anterior para detectar la tendencia han eliminado la estacionalidad de la serie, por lo que no se repiten pero al extraer la tendencia de una serie precisamente se ha eliminado de la misma la estacionalidad dentro de los periodos.

La estacionalidad es un cambio en la media de la serie que se repite de forma periódica cada S estaciones (mensual, trimestral, semestral, bimestral, etc.)

Para detectar la estacionalidad se usan distintos métodos igualmente. En primer lugar, en la gráfica si se ven ciclos dentro de cada uno de los años, se puede entender que existe estacionalidad.

Para detectar la misma de forma analítica:

Método de la razón a la media móvil

$$Y_t / \bar{Y}_t$$

Índices de variación estacional - IVE

Se calcula la media para cada mes (o cada trimestre o cada cuatrimestre, dependiendo del comportamiento estacional).

Una forma de desestacionalizar es precisamente dividir la serie original entre el IVE

Dummies

Una forma de captar la estacionalidad consiste en crear dummies (variables dicotómicas) para cada periodo. Si la serie es trimestral, por ejemplo, se crearían 4 dummies una para cada trimestre. La variable dicotómica expresaría la presencia o ausencia de estacionalidad en dicho trimestre

$$S_t = \beta_0 + \beta_1 S1_t + \beta_2 S2_t + \beta_3 S3_t + \beta_4 S4_t + \varepsilon_t$$

Transformación logarítmica

Una forma de conseguir que la dispersión sea más o menos constante a medida que crece la media es realizar una transformación logarítmica sobre la serie.

La tasa logarítmica indicará la variación de una variable y es un indicador de crecimiento relativo

$$\ln Y_t - \ln Y_{t-1} \cong \frac{Y_t - Y_{t-1}}{Y_{t-1}}$$

Diferenciación

Diferencia regular:

Para observar la estacionalidad eliminando la tendencia, esto es, induciendo estacionariedad en media, podemos usar el método de diferenciación. Se toma una diferencia regular y consiste en calcular la diferencia entre cada dato (por ejemplo, trimestral) y el anterior. Como en las medias móviles, se pierde el primer dato de la serie.

$$\nabla Y_t = Y_t - Y_{t-4}$$

Diferencia estacional:

Para desestacionalizar la serie entonces lo que se hará es obtener la diferencia de un valor de la serie con un valor de la serie de un periodo anterior (por ejemplo, el primer semestre de este año con respecto al anterior y así sucesivamente)

$$\nabla_s Y_t = Y_t - Y_{t-s}$$

Siendo s el periodo estacional.

Diferenciación en la tasa logarítmica:

Es un indicador de la aceleración de la tasa de crecimiento relativo de una variable.

$$\nabla(\ln Y_t - \ln Y_{t-1})$$

Lo más recomendable es tomar todas las transformaciones que induzcan la estacionariedad, la transformación logarítmica para conseguir que la varianza sea constante, la diferencia regular, para eliminar la tendencia, y la estacional, para eliminar el componente estacional. Con ello se estabiliza la media y la varianza de una serie temporal.

Componente cíclica.

El ciclo se define como oscilaciones en torno a la tendencia que se deben a la alternancia entre periodos de crisis y prosperidad, esto es son cambios significativos en la tendencia que se repiten cada cierto tiempo, a medio y largo plazo. Incluye cualquier característica distinta al resto de los componentes incluido el error o ruido.

La forma de eliminar el ciclo es observar la serie como distintas series en función de la variación de la tendencia, por ejemplo, si cada 20 años se produce un cambio de ciclo de alcista a bajista se pueden tomar las dos series como series separadas con inflexión cada 20 años en promedio.