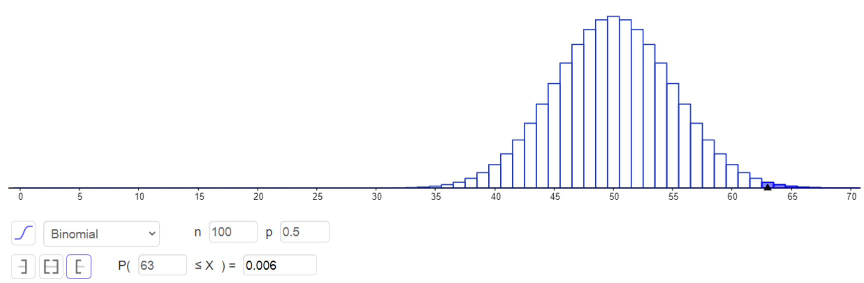
<u>**Examen corto**</u>

1. Un superintendente de producción dice que no hay diferencia entre los porcentajes de accidentes de empleados en turnos de día o de noche en una gran planta manufacturera. El número diario de accidentes se registran para los turnos de día y de noche durante n 100 días. Se encuentra que el número diario de accidentes en el turno de noche (x_E) excedió al número correspondiente de accidentes en el turno de día (X_D) en 63 de los 100 días. ¿Estos resultados dan suficiente evidencia para indicar que más accidentes tienden a ocurrir en un turno que en el otro?

Datos relevantes:

- Aquí la media podría ser falaz, por lo que resultará más revelador una prueba de medianas p o de probabilidad de éxito, como ventaja tomaremos esta medida no paramétrica porque no debe cumplir con el requisito de la normalidad.
- n=100;
- x_{E} = 63;
- x_D = 37;

1. **Parámetro de interés:** p.
2. **Hipótesis:**
   a. H0: p = 0.5;
   b. Ha: p ≠0.5; # alineada con lo que queremos probar
3. **Significancia:** alfa = 0.05
4. **Estadístico de prueba (definir un signo):**
   - Éxitos=37; # tomaremos los éxitos como accidentes en la noche. x_E
   - Fracasos=63; # tomaremos el complemento de n con los éxitos.
   - GeoGebra: ingresamos n=100, p=0.5, estamos interesados en todas las probabilidades conjuntas que sean mayor a 63 es decir $P(67 \leq X) = P(X = 67) + P(X = 68) + \cdots + P(X = 100)$. (Vale la pena recalcar el hecho que se usa la distribución binomial discreta y se puede aproximar usando distribuciones normales continuas, para este ejercicio me apegaré a la distribución binomial).



     - Como podemos observar nos sale que $P(67 \leq X) = 0.006$ puesto a que esta es una prueba de dos colas multiplicaremos el resultado por dos; no queda $2P(67 \leq X) = 0.012$.
     - Esto significa que el *valor-p* es 0.012.

- Establecer criterio de rechazo: rechazar H0 si: *valor-p* ≤ *significancia.*
   - 0.012 < 0.05; esta condición se cumple.

5. **Conclusión:**

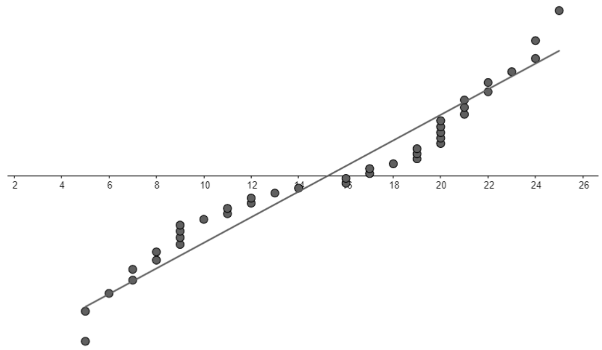
- Con una significancia de α =0.05 hay suficiente evidencia para rechazar la hipótesis nula y afirmar que sí hay diferencia entre las medianas de las personas que sufren accidentes en el turno nocturno y la mediana de personas que sufren accidentes de día no son iguales.

2. **Imágenes y recordar palabras** Un grupo de psicología realizó un experimento para determinar existe variación en la cantidad de palabras que un estudiante logra recordar.
Al primer grupo se les pidió que, con cada palabra, trataran de formar una imagen en su mente. Al segundo grupo no se les dio esta instrucción y ellos solamente intentaron memorizar las palabras. Veinte estudiantes participaron en el experimento, con los resultados que se indican en la tabla siguiente. ¿existe diferencia estadísticamente significativa en la cantidad de palabras recordadas utilizando estos dos procedimientos?

Estudiante	con imágenes	Sin imágenes
1	20	5
2	24	9
3	20	5
4	18	9
5	22	6
6	19	11
7	20	8
8	19	11
9	17	7
10	21	9
11	17	8
12	20	16
13	20	10
14	16	12
15	24	7
16	22	9
17	25	21
18	21	14
19	19	12
20	23	13

Datos relevantes:

- Puesto a que hay menos de 30 datos en cada muestra, no se puede asegurar normalidad. Es semi-normal, sin embargo, para esta prueba no nos dirá mucho la media puesto a que pueden haber valores atípicos que no describan o respondan la pregunta que estamos planteando. Como podemos ver en la figura presentada a continuación los datos se desvían un tanto de la línea de normalidad por lo que no se puede asegurar normalidad y por lo tanto no será confiable ni pertinente aplicar una prueba de hipótesis pertinente.

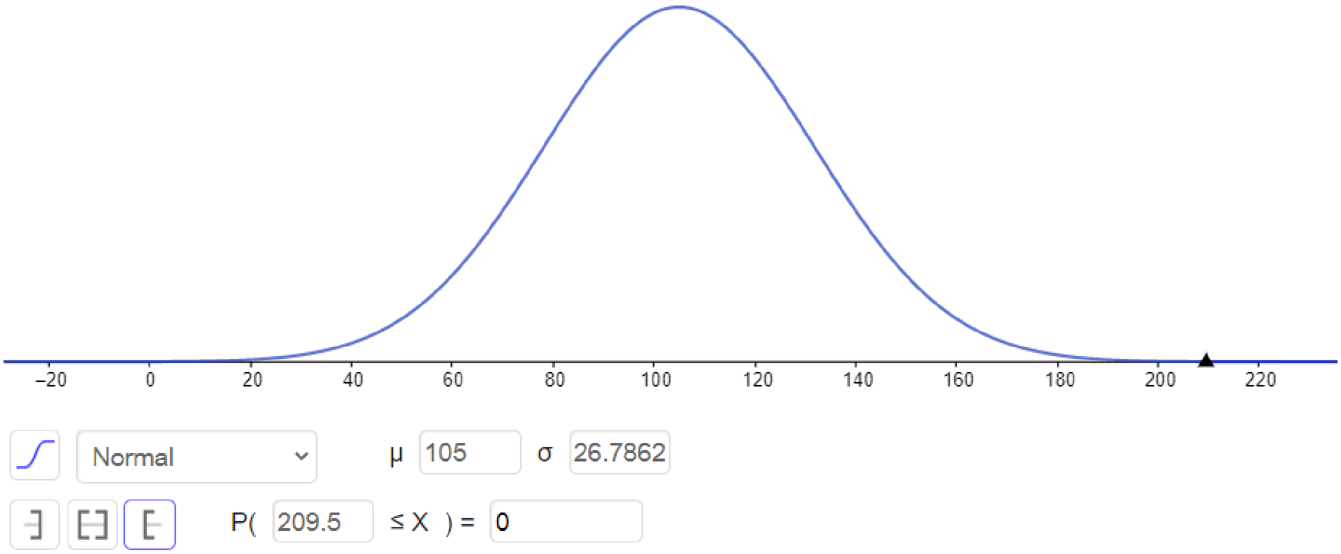


-

1. **Parámetro de interés:** diferencia de medianas (μ con ~ encima).
2. **Hipótesis:**
   - H0: medianas = 0; $\rightarrow$ p = 0.5;
   - Ha: medianas $\neq$ 0 $\rightarrow$ p $\neq$ 0.5;
3. **Significancia:** $\alpha = 0.05$;
4. **Estadístico de prueba:**
   - En Excel se sacan los rangos con Wilcoxon. Para sacar T^+.

						Rangos con signo	
Estudiante	con imágenes	Sin imágenes	Diferencia	Diferencias absol	Rango	"-"	"+"
1	20	5	15	15	2		2
2	24	9	15	15	2		2
3	20	5	15	15	2		2
4	18	9	9	9	4.5		4.5
5	22	6	16	16	4.5		4.5
6	19	11	8	8	6.5		6.5
7	20	8	12	12	6.5		6.5
8	19	11	8	8	8.5		8.5
9	17	7	10	10	8.5		8.5
10	21	9	12	12	11		11
11	17	8	9	9	11		11
12	20	16	4	4	11		11
13	20	10	10	10	13.5		13.5
14	16	12	4	4	13.5		13.5
15	24	7	17	17	15		15
16	22	9	13	13	17		17
17	25	21	4	4	17		17
18	21	14	7	7	17		17
19	19	12	7	7	19		19
20	23	13	10	10	20		20
						T^+:	210

   - Como podemos observar T^+ = 210.
   - A continuación, procedemos a sacar μ_{T^+}= 105.
     - En este caso (n(n+1)) / 4, esto resulta en (20(20+1))/4 = 105.
   - A continuación, procedemos a sacar σ_{T^+}= 26.78619047
     - En este caso sqrt((n(n+1)(2n+1) / (24)) = sqrt((20*(20+1)*(2*20+1)) / 24) = 26.78619047.
   - A continuación usaremos una distribución continua para aproximar una discreta.
     - Tomamos la probabilidad de X $\geq T^+$ - 0.5. En este caso únicamente restamos 0.5.
       - P(X>=T^{+}-0.5) = P(X >= 210-0.5) = P(X >= 209.5) = 0.
     - En GeoGebra vemos que esta probabilidad es 0 o demasiado pequeña.



- Establecer criterio de rechazo: rechazar si: valor-p <= significancia.
  - 2*0 <= 0.05, esta condición se cumple, por lo que rechazo la hipótesis nula.

5. Conclusión:

- Con una significancia de 0.05 se puede rechazar la hipótesis nula y afirmar que las medianas no son iguales por lo que podemos afirmar que si hay existe diferencia estadísticamente significativa en la cantidad de palabras recordadas utilizando estos dos procedimientos.