

# Parcial computacional sobre Test de Hipótesis

Métodos Estadísticos en Física Experimental

2021

## 1 Generalidades

1. Indicar nombre, apellido, libreta o DNI en el informe.
2. Dar una descripción clara y precisa de la metodología utilizada.
3. Incluir todos los gráficos como figuras con sus correspondientes leyendas.
4. Justifique las hipótesis en la que se sustenten sus resultados y discuta los resultados obtenidos.
5. La fecha límite para la entrega es el martes 13 de julio a las 11 hs.
6. Enviar por mail a [dariorodriguesfm@gmail.com](mailto:dariorodriguesfm@gmail.com). Nombrar al archivo de la siguiente manera: TH-SuApellido.pdf y adjuntar en el mismo mail los códigos desarrollados. Utilizar como asunto del mail: "Parcial TH".
7. Por consultas sobre interpretación de los enunciados escribir al grupo: [mefe2021@googlegroups.com](mailto:mefe2021@googlegroups.com), así todos tienen acceso a todas las respuestas.

## 2 Estadística del lenguaje

Empezamos la materia leyendo un cuento y calculando probabilidades en la biblioteca de Babel, por qué no cerrarla usando todo lo que aprendimos para testear hipótesis utilizando libros.

Cada idioma tiene sus características, una de ellas es la distribución de longitudes de las palabras que lo conforman. Sin hacer ningún estudio nos podríamos animar a decir, por ejemplo, que es más probable encontrar palabras largas en un libro escrito en alemán que en uno escrito en español. ¿Podríamos utilizar esa característica para identificar en que idioma está escrito un libro sin saber siquiera leerlo?

### 3 Consignas

1. **Distribución:** Elegí dos libros en formato digital, uno en español y otro en alemán<sup>1</sup>, y a partir de ellos construí la distribución de la variable aleatoria  $t_1$ : *número de letras en una palabra* para esos dos idiomas. Calculá su media, moda y mediana. Para este ítem podés usar el script *palabras.py* que está disponible en la solapa *Material Adicional* de la página de la materia.
2. **El estadístico:** A partir de esas distribuciones calculá, para los dos idiomas, la distribución del estadístico  $t_{20}$ : *número de letras en la palabra más larga en una muestra de  $n=20$  palabras*.
3. **El test:** Determiná cual es el valor crítico  $t_{20}^C$  del estadístico  $t_{20}$  a partir del cual rechazarías la hipótesis nula  $H_0$ : *el libro está escrito en español* con una significancia de a lo sumo  $\alpha=0.05$ .
4. **Potencia:** Calculá la potencia del test propuesto con  $t_{20}$  cuando la hipótesis alternativa es  $H_1$ : *el libro está escrito en alemán* ¿Como cambia el resultado si utilizas directamente  $t_1$  como estadístico? ¿Y si usas  $t_{100}$ ?
5. **Rachas** Para cada idioma, calculá la distribución del estadístico propuesto en el test de *runs* aplicado a la variable aleatoria  $t_1$ , para una muestra de tamaño 30. En este caso las rachas serán estar por encima o por debajo de la mediana.  
  
Si el valor coincide con la mediana entonces tomá uno de los siguientes tres criterios según tu fecha de nacimiento: si cumplís años en enero, febrero, marzo o abril: excludo de la muestra; si cumplís en mayo, junio, julio o agosto: asumí que está por encima y en otro caso: asumí que está por debajo.
6. **Quedate en casa:** Tomá una muestra de 30 palabras de un libro que tengas en tu casa y aplicale el test del ítem anterior para testear la hipótesis nula de que el libro está escrito en alemán. Para que la elección de las 30 palabras sea efectivamente al azar, usá la que aparece en la posición  $X$  de las primeras 30 carillas, donde  $X$  es justamente el número de página. Por ejemplo, si mirás la página 9, contá cuantas letras conforman la novena palabra de esa carilla. Calculá el p-valor. ¿Con que nivel de confianza dirías que el libro que agarraste no está escrito en alemán? sin intentar leerlo!
7. *Opcional* Si después de aplicar en el ítem 6 el Test desarrollado en el ítem 5, te pareció que podría mejorarse la sensibilidad del estadístico, entonces animate a proponer otro que incremente la potencia del test y rehacé el ítem 6.

---

<sup>1</sup><https://www.gutenberg.org/ebooks/results/>