

---

# **Resumen Estadística Documentation**

***Publicación 0***

**Roberto Rojas R.**

29 de November de 2015

|   |          |
|---|----------|
| <b>1. Introducción</b>                        | <b>1</b> |
| <b>2. Introducción a la estadística</b>       | <b>2</b> |
| 2.1. Qué es la estadística . . . . .          | 2        |
| 2.2. Método científico . . . . .              | 2        |
| 2.3. Método estadístico . . . . .             | 2        |
| 2.4. Elementos básicos . . . . .              | 3        |
| <b>3. Estadística Descriptiva</b>             | <b>4</b> |
| 3.1. Conceptos básicos . . . . .              | 4        |
| 3.2. Probabilidades . . . . .                 | 6        |
| 3.3. Variables aleatorias DISCRETAS . . . . . | 6        |
| 3.4. Variables aleatorias CONTINUAS . . . . . | 6        |
| <b>4. Estadística inferencial</b>             | <b>7</b> |
| 4.1. Estadística paramétrica . . . . .        | 7        |
| 4.2. Estadística NO paramétrica . . . . .     | 7        |
| <b>Índice</b>                                 | <b>8</b> |

---

## Introducción

---

La idea de este documento es resumir los apuntes y aglomerar información que puede estar dispersa, abarca más o menos lo encontrado en los apuntes combinando información de diapositivas y documentos.

**Advertencia:** El presente documento no pretende ser un texto de referencia ni ser infalible, es posible que se me pase algún errorcillo, por eso se recomienda usar de guía para complementar los apuntes y bibliografía del ramo, además este ramo lo hice hace como 10 años, la memoria es frágil y mucho ha pasado por esta azotea.

La estadística se compone de:

**Estadística Descriptiva** Recolectar, ordenar, resumir y representar. Solo muestra lo que hay, solo que más bonito, legible y conciso. En esencia tomar una tabla detallada y monstruosa de Excel y convertirla en unas lindas tablas resumidas y gráficos para el artículo, paper o presentación de PowerPoint.

**Estadística Inferencial** El inferir algo es proyectar lo general basándose en algo local, esta rama de la estadística se encarga de hacerlo bien, no basta con decir que porque soy el más alto de mi familia soy el más alto de Chile. Es labor de la estadística inferencial, por ejemplo, saber si el promedio obtenido con estadística descriptiva es representativo o no, si una muestra es más dispersa que otra, etc.

**Paramétrica** Los datos pueden distribuirse regularmente (generalmente lo están), estas distribuciones pueden definirse matemáticamente y como son regulares basta con conocer a qué distribución se asemejan y con unos pocos (en términos relativos) parámetros se puede describir el conjunto de datos e incluso proyectar y predecir futuras incidencias.

**NO paramétrica** Cuando no conocemos a qué distribución matemática se ajustan los datos, no se ajusta bien a ninguna (raro) o ni se parece a alguna distribución (muuuuy raro) se utiliza la estadística NO paramétrica (no conocemos los parámetros). Estas casi nunca se hacen a mano, así que pese a que suelen ser más complejas (no siempre y no todas) suelen hacerse por un computador. Solo es necesario saber cuando aplicar cada una y como interpretar los resultados. El programa que se suele usar para estos menesteres es el SPSS y su versión gratuita y OpenSource PSPP.

Esto está en los apuntes, pero no me pareció muy claro, estas clasificaciones se inducen pero no se explicitan (puede que lo haya hecho en clases)

---

## Introducción a la estadística

---

### 2.1 Qué es la estadística

### 2.2 Método científico

- 1. Detección y enunciado** Definición clara y específica, que quede claro de que se está hablando y no quede lugar para ambigüedades.
- 2. Formulación de hipótesis** Una posible explicación, si tiene algún fundamento es menos probable perder el tiempo comprobando tonteras.
- 3. Consecuencia verificable** Definir el como se va a comprobar la hipótesis, definiendo variables comprobables y que eniquívocamente confirmen o descarten la hipótesis. si estas no están bien definidas voy a hacer mucho trabajo en los pasos siguientes sin haber demostrado o desmentido nada.
- 4. Verificación** Recolectar datos, procesarlos y analizarlos para, basado en el punto anterior, confirmar o refutar la hipótesis.
- 5. Conclusión** Presentar resultados, responder si se cumple o no lo definido en el primer punto.

### 2.3 Método estadístico

#### A. Planificación

- 1. Definición de objetivos** Definición clara del objetivo, detalle del qué, cómo, dónde, cuándo y por qué, sin lugar a interpretaciones ni mal entendidos.
- 2. Universo de estudio** De donde se va a sacar la información.
- 3. Diseño de la muestra** Como seguramente es imposible, al menos en la práctica, obtener TODA la información existente, debemos tomar parte de ella, cuidando que represente lo mejor posible el universo definido en el punto anterior.
- 4. Definición de las unidades** Definir características del qué vamos a observar, unidades de medida, escalas de clasificación.
- 5. Preparación del plan** Establecer el plan de tabulación y análisis, el cómo se presentará la información y como se hará el análisis.

#### B. Ejecución

1. **Recolección** La parte *entretenida*, pararse en la calle a encuestar, pararse en un recinto a contar gente o cronometrar cosas una y otra y otra vez.
2. **Elaboración** Tomar los benditos datos del punto anterior, los datos que tomaron horas y horas capturar y convertirlo en gráficos y tablas que hagan a los demás fácil el visualizar y entender la información. Aquí es donde pasan de datos a información.
3. **Análisis** Teniendo la información del punto anterior ahora se procesa y analiza para, con ellos, confirmar o descartar las aseveraciones del punto 1.

## 2.4 Elementos básicos

**Unidad de análisis\*\*** Teniendo listo el punto 1 de la planificación se define qué es lo que sera observado.

**Atributos** Definir las características relevantes de las unidades antes establecidas, si vamos a medir inteligencia en personas de cierto lugar, salvo que sea una hipótesis muy *exótica*, el tamaño de las manos seguramente no es relevante, aquí se establecen cuales características si lo son.

**Variables** El qué y como voy a expresar los mencionados atributos, que posibles valores van a tener, las unidades de medida, etc.

**Escalas de medida** En esencia el tipo de variable, entre las que se distinguen 3 tipos:

**Nominal** Es solo una etiqueta no relacionable con otros valores solo sirve para diferenciar estados de la variable.

**Ordinal** A diferencia de las Nominales, las variables Ordinales son ordenables, conllevan una relación con respecto a otras variables cada valor lo hace mas o menos que otras.

**Intervalar** Además de ser ordenables pueden operarse matemáticamente, tienen una unidad de medida y todos los elementos la comparten, estas pueden ser:

**Discretas** Números enteros, cantidad de huevos, cantidad de personas, etc. Seguramente podemos considerar fracciones de personas pero seria poco ético andar partiendo personas.

**Continuas** Consideran todos los posibles valores dentro de un intervalo, formalmente números reales y complejos pero salvo que midamos el spin de partículas subatómicas o las funciones de probabilidad del orbital de un electrón nos basta con saber que son números decimales, entendiendo que entre el 0 y el 1 hay infinitos valores (los intervalos siempre pueden partirse en 2).

**Población** El conjunto de todos los elementos que nos interesan, el conjunto universo.

**Muestra** Por razones prácticas no siempre puede usarse la información de cada elemento de la población, por ejemplo para hacer un control de calidad de fósforos o cuando queremos saber cuanto miden los Chilenos; en el primer caso la única forma de saber si un fósforo esta bueno es póstumo, lo que hace ridículo probarlos todos y en el segundo, seria muy caro y difícil medir a cada chileno. Lo importante es tratar de que la muestra represente a la población lo mejor posible. Si pretendo analizar la altura de los chilenos entre los jinetes en el hipódromo, seguramente no voy a tener una muestra representativa.

---

## Estadística Descriptiva

---

Es la que consolida los datos, ordena, resume y representa. O sea describe entrega ciertas características de la muestra sin decir mucho de que tan buen trabajo está haciendo con respecto a la población, no hace predicciones ni proyecciones, solo da información sobre los datos de la muestra.

### 3.1 Conceptos básicos

#### 3.1.1 Graficar y tabular

Graficar y tabular como todo tiene su ciencia y su arte, puede ser la diferencia entre auditores interesados que aprenden y auditores que no entienden nada y se duermen. Además si no se hace con cuidado puede distorsionar la información percibida. Lo dicen 8 de cada 10 dentistas.

Hay muchas formas de tabular y graficar datos, cada una tiene sus usos, gracias y desgracias, lo primero es identificar que tipo de variable (*escalas de medida*) vamos a representar:

#### Nominales (no ordenables)

##### Tablas

Basta con cuantas veces se repite cada valor y eso seria, también se puede agregar el % de cada uno.

##### Gráficos

Son preferibles los circulares, se pueden usar tambien de barras pero las variables al no tener orden no se gana nada con ponelas en ejes y se pierde la visualización de la proporción con el total. Con el circular el orden no importa (es un circulo) y no solo se representan las relaciones entre las variables, también se representan sus relaciones con respecto al total de yapa.

#### Ordinales (ordenables)

##### Tablas

Aquí como se pueden y deben ordenar hace sentido agregar la columna de acumulado, en esta se van sumando las frecuencias de las variables menores (recordar que siempre se ordena de menor a mayor), así se tiene la gracia de

poder ver rápidamente por ejemplo todos los que miden menos de 1,75m sin necesidad de tomar una calculadora y sumar todas las frecuencias.

### Gráficos

Aquí se pueden usar de barra o circulares, pero suelen ser mejores los de barra por que:

- En el circular se pierde el orden (que con las nominales no teníamos pero ahora si), podriamos acordar que partimos arriba y seguimos en sentido horario pero es trabajo adicional innecesario y nos carga ese tipo de trabajo.
- Si tenemos muchos posibles valores para las variables o simplificamos los datos y agrupamos por rango, perdiendo información o nos va a quedar un gráfico circular ilegible, diferenciar el ancho de una tajada de un 1 % y una de un 3 % es bastante difícil y hasta los colores y su posición relativa a los vecinos alterna la percepción del área de dicha tajada. Con un grafico de barras basta con ensanchar el gráfico y adelgazar las barras (acá es el alto el que nos da la información no el ancho).

### Intervalares (ordenable Y operables)

#### Tablas

Aquí se pueden usar tablas parecidas a las Ordinales pero mejor usamos lo último de la tecnología en tablas con el diagrama de tallo y hoja inventado en 1977 por Turkey. Con esta maravilla de la tecnología se separan los valores por orden y se enlistan las unidades o sea para:

{ 12, 13, 13, 14, 15, 23, 23, 23, 45, 46, 46, 47 }

queda

1 | 2 3 3 4 5 2 | 3 3 3 4 | 5 6 6 7

y con:

{ 123, 123, 124, 134, 135, 135, 136, 137, 141, 141, 142 }

queda

1 | 2 | 3 3 4

3 | 4 5 5 6 7

4 | 1 1 2

Notese que para hacer la vida más placentera ordenamos los datos antes de hacer cualquier cosa.

La gracia de este tabulado es que no se pierde información, es fácil de interpretar y si torcemos nuestras cabezas 90° en el sentido del reloj y voilá tenemos un grafico de frecuencia agrupado por decenas.

### Gráficos

Se pueden mostrar:

**Frecuencias:** cada barra muestra la frecuencia de dado valor en el valor de altura de 1.70m va la frecuencia de las personas que miden 1.70.

**Frecuencias acumuladas:** cada barra indica su frecuencia y las frecuencias de los valores más chicos por eso acumulada así en la altura de 1.70m van todas las personas que miden menos de 1.70m

Y para eso se usan los gráficos:

**Gráfico de barras:** se usan para variables discretas y se dibujan con sus barras separadas entre si.

**Histogramas:** se usan para variables continuas y se dibujan con las barras pegadas a sus vecinos.

**Gráfico de línea:** en vez de hacer rectángulos se tira una línea en el valor dado, cuando es de frecuencias acumuladas se llama OJIVA, ni idea de porque se llama así. No se suelen usar si son de frecuencias, se usa el de los anteriores que corresponda.

---

**Nota:** Si al hacer la tabla se agrupan los datos en distintos rangos ej:

| edad   | freq | largo | freq* |
|--------|------|-------|-------|
| 0 - 3  | 30   | 3     | 10    |
| 3 - 5  | 50   | 2     | 25    |
| 5 -10  | 40   | 5     | 8     |
| 10 -30 | 60   | 20    | 3     |

Vemos que en el primer rango  $3 - 0 = 3$ , y el tercero  $5 - 10 = 5$ , lo que se refleja en la tercera columna, para ajustar las frecuencias y hacer que los gráficos que salgan de aquí se vean representativos se divide la frecuencia por el largo y ESO se grafica.

---

### 3.1.2 Estadígrafos y estadísticos

## 3.2 Probabilidades

### 3.3 Variables aleatorias DISCRETAS

### 3.4 Variables aleatorias CONTINUAS



---

## Estadística inferencial

---

### 4.1 Estadística paramétrica

### 4.2 Estadística NO paramétrica

distribución, función, integral, sumatoria, exponencial (e), factorial (!)

## D

distribución, [7](#)

## E

exponencial (e), [7](#)

## F

factorial (

), [7](#)

función, [7](#)

## I

integral, [7](#)

## S

sumatoria, [7](#)