

Estadística I

Sección 2: Capítulo I

Felipe López S

Presentación

- Profesor: Felipe López S.
 - lopezsabalaga@ gmail.com
 - LinkedIn
- Ayudante:
- Clases martes y miércoles 17:30 - 18:50

¿Por qué estudiar Estadística?

La Estadística es una ciencia que se encarga de estudiar la incertidumbre basada en las leyes de la probabilidad.

Viene del término alemán *statistik*, la cual hace referencia a las ciencias del Estado. Surge debido al origen de los Estados modernos en el siglo 16, que debían coleccionar datos de los habitantes y la economía.

Una de las herramientas de recolección de datos es el *Censo* de población.

Avances históricos

En el siglo 19 la teoría de la probabilidad fue desarrollada por matemáticos, entre ellos, Jacob Bernoulli, Karl Friederich Gauss y Pierre Simon Laplace.

Al final de este siglo empezaron los primeros avances en inferencia estadística.

El objetivo es obtener conclusiones, mediante métodos matemáticos, a partir de una **muestra** de datos.

En el siglo 20 fueron múltiples las áreas donde se desarrolló y aplicó el conocimiento estadístico: Agricultura, medicina, marketing, encuestas, etc...



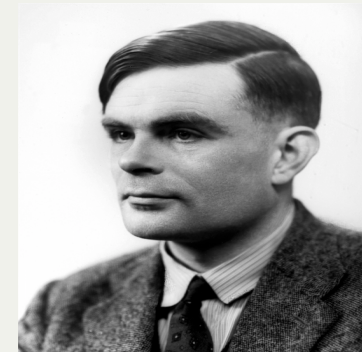
Karl Friederich Gauss

En la actualidad

En la actualidad los campos aplicados de la ciencia estadística son cada vez más relevantes.

Uno de ellos es la Ciencia de Datos, la cual comprende un paradigma en la utilización de datos para la toma de decisiones en el ámbito privado o público.

Adicionalmente, la Inteligencia Artificial (IA) se ha establecido como un campo interdisciplinario fundamentado en el uso de datos masivos. Cuyo usos van desde la generación de imágenes y texto hasta la detección de enfermedades.



Alan Turing, pionero de las ciencias de la computación

Datos Tabulares

Existen muchas formas de almacenar y presentar datos. La manera más utilizada es la presentación en forma tabular, esta consiste en una presentación bidimensional. Generalmente, las filas representan unidades observadas y las columnas sus características. Por ejemplo:

id_diamante	precio	quilate	corte
1	326	0.23	Ideal
2	326	0.21	Premium
3	327	0.23	Bueno
4	334	0.29	Premium
5	335	0.31	Bueno
6	336	0.24	Muy bueno

Descripción Gráfica de los Datos

“The simple graph has brought more information to the data analyst’s mind than any other device.”— John Tukey

Los datos y resultados de un experimento deben ser presentados de manera clara, concisa y debe facilitar el trabajo de quién los observa.

A lo largo del tiempo se han desarrollado técnicas y consejos para visualizar datos. Usted puede visitar los siguientes recursos adicionales:

Fundamentals of Data Visualization, Claus Wilke

Capítulo 03, R para Ciencia de Datos (Wickham & Grolemund, 2017)

Datos Cualitativos

Los datos cualitativos son medidas de los “tipos” o características de un fenómeno observado, pueden representar un nombre, símbolo o un código numérico. Ej: sexo de la persona, color del cabello, mes de nacimiento.

En estadística son referidas como variables categóricas, porque generan categorías.

Tabla de Frecuencias

Un conjunto de datos categóricos o cualitativos que tenga un rango de valores reducido puede ser representado en una tabla de frecuencias. Indica cuantas veces aparece un dato dentro del conjunto de observaciones.

Corte del Diamante	Frecuencia
Regular	1.610
Bueno	4.906
Muy bueno	12.082
Premium	13.791
Ideal	21.551
Total	53.940

Gráfico de Frecuencias

Esta tabla se puede representar en un plano cartesiano, indicando en el eje Y la frecuencia y en el eje X el rango de los datos.

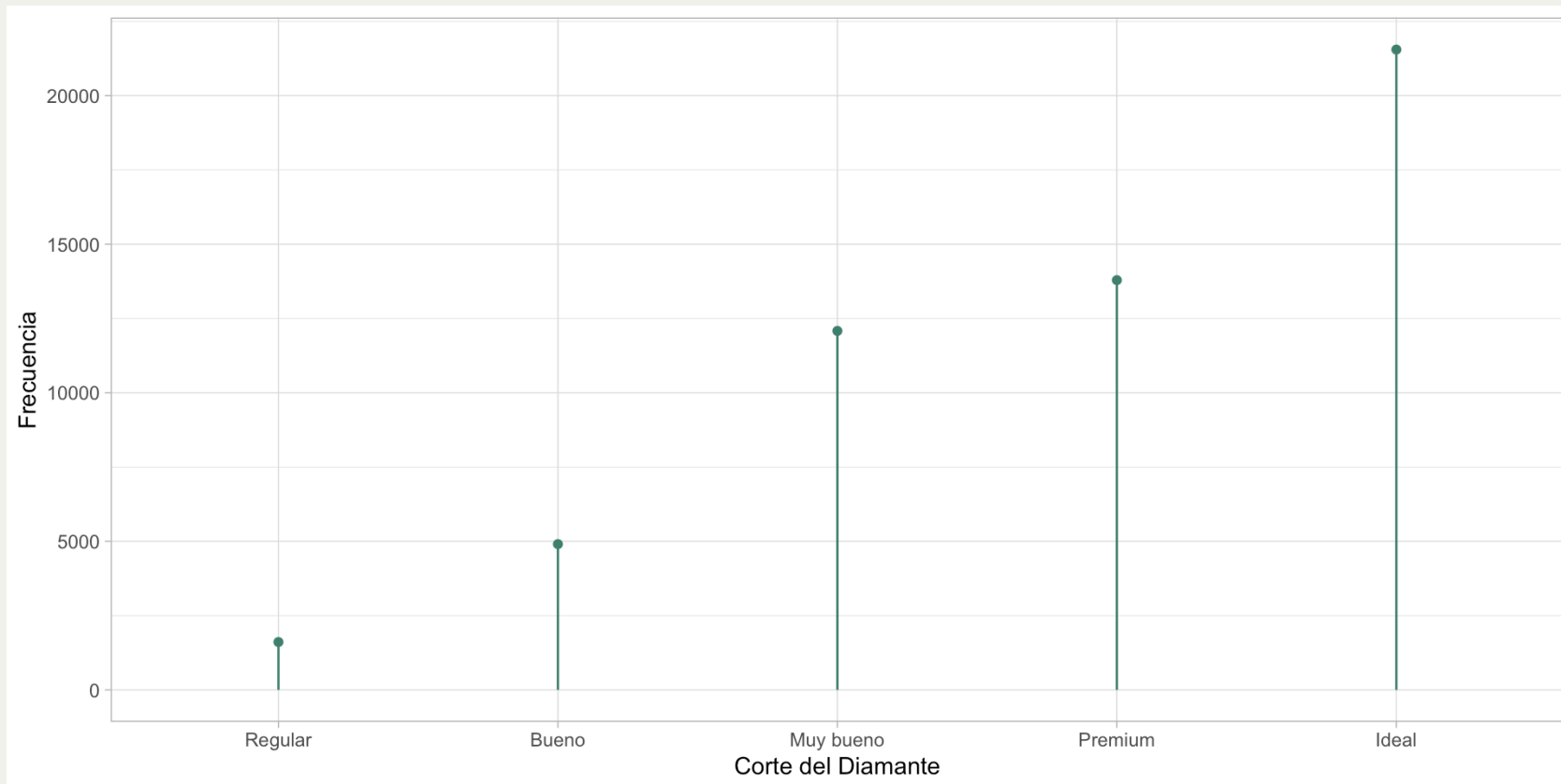


Gráfico de Frecuencias

También podemos utilizar un gráfico de barras:

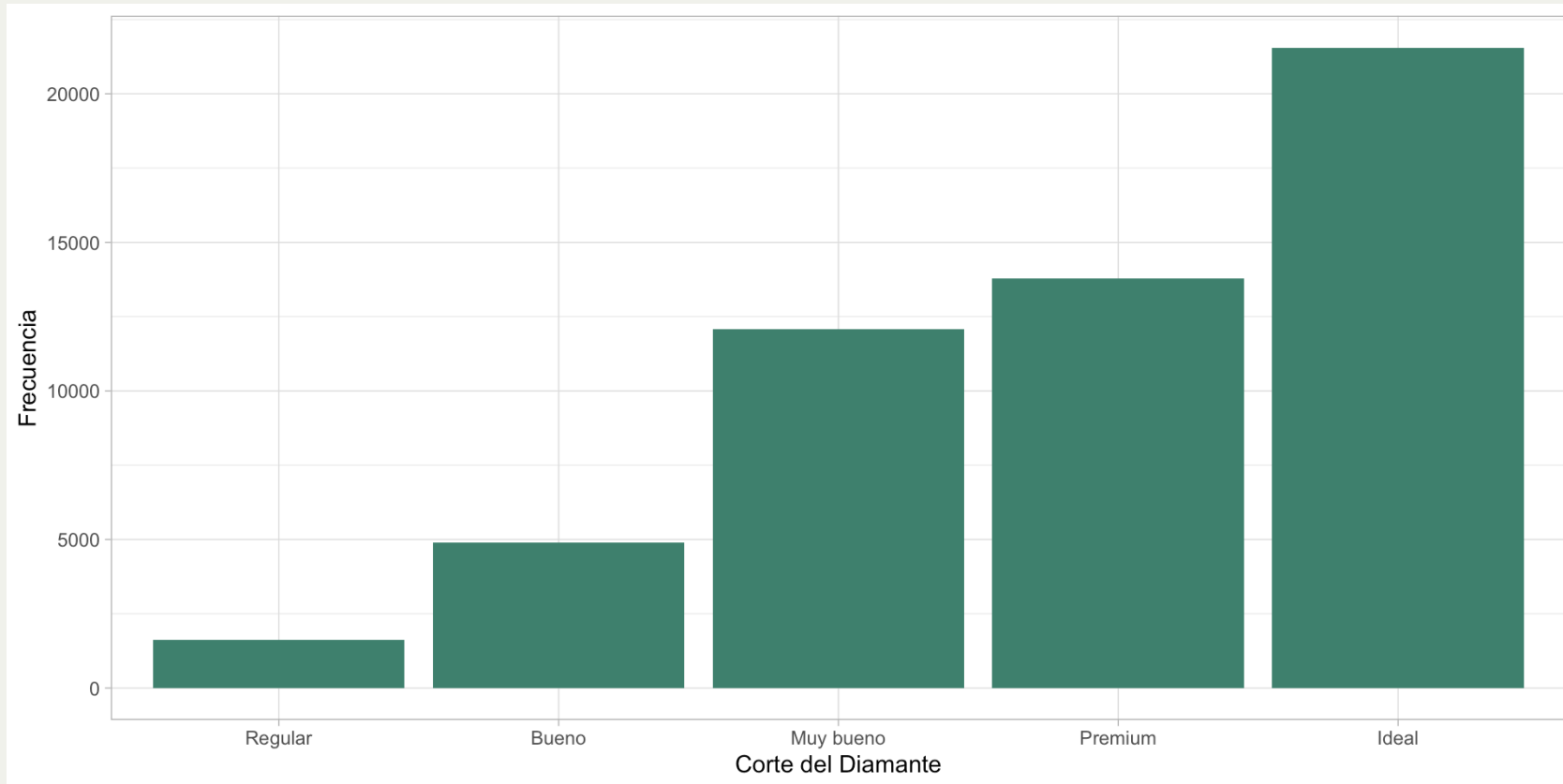


Tabla de Frecuencias Relativas

La Frecuencia relativa indica la proporción de los datos que tienen el valor indicado. Esto se calcula dividiendo la frecuencia del valor por el total de los datos: f/n . En la tabla de diamantes, obtendríamos:

Corte del Diamante	Frecuencia	frec_relativa	frec_relativa_porcentaje
Regular	1.610	1610/53.940	0,03
Bueno	4.906	4906/53.940	0,09
Muy bueno	12.082	12082/53.940	0,22
Premium	13.791	13791/53.940	0,26
Ideal	21.551	21551/53.940	0,40
Total	53.940	-	1,00

Gráfico de Torta

El gráfico consiste en dividir una circunferencia respecto a la frecuencia relativa de cada categoría.

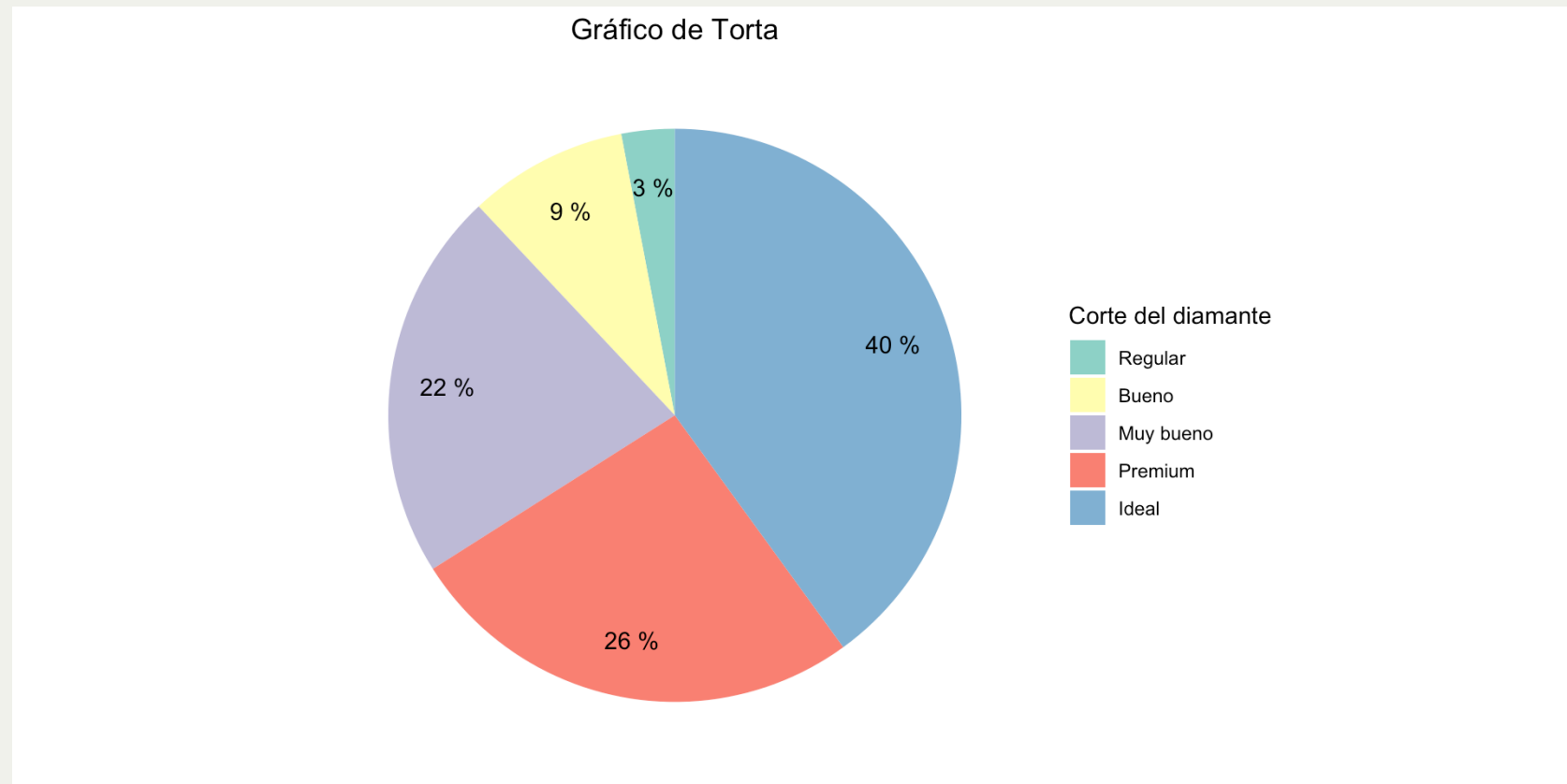
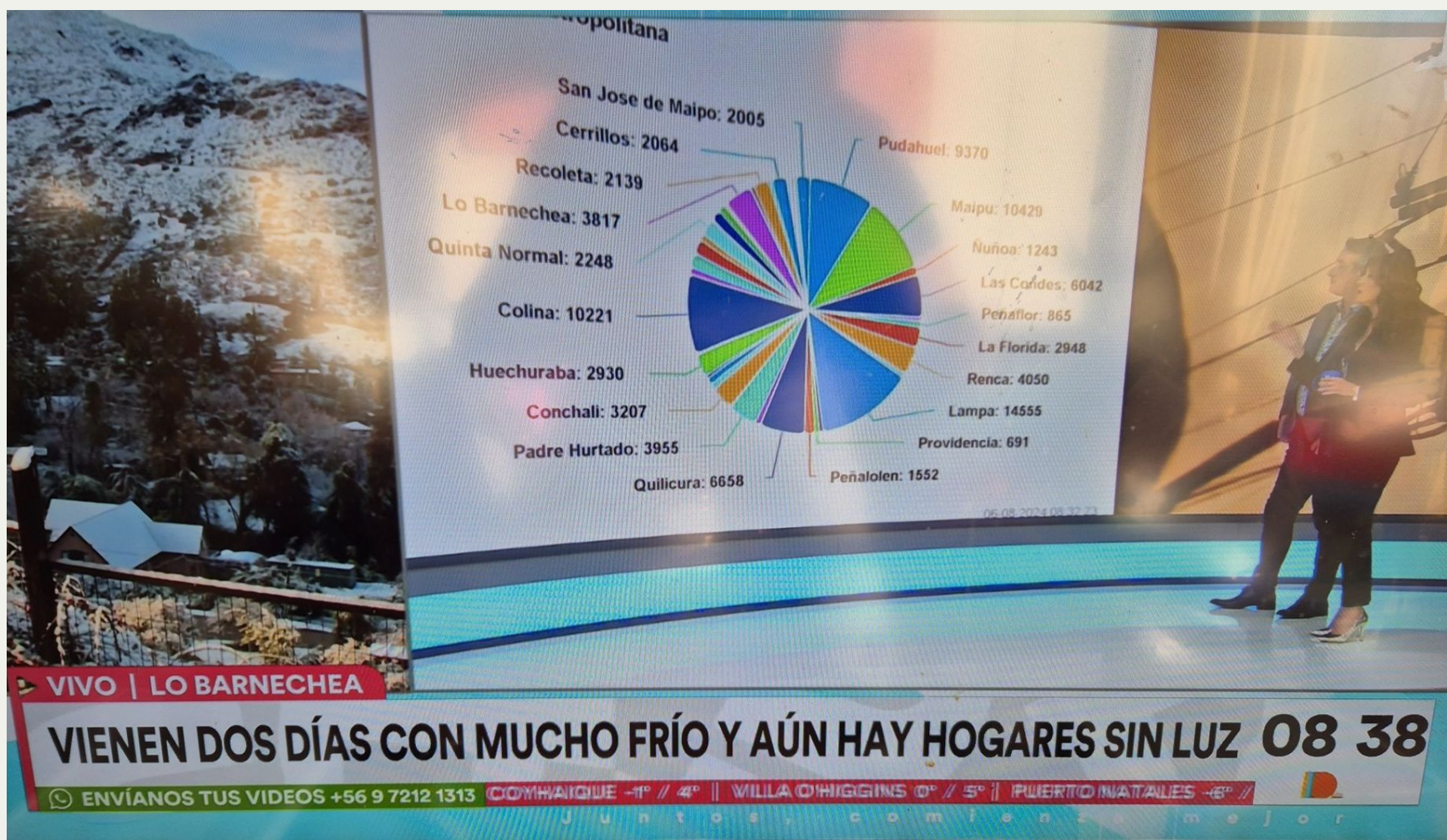


Gráfico de Torta II

Sin embargo, su uso no es recomendable si existen muchas categorías ya que no existe punto de referencia y es difícil comparar.



Usted no lo haga