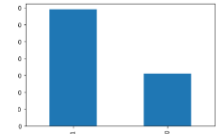


CHEATSHEET ESTADÍSTICA INFERENCIAL

DISTRIBUCIONES

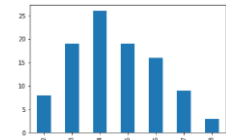
- Bernoulli - experimento aleatorio en que solo pueden ocurrir dos sucesos que además son mutuamente excluyentes (con Scipy):

- `from scipy.stats import bernoulli`
- `bernoulli.rvs(size=100, p=0.7)`



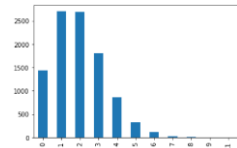
- Binomial - suma de los "éxitos" (los unos) que se obtienen en n repeticiones de un experimento de Bernoulli donde la probabilidad (p) de unos se mantiene constante (con Scipy):

- `from scipy.stats import binom`
- `monedas = binom.rvs(n=10, p=0.5, size=100)`



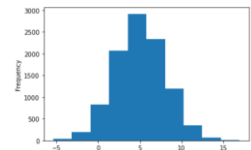
- Poisson - número de veces que un evento ocurrirá en un intervalo de tiempo (con Scipy):

- `from scipy.stats import poisson`
- `poisson.rvs(mu=2, size=10000)`



- Normal - cuando una variable está compuesta de otras muchas variables (con Scipy):

- `from scipy.stats import norm`
- `normal = norm.rvs(size=10000, loc=5, scale=3)`



TEOREMA DEL LIMITE CENTRAL

- Si hacemos medias sobre muchas muestras aleatorias de una población, la distribución resultante de todas esas muestras se va a distribuir según una normal, da igual cual fuera forma de la distribución original en la población
- La media de la distribución muestral tiende a converger a media de la población. Por tanto podemos usar el dato de la distribución muestral como válido en la población
- La variabilidad de la distribución muestral, que se llama error típico (o también lo verás como error estándar), va a ser igual a la desviación típica de la población dividida por la raíz cuadrada del tamaño de la muestra

CHEATSHEET ESTADÍSTICA INFERENCIAL

INTERVALOS DE CONFIANZA EN MEDIAS

1. Calcular el error típico: $np.array(muestra).std() / \text{math.sqrt}(n)$
2. Desv típicas según Nivel de Confianza: 2 para 95.5%, 3 para 99%
3. Calcular margen de error: $error_99 = error_típico * desv\ tip$
4. Calcular Intervalo de Confianza:
 $ic_99 = [media_muestra - error_99, media_muestra + error_99]$

INTERVALOS DE CONFIANZA EN PROPORCIONES

1. Calcular el error típico: $\text{math.sqrt}((p * q)/n)$
2. Desv típicas según Nivel de Confianza: 2 para 95.5%, 3 para 99%
3. Calcular margen de error: $error_99 = error_típico * desv\ tip$
4. Calcular Intervalo de Confianza: $ic_99 = [p - error_99, p + error_99]$

CALCULAR EL TAMAÑO DE LA MUESTRA (Poblaciones Infinitas)

- **Fórmula:** $muestra = ((z^{**2}) * (50*50)) / (margen\ error^{**2})$
- **Función para copiar y pegar:**

```
def tamaño(z,me):  
    tamaño = ((z**2) * (50*50)) / (me ** 2)  
    return(tamaño)
```

ALPHA Y PVALOR

- **Alpha:** no es más que 1-NC. Es decir, si trabajamos a un NC del 95% entonces Alpha es 5%. Aunque normalmente nos vendrá en tanto por uno, es decir 0.05
- **Pvalor:** es la probabilidad de obtener un valor tan extremo o más que el obtenido en la muestra asumiendo que la hipótesis nula es cierta
- **Aplicación:** si el pvalor es menor al Alpha entonces tenemos que rechazar la hipótesis nula, es decir, el patrón SI es estadísticamente significativo

CHEATSHEET ESTADÍSTICA INFERENCIAL

CONTRASTE DE MEDIAS EN LA POBLACION

1. Definir las hipótesis H_0 y H_1 : Ej H_0 "el valor es 25 en la población"
2. Elegimos un nivel de confianza que nos da un α : Ej 95% y 0.05
3. Calculamos el pvalor con Scipy (t sobre una muestra):
`sp.stats.ttest_1samp(a = df.total_bill, popmean = h0)[1]`
4. Si $p\text{valor} \leq \alpha$ rechazamos H_0 y aceptamos H_1

CONTRASTE DE MEDIAS ENTRE 2 MUESTRAS

1. Definir las hipótesis H_0 y H_1 : Ej H_0 "las medias son iguales en la población"
2. Elegimos un nivel de confianza que nos da un α : Ej 95% y 0.05
3. Calculamos el pvalor con Scipy (t sobre dos muestras independientes):
`sp.stats.ttest_ind(a = hombres, b = mujeres, equal_var = False)[1]`
4. Si $p\text{valor} \leq \alpha$ rechazamos H_0 y aceptamos H_1

CONTRASTE DE PROPORCIONES EN LA POBLACION

1. Definir las hipótesis H_0 y H_1 : Ej H_0 "el valor es 0.7 en la población"
2. Elegimos un nivel de confianza que nos da un α : Ej 95% y 0.05
3. Calculamos el pvalor con statsmodels (z sobre una muestra):
`proportions_ztest(count = exitos, nobs = muestra, value = h0)[1]`
4. Si $p\text{valor} \leq \alpha$ rechazamos H_0 y aceptamos H_1

CONTRASTE DE PROPORCIONES ENTRE 2 MUESTRAS

1. Definir las hipótesis H_0 y H_1 : Ej H_0 "las proporciones son iguales en la población"
2. Elegimos un nivel de confianza que nos da un α : Ej 95% y 0.05
3. Calculamos el pvalor con statsmodels (z sobre dos muestras):
`proportions_ztest(count = array_exitos, nobs = array_muestras)[1]`
4. Si $p\text{valor} \leq \alpha$ rechazamos H_0 y aceptamos H_1