(#438).TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIIe)

[MONOTEMA] Avanzamos con el quinto apartado del tercer capítulo de Probability Theory and Statistical Inference, de Aris Spanos.

Parámetros y momentos

Además del histograma de la distribución de datos observados, también disponemos de ciertos números que caracterizan la distribución como la media o la varianza. Esos valores numéricos están relacionados con los momentos de la distribución, que son esperanzas matemáticas de ciertas funciones de la variable aleatoria X, genéricamente denotadas por $h(X)$.

$$E[h(X)] = \int_{-\infty}^{\infty} h(x) f_x(x;\theta)dx = g(\theta)$$

Si escogemos diferentes funciones $g(\theta)$ obtendremos diferentes momentos de la distribución. Por ejemplo:

Media μ

$h(X) = X, x \in \mathbb{R}_X$

Para variables continuas:

$$\mu = E[h(X)] = E(X) = \int_{-\infty}^{\infty} x f_x(x;\theta)dx$$

Para variables discretas:

$$\mu = E(X) = \sum_{x_i \in \mathbb{R}_X} x_i f_x(x_i;\theta)$$

Varianza σ^2

$$h(X) = [X - E(X)]^2, x \in \mathbb{R}_X$$

$$\sigma^2 = h(X) = [X - E(X)]^2 = \int_{-\infty}^{\infty} [x-\mu]^2 f_x(x;\theta)dx$$

Una forma conveniente de calcular los momentos de una distribución es a través de la función generatriz de momentos (mgf), donde $h(X) = e^{tX}$

$$m_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tX} f(x)dx, t \in (-h, h), h > 0$$

Para variables aleatorias discretas las integrales se vuelven sumatorios.

Por ejemplo, para una variable aleatoria X que sigue una distribución de Poisson:

$$m_X(t) = E(e^{tX}) = \sum_{r=0}^{\infty} e^{tX} f(x)dx$$

Dado que: $f_x = \frac{e^{-\theta}\theta^r}{r!}$

Entonces:

$$m_X(t) = E(e^{tX}) = \sum_{r=0}^{\infty} e^{tX} \left[\frac{e^{-\theta}\theta^r}{r!}dx\right] = e^{-\theta}\sum_{r=0}^{\infty} \frac{(e^t e^{\theta})^r}{r!} = e^{\theta(e^t-1)}$$

A partir de los momentos de la distribución se puede estudiar la asimetría y el apuntamiento. De este modo, podemos caracterizar la forma de la distribución a partir de los momentos.

Todos los posts relacionados

(#449). AUMENTO DE CASOS DE GLIOBLASTOMA TAMBIÉN EN FRANCIA

(#448). INCREMENTO DE LA INCIDENCIA DE GLIOBLASTOMA

(#446) RENDIMIENTO EN BASE A VICTORIAS PRODUCIDAS POR LOS JUGADORES

(#445) LOS JUGADORES MÁS PRODUCTIVOS DE LA ACB 2019/20

(#443). LOS JUGADORES MÁS PRODUCTIVOS DE LA NBA 2019/20

(#438).TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIIe)

(#437). TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIId)

(#436). TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIIc)

(#435).TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIIb)

(#434).TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIIa)

(#428) TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIc)

(#427) TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIb)

(#426) TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIa)

(#424) TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (I)

(#422). JUGADORES ACB AL NIVEL DE NBA EN LA TEMPORADA 2018/19

(#421). LOS JUGADORES MÁS PRODUCTIVOS DE LA ACB 2018/19

(#420). LA PARADOJA DE CHEVALIER DE MERE Y LOS ESTUDIANTES DE MARKETING

(#412). LOS JUGADORES MÁS PRODUCTIVOS DE LA NBA 2018/19

(#411). LOS ROOKIES MÁS PRODUCTIVOS DE LA NBA 2018/19

(#410) ¿QUIÉN HA SIDO EL MEJOR ANOTADOR DE LA FASE REGULAR DE LA EUROLIGA 2018/19?

(#409). ALGUNAS CONSIDERACIONES EN LA ESTIMACIÓN DEL RENDIMIENTO POR MINUTO EN BALONCESTO (II)

(#408). ALGUNAS CONSIDERACIONES EN LA ESTIMACIÓN DEL RENDIMIENTO POR MINUTO EN BALONCESTO (I)

(#407). AVANCE DE LOS NOVATOS MÁS PRODUCTIVOS DE LA NBA 2018/19

(#398). AJUSTE DE FUNCIONES (II): CAPACIDAD EXPLICATIVA Y PARTICIÓN DE LA FUNCIÓN

(#397). AJUSTE DE FUNCIONES (I): SPLINES VS MÍNIMOS CUADRADOS

(#396). INTERPOLACIÓN (IV): MÉTODO DE SPLINE CÚBICO

(#395). INTERPOLACIÓN (III): MÉTODO DE DIFERENCIAS DIVIDIDAS

(#394). INTERPOLACIÓN (II): MÉTODO DE NEVILLE

(#393). INTERPOLACIÓN (I): POLINOMIO DE LAGRANGE

(#391). BÚSQUEDA DE SOLUCIONES (V): MÉTODO DE LA POSICIÓN FALSA

(#390). BÚSQUEDA DE SOLUCIONES (IV): MÉTODO DE LA SECANTE

(#389). BÚSQUEDA DE SOLUCIONES (III): METODO DE NEWTON

(#388). BÚSQUEDA DE SOLUCIONES (II): METODO DEL PUNTO FIJO

(#387). BÚSQUEDA DE SOLUCIONES (I): MÉTODO DE LA BISECCIÓN

(#384). MEJOR UN SÓLO ITEM QUE VARIOS PARA MEDIR ACTITUDES

(#381). TABLAS DE TUMORES TOTALES EN EL ESTUDIO DEL NTP SOBRE EFECTOS DE LA RADIACIÓN DE TELÉFONO MÓVIL

(#380). ÍNDICES APROXIMADOS FLEXIBLES EN ECUACIONES ESTRUCTURALES

(377). AVANZANDO EN LA COMPLEJIDAD: DOBLE DEPENDENCIA

(#376). CAMPOS DIRECCIONALES, COMENZANDO A REPRESENTAR LA COMPLEJIDAD

(#373). PREDICCIONES CON BUENAS APROXIMACIONES

(#366). MODELO DE HOMOGENEIDAD EN TABLAS DE 2X2

(#365). BÁSICOS DE ECUACIONES ESTRUCTURALES (V): MATRIZ DE DATOS BRUTOS

(#358). BÁSICOS DE ECUACIONES ESTRUCTURALES (IV): REGRESIÓN LINEAL SIMPLE CON FIABILIDADES DIVERSAS

(#354). BÁSICOS DE ECUACIONES ESTRUCTURALES (III): COVARIANZA Y CORRELACIÓN ENTRE VARIABLES LATENTES

(#353). BÁSICOS DE ECUACIONES ESTRUCTURALES (II): VARIABLES LATENTES Y FIABILIDAD

(#352). TESTOSTERONA Y CONSUMO CONSPICUO EN HOMBRES

(#350). BÁSICOS DE ECUACIONES ESTRUCTURALES (I): COVARIANZAS Y DESVIACIONES SOBRE LA MEDIA

(#346). MODELO NO LINEAL EN LAS VARIABLES PERO LINEAL EN LOS PARÁMETROS CON STATA Y MAXIMA

(#337). EXPLORANDO LA PREDICCIÓN LINEAL EN MARKETING CON MAXIMA

(#336) EXPLORANDO EL DETERMINISMO Y EL AZAR CON C Y MAXIMA

(#335). EXPLORANDO EL COMPORTAMIENTO CAÓTICO CON MAXIMA

(#334). EXPLORANDO LA NO LINEALIDAD CON MAXIMA

(#331). LOS CONSTRUCTOS FORMATIVOS NO REFLEJAN ESTADOS PSICOLÓGICOS

(#314). REDUCIR EL SESGO EN ESTUDIOS SOBRE EFECTIVIDAD DE VACUNAS

(#313). EL SISTEMA PLANETARIO DE TRAPPIST-1

(#310). MÁS ALLÁ DEL REDUCCIONISMO EN BIOLOGÍA

(#307). LA CRISIS DE LA INVESTIGACIÓN EN PSICOLOGÍA

(#302). CONSTRUCTOS EN PSICOLOGÍA E IMPLICACIONES ESTADÍSTICAS

(#301). CENSURA EN EL LENGUAJE CIENTÍFICO

(#298). ¿CAMBIAR DE ENTRENADOR A MITAD DE TEMPORADA?

(#294). NEUROLOGÍA DE LA MALDAD: PISTAS SOBRE LAS RAÍCES DE LA MALDAD INDIVIDUAL

(#287). MEDIAS VERDADES Y OCULTACIÓN: ACEPTACIÓN DEL ENGAÑO DE CORPORACIONES

(#283). INTERPRETACIÓN FRECUENTISTA DE LA PROBABILIDAD: EL PAPEL DE LOS MODELOS

(#276). LA IRRACIONALIDAD, SCHRÖDINGER, Y EL EFECTO TÚNEL

(#265). DEFENDIENDO EL P-VALOR A TRAVÉS DE SU CORRECTA INTERPRETACIÓN

(#264). MÁS ALLÁ DE LOS ENSAYOS CONTROLADOS ALEATORIZADOS

(#263). CAUSALIDAD EN CASOS INDIVIDUALES: ALGUNAS RECOMENDACIONES PARA SU ESTUDIO

(#260). PROPONEN SER 10 VECES MÁS EXIGENTES CON EL P-VALOR

(#257). ERRORES DE CODIFICACIÓN DE EMBOLIAS PULMONARES AL USAR EL ICD-10

(#252). LA PREGABALINA NO SE ASOCIA A MALFORMACIONES EN EL EMBARAZO

(#243). LOS DEPORTISTAS DE ÉLITE VIVEN MÁS, PERO CON MATICES

(#232). LA OTRA HISTORIA DE NIKE

(#191). LA TRIADA OSCURA, LOS ANTICONCEPTIVOS Y LA PREFERENCIA SEXUAL

(#163). CRECEN LAS DUDAS SOBRE PLS (PARTIAL LEAST SQUARES)

(#159). ENTENDER LA NATURALEZA DE LOS DATOS Y DE LOS TEST ESTADÍSTICOS

(#128). INDICADORES FORMATIVOS EN ECUACIONES ESTRUCTURALES

(#127). AJUSTE EXACTO EN ECUACIONES ESTRUCTURALES

(#104). CORRUPCIÓN Y DESIGUALDAD ENTRE LOS MUY RICOS

(#100). DESCREÍDOS DE LA EPIDEMIOLOGÍA

(#90). REDES EN PSICOPATOLOGÍA Y EN MARKETING

(#76). EL 85-90% DE LOS NUEVOS MEDICAMENTOS SON INNECESARIOS

(#33). CLUSTERS DE CÁNCER PEDIÁTRICO

(#31). TRÁFICO Y CÁNCER DE PULMÓN

(#26). FOTOVOLTAICAS HIPERGEOMÉTRICAS

(#24). ROMPER EL CRISTAL

(#22). COMPLEJIDAD DE LOS FENÓMENOS SOCIALES

(#20). DECISIONES HEURÍSTICAS

(#17). RADICALES DE LAS VACUNAS

(#13). 24755 CACAS EN EL PARQUE

(#12). ESTADÍSTICAS SIN ESTADÍSTICA

(#10). EL DESCUBRIMIENTO DEL ARGÓN

(#6). ¿VIVIMOS MÁS Y CON MEJOR SALUD?

(#5). ¿ES LA ECONOMÍA UNA CIENCIA?

(#3). HEZONJA Y LAS PREDICCIONES ESTADÍSTICAS

(#1) ¿CÓMO JUGARÁ MARIO HEZONJA SU PRÓXIMO PARTIDO?

(#437). TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIId)

[MONOTEMA] En esta cuarta entrega del tercer capítulo de Probability Theory and Statistical Inference, de Aris Spanos, seguimos profundizando en la relación entre espacio de probabilidad y modelo de probabilidad.

Cuando las probabilidades son funciones conocidas de ciertos parámetros desconocidos θ, entonces podemos transformar el espacio probabilístico en un modelo de probabilidad definido por:

$$\Phi = \{f(x;\theta), \theta \in \Theta, x \in \mathbb{R}_X\}$$

donde Φ es una colección de funciones de densidad que dependen de un conjunto de parámetros θ en el espacio paramétrico Θ.

Podríamos usar también la función de distribución:

$$\Phi_F = \{F(x;\theta), \theta \in \Theta, x \in \mathbb{R}_X\}$$

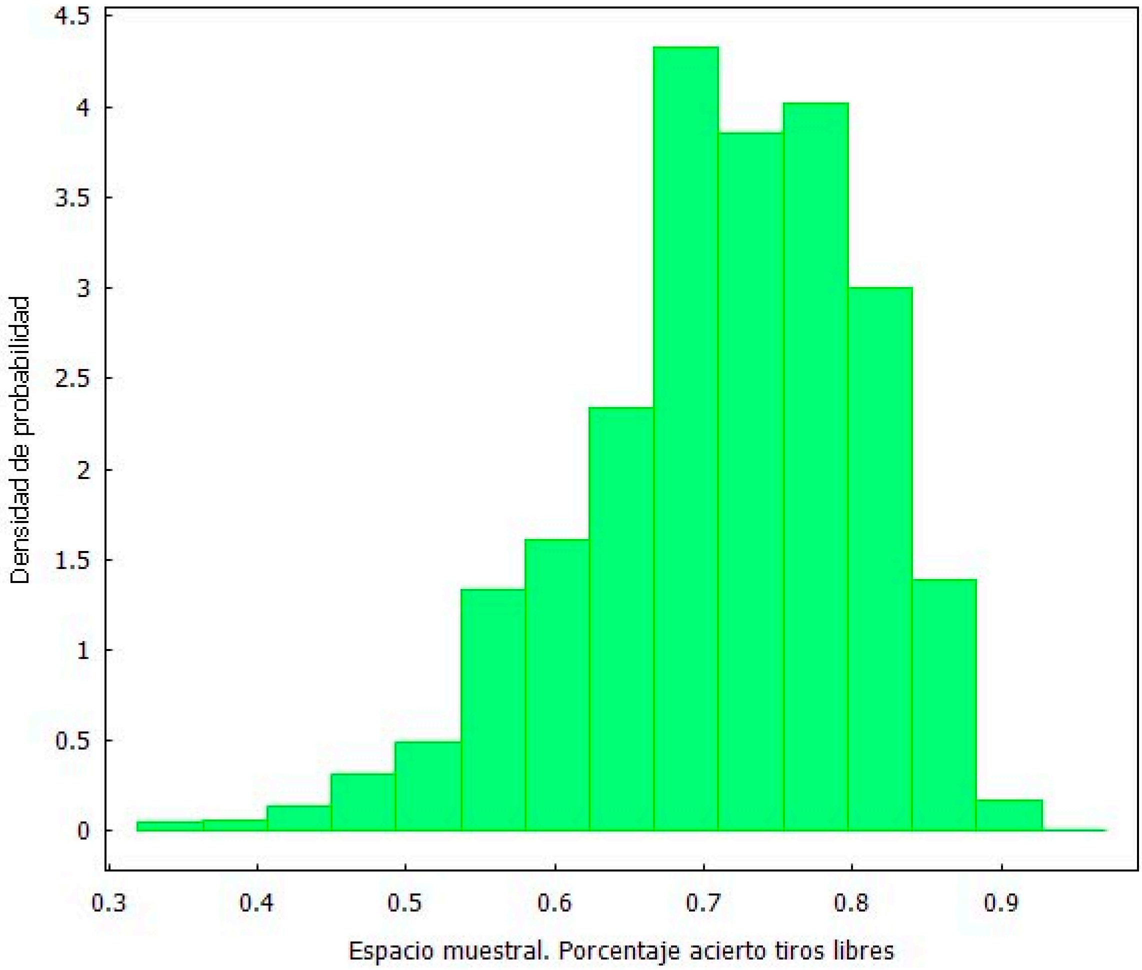
Pongamos un ejemplo usando la distribución Beta como modelo de probabilidad:

$$\Phi = f_x(x;\theta) = \left\{ \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B[\alpha,\beta]}, \theta := (\alpha,\beta) \in \mathbb{R}_+^2, 0 < x < 1 \right\}$$

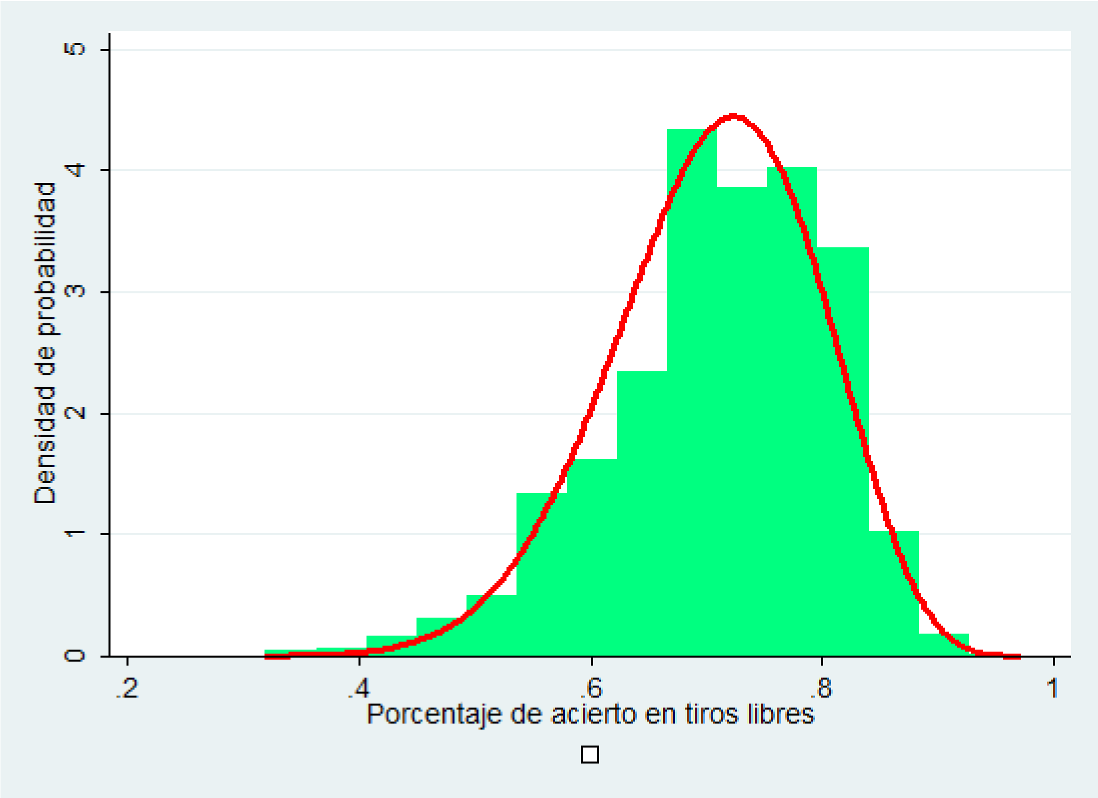
Podemos analizar el porcentaje de acierto en los tiros libres de los jugadores NBA hasta 2015 (el acumulado en sus respectivas carreras), para aquellos que hubieran lanzado al menos 30 tiros libres.

El histograma de la distribución es el siguiente:

```
data:read_list(file_search("RUTADELARCHIVO.txt"));
datatranspose:transpose(data);
estatura:datatranspose;
histogram (
estatura,
nclasses=15,
frequency=density,
xlabel="Espacio muestral. Porcentaje acierto tiros libres",
ylabel="Densidad de probabilidad",
fill_color=green,
fill_density=0.5);
```



Para ello nos ayudamos de nuevo de Stata 13.0, y estipulamos una distribución Beta de parámetros (18, 7.5).



Es decir, para la modelización empírica debemos postular a priori una familia de densidades que refleje el mecanismo estocástico que da origen a los datos observados. Para ello, tiene espacial relevancia el rango de valores de la variable aleatoria.

Estamos todavía al comienzo, pero ya hemos intuido cómo se plantea un modelo de probabilidad.

Todos los posts relacionados

(#449). AUMENTO DE CASOS DE GLIOBLASTOMA TAMBIÉN EN FRANCIA

(#448). INCREMENTO DE LA INCIDENCIA DE GLIOBLASTOMA

(#446) RENDIMIENTO EN BASE A VICTORIAS PRODUCIDAS POR LOS JUGADORES

(#445) LOS JUGADORES MÁS PRODUCTIVOS DE LA ACB 2019/20

(#443). LOS JUGADORES MÁS PRODUCTIVOS DE LA NBA 2019/20

(#438).TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIIe)

(#437). TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIId)

(#436). TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIIc)

(#435).TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIIb)

(#434).TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIIa)

(#428) TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIc)

(#427) TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIb)

(#426) TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIa)

(#424) TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (I)

(#422). JUGADORES ACB AL NIVEL DE NBA EN LA TEMPORADA 2018/19

(#421). LOS JUGADORES MÁS PRODUCTIVOS DE LA ACB 2018/19

(#420). LA PARADOJA DE CHEVALIER DE MERE Y LOS ESTUDIANTES DE MARKETING

(#412). LOS JUGADORES MÁS PRODUCTIVOS DE LA NBA 2018/19

(#411). LOS ROOKIES MÁS PRODUCTIVOS DE LA NBA 2018/19

(#410) ¿QUIÉN HA SIDO EL MEJOR ANOTADOR DE LA FASE REGULAR DE LA EUROLIGA 2018/19?

(#409). ALGUNAS CONSIDERACIONES EN LA ESTIMACIÓN DEL RENDIMIENTO POR MINUTO EN BALONCESTO (II)

(#408). ALGUNAS CONSIDERACIONES EN LA ESTIMACIÓN DEL RENDIMIENTO POR MINUTO EN BALONCESTO (I)

(#407). AVANCE DE LOS NOVATOS MÁS PRODUCTIVOS DE LA NBA 2018/19

(#398). AJUSTE DE FUNCIONES (II): CAPACIDAD EXPLICATIVA Y PARTICIÓN DE LA FUNCIÓN

(#397). AJUSTE DE FUNCIONES (I): SPLINES VS MÍNIMOS CUADRADOS

(#396). INTERPOLACIÓN (IV): MÉTODO DE SPLINE CÚBICO

(#395). INTERPOLACIÓN (III): MÉTODO DE DIFERENCIAS DIVIDIDAS

(#394). INTERPOLACIÓN (II): MÉTODO DE NEVILLE

(#393). INTERPOLACIÓN (I): POLINOMIO DE LAGRANGE

(#391). BÚSQUEDA DE SOLUCIONES (V): MÉTODO DE LA POSICIÓN FALSA

(#390). BÚSQUEDA DE SOLUCIONES (IV): MÉTODO DE LA SECANTE

(#389). BÚSQUEDA DE SOLUCIONES (III): METODO DE NEWTON

(#388). BÚSQUEDA DE SOLUCIONES (II): METODO DEL PUNTO FIJO

(#387). BÚSQUEDA DE SOLUCIONES (I): MÉTODO DE LA BISECCIÓN

(#384). MEJOR UN SÓLO ITEM QUE VARIOS PARA MEDIR ACTITUDES

(#381). TABLAS DE TUMORES TOTALES EN EL ESTUDIO DEL NTP SOBRE EFECTOS DE LA RADIACIÓN DE TELÉFONO MÓVIL

(#380). ÍNDICES APROXIMADOS FLEXIBLES EN ECUACIONES ESTRUCTURALES

(377). AVANZANDO EN LA COMPLEJIDAD: DOBLE DEPENDENCIA

(#376). CAMPOS DIRECCIONALES, COMENZANDO A REPRESENTAR LA COMPLEJIDAD

(#373). PREDICCIONES CON BUENAS APROXIMACIONES

(#366). MODELO DE HOMOGENEIDAD EN TABLAS DE 2X2

(#365). BÁSICOS DE ECUACIONES ESTRUCTURALES (V): MATRIZ DE DATOS BRUTOS

(#358). BÁSICOS DE ECUACIONES ESTRUCTURALES (IV): REGRESIÓN LINEAL SIMPLE CON FIABILIDADES DIVERSAS

(#354). BÁSICOS DE ECUACIONES ESTRUCTURALES (III): COVARIANZA Y CORRELACIÓN ENTRE VARIABLES LATENTES

(#353). BÁSICOS DE ECUACIONES ESTRUCTURALES (II): VARIABLES LATENTES Y FIABILIDAD

(#352). TESTOSTERONA Y CONSUMO CONSPICUO EN HOMBRES

(#350). BÁSICOS DE ECUACIONES ESTRUCTURALES (I): COVARIANZAS Y DESVIACIONES SOBRE LA MEDIA

(#346). MODELO NO LINEAL EN LAS VARIABLES PERO LINEAL EN LOS PARÁMETROS CON STATA Y MAXIMA

(#337). EXPLORANDO LA PREDICCIÓN LINEAL EN MARKETING CON MAXIMA

(#336) EXPLORANDO EL DETERMINISMO Y EL AZAR CON C Y MAXIMA

(#335). EXPLORANDO EL COMPORTAMIENTO CAÓTICO CON MAXIMA

(#334). EXPLORANDO LA NO LINEALIDAD CON MAXIMA

(#331). LOS CONSTRUCTOS FORMATIVOS NO REFLEJAN ESTADOS PSICOLÓGICOS

(#314). REDUCIR EL SESGO EN ESTUDIOS SOBRE EFECTIVIDAD DE VACUNAS

(#313). EL SISTEMA PLANETARIO DE TRAPPIST-1

(#310). MÁS ALLÁ DEL REDUCCIONISMO EN BIOLOGÍA

(#307). LA CRISIS DE LA INVESTIGACIÓN EN PSICOLOGÍA

(#302). CONSTRUCTOS EN PSICOLOGÍA E IMPLICACIONES ESTADÍSTICAS

(#301). CENSURA EN EL LENGUAJE CIENTÍFICO

(#298). ¿CAMBIAR DE ENTRENADOR A MITAD DE TEMPORADA?

(#294). NEUROLOGÍA DE LA MALDAD: PISTAS SOBRE LAS RAÍCES DE LA MALDAD INDIVIDUAL

(#287). MEDIAS VERDADES Y OCULTACIÓN: ACEPTACIÓN DEL ENGAÑO DE CORPORACIONES

(#283). INTERPRETACIÓN FRECUENTISTA DE LA PROBABILIDAD: EL PAPEL DE LOS MODELOS

(#276). LA IRRACIONALIDAD, SCHRÖDINGER, Y EL EFECTO TÚNEL

(#265). DEFENDIENDO EL P-VALOR A TRAVÉS DE SU CORRECTA INTERPRETACIÓN

(#264). MÁS ALLÁ DE LOS ENSAYOS CONTROLADOS ALEATORIZADOS

(#263). CAUSALIDAD EN CASOS INDIVIDUALES: ALGUNAS RECOMENDACIONES PARA SU ESTUDIO

(#260). PROPONEN SER 10 VECES MÁS EXIGENTES CON EL P-VALOR

(#257). ERRORES DE CODIFICACIÓN DE EMBOLIAS PULMONARES AL USAR EL ICD-10

(#252). LA PREGABALINA NO SE ASOCIA A MALFORMACIONES EN EL EMBARAZO

(#243). LOS DEPORTISTAS DE ÉLITE VIVEN MÁS, PERO CON MATICES

(#232). LA OTRA HISTORIA DE NIKE

(#191). LA TRIADA OSCURA, LOS ANTICONCEPTIVOS Y LA PREFERENCIA SEXUAL

(#163). CRECEN LAS DUDAS SOBRE PLS (PARTIAL LEAST SQUARES)

(#159). ENTENDER LA NATURALEZA DE LOS DATOS Y DE LOS TEST ESTADÍSTICOS

(#128). INDICADORES FORMATIVOS EN ECUACIONES ESTRUCTURALES

(#127). AJUSTE EXACTO EN ECUACIONES ESTRUCTURALES

(#104). CORRUPCIÓN Y DESIGUALDAD ENTRE LOS MUY RICOS

(#100). DESCREÍDOS DE LA EPIDEMIOLOGÍA

(#90). REDES EN PSICOPATOLOGÍA Y EN MARKETING

(#76). EL 85-90% DE LOS NUEVOS MEDICAMENTOS SON INNECESARIOS

(#33). CLUSTERS DE CÁNCER PEDIÁTRICO

(#31). TRÁFICO Y CÁNCER DE PULMÓN

(#26). FOTOVOLTAICAS HIPERGEOMÉTRICAS

(#24). ROMPER EL CRISTAL

(#22). COMPLEJIDAD DE LOS FENÓMENOS SOCIALES

(#20). DECISIONES HEURÍSTICAS

(#17). RADICALES DE LAS VACUNAS

(#13). 24755 CACAS EN EL PARQUE

(#12). ESTADÍSTICAS SIN ESTADÍSTICA

(#10). EL DESCUBRIMIENTO DEL ARGÓN

(#6). ¿VIVIMOS MÁS Y CON MEJOR SALUD?

(#5). ¿ES LA ECONOMÍA UNA CIENCIA?

(#3). HEZONJA Y LAS PREDICCIONES ESTADÍSTICAS

(#1) ¿CÓMO JUGARÁ MARIO HEZONJA SU PRÓXIMO PARTIDO?

(#436). TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIIc)

[MONOTEMA] Continuamos con el tercer apartado del tercer capítulo de Probability Theory and Statistical Inference, de Aris Spanos.

Si vemos $P_X(.)$ como sólo una función del punto final del intervalo $(-\infty, x]$, entonces podemos definir la distribución acumulada (cdf):

$$F_X(.) : \mathbb{R} \to [0,1], F_X(x) = \mathbb{P}\{s : X(s) \leqslant x\} = P_X((-\infty, x)])$$

Ahora sí hemos generado una función que relaciona el número real asignado a cada posible evento con su probabilidad de ocurrencia. Pero en este caso es la probabilidad acumulada.

Para el caso simple (discreto) tenemos la función de densidad:

$$f_x(.) : \mathbb{R} \to [0,1], f_x(x) = \mathbb{P}(X = x), x \in \mathbb{R}_X$$

Por tanto, los espacios probabilísticos pueden simplificarse en el caso de variables aleatorias discretas y continuas a los siguientes:

$$(S, \Im, \mathbb{P}(.)) \overset{X(.)}{\to} (\mathbb{R}, f_x(.))$$

$$(S, \Im, \mathbb{P}(.)) \overset{X(.)}{\to} (\mathbb{R}, F_X(.))$$

Spanos se plantea en este punto si se pueden definir funciones de densidad para variables continuas y funciones de distribución para variables discretas, y la respuesta es que sí.

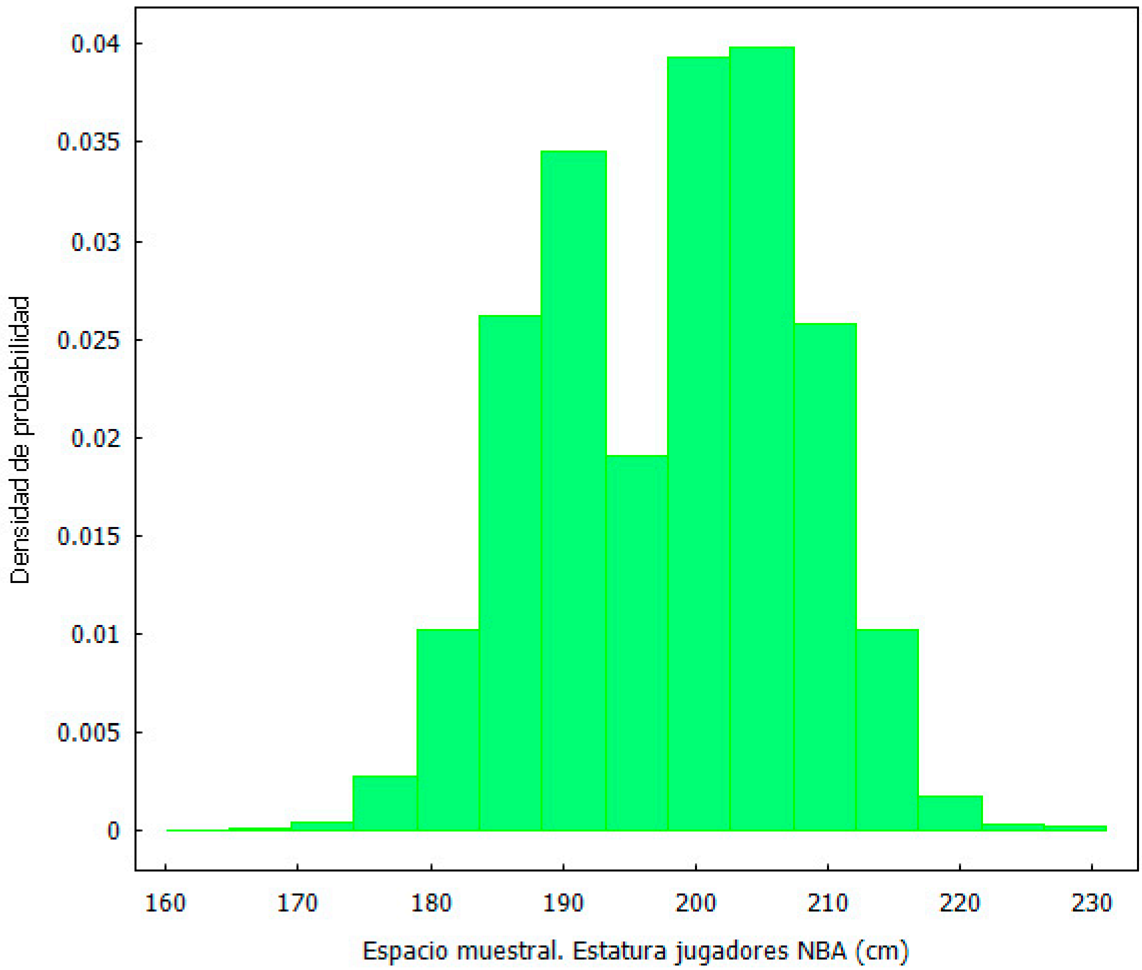
$$F_X(x) = \int_{-\infty}^{x} f_x(u)du, f_x(u) \geq 0$$

$$\int_{a}^{b} f_x(u)du = F_x(b) - F_x(a)$$

La estatura de los jugadores de la NBA se puede considerar como una variable continua. Desde el inicio de la NBA hasta el año 2015, hay 3984 jugadores cuya estatura se muestra en este archivo.

El histograma de la distribución es el siguiente:

```
data:read_list(file_search("RUTADELARCHIVO.txt "));
datatranspose:transpose(data);
estatura:datatranspose;
histogram (
estatura,
nclasses=15,
frequency=density,
xlabel="Espacio muestral. Estatura jugadores NBA (cm)",
ylabel="Densidad de probabilidad",
fill_color=green,
fill_density=0.5);
```



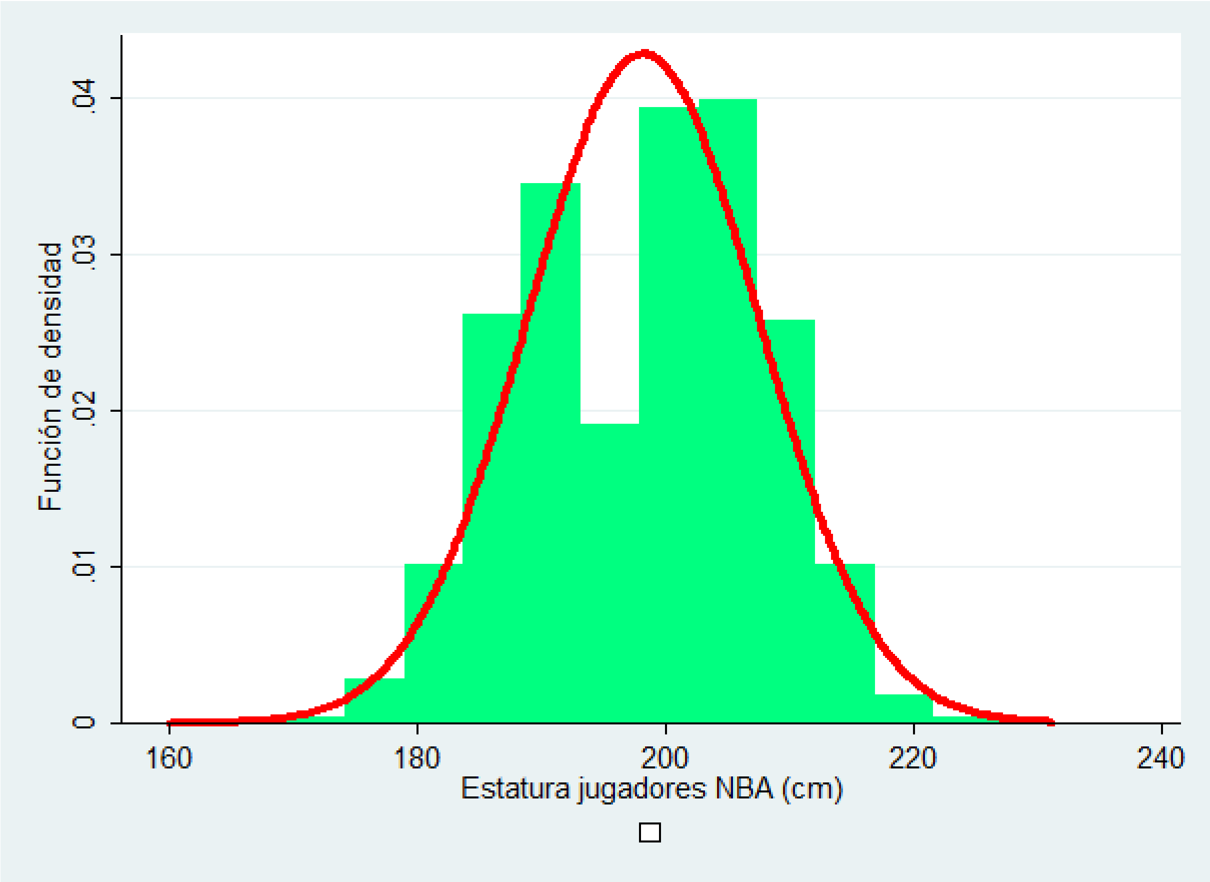
Es una distribución que se aproxima a una Normal, pero que no sabemos realmente si lo es. Recordemos que una distribución Normal tiene como función de densidad:

$$f_x(x;\theta) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \theta := (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+, x \in \mathbb{R}$$

De este modo, podemos tomar como media y desviación típica la de la muestra, como una estimación de los parámetros poblacionales.

$$f_x(x;\theta) = \frac{1}{9.3\sqrt{2\pi}}e^{-\frac{(x-198.2)^2}{2\cdot 9.3^2}}, \theta := (198.2, 9.3^2) \in \mathbb{R} \times \mathbb{R}_+, x \in \mathbb{R}$$

El resultado, tras emplear Stata 13.0, es el mostrado en el gráfico siguiente:



Sin embargo, otras distribuciones también podrían ajustarse a los datos. Por ejemplo, la distribución Weibull: