UNIVERSIDAD POLITÉCNICA DE MADRID

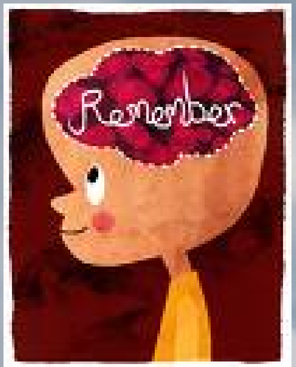
conento

Aprendiendo idiomas II

María Jesús Vázquez Gallo

Formación Conento: octubre 2008

1. Introducción: Inferencia estadística.
2. Estimación paramétrica.
3. Esperanza matemática.
4. Continuará…



En Aprendiendo idiomas I hemos realizado un análisis descriptivo de los datos de la variable aleatoria: número de latas de cerveza que bebe anualmente cada habitante de Dakota del Norte, a partir de una muestra extraída de la población de Dakota del Norte.

Usando medidas de tendencia como media, mediana y moda, afirmamos:

→Los de Dakota del Norte en promedio consumen 353 latas al año.

→ Hay tantos consumos por encima de 360 como por debajo.

→ 360 es el número preferido de latas de cerveza por año.

Usando medidas de dispersión como la desviación típica, vimos:

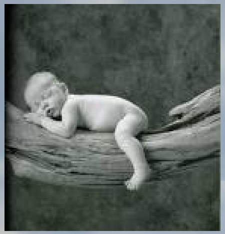
→En promedio un individuo se aleja 22.86 del número medio de latas.

Volviendo al diccionario, nuestra variable aleatoria: número de latas que bebe anualmente cada habitante; sería una variable aleatoria discreta mientras que, por ejemplo, la altura de cada habitante, sería continua puesto que puede tomar como valor cualquier número (real) en un intervalo entre la altura mínima y la máxima.

Variable aleatoria discreta

315 316 …

Variable aleatoria continua



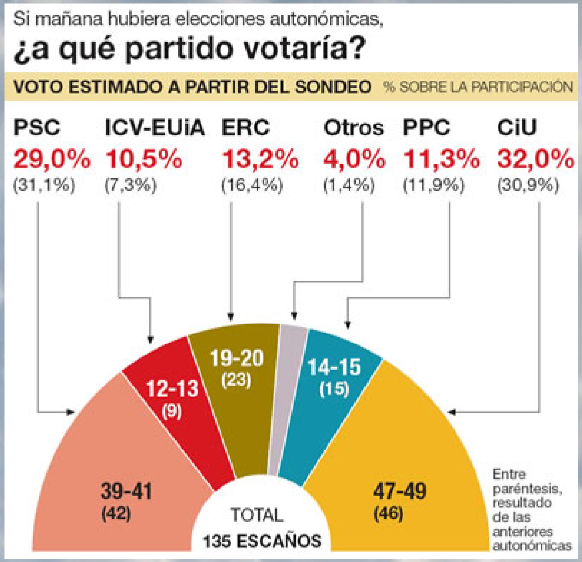
50 cm 203 cm

Pasamos a conceptos relacionados con la INFERENCIA ESTADÍSTICA, es decir, la extracción de conclusiones, modelos y predicciones de una población a partir de la muestra.

- Dos problemas típicos de la inferencia:

1. Estimación: aproximar datos numéricos desconocido de la población a partir de la muestra.

Ejemplo: aproximar el porcentaje de población que votará a un partido a partir de una encuesta.



El Periódico, 24 de octubre de 2006

2. Contraste de hipótesis: comprobar alguna información sobre la población a partir de la muestra.

Ejemplo: comprobar que el partido “Políticos reunidos” obtendrá más del 50% de los votos a partir de una encuesta.

Empezamos con la estimación y trataremos la próxima vez el contraste.

Estimación:

- **Paramétrica**

  Se busca aproximar o estimar características numéricas que describan la población: los parámetros.

- **No Paramétrica**

  Interesa la distribución de los valores de la población (función de probabilidad –caso discreto- o de densidad –variables continuas-) para saber a qué modelo se ajusta.

Estimación: Pensemos primero en la **estimación paramétrica,**

Pero ¿qué es un parámetro?

Imagina que queremos estudiar el precio medio de un producto concreto que se vende en España.
Parece natural recabar el precio del producto en diversas tiendas y calcular la media aritmética de todos ellos. Ese número aproximará el precio medio.

Aquí:

- la población la constituyen todos los precios a los que se vende el producto en España;
- la muestra son los precios que hemos recogido;
- el parámetro es el verdadero precio medio (desconocido) y
- el estimador es la media de los precios recogidos (estos serían las observaciones de la muestra).

En general:

Parámetro: número que se calcula a partir de la población y que sirve para describirla.

Ejemplo: la media de la población, su varianza, etc.

Estimador: función de la muestra (o estadístico) que se usa para aproximar o estimar un parámetro.

Ejemplo: la media muestral , la varianza muestral, etc.

muestra es aleatoria $\longrightarrow$ estimador es una variable aleatoria

Uff… me surgen dudas…

¿Para cada parámetro sólo hay un posible estimador?

Y si hay varios ¿con cuál nos quedamos?

Elegiremos entre estimadores según sus propiedades:

1. Sesgo: el sesgo de un estimador es la diferencia entre su “valor esperado” (esperanza) y el valor del parámetro que estima.

Ejemplo: la media muestral como estimador de la media de la población es insesgado (sesgo=0) pero la varianza muestral es un estimador sesgado de la varianza de la población.

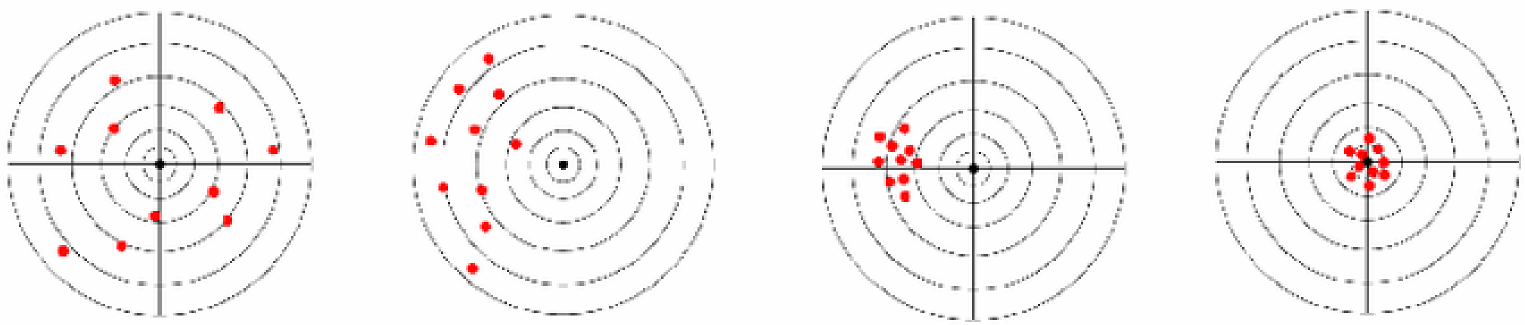
2. Eficiencia: un estimador es más eficiente cuanto menor sea su varianza.

Ejemplo: la media muestral es más eficiente que la mediana muestral, cuando ambos estiman la media de la población (porque la mediana fluctúa más).

Observa: en Aprendiendo idiomas I, los números que calculamos eran la media, mediana, moda, varianza y desviación típica muestrales.
Nos sirvieron como estimadores de la media, mediana, etc. de la población de Dakota del Norte

Propiedades de los estimadores: un ejemplo

Cuatro tiradores han efectuado 10 disparos sobre una diana. Si traducimos cada disparo en una estimación, efectuada por un determinado estimador, sobre una muestra, podemos interpretar las propiedades de los estimadores de la siguiente forma:



Tirador A	Tirador B	Tirador C	Tirador D
Estimador insesgado y no eficiente	Estimador sesgado y no eficiente	Estimador sesgado y eficiente	Estimador insesgado y eficiente

http://maralboran.org/wikipedia/index.php/Estad%C3%ADsticos

Insesgado: centro de masas de disparos (esperanza) es el centro

Eficiente: la fluctuación (varianza) es baja

Preferiremos en general estimadores insesgados y eficientes

Otras propiedades de los estimadores:

Consistencia: al crecer el tamaño de la muestra, el valor del estimador se acerca al del parámetro (el sesgo tiende a 0)

Ejemplo: la varianza muestral (s^2) es un estimador sesgado pero consistente de la varianza de la población (σ^2). Lo mismo con su raíz cuadrada (s) respecto a la desviación típica (σ).

Robustez: al variar ligeramente las hipótesis en las que se basa la estimación, no se alteran significativamente los resultados que da el estimador. Eliminan en cierto modo la influencia de observaciones extremas y por tanto son útiles cuando se tienen muestras pequeñas.

Ejemplo: La media muestral no es un estimador robusto.
La mediana muestral sí.

¿Qué es exactamente la esperanza de una variable aleatoria?

El concepto lo introdujo Huygens, astrónomo, físico y matemático holandés del S. XVII, en su obra *Libellus de Ratiotiniis in Ludo Aleae*.

Si esperamos ganar *x o y*, con igual probabilidad, entonces la expectativa es $(x+y)/2$. Y si hay *n* posibles resultados con la misma probabilidad, el valor esperado es su suma dividida por n.

Si hay *p* oportunidades de ganar *x*, y *q* oportunidades de ganar *y*; y además, cada uno de los *p+q* resultados ocurre con igual probabilidad, el valor esperado o esperanza será igual a

$$x\frac{p}{p+q}+y\frac{q}{p+q}$$

A Huygens le siguió Jacob Bernoulli.
Su razonamiento utiliza la noción de frecuencia: en una baza concreta el resultado es incierto pero, basándose en la experiencia de pasadas partidas, se puede valorar a priori los porcentajes de veces en que se gana o se pierde.
Estas proporciones posteriormente fueron asimiladas a probabilidades.

El valor esperado del juego sería entonces:

Ganancia x (proporción en que se gana)
- Pérdida x (proporción en que se pierde)

¿Sabías que hubo varias docenas de personas en la familia del famoso Bach que fueron músicos eminentes del siglo XVI al XIX? Hay regiones de Alemania donde la palabra “bach” significa músico.
Lo mismo ocurrió con la familia Bernoulli para las Matemáticas y las Ciencias. En tres generaciones, esa notable familia suiza produjo ocho matemáticos -tres de ellos extraordinarios- que, a su vez, tuvieron un buen lote de descendientes que se distinguieron en muchos campos.

La aplicación del enfoque de la probabilidad llevó, para el caso general de un juego que presentara más de dos posibles resultados, a que el valor esperado del juego si se jugara un número infinito de veces, sería una suma infinita en la que cada sumando es uno de los posible resultados por la probabilidad de que suceda.

Para una variable aleatoria discreta X que pueda tomar infinitos valores, su esperanza es

$$E(X) = \sum_{i=1}^{\infty} x_i P(X = x_i)$$

Decir valor esperado a veces confunde porque la esperanza no proporciona un valor que realmente podamos esperar para nuestra variable. ¿Se puede afirmar, por ejemplo, que la esperanza matemática de un juego proporciona el valor que se espera obtener del mismo?

No, por ejemplo, el valor esperado del juego proporciona el valor del juego para un infinito número de partidas.
Por este motivo la mayoría de los juegos de casino o de azar son injustos (ya sabes para quién), porque la cantidad apostada supera la ganancia esperada en la práctica, ya que ningún jugador dispone de infinito dinero para apostar, ni de tiempo para jugar infinitas partidas.


Continuará...