Resumen de teorico para el final de Probabilidad y Estadística

Agustin Curto, agucurto95@gmail.com

2015

Índice general

Capítulo 1

Estadstica Descriptiva

La estadística está dividida en dos grandes ramas:

- **Estadistica Descriptiva:** Organizar y resumir la información.
- **Estadistica Inferencial:** Extraer conclusiones acerca de hipótesis planteadas.

Población: Colección de sujetos o elementos de interés, puede ser finita o infinita.

Muestra: Subconjunto de la población elegido al azar, de tamaño muestral n.

Tipos de datos de la muestra:

- **Numéricos:** Discretos (determinados valores), Continuos (valores en un intervalo).
- **Categóricos:** Ordinal (orden), Nominal (no orden).

Intervalos de Clase: Datos $\{x_1 \leq x_2 \dots x_n\} \subset [a, b]$, con $a < b$.

$$\underbrace{[\dots]}_{I_1}\underbrace{(\dots]}_{I_2}\underbrace{(\dots]}_{I_3}\underbrace{(\dots]}_{I_4}$$

Tabla de Distribución de Frecuencia

Iintervalos de Clase	Frecuencia Absoluta	Frecuencia Relativa
I_1	3	3/8
I_2	+ 2	+ 2/8
I_3	+ 1	+ 1/8
I_4	+ 2	+ 2/8
Total	8	1

Referencias:

- MC: Marca de clase. Es el punto medio del intervalo.
- FAA: Frecuencia absoluta acumulada.
- FRA: Frecuencia relativa acumulada.

Histogramas: Es el gráfico de mayor difusión y la representación gráfica de la distribución de frecuencia.

Como hacerlo:

- En una recta horizontal marcar los k intervalos. ($k = \sqrt{n}$).
- En cada intervalo trazar un rectángulo cuya área sea proposicional al número de observaciones en el mismo.
- Altura de los triángulos: $h_i = \frac{FR_i}{long(IC_i)}$
- Suma de las áreas de los rectángulos igual a 1.

Media muestral: Es el promedio de los $x_1, x_2 \dots x_n$ puntos, denotada $\overline{x}$:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Mediana muestral: Es el valor medio de las observaciones:

$$\widetilde{x} = \begin{cases} n \text{ impar} \rightarrow \frac{n+1}{2} \\ n \text{ par} \rightarrow \text{promedio}\left(\frac{n}{2}, \frac{n}{2}+1\right) \end{cases}$$

Primer Cuartil (Q_1): Es la mediana de la primera mitad de las observaciones.

Tercer Cuartil (Q_3): Es la mediana de la segunda mitad de las observaciones.

Media Recortada: Es un termino medio entre $\overline{x}$ y $\widetilde{x}$. Una media recortada 10 % sería quitar el 10 % más pequeño y el 10 % más grande de la muestra para luego calcular la media.

Rango: Direfencia entre la máxima y la mínima ($x_n - x_1$).

Rango Intercuartil (RIC): Direfencia entre el tercer y primer cuartil ($Q_3 - Q_1$).

Varianza Muestral:

$$\sigma^2 = S^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}$$

Desviación Estándar Muestral:

$$\sigma = S = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}}$$

Fórmula para calcular S^2:

$$S^2 = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

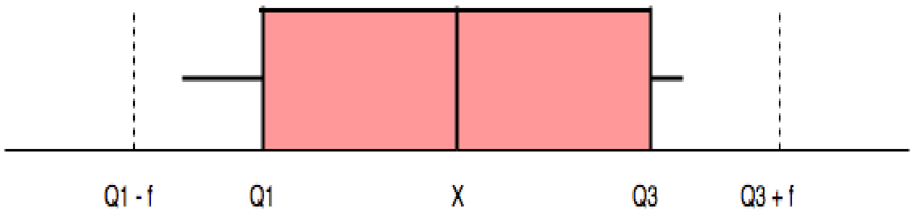
Proposición: Sea $x_1, x_2 \dots x_n$ una muestra y c cualquier constante distinta de cero, entonces:

- Si $y_1 = x_1 + c$, $y_2 = x_2 + c \dots\ y_n = x_n + c \Rightarrow S_y^2 = S_x^2$
- Si $y_1 = c\, x_1$, $y_2 = c\, x_2 \dots\ y_n = c\, x_n \Rightarrow S_y = |c|\ S_x$

Coeficiente de Variación: $\frac{S}{\overline{x}}\, 100\,\%$

Diagramas de caja (Box Plot):

1. Ordenar los datos de menor a mayor.
2. Calcular $\widetilde{x}, Q_1, Q_3$ y RIC.
3. Sobre un eje vertical u horizontal marcar los valores extremos (x_1, x_n) y los cuartiles (Q_1 y Q_3).
4. Sobre este eje dibujar una caja, cuyo borde izquierdo sea Q_1 y el derecho Q_3.
5. Dentro de la caja trazar una línea sobre la mediana.
6. Trazar segmentos desde cada extremo de la caja hasta las observaciones más alejadas que no superen 1,5 RIC de los brodes correspondientes.
7. Marcar con circunferencias aquellos puntos comprendidos entre 1,5 RIC y 3 RIC respecto del borde más cercano, esos son los puntos llamados *anómalos suaves* y con asteríscos aquellos puntos que superen 3 RIC, estos puntos son llamados puntos *anómalos extremos*.



Capítulo 2

Probabilidad

2.1. Modelo Probabilístico

Espacio Muestral: Es el conjunto de todos los resultados posibles del experimento. Se lo denota S.

Evento: Es cualquier subconjunto de S. Si el evento tiene un solo elemento se llama evento simple, si no, es un evento compuesto.

Definición: Cuando A y B (eventos) no tienen resultados en común, se dice que son eventos mutuamente excluyentes o disjuntos. Además $P(A \cap B) = \emptyset$.

A se dice **familia de eventos** si:

- $S \in A$
- Si $a \in A \Rightarrow \overline{a} \in A$
- $\{a_i\}_{i=1}^{\infty}$ tal que $a_i \in A \Rightarrow \cup_{i=1}^{\infty} a_i \in A$

Medida de Probabilidad: Diremos que $P : A \to [0, 1]$, con A evento, es medida de probabilidad si:

- $0 \leq P(a) \leq 1 \;\; \forall a \in A$
- $P(S) = 1$
- $\{a_i\}_{i=1}^{\infty}$ con $a_i \in A \; \forall i$ y mutuamente disjuntos.

Modelo Probabilístico: Es una terna compuesta (S, A, P), espacio muestral, familia de eventos y medida de probabilidad respectivamente.

Propiedades: Dado un experimento, tenemos (S, A, P) entonces se puede probar que:

- Si $A \subset B \Rightarrow P(B - A) = P(B) - P(A)$ y $P(B) \geq P(A)$
- $P(\overline{A}) = 1 - P(A)$ $\qquad P(\emptyset) = 0$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

2.2. Técnicas de Conteo

Reglas del producto: Suponga que un conjunto consiste en colecciones ordenadas de k elementos y que hay n_1 opciones posibles para el primer elemento; para cada elección del primer elemento, hay n_2 elecciones posibles del segundo elemento; ... para elección posible de los $k-1$ elementos, hay n_k elecciones del *k-esimo* elemento. Entonces hay $n_1 n_2 \dots n_k$ *k-tuplas* posibles.

Definición: Para cualquier secuencia ordenada de k objetos tomada de un conjunto de n objetos, el número de *permutaciones* de tamaño k que se pueden construir a partir de n objetos, se denota $P_{k,n}$ y se define:

$$P_{k,n} = \frac{n!}{(n-k)!}$$

Definición: Dado un conjunto de n elementos distintos, cualquier subconjuto no ordenado de tamaño k de los objetos, se llama *combinación*. El número de combinaciones de tamaño k que se puede formar a partir de n objetos, se denota $\binom{n}{k}$, y se define:

$$C_{n,k} = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

2.3. Probabilidad Condicional

Sean A y B eventos tal que $P(B) > 0$, llamamos **probabilidad condicional** de A dado B, y denotamos $P(A|B)$, porbabilidad de A dado que ocurrio B, al evento:

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

Definición: Diremos que A y B son *eventos independientes*, si:

$$P(A \cap B) = P(A)\ P(B) \qquad \text{con A, B} \in \mathbf{a}$$

Proposición: Sean A y B eventos:

- $\overline{A}$ y $\overline{B}$ son independientes.
- $\overline{A}$ y B son independientes.
- A y $\overline{B}$ son independientes.

Definición: Diremos que $A_1, A_2 \dots A_n$ son *mutuamente independientes* si $\forall I \in \{1, 2, \dots n\}$ resulta que:

$$P(\cap_{i,j\in I} A_{i,j}) = \prod_{i,j\in I} P(A_{i,j})$$

Ley de la multiplicación: Sean $\{A_i\}_{i=1}^{\infty}$ eventos en $\mathbf{a}$, entonces:

$$P(\cap_{i=1}^{n} A_i) = P(A_1)\ P(A_2|A_1) P(A_3|A_1 \cap A_2) \dots P(A_n| \cap_{i=1}^{n-1} A_i)$$

Ley de Probabilidad Total: Si $\{A_i\}_{i=1}^{\infty}$ son eventos disjuntos en $\mathbf{a}$ tal que $S = \cup_{i=1}^{n} A_i$ entonces $\forall B \in \mathbf{a}$:

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

Teorema de Bayes: Si $\{A_i\}_{i=1}^{\infty}$ son eventos disjuntos en **a** y $P(A_i) > 0 \; \forall i$ tal que $S = \cup_{i=1}^{n} A_i$, entonces para cualquier otro evento B tal que $P(B) > 0$:

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

donde $P(B) = \sum_{i=1}^{n} P(A_i)\; P(B|A_i)$ como se dijo en la ley de probabilidad total.

Capítulo 3

Variables Aleatorias Discretas

Capítulo 4

Variables Aleatorias Continuas

Capítulo 5

Distribución de Probabilidad Conjunta

Capítulo 6

Estimación Puntual

Capítulo 7

Intervalos de Confianza

Capítulo 8

Prueba de Hipótesis

Bibliografía

[1] Agustín Curto, «Carpeta de Clase, 2015», *FaMAF, UNC.*