

CHEATSHEET ESTADÍSTICA DESCRIPTIVA

ESCALAS DE MEDIDA

- **Categórica nominal:** Tiene la propiedad de identidad (igual/diferente) pero NO la de orden. Ejemplos: sexo, colores, ...
- **Categórica ordinal:** Añade la propiedad de orden, pero NO tiene una igual distancia entre elementos. Ejemplos: clase social, nivel formativo, ...
- **Numérica discreta:** añade la propiedad de igualdad de distancias, pero no permite infinitos valores entre sus elementos. Ejemplos: número de hijos, número de siniestros, ...
- **Numérica continua:** Puede tener (teóricamente) infinitos valores entre sus elementos. Ejemplos: precio de un producto, distancias, ...

METRICAS PARA VARIABLES CATEGORICAS

- Conteo de frecuencias (con Pandas): `df.smoker.value_counts()`
- Moda (con statistics): `statistics.mode(vector)`
- Tablas cruzadas (con Pandas): `pd.crosstab(df.sex,df.smoker, margins = True)`
- Tablas cruzadas en porcentaje (con Pandas): `pd.crosstab(df.sex,df.smoker, margins = True, normalize = 'all')`
- Chi-cuadrado (con Scipy):
 - `tabla = pd.crosstab(df.sex,df.smoker)`
 - `chi, pvalor, gl, experado = sp.stats.chi2_contingency(tabla)`

METRICAS PARA VARIABLES NUMERICAS: CENTRALIZACION

- Media aritmética (con Statistics): `statistics.mean([3,4,5])`
- Media winsorizada (con Scipy): `sp.stats.mstats.winsorize(vector, limits = [0, 0.2])`
- Media geométrica – para porcentajes (con Statistics):
`statistics.geometric_mean([0.3,0.4,0.5])`
- Media armónica – para media de medias (con Statistics):
`statistics.harmonic_mean([30.3,40.8,50.1])`
- Mediana (con Statistics): `statistics.median(vector)`

METRICAS PARA VARIABLES NUMERICAS: DISPERSION

- Varianza (con Numpy): `vector1.var()`
- Desviación típica (con Numpy): `vector1.std()`
- Coeficiente de variación: `vector1.std() / vector1.mean() * 100`

CORRELACION

- Pearson – para lineal (con Scipy): `sp.stats.pearsonr(var1,var2)`
- R cuadrado (con Scipy): `sp.stats.pearsonr(var1,var2) ** 2`
- Spearman – para no lineal (con Scipy): `sp.stats.spearmanr(var1,var2)`