

Ejercicios

Analítica de Datos
Pontificia Universidad Javeriana

1 Distribución Normal

- Encuentre cada una de las siguientes probabilidades para la distribución normal.
 - $P(z > 1.25)$ **p=0.105**
 - $P(z > -0.6)$ **p=0.725**
 - $P(z < 0.7)$ **p=0.758**
 - $P(z < -1.3)$ **p=0.096**
- ¿Qué proporción de la distribución normal se ubica entre cada uno de los siguiente intervalos?
 - $z = -0.25$ y $z = 0.25$ **p=0.197**
 - $z = -0.67$ y $z = 0.67$ **p=0.497**
 - $z = -1.20$ y $z = 1.20$ **p=0.769**
- Para un examen que se distribuye normal con media $\mu = 3$ y desviación estándar $\sigma = 0.5$, encuentre la proporción de la población que corresponde a cada uno de lo siguiente:
 - Notas mayores a 3.25 **p=0.308**
 - Notas menores a 3.4 **p=0.788**
 - Notas entre 2.5 y 3.5 **p=0.682**
- La distribución de ingresos para Colombia es normal con $\mu = 500$ (en miles de pesos) y $\sigma = 100$ (en miles de pesos).
 - ¿Qué valor de ingreso, separa el 15% más rico de la distribución del resto? **RTA=603.64**
 - ¿Qué valor de ingreso, separa el 10% más rico de la distribución del resto? **RTA=628.15**
 - ¿Qué valor de ingreso, separa el 2% más rico de la distribución del resto? **RTA=705.37**
- Las personas en Colombia sonríen en promedio $\mu = 62$ veces por día. Asumiendo que la distribución de sonrisas es aproximadamente normal con una desviación estándar $\sigma = 18$, encuentre cada uno de los siguientes valores:
 - ¿Qué proporción de los colombianos sonríe más de 80 veces al día? **p=0.159**
 - ¿Qué proporción de los colombianos sonríe menos de 50 veces al día? **p=0.255**
 - ¿Qué proporción de los colombianos sonríe entre 40 veces y 75 veces al día? **p=0.653**
 - ¿Cuántas sonrisas al día corresponden al percentil 30 de la distribución? **z=-0.53 - 52.46 sonrisas**
 - Si el enunciado se basa en un estudio realizado a 2000 colombianos, ¿cuántas personas del estudio aproximadamente sonríen menos de 30 veces al día? **p=0.038 - 76 personas**

6. La Secretaría de Transporte de Bogotá reporta que la edad promedio de los conductores bogotanos es $\mu = 45.7$ años con una desviación estándar $\sigma = 12.5$ años. Asumiendo que la distribución de la edad de los conductores es aproximadamente normal.
- (a) ¿Qué proporción de los conductores es menor a 30 años? **p=0.104**
 - (b) ¿Qué proporción de los conductores tiene entre 30 y 40 años? **p=0.22**
 - (c) ¿Qué edad corresponde al percentil 60 de la distribución? **z=0.25 - 48.82 años**
 - (d) Suponga que en Bogotá hay 3 millones de conductores, ¿cuántos conductores aproximadamente tienen entre 60 y 70 años? **p=0.101 - 303000 conductores**
7. Una encuesta a consumidores indica que los hogares gastan en promedio $\mu = \$185000$ en productos a la semana. La distribución del dinero gastado es aproximadamente normal con una desviación estándar de $\sigma = \$25000$. Basándose en esta distribución:
- (a) ¿Qué proporción de la población gasta entre 120000 y 150000 pesos a la semana? **p=0.076**
 - (b) ¿Qué nivel de gasto corresponde al percentil 25? **z=-0.67 - 168.250 pesos**
 - (c) ¿Qué nivel de gasto corresponde al percentil 85? **z=1.04 - 211.000 pesos**
 - (d) Asumiendo que en la encuesta participaron 1000 personas, ¿cuántas personas gastan menos de 210000 pesos? **p=0.841 - 841 personas**
 - (e) Asumiendo que en la encuesta participaron 1000 personas, ¿cuántas personas gastan entre 200000 y 240000 pesos? **p=0.26 - 260 personas**

2 Intervalos de Confianza

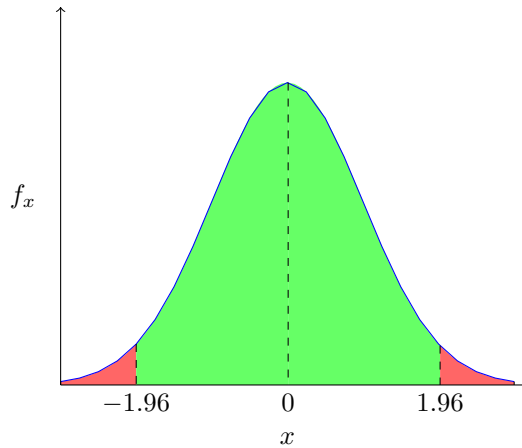
2.1 Ejercicio Resuelto

- Se quiere conocer la permanencia media de los pacientes de un hospital, con el fin de estudiar una posible ampliación del mismo. Se tienen datos referidos a la estancias, expresada en días, de 800 pacientes, obteniéndose los siguientes resultados: $\bar{X} = 8,1$ días; $s = 9$ días. Se pide obtener un intervalo de confianza del 95% para la estancia media
- **Solución:** Lo primero es calcular el nivel de significancia (error esperado) dado el nivel de confianza que presenta el enunciado. El nivel de significancia se denota con la letra α . Por lo tanto:

$$\alpha = 100\% - \text{nivel de confianza}$$

$$\alpha = 100\% - 95\% = 5\%$$

Dado este error esperado, ahora se debe encontrar el valor Z crítico, simbolizado Z_α . En este caso tenemos que encontrar el Z_α para un nivel de significancia igual al 5%. Por lo tanto, buscamos en la tabla de la normal el valor Z para el cual 97.5% de las observaciones caigan debajo de ese valor. ¿Por qué el 97.5%? Al observar la distribución de abajo, se observa que ambas áreas rojas deben sumar 5%, por lo tanto, cada una debe sumar la mitad. Si el área roja de la derecha es igual al 2.5% del área total, el resto es igual a $100 - 2.5 = 97.5\%$. Al buscar este valor en la tabla de la distribución normal, se encuentra que el valor Z es igual a 1.96. Considerando la simetría de la curva normal, se sabe que el valor que divide el 2.5% del área roja de la izquierda con respecto al área verde es -1.96 .



Encontrado el valor Z crítico, se calcula el error estándar para la distribución de medias muestrales. En este caso:

$$s_{\bar{X}} = \frac{s_X}{\sqrt{n}} = \frac{9}{\sqrt{800}} = 0,318$$

Una vez calculado el error estándar, este se multiplica por Z_α para obtener el término o margen del error:

$$\text{Margen de error} = Z_\alpha * s_{\bar{X}} = 1,96 * 0,318 = 0,62$$

El intervalo de confianza de una media poblacional es una media muestral más y menos un término de error, la fórmula general para calcular el intervalo de confianza es la siguiente:

$$IC \text{ Superior} = \bar{X} + Z_\alpha * s_{\bar{X}} = 8,1 + 0,62 = 8,72$$

$$IC \text{ Inferior} = \bar{X} - Z_\alpha * s_{\bar{X}} = 8,1 - 0,62 = 7,48$$

Estoy 95% seguro que la estancia media de los pacientes está entre 7,4 y 8,72 días.

2.2 Ejercicios

- Una muestra aleatoria simple de 25 estudiantes responde a un test de inteligencia, obteniendo una media de 100 puntos. Se sabe por experiencia que la variable “inteligencia de todos los estudiantes” es normal con una desviación estándar igual a 10, pero se desconoce la media. ¿Entre qué límites se hallará la verdadera inteligencia media de todos los estudiantes, con un nivel de confianza de 99%? **Rta: ICS=105,1 ICI=94,9**
- Se asume que la recaudación diaria de las tiendas de un barrio determinado es una variable aleatoria que se aproxima a una distribución normal con una desviación estándar de 328000 pesos. Se ha extraído una muestra de 100 tiendas de dicho barrio, obteniéndose que la recaudación diaria media asciende a 1'248.000 pesos. Calcular:
 - El intervalo de confianza para la recaudación media con un nivel de confianza del 99%. **Rta: ICS=1'331.640 ICI=1'164.360**
 - El tamaño muestral mínimo necesario para conseguir, con un nivel de confianza del 95%, un error en la estimación de la recaudación diaria media menor de 127.000 pesos. **Rta: $n > 26$**
- En una encuesta se pregunta a 10.000 personas cuántos libros lee al año, obteniéndose una media de 5 libros. Se sabe que la población tiene una distribución normal con desviación estándar de 2. Hallar un intervalo de confianza al 80% para la media poblacional. **Rta: ICS=5,0256 ICI=4,9744**

4. Se quiere estimar el sueldo medio de un trabajador de transporte público. Se toma para ello una muestra de 625 de estos trabajadores y se obtiene un sueldo medio muestral de 1'480.000 pesos. Si la desviación estándar es igual a 250.000 pesos:
- (a) Con un nivel de confianza del 90%, determine el intervalo de confianza para el sueldo medio de un trabajador del transporte público. **Rta: ICS=1'496.500 ICI=1'463.500**
 - (b) Si se quiere que el error máximo de la estimación sea de 10.000 pesos, hallar el tamaño de la muestra que se debe tomar considerando un nivel de confianza del 99%. **Rta: $n > 4065$**
5. Para hacer un estudio sobre el precio/día de una habitación doble en hoteles de cuatro estrellas en Cartagena, se elige una muestra de 64 de estos hoteles y se obtiene un precio/día medio de 56 mil pesos con una desviación típica de 6 mil pesos. Determine el intervalo de confianza para el precio/día medio de una habitación doble en un hotel de cuatro estrellas en Cartagena con un nivel de confianza del 97%. **Rta: ICS=57,62 ICI=54,38**
6. Se ha obtenido una muestra al azar de 150 vendedores de una Editorial para estimar la proporción de vendedores en la Editorial que no alcanza un límite de ventas mínimo establecido por la dirección. De entre los seleccionados, 50 no han conseguido llegar al límite de ventas mínimo establecido.
- (a) Calcule el intervalo de confianza para la proporción de trabajadores en la Editorial que no alcanza el límite al 80%. **Rta: ICS=0,38 ICI=0,28**
 - (b) Calcule el intervalo de confianza para la proporción de trabajadores en la Editorial que no alcanza el límite al 99%. **Rta: ICS=0,43 ICI=0,23**