

Estadística Descriptiva

Introducción a la inferencia estadística

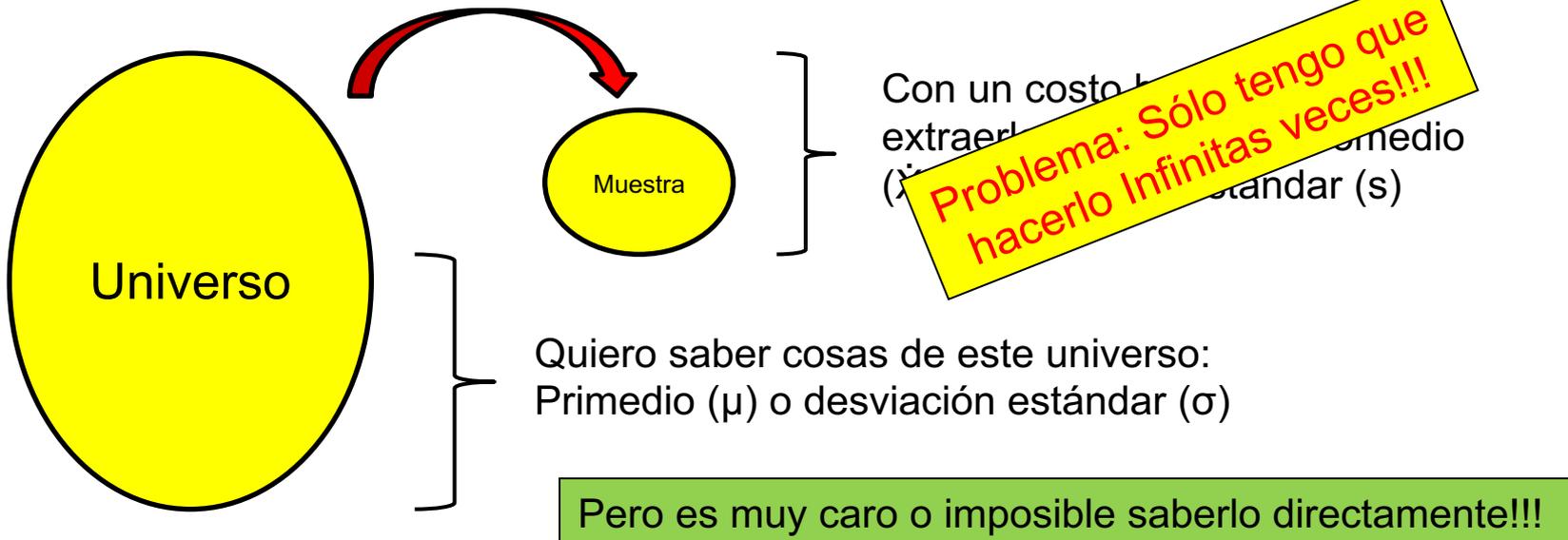
Rodrigo Asun Inostroza

Temas a tratar

- Aplicación de la **distribución de muestreo** a la inferencia estadística. Concepto de **inferencia estadística**.
- Concepto de estimación de **punto y de intervalo**.
- Concepto de **error estándar**.
- Estimación de **intervalos de confianza** para medias y proporciones.
- Definición de **hipótesis nula y estadística**.
- **Prueba de hipótesis** para una variable.

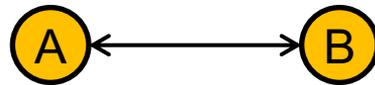
¿Para que sirve la Distribución de Muestreo?

- Si extraigo **infinitas muestras de tamaño grande (o de una variable con distribución normal)** desde un universo, podré saber el **promedio y desviación estándar** del Universo.
- Es decir, a partir de datos muestrales, podré saber cosas de un universo.
- Eso se llama: **Inferencia Estadística**.



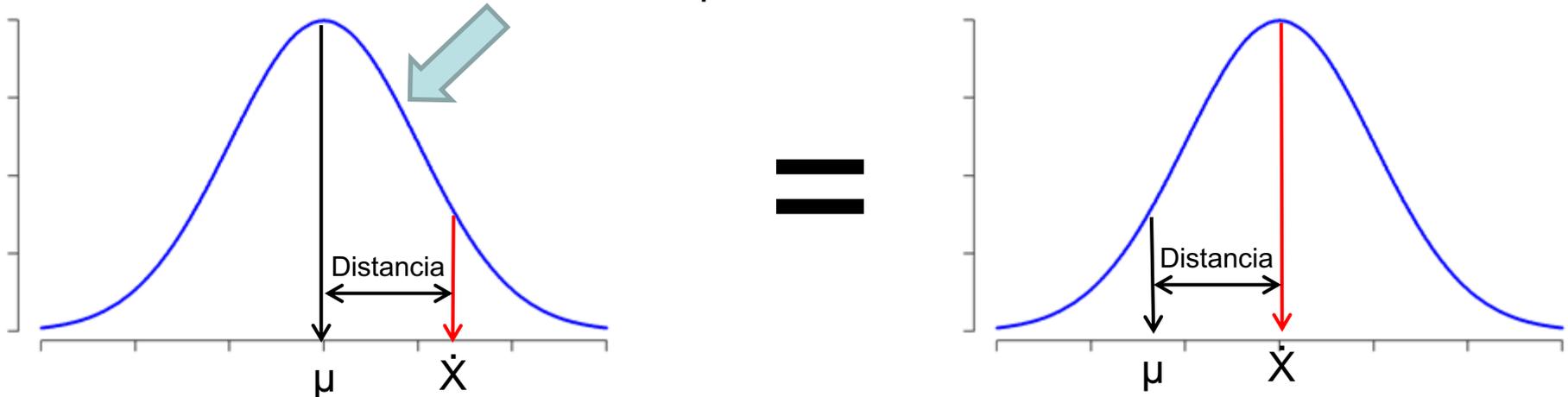
Truco... equidistancias en el espacio euclidiano

- En un espacio euclidiano, la distancia de A hasta B es igual a la distancia de B hasta A.

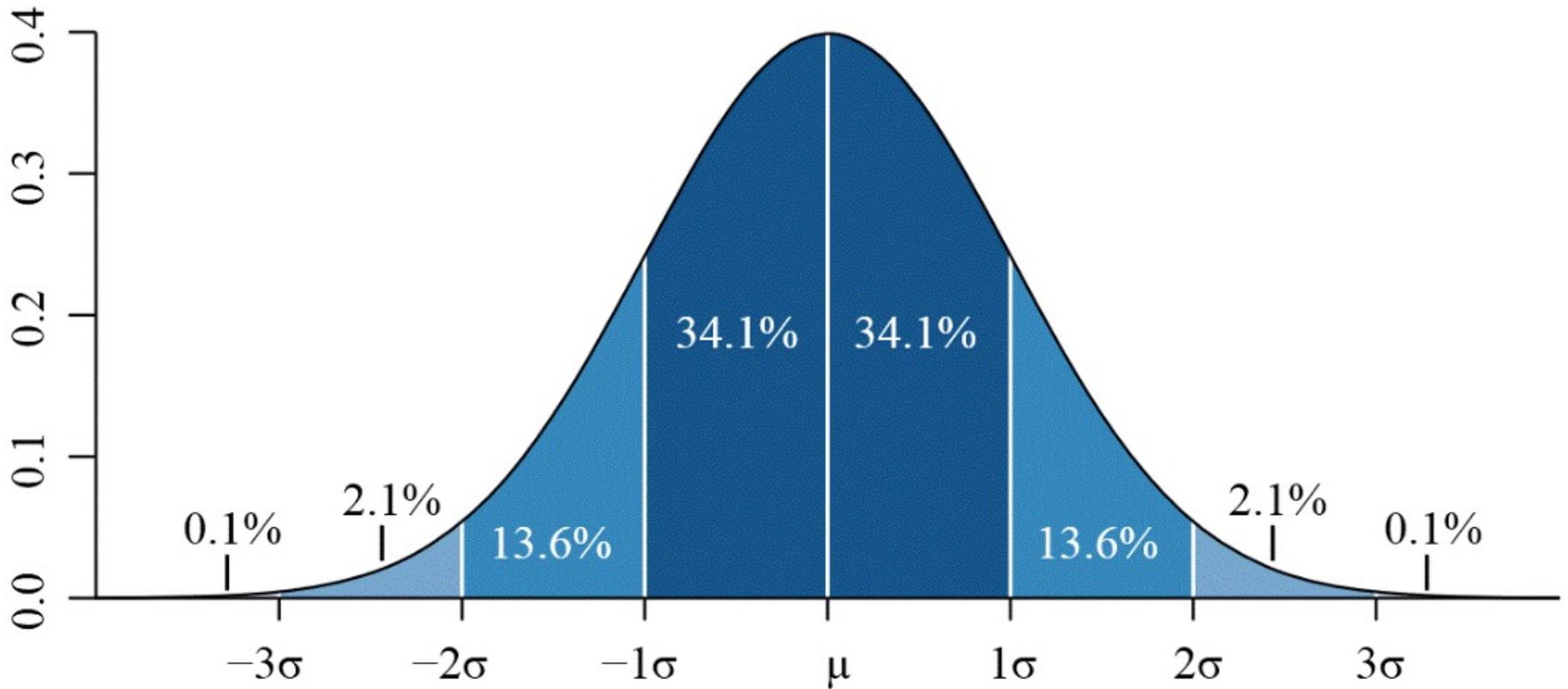


- Entonces... distancia desde μ a \bar{X} es igual a la distancia entre \bar{X} a μ .

Distribución de muestreo hipotética



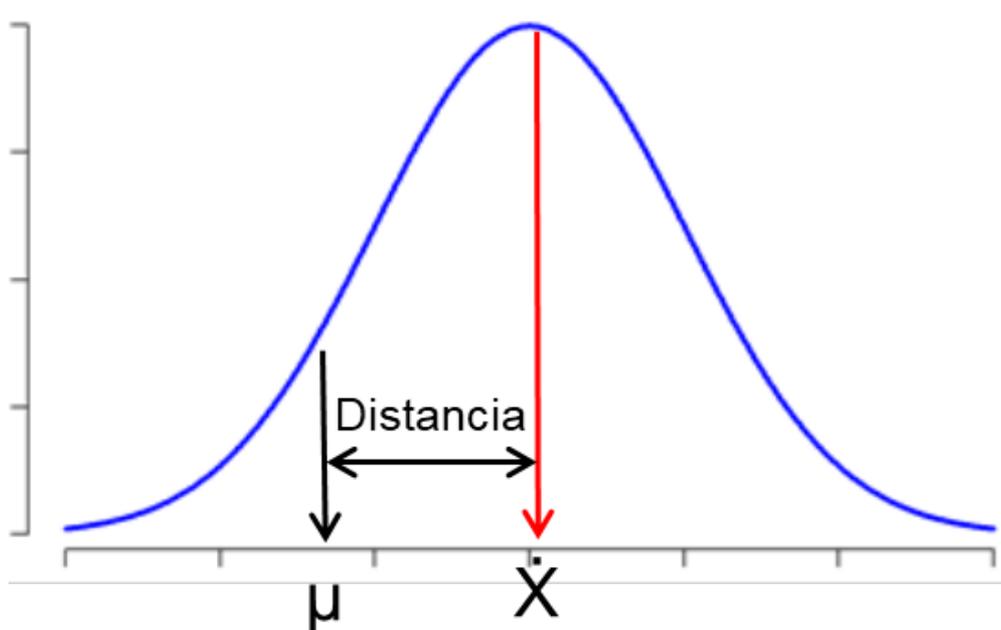
Recordemos...



Por la distribución de muestreo yo se que...

- Si extraigo infinitas muestras grandes de un Universo y calculo en cada una de ellas su promedio:
 - Los promedios se distribuirán con **función $N(\mu, \sqrt{n})$** .
 - Ello implica que el **68,2%** de los promedios de las muestras estarán a sólo **$1\sigma/\sqrt{n}$** del promedio del universo.
 - También, el **95,4%** de los promedios de las muestras estarán a sólo **$2\sigma/\sqrt{n}$** del promedio del universo.
 - Además, el **99,6%** de los promedios de las muestras estarán a sólo **$3\sigma/\sqrt{n}$** del promedio del universo.
- Por ejemplo: si el promedio de sueldos en Chile fuera de 600.000 pesos, con σ de 1.000 pesos.
 - Yo sabía que el **99,6%** de las muestras que yo extrajera de ese universo (siempre que fueran muestras grandes, por ejemplo 100 casos), **obtendrían resultados entre 580.000 y 620.000 pesos.**
 - Por ello, si una muestra me diera una media de **350.000 pesos**, sería una muestra **muy improbable o sesgada.**

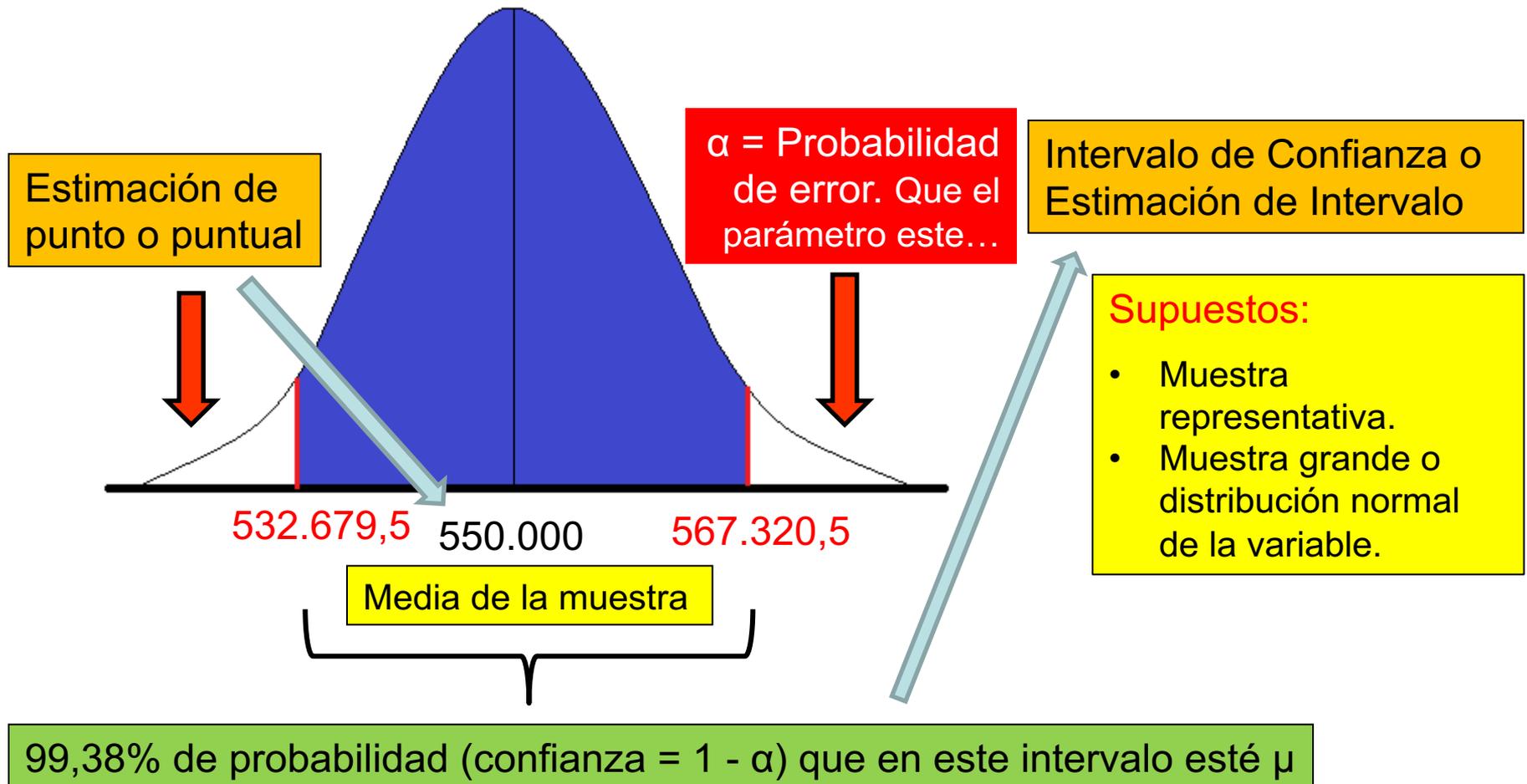
- **El problema** es que lo que yo conozco es **sólo mi única muestra...** y no se **nada del universo**.
- Pero... Si mi muestra es grande, puede ser considerada una del infinito número de muestras posibles de extraer del universo.
- Y ello implica que (por la propiedad de equivalencia de las distancias euclidianas)...



- Existe una probabilidad de **68,2%** que el promedio del universo este a $1\sigma/\sqrt{n}$ del promedio de mi muestra.
- Existe una probabilidad del **95,4%** que el promedio del universo este a $2\sigma/\sqrt{n}$ del promedio de mi muestra.
- Existe una probabilidad del **99,6%** que el promedio del universo este a $3\sigma/\sqrt{n}$ del promedio de mi muestra.

No conozco σ , pero puedo sustituirla por S

- Ello implica que si extraigo una muestra de 300 casos y obtengo un promedio de ingresos de 550.000 pesos y una desviación estándar de 100.000 pesos.
 - Existirá una probabilidad de 99,6% de que el promedio del universo esté entre 532.679,5 pesos y 567.320,5 pesos.



Cálculo de Intervalos de confianza

- Repetimos los Supuestos:
 - Muestra representativa = equiprobabilidad.
 - Muestra grande o variable con distribución normal.

- Formula para promedios:

$$\bar{x} \pm Z_{\alpha/2} SE$$

Siendo:

- $Z_{\alpha/2}$ = Puntuación Z para la Prob deseada.
- SE = **Error Estándar** = $S_{\bar{x}} = s/\sqrt{n}$

- Fórmula para proporciones:

$$\bar{p} \pm Z_{\alpha/2} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Siendo:

- $Z_{\alpha/2}$ = Puntuación Z para la Prob deseada.
- SE = Error Estándar = σ^2/n
- P = Proporción de casos positivos.
- Q = Proporción de casos negativos.

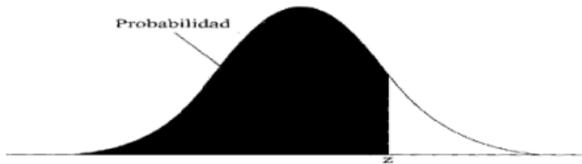
En sociología se usa una confianza de 95%, por lo que $Z_{\alpha/2} = 1,96 \approx 2$.

Ejemplos:

- En un estudio ($n=400$) se encontró que la media de amigos cercanos que tienen las personas es de 4,5, con una desviación de 1,2.
 - ¿Cuál sería la mejor estimación puntual que podríamos hacer respecto del número de amigos del universo?
 - Si quisiéramos tener un 95% de certeza ($1-\alpha=.95$) sobre nuestra afirmación, ¿entre que valores apostaríamos que está el valor del Universo?
 - ¿Y si quisiéramos una certeza del 99%?
- En una encuesta ($n=850$) se encontró que actualmente el 46% de las personas señala que sería partidaria de rechazar la nueva propuesta de Constitución.
 - ¿Se podría afirmar que la encuesta no puede descartar que gane el rechazo con un 95% de confianza ($1-\alpha=.95$)?
 - ¿Y que pasaría con un 99% de confianza ($1-\alpha=.99$)?

Cálculo de probabilidades bajo la curva normal

- Tabla Z



z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
+0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
+0.1	.53983	.54380	.54776	.55172	.55567	.55966	.56360	.56749	.57142	.57535
+0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
+0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
+0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
+0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
+0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
+0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
+0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
+0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
+1	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
+1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
+1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
+1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91308	.91466	.91621	.91774
+1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
+1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
+1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
+1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
+1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
+1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
+2	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
+2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
+2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
+2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
+2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
+2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
+2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
+2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
+2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
+2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861

Aplicación a las pruebas de hipótesis

- Definiciones:

- Hipótesis nula (H_0):

- Hipótesis que usualmente se quiere descartar.
 - Hipótesis que señala la no asociación entre variables.
 - Hipótesis que señala cuál podría ser el valor del parámetro.

- Hipótesis alternativa (H_1):

- Es la que complementa todo lo que no sea la hipótesis nula.

- Conceptualmente:

- H_0 : El rechazo obtendrá una mayoría de votos en Diciembre.
 - H_1 : El rechazo empatará u obtendrá menos votos en Diciembre.

- Matemáticamente:

- $H_0: Q > 0.50$
 - $H_1: Q \leq 0.50$

Prueba de hipótesis de una variable

- Son pruebas donde se pone a prueba el valor propuesto (hipotetizado) para un parámetro.
 - Ejemplo. Hipótesis: El ingreso promedio de los chilenos es de 600.000 pesos.
 - $H_0: \mu = 600.000$
 - $H_1: \mu \neq 600.000 \rightarrow \mu > 600.000$ o $\mu < 600.000$
 - Para resolverlo, obtenemos una muestra representativa de 700 casos, encontrando un promedio de sueldos de 685.000 pesos y una $s=135.000$ pesos.
 - ¿Se puede rechazar la H_0 con un 95% de confianza (es decir con una proporción de error de $\alpha=0.05$).
 - Solución: construir un **intervalo de confianza** alrededor de la media muestral.

¿Nos podemos equivocar?

- **Si rechazamos la H_0 :**
 - Tenemos una probabilidad de equivocarnos igual a $1 - \alpha$, lo que se llama:
 - Error Tipo 1 o error α .
 - Su inverso es la confianza.
- **Si NO rechazamos la H_0 :**
 - Tenemos una probabilidad de equivocarnos igual a $1 - \alpha$, lo que se llama:
 - Error Tipo 2 o error β .
 - Su inverso es la Potencia.
- **Siempre podemos estar equivocados!!!**
 - De hecho... si hacemos muchas pruebas de hipótesis, **un cierto % de ellas estarán erradas.**

Para finalizar...

- El concepto de **valor p**.
 - Es el **α real de la hipótesis que se ha puesto a prueba**.
 - Implica (a modo poco ortodoxo) la probabilidad de error real que se comete al rechazar una hipótesis nula.
 - No se obtiene de cálculos a mano, sino que lo calculan los softwares.
 - Y lo que uno hace es **ver si es menor que el α teórico que uno emplea**.
 - Usualmente el alfa teórico que se usa en sociología es:
 - **$\alpha = 5\%$** (confianza 95%). 5 probabilidades en 100 de equivocarnos, es decir, cometer un error tipo 1.
 - **$\alpha = 1\%$** (confianza 95%). 1 probabilidad en 100 de equivocarnos, es decir, cometer un error tipo 1.
 - **$\alpha = 0,1\%$** (confianza 99,9%) 1 probabilidad en 100 de equivocarnos, es decir, cometer un error tipo 1.

- Vamos a ver **como se hace todo esto en SPSS:**
 - Gráficos (caja, barras, líneas, torta, entre otros).
 - Asimetría y curtosis.
 - Selección de casos.
 - Intervalos de confianza
 - Pruebas de hipótesis.
- Y luego...

FIN