**Capítulo 1:** tipos de variables, como crear las matrices de datos, filtrar, etc. (lo básico)

**Capítulo 2:** media, mediana, moda, varianza, desviación estándar, cuantiles, rango intercuartil, GRÁFICOS (histograma, gráfico de cajas, etc.)

**Capítulo 3:** valor esperado, distribución normal, gráfico QQ (*para verificar que la muestra tiene distribución normal),*

**DISTRIBUCIONES CONTINUAS:**

* distribución Z *(qué tan raro es un número en una escala)*,
* distribución chi-cuadrado *(cantidad de valores que pueden cambiar libremente en un conjunto de datos),*
* distribución t,
* distribución F (comparar varianzas),

**DISTRIBUCIONES DISCRETAS:**

* distribución de Bernoulli *(cada intento individual tiene solo dos resultados posibles: éxito y fracaso),*
* distribución geométrica *(cantidad de intentos que debemos realizar hasta obtener un éxito),*
* distribución binomial *(probabilidad de tener k éxitos en n intentos independientes),*
* distribución binomial negativa *(probabilidad de encontrar el k-ésimo éxito al n-ésimo intento)*
* distribución de Poisson *(cantidad de eventos en una población grande en un tiempo dado)*

**Capítulo 4:** distribución muestral, teorema del límite central, error estándar, intervalos de confianza *(rango de valores plausibles para el parámetro estimado),* pruebas de hipótesis (nula y alternativa), valor p, estadístico de prueba (α > p => Ha, α < p => H0)

*Prueba de hipótesis usando el modelo normal:*

*1. Formular la hipótesis nula (H0) y alternativa (HA) en lenguaje llano y luego en notación*

*matemática.*

*2. Identificar un estimador puntual (estadístico) adecuado e insesgado para el parámetro de*

*interés.*

*3. Verificar las condiciones para garantizar que la estimación del error estándar sea razonable y*

*que la distribución muestral del estimador puntual siga aproximadamente una distribución*

*normal.*

*4. Calcular el error estándar. Luego, graficar la distribución muestral del estadístico bajo el*

*supuesto de que H0 es verdadera y sombrear las áreas que representan el valor p.*

*5. Usando el gráfico y el modelo normal, calcular el valor p para evaluar las hipótesis y escribir*

*la conclusión en lenguaje llano.*

Valor z => muestras grandes (> 30)

Valor t => muestras pequeñas (<= 30)

**Capítulo 5:**

* Prueba Z *(asegurar o descartar que la media de la población tiene un cierto valor hipotético, debe conocerse la desviación estándar)*
* Shapiro-Wilk *(comprobar que una muestra tiene distribución normal)*
* Prueba t *(se usa cuando no se conoce la desviación estándar)*
* Prueba t una muestra
* Prueba t dos muestras pareadas *(cada observación de un conjunto tiene una correspondencia con exactamente una observación del otro, se debe sacar la media de las diferencias junto con la desviación estándar)*
* Prueba t dos muestras independientes (comparar las medias de dos poblaciones en que las observaciones con que se cuenta no tienen relación con ninguna de las otras observaciones, se usa la diferencia de las medias muestrales)

**Capítulo 6:**

* Tamaño del efecto *(valor de la diferencia entre dos grupos, o del valor observado con respecto al valor nulo)*
* d de Cohen *(prueba tamaño del efecto)*
* Poder estadístico *(probabilidad de no cometer un error de tipo II, o sea, rechazar Ha cuando es verdadera),* también qué tan propensa es una prueba estadística para distinguir un efecto real de una simple casualidad *(o determinar el tamaño adecuado de la muestra para detectar un cierto tamaño del efecto)*
* Prueba bilateral (2 colas): se desconoce el signo de la diferencia de medias

**Ha: media(A) ≠ media (B)**

* Prueba unilateral (1 cola): se conoce el signo de la diferencia de medias

**Ha: media(A) > media(B)** o **media(A) < media(B)**

Existe una ventaja de las pruebas unilaterales: cuando el tamaño del efecto aumenta en el sentido de la hipótesis alternativa, el poder es mayor que para una prueba bilateral.

**Capítulo 7:** métodos de Wald y de Wilson para inferir acerca de una y dos proporciones según una distribución binomial

* Wald para 1 proporción *(construir intervalos de confianza y contrastar hipótesis bajo el supuesto de normalidad)*
* Wald para 2 proporciones *(diferencia entre las proporciones de dos poblaciones)*
* Wilson *(sirve para ambos casos, pero no permite poner un valor nulo ≠ 0)*
* Poder estadístico para pruebas de proporciones

**Capítulo 8:** pruebas con hipótesis que no mencionan parámetro ni distribución de la población, estudiar variables categóricas

* Entregan menos información
* Tienen menor poder estadístico, suelen necesitar muestras de mayor tamaño para detectar diferencias significativas que pudieran existir entre las poblaciones comparadas
* Prueba chi cuadrado de Pearson *(inferir con proporciones cuando hay dos variables categóricas y una es dicotómica => 2 niveles), deben ser a lo menos 5 muestras*
  + Chi cuadrado de homogeneidad *(determinar si dos poblaciones presentan las mismas proporciones en los diferentes niveles de una variable categórica)*
  + Chi cuadrado de bondad de ajuste *(comprobar si una distribución de frecuencias observada se asemeja a una distribución esperada)*
  + Chi cuadrado de independencia (determinar si dos variables categóricas, de una misma población, son estadísticamente independientes o si, por el contrario, están relacionadas)
* Pruebas para muestras pequeñas (chi cuadrado)
  + Prueba exacta de Fisher (cuando ambas variables sean dicotómicas => 2 niveles)
  + Prueba de mcNemar (una misma característica, con respuesta dicotómica, se mide en dos ocasiones diferentes para los mismos sujetos [muestras pareadas] y queremos determinar si se produce o no un cambio significativo entre ambas mediciones)
    - **H0: no hay cambios significativos en las respuestas**
    - **Ha: sí hay cambios significativos en las respuestas**
  + Prueba Q de Cochran (extensión de la prueba de mcNemar, adecuada cuando la variable de respuesta es dicotómica y la variable independiente tiene más de dos observaciones pareadas)
    - **H0: la proporción de “éxitos es la misma para todos los grupos”**
    - **Ha: la proporción de “éxitos” es distinta para al menos un grupo**

**Pruebas post-hoc sólo si gana Ha**

- corrección de Bonferroni (no se recomienda si el número de grupos es alto)

- corrección de Holm (mayor poder estadístico)

* Chi cuadrado => **H0: las variables son independientes**

**Ha: las variables están relacionadas**