Inferencia Estadística

Autor: Luis Fernando Torres Quitora
Fecha de actualización: Enero del 2021

TABLA DE CONTENIDO

INTRODUCCIÓN

Este curso va a dirigido a profesionales de los pregrados de la UMB Virtual que requieren el empleo de herramientas estadísticas básicas, que le permitirán adquirir habilidades analíticas para abordar las diferentes situaciones que se presenten en su entorno laboral. La estadística tiene múltiples aplicaciones en las áreas de la salud, la administración, la logística, la ingeniería, entre otras, donde se hace indispensable la aplicación de procedimientos estadísticos que le posibiliten la formulación de predicciones a partir de la recolección, organización y análisis de datos. Con los anteriores recursos se "optimiza" la toma de decisiones en los procesos de planeación y ejecución de proyectos a nivel empresarial, investigativo y académico.

La estadística inferencial abarca un grupo de conceptos, procedimientos y metodologías que permiten predecir información sobre el comportamiento de diferentes variables en una población de estudio. Los resultados obtenidos al aplicar estas técnicas de recolección y análisis de datos cuentan con sus respectivos grados de confianza. En la toma de decisiones, el profesional requiere de herramientas estadísticas y probabilísticas que le permitan dar un adecuado tratamiento a los datos relacionados con problemáticas específicas de su campo de trabajo, así, esta asignatura brinda al estudioso los conceptos fundamentales para el análisis de la información y la implementación de planes de acción teniendo en cuenta las necesidades propias de su campo de acción

1. ESTADÍSTICA INFERENCIAL

1.1. Distribuciones de probabilidad variables aleatorias continuas

Las probabilidades es una ciencia que se determina la posibilidad de que ocurran los sucesos aleatorios utilizando una medida designada cómo "probabilidad". La importancia de esta ciencia dentro de la estadística y áreas transversales del conocimiento radica en que proporciona los recursos para modelar y predecir situaciones relacionadas con la incertidumbre, por ejemplo "la probabilidad de ganar las elecciones de un determinado candidato a la presidencia", "la probabilidad de ganar la copa del mundo por el equipo de Colombia" o "la probabilidad de que este año se arregle la huelga de una empresa estatal".

1.2. Distribución de probabilidad

La probabilidad son sucesos o eventos que permiten preceder situaciones relacionadas con alguna situación de la sociedad. Esto indica que se puede utilizar la ley Laplace. Que consiste en la proporción del número de casos favorables al número de casos posibles, siempre que todos los resultados sean igualmente probables.

$$P(E) = \frac{Casos\ favorables}{Casos\ posibles}$$

1.2.1. Concepto de variable aleatoria.

Una variable X es una variable aleatoria, si los valores que toma corresponden a los distintos resultados posibles de un experimento, y por ello el hecho de que tome un valor particular de un evento aleatorio.

Por ejemplo, considérese el muestreo de 20 deportistas a los que se le pregunta su preferencia por el refresco A o B. el número de deportistas que prefiera el producto A puede considerarse como un variable aleatoria X, puede tomar valores 0, 1, 2, ……20. Cada uno de estos valores corresponde a un resultado posible del experimento consistente en la extracción de una muestra de 20 deportistas y el consiguiente registro del número de ellos que prefiere el producto A.

Ejemplo: Se puede considerar el evento de lanzar un dado. Entonces la variable aleatoria $X = 1,2,3,4,5,6$. La probabilidades asignadas a cada valor de X son:

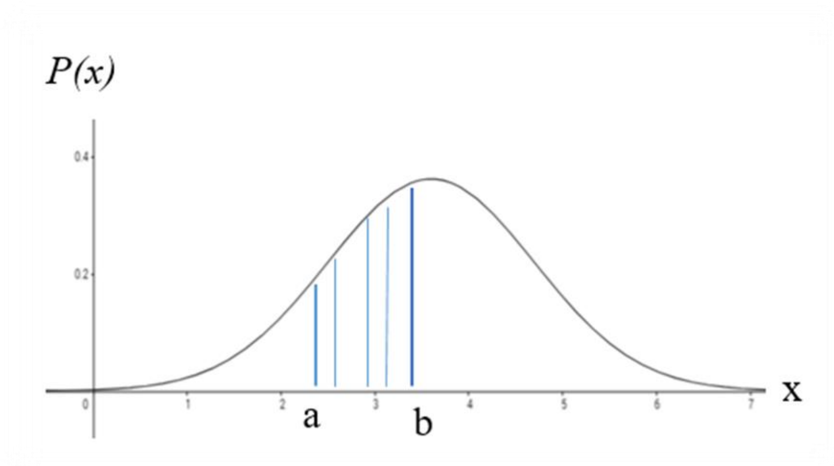
$P(1) = 1/6$; $P(2) = 1/6$; $P(3) = 1/6$; $P(4) = 1/6$; $P(5) = 1/6$; $P(6) = 1/6$

Estas probabilidades forman una distribución de probabilidad, puesto que $P(1)\ + P(2) + P(3) + P(4) + P(5) + P(6)$

Con las ideas anteriores se extienden a variables X que pueden tomar un conjunto continuo de valores, el polígono de frecuencias relativas de una muestra se convierte, en el caso teórico o límite de una población, en una curva continua (como se nuestra en la figura 1) de ecuación $y =\ P(X)$. el área total bajo la curva y sobre el eje X es 1, y el área entre $X =\ a\ y\ X =\ b$, da la probabilidad de que X esté entre a y b, que se denota por $P(a < X < b)$

Entonces si X es una variable aleatoria continua las principales distribuciones son: uniforme, Normal, T-student, Chi-cuadrado.

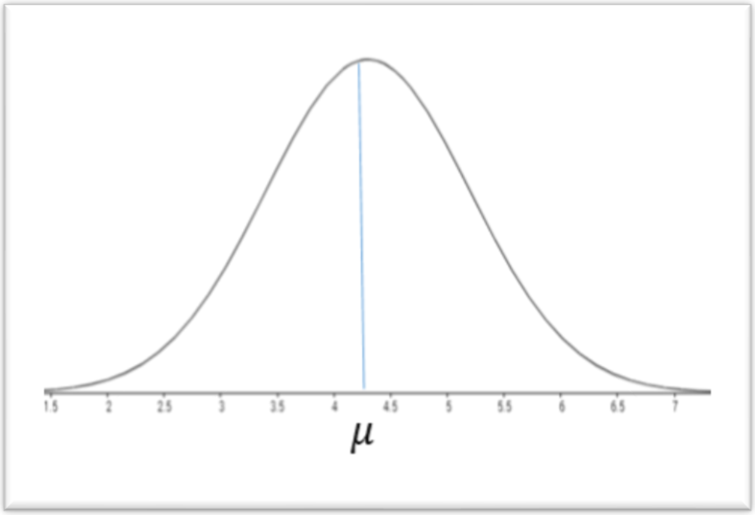
Figura 1: Una distribución de probabilidad



Fuente, propia del autor

1.2.1 Normal
Muchas distribuciones de mediciones que se hacen tanto en las ciencias naturales como el programa de Psicologia tienen a tener un polígono de frecuencias como una forma que se asemeja al corte transversal de una campana. Ver figura 2.

Figura 2: Representación una campana de Gauss



Fuente: propia del autor

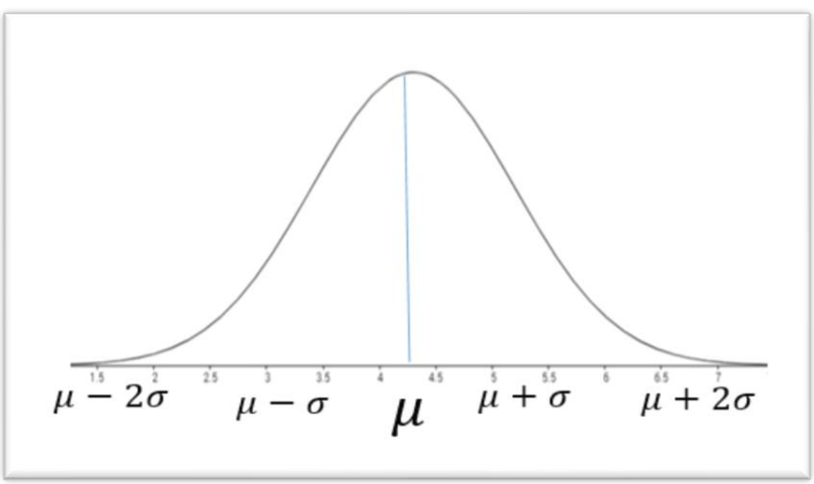
Esta distribución se observa más cuando el número de observaciones es grande y cuando en muchos casos las investigaciones se realizan con muestras de poblaciones grandes; en la mayoría de los casos las distribuciones tienden aproximarse a la curva en forma de campana ya mencionada.

El modelo de distribución de probabilidad para variables continuas más importante es esta cuya función de densidad es:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(X - \mu)^{2/\sigma 2}}$$

Donde μ es la media aritmética, σ es la desviación estándar cuya grafica se muestra en la figura 3.

Figura 3: Distribución normal



Fuente, propia del autor

Una distribución normal juego un papel fundamental en la estadística. Sin exagerar se puede decir es la más importante. Porque:

- Muchas variables que aparece en experimentos prácticos están distribuidas normalmente.
- Otras variables están distribuidas aproximadamente normal.
- Ciertas distribuciones más complicadas se pueden aproximar a una normal.
- Algunas variables que son básicas para justificar pruebas estadísticas están distribuidas normalmente.

El área total limitada por la curva y el eje X es uno (1); de aquí que el área bajo la curva entre dos puntos de X , $a < b$, representa la probabilidad de que X se encuentra entre a y b, se denota $P(a < X < b)$

En la tabla de área bajo la curva normal tipificada Z (anexa al final del contenido), están las áreas o probabilidades correspondientes a las variables tipificadas, entonces una variable tipificada está definida así:

$$Z = \frac{X - \mu}{\sigma}$$

Para una distribución normal se utiliza la tabla Z, con la finalidad de establecer las probabilidades para un suceso o evento determinado ver tabla 1.

Tabla 1: Tabla Z, Distribución normal.

Z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
-3	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,001	0,001
-2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
-2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,002	0,0019
-2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,003	0,0029	0,0028	0,0027	0,0026
-2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,004	0,0039	0,0038	0,0037	0,0036
-2,5	0,0062	0,006	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
-2,4	0,0082	0,008	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
-2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
-2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,011
-2,1	0,0179	0,0174	0,017	0,0166	0,0162	0,0158	0,0154	0,015	0,0146	0,0143
-2	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
-1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,025	0,0244	0,0239	0,0233
-1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
-1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1,5	0,0668	0,0655	0,0643	0,063	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
-1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
-1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,102	0,1003	0,0985
-1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,123	0,121	0,119	0,117
-1	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
-0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,166	0,1635	0,1611

-0,8	0,2119	0,209	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
-0,7	0,242	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
-0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
-0,5	0,3085	0,305	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,281	0,2776
-0,4	0,3446	0,3409	0,3372	0,3336	0,33	0,3264	0,3228	0,3192	0,3156	0,3121
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,352	0,3483
-0,2	0,4207	0,4168	0,4129	0,409	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0	0,5	0,496	0,492	0,488	0,484	0,4801	0,4761	0,4721	0,4681	0,4641
0	0,5	0,504	0,508	0,512	0,516	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,591	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,648	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,67	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,695	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,719	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,758	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,791	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,834	0,8365	0,8389
1	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,877	0,879	0,881	0,883
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,898	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,937	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,975	0,9756	0,9761	0,9767
2	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,983	0,9834	0,9838	0,9842	0,9846	0,985	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,989
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,992	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,994	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,996	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,997	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,998	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

Fuente: propia del autor

Ejemplo: los resultados de un examen de admisión en un colegio tienen una distribución normal con media 75 y desviación estándar 10. ¿Qué fracción de resultado quedan? Cuando ¿superior a 90? ¿inferior a 60?, ¿entre 70 y 90?

Solución.

Para aplicar la distribución normal a datos discretos es necesario tratar los datos como si fuera continuos, así, un puntaje superior a 90 considera 90.5

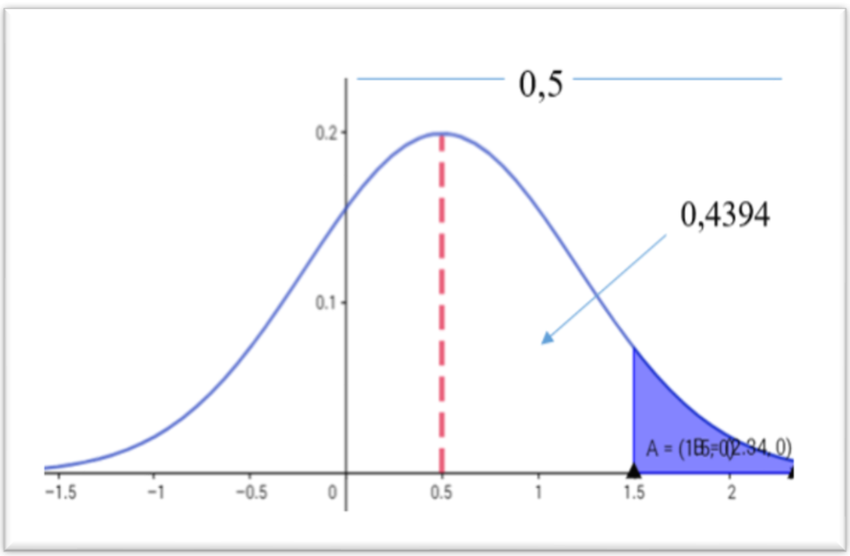
¿superior a 90?

Es necesario tipificar la variable $X = 90.5$

$$Z = \frac{X-\mu}{\sigma} = Z = \frac{90.5-75}{\sigma} = Z = \frac{15.5}{10} = 1.55$$

Se utiliza la tabla Z, ubicando 1.55, por lo tanto, equivale a 0.4394, en la figura 4, se muestra el área sombreada.

Figura 4: Representación de una distribución normal



Fuente, propia del autor

La proporción o fracción pedida corresponde al área sombreada y es igual a: 0.5 -0.4394= 0.0606= 6%. 0.5 es el valor correspondiente a la mitad de la curva; 0.4394 es el valor tomado de la tabla, área bajo la curva. Lo que significa que el 6% de los aspirantes obtuvieron una nota superior al 90 en el examen de admisión.

En la siguiente pregunta inferior a 60, entonces obtenemos

$$Z = \frac{X-\mu}{\sigma} = Z = \frac{59.5-75}{10} = Z = \frac{-15,5}{10} = -1.55$$

Se emplea la tabla Z, para ubicar el valor 1.55, cumple el mismo valor 0.4394, ver figura 5.

Lo mismo se establece con la pregunta siguiente

$$Z = \frac{X - \mu}{\sigma} = Z = \frac{69.5 - 75}{10} = Z = \frac{-5,5}{10} = -0.55$$

El área que hay de 0 a -0.55 es igual al área que hay entre 0 a 0.55, puesta que la gráfica es simétrica sobre el eje y. en la tabla anterior corresponde a 0.2088.ver figura 5.

Figura 5: Representación de una distribución normal

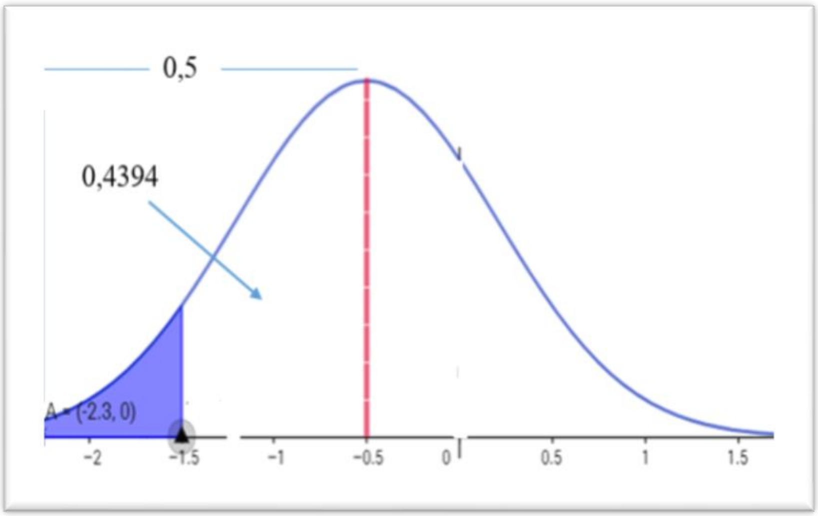


Figura: propia del autor

Superior a 90, se cumple la condición de

$$Z = \frac{X - \mu}{\sigma} = Z = \frac{90.5 - 75}{10} = Z = \frac{15.5}{10} = 1.55$$, en la tabla Z, equivale a 0.4394.

La proporción o fracción pedida corresponde al área sombreada y es igual a 0.2088 +0.438494= 0.6482 = 64.82%. lo se quiere decir es que el 64.82% de los aspirantes obtuvieron una nota entre 70 y 90 en el examen de admisión.

Ejemplo: Las ventas diarias de una empresa de gaseosas tienen una media igual a 5.5 de pesos con una desviación estándar de 2 millones de pesos. ¿Cuál es la probabilidad de que en un día se vendan menos de 5 millones?

$x: media\ muestral;\ \mu: Media\ poblacional;\ \sigma: desviacion\ estandar\ poblacional;$

$x = 5{,}5\ millones\ ;\ \mu = 5\ millones\ ;\ \sigma = 2\ millones$

Realizamos la variable Z:

$$Z = \frac{X - \mu}{\sigma} = Z = \frac{5.5 - 5}{2} = Z = \frac{0.5}{2} = 0.25$$

Buscamos en la tabla p(z<0,25). Ver tabla 2.

Tabla 2. Distribución normal

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	Z
0.0	0.0000000	0.0039894	0.0079783	0.0119665	0.0159534	0.0199388	0.0239222	0.0279032	0.0318814	0.0358564	0.0
0.1	0.0398278	0.0437953	0.0477584	0.0517168	0.0556700	0.0596177	0.0635595	0.0674949	0.0714237	0.0753454	0.1
0.2	0.0792597	0.0831662	0.0870644	0.0909541	0.0948349	0.0987063	0.1025681	0.1064199	0.1102612	0.1140919	0.2
0.3	0.1179114	0.1217195	0.1255158	0.1293000	0.1330717	0.1368307	0.1405764	0.1443088	0.1480273	0.1517317	0.3
0.4	0.1554217	0.1590970	0.1627573	0.1664022	0.1700314	0.1736448	0.1772419	0.1808225	0.1843863	0.1879331	0.4
0.5	0.1914625	0.1949743	0.1984682	0.2019440	0.2054015	0.2088403	0.2122603	0.2156612	0.2190427	0.2224047	0.5
0.6	0.2257469	0.2290691	0.2323711	0.2356527	0.2389137	0.2421539	0.2453731	0.2485711	0.2517478	0.2549029	0.6
0.7	0.2580363	0.2611479	0.2642375	0.2673049	0.2703500	0.2733726	0.2763727	0.2793501	0.2823046	0.2852361	0.7
0.8	0.2881446	0.2910299	0.2938919	0.2967306	0.2995458	0.3023375	0.3051055	0.3078498	0.3105703	0.3132671	0.8
0.9	0.3159399	0.3185887	0.3212136	0.3238145	0.3263912	0.3289439	0.3314724	0.3339768	0.3364569	0.3389129	0.9
1.0	0.3413447	0.3437524	0.3461358	0.3484950	0.3508300	0.3531409	0.3554277	0.3576903	0.3599289	0.3621434	1.0

Fuente: modulo 2 Estadística y Probabilidad

1.2.2. Gamma

Varias densidades de probabilidad importantes, cuyas aplicaciones se explican más adelante, son casos especialmente de la distribución gamma. La distribución gamma es un modelo de la exponencial ya que en ocasiones se emplea para determinar tiempos en un determinando suceso. En conclusión, la distribución gamma es una ampliación de la distribución exponencial.

La función de densidad parte de:

$$f(x) = \begin{cases} \frac{k^{\lambda}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, para\ x > 0. \alpha > 0, \beta > 0 \\ 0, \quad es\ otra\ parte \end{cases}$$

Es importante tener en cuenta que la distribución toma valores solamente positivos.

Donde $T(\alpha)$ es el valor de la función de gamma. definida por la integrar, pero se parte de la integral.

$\int_{-\infty}^{\infty} f(x) d(x) = T(\alpha) = \int_0^{\infty} \frac{(\lambda x^{\alpha-1}}{\Gamma(\alpha)} \lambda e^{-\lambda x} dx$

Su integral es $\int_0^{\infty} \frac{(\lambda x^{\alpha-1}}{\Gamma(\alpha)} \lambda e^{-\lambda x} dx = \frac{1}{\Gamma(\alpha)} \lambda^{\alpha} \int_0^{\infty} x^{\alpha-1} e^{-\lambda x}\, dx$. Se hace cambio de variable para $y = \lambda x$, sustituimos $\frac{1}{\Gamma(\alpha)} \cancel{\lambda^{\alpha}} \int_0^{\infty} \cancel{\left(\frac{y}{\lambda}\right)}^{\alpha-1} e^{\cancel{y\frac{1}{\lambda}}} dy$)= $\frac{1}{\Gamma(\alpha)} \int_0^{\infty} y^{\alpha-1} e^{y} dy$, de esta forma es la distribución gamma por lo tanto es 1.

Para toda $\alpha > 1$, y por ende $T(\alpha) = (\alpha - 1)!$, cuando α es un numero entero positivo. En la figura 8, se representa una distribución gamma.

Ejemplo: El tiempo en horas que semanalmente requiere una máquina para mantenimiento, es una variable aleatoria de distribución gamma con parámetros de $\alpha = 3$ y $\lambda = 2$, ¿cuál es la probabilidad que en alguna semana que tiempo de mantenimiento de la maquina sea mayor a 8 horas?.

Solución.

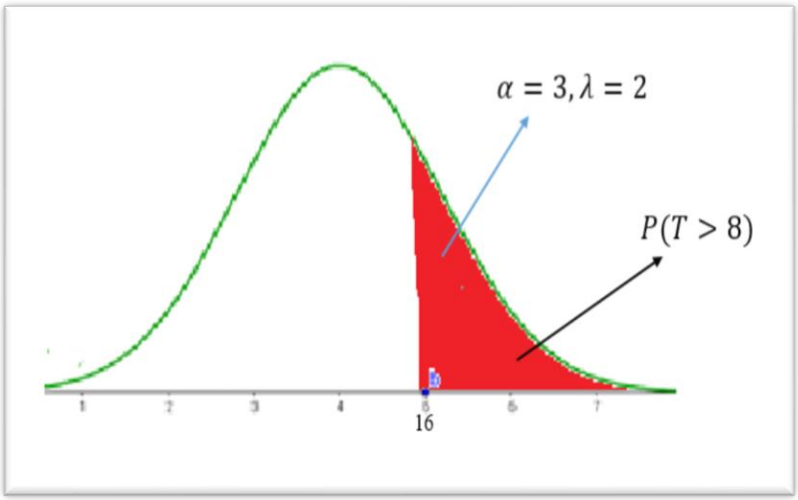
Su función de densidad de probabilidad es

$f(x) = \frac{1}{\lambda^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{\frac{-x}{\lambda}} = \frac{1}{2^{3}\Gamma(3)} x^{3-1} e^{\frac{-x}{2}} = \frac{1}{16(3-1)!} x^{2} e^{\frac{-x}{2}} = \frac{1}{16} x^{2} e^{\frac{-x}{2}} =$

Se emplea la distribución binomial, que indica de la probabilidad sea mayor a 8 horas

$P(T > 8) = e^{-\lambda t}$, o $P(T < 8) = \left(1 - e^{-\lambda t}\right) = \frac{1}{16(2!)} x^{2} e^{\frac{-x}{2}}$, en la figura 6 se muestra la distribución.

Figura 6: Distribución gamma



Fuente, propia del autor

En la distribución gamma se utiliza la fórmula de integración

$\frac{1}{\Gamma(\alpha)} \int_{0}^{\infty} y^{\alpha-1} e^{y} dy = P(T > 8) = 1 - \frac{1}{16} \int_{0}^{8} x^{2} e^{\frac{-x}{2}} dx.$

Para resolver esta función se utiliza la integración por partes, entonces sustituyendo los resultados obtenidos se tiene que:

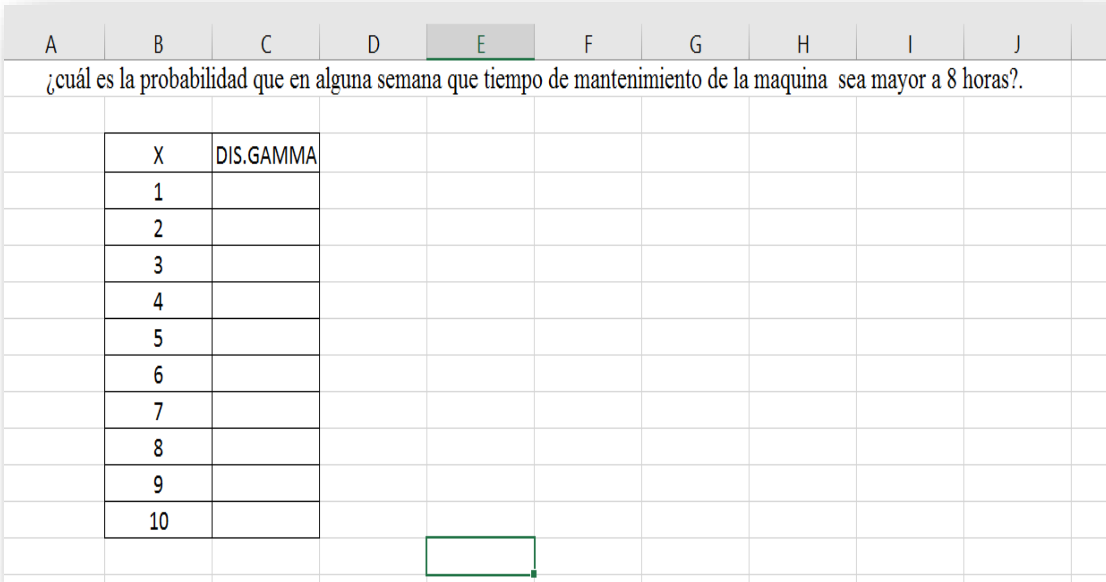
$P(T > 8) = 1 - \left[-2x^{2} e^{\frac{-x}{2}} + 4\left(2xe^{\frac{-x}{2}}\right) + 2\left(2e^{\frac{-x}{2}}\right)\right]_{0}^{8} = \mathbf{0.2381}$

Significa que 23.8% de probabilidad que en alguna semana el tiempo de mantenimiento sea mayor de 8 horas.

Si se emplea la herramienta Excel, para representar una distribución gamma, entonces

1. Se registran los datos en una hoja de Excel, para determinar la distribución gamma. ver tabla 3.

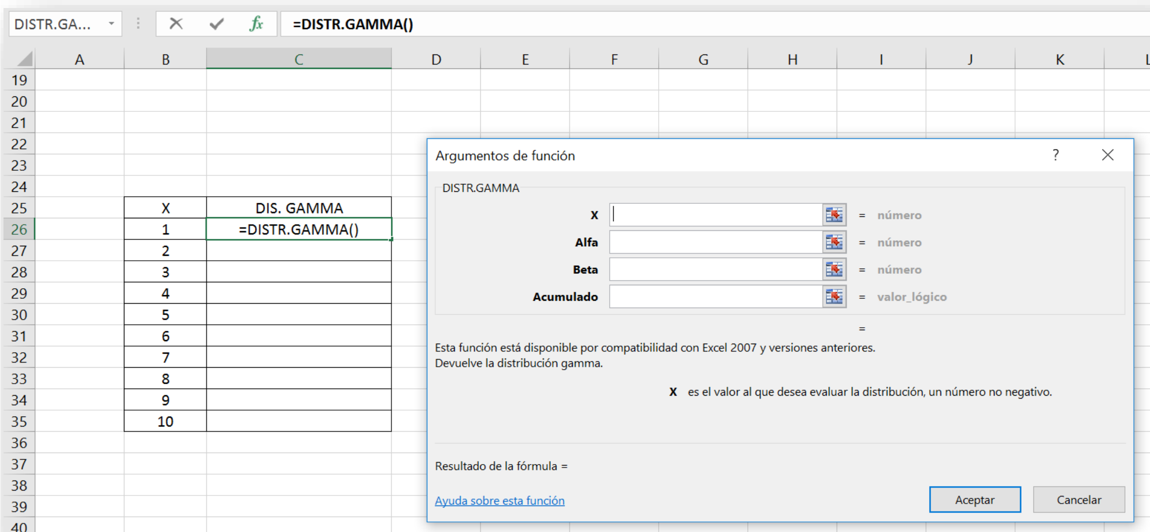
Tabla 3. Distribución gamma.



Fuente: propia del autor

2.Se da clic en formulas DISTR. GAMM, para ingresar los valores como lo muestra la tabla 4

Tabla 4. Distribución gamma



Fuente: propia del autor.

3.Se seleccionan los datos para el cálculo de la Distribución Gamma, teniendo en cuenta que se le resta la unidad, ver tabla 5

Tabla 5. Distribución gamma

C23 =1-DISTR.GAMMA(B23:B32;3;2;VERDADERO)

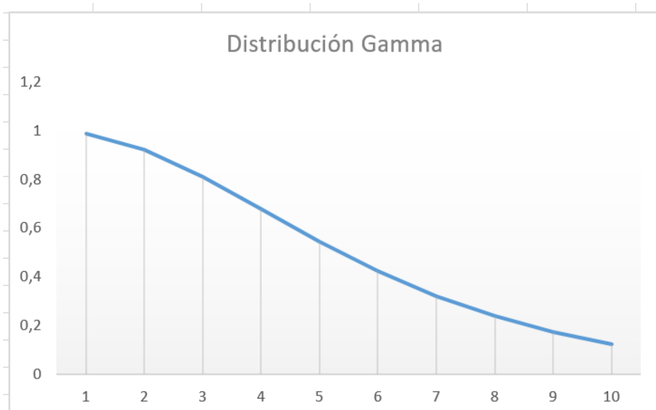
¿cuál es la probabilidad que en alguna semana que tiempo de mantenimiento de la maquina sea mayor a 8 horas?.

X	DIS. GAMMA
1	0,985612322
2	0,919698603
3	0,808846831
4	0,676676416
5	0,543813116
6	0,423190081
7	0,320847199
8	0,238103306
9	0,173578071
10	0,124652019

Fuente: propia del autor

En la figura 7 se puede observar la gráfica de la distribución gamma y la relación con la distribución exponencial.

Figura 7: Distribución Gamma



Fuente: propia del autor

1.2.3. Exponencial

La distribución exponencial es el equivalente continuo de la distribución geométrica discreta.

La variable aleatoria exponencial t, tiene como función de densidad

Donde λ es el número medio de llegadas independientes por unidad de tiempo, t es el número de unidades de tiempo hasta la siguiente llegada y $e = 2{,}71828$.

Probabilidad acumulada

$$P(T \leq t) = \left(1 - e^{-\lambda t}\right)$$

$$P(T > t) = 1 - P(T \leq t) = e^{-\lambda t}$$

Probabilidad de un intervalo

$$P(t_a \leq T \leq t_b) = e^{-\lambda t_a} - e^{-\lambda t_b}$$

Es importante tener en cuenta que para identificar una distribución exponencial es revisar tiempo en función de una unidad. O simplemente un tiempo medio de espera.

Ejemplo: el tiempo que se dedica a atender al público en el mostrador de información de una biblioteca puede representarse por medio de una distribución exponencial que tiene tiempo medio de atención de cinco minutos.

¿Cuál es la probabilidad de que el tiempo de atención al público sea de más de diez minutos?

¿Cuál es la probabilidad de que el tiempo de atención al público sea de menos de 8 minutos?

¿Cuál es la probabilidad de que el tiempo de atención al público sea de 4 a 7 minutos?

1. Para resolver la primera pregunta es importante tener en cuenta lamdar λ. Entonces se puede indicar que

$$\lambda = \frac{1}{tiempo\ promedio} = \frac{1}{5} = 0.2$$

Teniendo el promedio o λ, se parte de observar que el tiempo sea mayor $(T > t), o\ (T > 10)$, empleamos la función

$P(T > t) = 1 - P(T \leq t) = e^{-\lambda t}$

$P(T > 10)$= $e^{-\lambda t}$, entonces $2.71828^{-0.2(10)}$= 2.71828^{-2}= $\frac{1}{2.71828^2}$ = $\frac{1}{7.389046}$ = 0.13

Entonces $P(T > 10)$=0.13

Está indicando que el 13% de probabilidad que el tiempo de atención al público sea más de 10 minutos.

¿Cuál es la probabilidad de que el tiempo de atención al público sea de menos de 8 minutos?

Según la pregunta la probabilidad cuando es menos de 8 minutos, entonces se cumple la función. $P(T \leq t) = \left(1 - e^{-\lambda t}\right)$

$P(T < 8)$= $\left(1 - e^{-\lambda t}\right)$= $2.71828^{-0.2(8)}$= $2.71828^{-1.6}$= $\frac{1}{2.71828^{1.6}}$= $\frac{1}{4.953027}$ = 0.20, por lo tanto se resta la unidad $1 - 0.20 = 0.80$

De esta manera significa que el 80% de probabilidad que el tiempo de atención sea menor de 8 minutos.

En la pregunta tres, en donde se debe determinar cuál es la probabilidad del tiempo de atención sea entre 4 a 7 minutos.

Se utiliza la función de intervalo $P(t_a \leq T \leq t_b) = e^{-\lambda t_a} - e^{-\lambda t_b}$

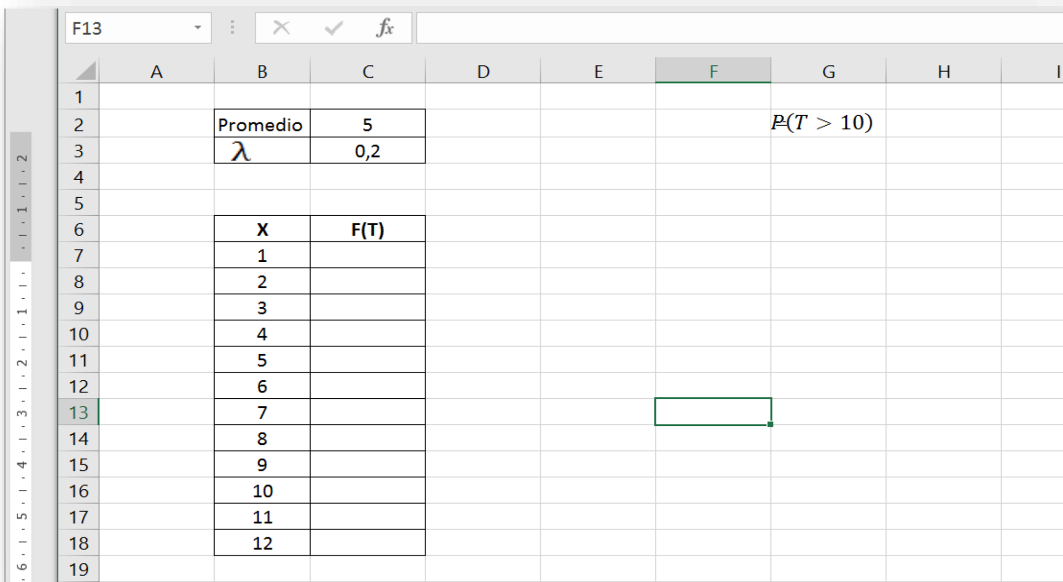
$P(t_a \leq T \leq t_b)$= $2.71828^{-0.2(4)}$- $2.71828^{-0.2(7)}$= $2.71828^{-0.8}$- $2.71828^{-1.4}$=$\frac{1}{2.71828^{0.8}}$ - $\frac{1}{2.71828^{1.4}}$= $\frac{1}{2.22539}$- $\frac{1}{4.0551961}$= $0.449329 - 0.24659 = 0.20$

Por lo tanto, significa que hay un 20% de probabilidad que la atención al público sea entre 4 a 7 minutos.

Si se emplea la herramienta Excel, para representar una distribución exponencial, entonces.

Se ingresa a una hoja de Excel los datos que se desea calcular, ver tabla 6.

Tabla 6. Proceso para el cálculo de la distribución exponencial



Promedio	5
λ	0,2

X	F(T)
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	

Fuente: propia del autor.

2. A continuación se muestra el procedimiento para el cálculo de la distribución exponencial (λ). En la hoja de Excel, ver tabla 7

Tabla 7. Proceso para el cálculo de la distribución exponencial

C7 =C3*2,71828^-(B7/C2)

	A	B	C	D	E	F	G	H
1								
2		Promedio	5				$P(T > 10)$	
3		λ	0,2					
4								
5								
6		X	F(T)					
7		1	0,163746173					
8		2						
9		3						
10		4						
11		5						
12		6						
13		7						
14		8						
15		9						
16		10						
17		11						
18		12						
19								
20								

Fuente: propia del autor

3. Extendemos el cursor para calcular los demás valores, anclando la formula con F4. Ver tabla 8.

C7 =$C$3*2,71828^-(B7/$C$2)

	A	B	C	D	E	F	G
1							
2		Promedio	5				$P(T > 10)$
3		λ	0,2				
4							
5							
6		X	F(T)				
7		1	0,163746173				
8		2	0,134064045				
9		3	0,109762372				
10		4	0,089865841				
11		5	0,073575938				
12		6	0,060238891				
13		7	0,049319439				
14		8	0,040379347				
15		9	0,033059818				
16		10	0,027067093				
17		11	0,022160664				
18		12	0,01814362				
19							

Fuente: propia del autor

4. La pregunta indica que cual es la probabilidad que el tiempo de atención sea mayor a 10 minutos, entonces ingresamos a icono de fórmulas y buscamos la DTRIS.EXP N. ver tabla 9

Tabla 9. Proceso para la distribución exponencial

DISTR.EXP.N =1-DISTR.EXP.N(10;0,2;VERDADERO)

Promedio	5
λ	0,2

$P(T > 10)$

=1-DISTR.EXP

X	F(T)
1	0,163746173
2	0,134064045
3	0,109762372
4	0,089865841
5	0,073575938
6	0,060238891
7	0,049319439
8	0,040379347
9	0,033059818
10	0,027067093
11	0,022160664
12	0,01814362

Argumentos de función

DISTR.EXP.N

X	10	= 10
Lambda	0,2	= 0,2
Acum	VERDADERO	= VERDADERO

= 0,864664717

Devuelve la distribución exponencial.

X es el valor de la función, un número no negativo.

Resultado de la fórmula = 0,135335283

Ayuda sobre esta función

Aceptar Cancelar

Fuente: propia del autor

5. Con la finalidad de analizar la pregunta de cuál es la probabilidad del tiempo de atenciones entre 4 y 7 minutos, ver la tabla 10

Tabla 10. Distribución exponencial

G7 =1-DISTR.EXP.N(F7:F18;0,2;VERDADERO)

Promedio	5
λ	0,2

$P(T > 10)$

X	F(T)
1	0,163746173
2	0,134064045
3	0,109762372
4	0,089865841
5	0,073575938
6	0,060238891
7	0,049319439
8	0,040379347
9	0,033059818
10	0,027067093
11	0,022160664
12	0,01814362

X	P(X)
1	0,81873075
2	0,67032005
3	0,54881164
4	0,44932896
5	0,36787944
6	0,30119421
7	0,24659696
8	0,20189652
9	0,16529889
10	0,13533528
11	0,11080316
12	0,09071795

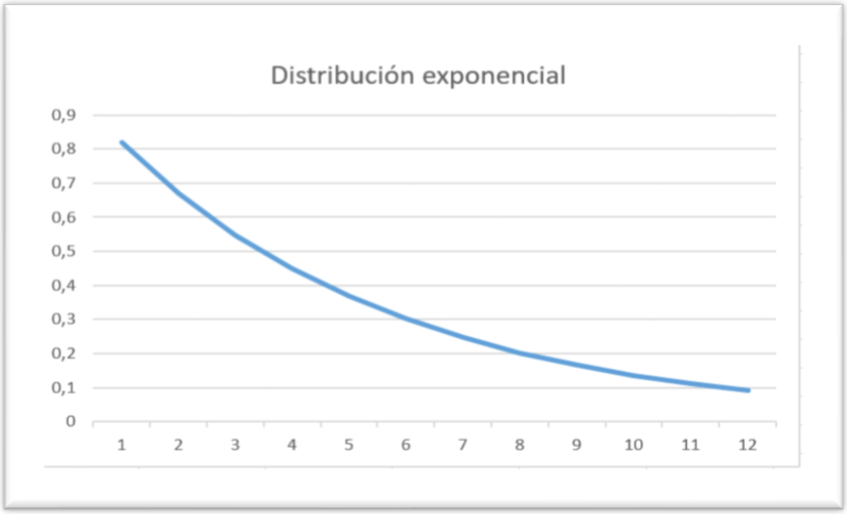
$P(t_a \le T \le t_b) = e^{-\lambda t_a} - e^{-\lambda t_b}$

Diferencia
0,202732

Fuente. Propia del autor

En la figura 8 se puede observar la gráfica de la distribución exponencial.

Figura 8: Distribución exponencial



Fuente: propia del autor

1.3. Distribución t de Student.

En el tema anterior se explica distribuciones normales cuando las estimaciones de una media poblacional son muestras grandes ($n \geq 30$), y la una muestra pequeña ($n \leq 30$),existe una distribución t- estudent, siempre y cuando no se conoce la desviación estándar. El uso de la distribución t para inferencias sobre la media poblacional es adecuado cuando σ es desconocido y la distribución de la media es normal, aunque es importante tener en cuenta que las distribuciones t se emplea tanto para muestras grandes como pequeñas.

De igual manera es importante conocer el concepto de distribución t que aparece del estadístico Williams Gosset en 1908, quien era empleado de la compañía Guinness. Las características de la distribución t- estudent parte en muchas ocasiones cuando no se conoce σ y el número de observaciones es menor a 30, en este se puede utilizar la desviación media como un estimador de la σ.

Por lo tanto, en las distribuciones t los grados de libertad se define como el número de valores que se puede elegir libremente de una población.

La función de densidad es:

$$f(x) = \frac{r(\frac{n-1}{2}}{r\frac{n}{2}\sqrt{n\pi}}(1 + \frac{x^2}{n})^{\frac{n+1}{2}}$$

Ejemplo: El ciclo medio de vida operativa de una muestra" aleatoria" de $n = 10$ focos es $\bar{x}$= 4000 horas, con la desviación estándar muestral $s = 200hr$. Se supone que el ciclo de vida operativa de los focos en general tiene una distribución aproximadamente normal. Estimamos el ciclo medio de vida operativa de la población de focos de la que fue tomada esta muestra, aplicando un intervalo de confianza de 95% en esta forma:

Nota: el concepto de intervalo de confianza se estudia en el siguiente módulo.

Solución

Promedio: $\bar{x}$= 4000

$\boldsymbol{t = t_{n-1} = t = 2.262,\ s_x = \frac{s}{\sqrt{n}} = \frac{200}{\sqrt{10}} = \frac{200}{3.16} = 63.3}$

1.4. Intervalos de confianza para la media con varianza conocida y desconocida.

1.4.1. Estimadores

Es en donde se quiere determinar un valor numérico que nos sirva como aproximación.

Ejemplo: La estimación del puntaje promedio de las pruebas de SABER PRO obtenido por los estudiantes de cierta Universidad en el último año fue de 280, lo que significa que nuestra mejor estimación de la media poblacional es de 280.

La estimación por intervalo es la importante de no conocer un solo valor, sino que se pueda calcular dos valores dentro de los cuales podemos calcular la probabilidad dada.

Para construir los intervalos de estimación se requiere conocer la distribución de la cual proviene la muestra y el margen de error, de igual manera la estimación indica la exactitud de una estima y por lo tanto son estimadores puntuales.

1.5. Teorema de límite central.

Una distribución de medias muéstrales sigue una distribución normal si las muestras se toman de una población normal, pero en muchos casos la población no sigue una distribución normal, entonces debemos recurrir al teorema central del límite. Esta proposición esencial afirma que para cualquier población sea normal o no, la distribución de las medias muéstrales aproximará a la normalidad si el tamaño de la nuestra es grande, $(n \geq 30)$

El teorema central del límite y la hipótesis de una distribución normal de las medias muéstrales sólo se aplica si el muestreo se realiza con remplazamiento o la extracción se hace de una población infinita. Si la población es finita y el número n de miembros de la muestra no es una fracción muy pequeña del número N (supera el 10%) de la población, no se puede asumir que los valores individuales de la muestra se distribuyan independientemente. Si la extracción se hace sin sustitución el proceso de muestreo queda alterado, la probabilidad de elegir un elemento dado en cualquier extracción depende de la selección previa realizada anteriormente.

Ejemplo: Un ingeniero toma una muestra aleatoria de tamaño n= 36 de una población de 500 cuentas por pagar. El valor medio de las cuentas por cobrar para la población es $\mu = \$260$, con la disviacion estándar de la población $\sigma = \$45$. ¿Cuál es la probabilidad de que la media maestral se inferior a $250?.

Solución

$E(x) = \mu = 260$. Posteriormente se establece la función de densidad

$\sigma = \frac{\sigma}{\sqrt{n}} = \sigma = \frac{45}{\sqrt{36}} = \frac{45}{6} = 7.5$

Luego se calcula la estandarización o tipificada

$z = \frac{x-\mu}{\sigma} = z = \frac{250-260}{7.5} = \frac{-10}{7.5} = -1.33$

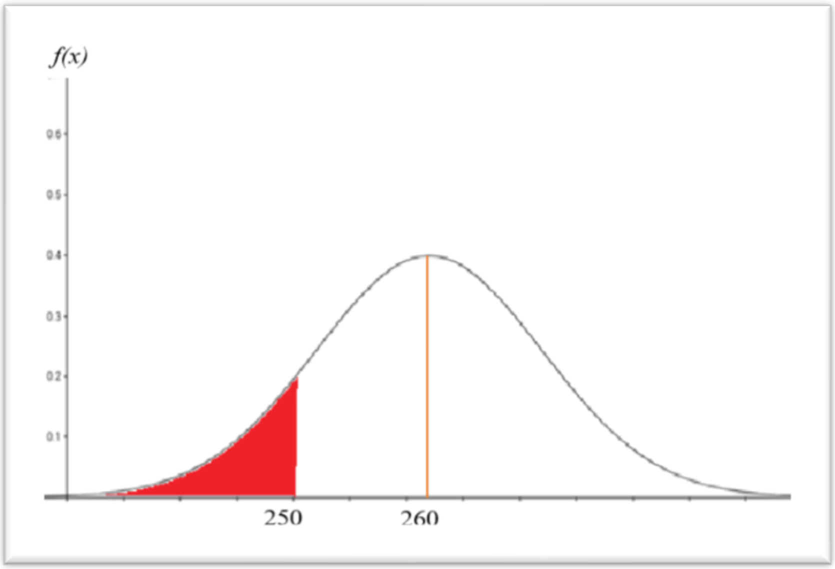
Por lo tanto:

$P(x < 250\mathrm{I}\mu = 260, \sigma = 7.5) = P(Z < -1.33)$

$P(z < -1.33) = 0.5 - p(-1.33 < z < 0) = 0.5 - 0.40 = 0.0918$

En la figura 9, aparece la curva de probabilidad. la distribución de muestreo es descrita por la media y el error estándar.

Figura 9: Curva de probabilidad de distribución de muestreo



Fuente, propia del autor

1.6. Ejercicios de reflexión.

- Un contador de una empresa de cosméticos desea saber cuántas facturas debe tomar como muestra para revisar el promedio de ventas de la empresa. Se sabe que en el mes de octubre de 350 facturas, se encontró un promedio de ventas de 20.000 pesos y una desviación estándar de 2100 pesos. Suponiendo que el margen de error es 500. Bajo estas condiciones. ¿Qué número de la muestra que se debe seleccionar?

- El promedio de la nota de 400 estudiantes de un colegio en la asignatura de física es 70 (en una escala de 0 a 10), con una varianza de 9. Suponiendo que las notas se distribuyen normalmente, ¿Cuántos estudiantes sacaron entre 60 y 75?

- El promedio de la nota de 500 estudiantes de un colegio en la asignatura de Matemáticas es 70 (en una escala de 0 a 10) con una varianza de 9. Suponiendo que las notas se distribuyen normalmente, ¿Cuántos estudiantes sacaron menos de 64?

- Con una prueba de coeficiente intelectual, un estudiante de psicología requiere encontrar la media del coeficiente intelectual en una población, la escala de la prueba es de 0 a 150. Para hacerlo el estudiante reúne los datos de las pruebas de 15 personas solamente y encuentra el promedio es 98.39 con una desviación estándar de 0.535. ¿Es cierta la aseveración de que la muestra de 15 personas viene de una población con promedio menor a 98?

1.7. Intervalos de confianza para:

1.7.1. La Proporción**.**

1.7.2. La diferencia de medias con varianza conocida y desconocida.

A menudo es necesario estimar la diferencia entre dos medias poblacionales, como la diferencia entre los niveles salariales de dos empresas. Como se indicó en la distribución t-student con muestras pequeñas, el estimador puntual insesgado de $(\mu_1 - \mu_2)$ es $(\overline{x_1} - \overline{x_2})$, el intervalo de confianza se elabora en forma similar usado para la estimación la media. El uso de la distribución normal se basa en las mismas condiciones que en el caso de las distribuciones muéstrales.

Para esta condición cuando se desconocen las desviaciones estándar de las poblaciones el error estándar estimado de la diferencia entre medias dado por el uso apropiado de la distribución normal es

$$s_x - x_2 = \sqrt{{s_1}^2 + {s_2}^2}$$

De acuerdo a esta situación se puede indicar también que cuando en un intervalo μ, la varianza σ desconocida, dado que n es pequeño no se puede utilizar la varianza, por lo tanto el nivel de confianza se determina a partir de $1 - \alpha$,

Varianza desconocida

Estimador: $\bar{x}$

Distribución de la muestra: $T = \frac{(\bar{x}-\mu)}{\frac{s}{\sqrt{n}}}$, entonces t_{n-1}

De este modo la cuasi desviación estándar es:

$$s = \sqrt{\frac{1}{n-1}\Sigma(\bar{x}-\mu)^2}$$

Por ende, se puede indicar que la probabilidad se puede marcar cuando

$P[-t_{\alpha/2;n-1} \leq T \leq t_{\alpha/2;n-1}] = 1-\alpha$

$P[-t_{\alpha/2;n-1} \leq \frac{(\bar{x}-\mu)}{\frac{s}{\sqrt{n}}} \leq t_{\alpha/2;n-1}] = 1-\alpha$

$P[x - t_{\alpha/2;n-1}\frac{s}{\sqrt{n}} \leq (\bar{x}-\mu) \leq t_{\alpha/2;n-1}\frac{s}{\sqrt{n}}] = 1-\alpha$

Ejemplo: Se hace un estudio del peso en kilogramos de una fruta producida por una planta. Para ello se escogió una muestra de 16 plantas observando los siguientes pesos, el registro tomado fue el siguiente:

$506, 508, 499, 503, 504, 510, 497, 512, 514, 505, 493, 496, 506, 502, 509, 496$. El peso de la fruta solo se conoce en una varianza normal. Calcular el intervalo de confianza del 90% para el peso medio de esta fruta.

Solución

Se determina la media

$$x = \frac{506 + 508 + 499 + 503 + 504, 510 + 497 + 512 + 514 + 505 + 493 + 496 + 506 + 502 + 509 + 496.}{16}$$

$x = 503.75$

Su desviación es: $s = \frac{(\bar{x}-\mu)}{\frac{s}{\sqrt{n}}} = 6.2022$

Muestra: $n = 16\, plantas$

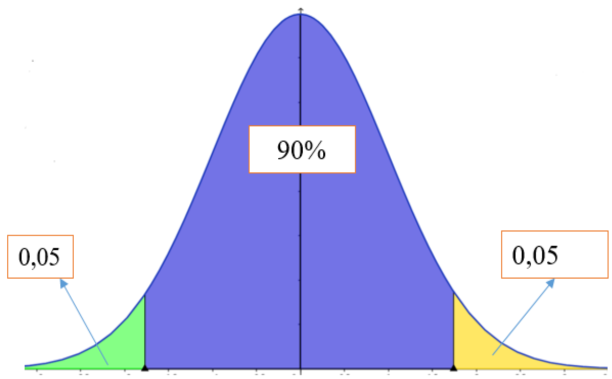
$P[x - t_{\alpha/2;n-1}\frac{s}{\sqrt{n}} \leq (\bar{x}-\mu) \leq t_{\alpha/2;n-1}\frac{s}{\sqrt{n}}] = 1-\alpha$

$P[503.75 - 1.75 * \frac{6.2022}{4}, 503.75 + 1.75\frac{6.2022]}{4}] = P[501.03, 506.46]$

El nivel de confianza parte de la condición $1-\alpha = 1 - 0.90 = 0.10$, entonces $\frac{\alpha}{2} = \frac{0.10}{2} = 0.05$

En una distribución norma se analiza el nivel de confianza del estudio hecho, ver figura 10

Figura 10: Distribución normal de una varianza desconocida

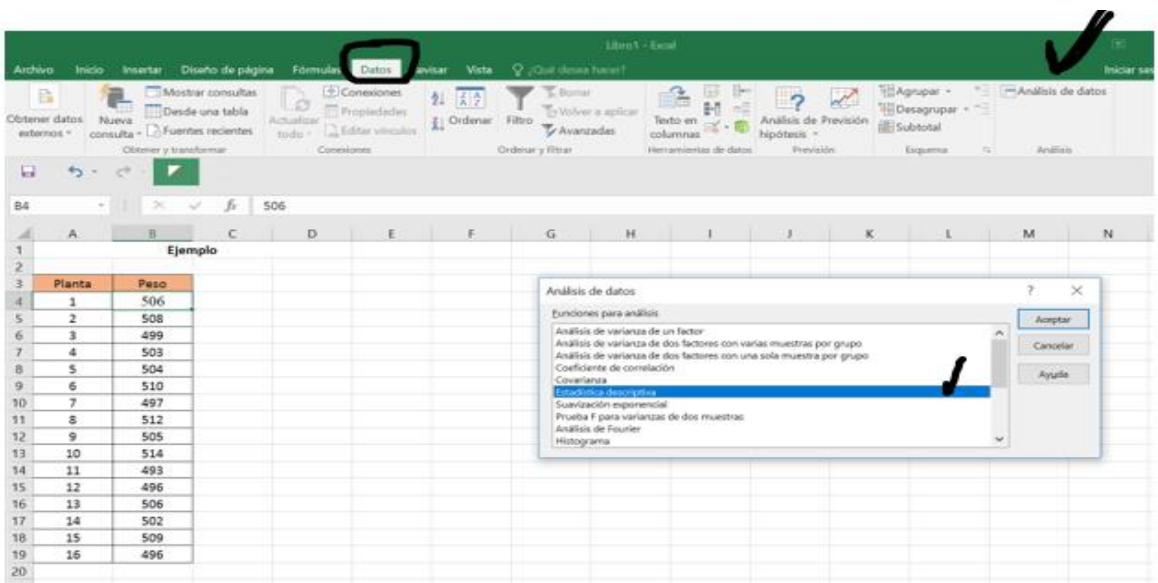


Fuente: propia del autor

Para el mismo ejemplo se emplea la herramienta Excel.

1. Se organiza los datos en una hoja de Excel. Posteriormente se da clic en datos icono (análisis de datos) como muestra la Tabla 11

Tabla11: Intervalo de confianza para varianzas desconocidas



Fuente: propia del autor

2.Aparece un cuadro de datos, especificando en estadística descriptiva, para el cálculo de los intervalos de confianza, en la siguiente tabla 12 se muestra el proceso del nivel de confianza del 95%.

Tabla 12: Intervalo de confianza para varianzas desconocidas

Ejemplo

Planta	Peso
1	506
2	508
3	499
4	503
5	504
6	510
7	497
8	512
9	505
10	514
11	493
12	496
13	506
14	502
15	509
16	496

Peso	
Media	503,75
Error típico	1,550537541
Mediana	504,5
Moda	506
Desviación estándar	6,202150165
Varianza de la muestra	38,46666667
Curtosis	-0,914374741
Coeficiente de asimetría	-0,148500122
Rango	21
Mínimo	493
Máximo	514
Suma	8060
Cuenta	16
Nivel de confianza(95,0%)	0,25

Intervalo de Confianza	
ME	3,039053581
Limite inferior	500,7109464
Limite superior	506,7890536
Z(0.025)	1,96

Fuente: propia del autor

1.7.3. La diferencia de Proporciones.

La diferencia de proporciones permite conocer el estadístico y hacer un contraste entre la hipótesis nula y alternativa, a partir de dos muestras aleatorias independientes, teniendo en cuenta que p es una estimación π es el total de observaciones. La expresión matemática es

$$z = \frac{P_1 - P_2}{\sqrt{\frac{P(1-P)}{n1} + \frac{p(1-p)}{n2}}}$$

De igual manera se considera que para las proporciones y medias se debe aplicar la prueba t para comparar medias poblacionales sin embargo los resultados no son fiables ya que la estimación del error típico que realiza el no coincide con la del estadístico de prueba. En la estadística se puede utilizar diferentes herramientas para el cálculo de las proporciones, en este caso se recomienda software SPSS.

1.8. Distribución Chi-Cuadrado:

Dada la población de valores con distribución normal, se puede demostrar que las distribuciones $x^2(ji\ cuadrada)$, son distribuciones de poblaciones adecuadas para razón de $(n-1)\,{s^2}/{\sigma^2}$, hay una distribución ji cuadrada diferente según el valor del grado de libertad $n-1$, de acuerdo a esta condición podemos decir que:

$${x_{gl}}^2 = \frac{(n-1)s^2}{\sigma^2}$$

De igual manera, se puede indicar que la varianza de la muestra a un estimador insesgado de la varianza poblacional, el valor esperado a largo plazo de la razón anterior es igual a los grados de libertad $n-1$. Sin embargo, en cualquier muestra dada por lo general la varianza muestral no es idéntica en el valor a la varianza poblacional.

La distribución ji cuadrada no son simétricas. En consecuencia un intervalo de confianza de dos extremos para una varianza o desviación estándar implica el uso de valores diferentes de $x^2(ji\ cuadrada)$. La fórmula para la elaboración de un intervalo de confianza para varianza de población es:

$$\frac{(n-1)s^2}{x^2{}_{gl,superior}} \leq \sigma^2 \leq \frac{(n-1)s^2}{x^2{}_{gl,inferior}}$$

Ejemplo: Una empresa en su planta de empleados contiene una tabla de salarial entonces el salario medio semanal de 50 empleados por hora de una empresa de zapatos es $\bar{x}$= \$280000, con una desviación estándar de $s = \$14000$. Se tiene en cuenta que los montos salariales tienen una distribución aproximadamente normal. Calcular el intervalo de confianza del 95% para estimar la desviación estándar de los salarios semanales de la población?

Solución

Para el cálculo del intervalo se emplea

$\frac{(n-1)s^2}{x^2{}_{gl,superior}} \leq \sigma^2 \leq \frac{(n-1)s^2}{x^2{}_{gl,inferior}} = \frac{(50-1)(196000000)}{x^2{}_{49,0.0975}} \leq \sigma^2 \leq \frac{(50-1)(196000000)}{x^2{}_{49,0025}}$

$\frac{(49)(196000000)}{45.72} \leq \sigma^2 \leq \frac{(49)(196000000)}{16.05} = \frac{(9604000000}{45.72} \leq \sigma^2 \leq \frac{9604000000}{16.05} = 21006124.3 \leq \sigma^2 \leq 12211838.01$, se le extrae la raíz

$\sqrt{21006124.3} \leq \sigma^2 \leq \sqrt{12211838.01} = 4583.24 \leq \sigma^2 \leq 3494.54$

Así; se puede afirmar que con un intervalo de confianza al 95% el nivel del salario en la empresa de zapatos está entre $4583.24 \leq \sigma^2 \leq 3494.54$

1.9. Distribución F

Durante el proceso del módulo se ha visto la importancia de algunas aplicaciones como es el caso de conocer la distribución de muestreo de la distancia en medias $(\bar{x}_1 - \bar{x}_2)$ de dos muestras. De esta manera podemos necesitar la distribución de muestreo de la diferencia en varianzas $({S_1}^2 - {S_2}^2)$, es importante tener en cuenta que la distribución F, cumple las mismas condiciones del cociente entre varianzas desconocidas, significa que si el cociente es cercano a 1, pues existe una pequeña diferencia entre la distribución de muestreo o simplemente es la diferencia entre los grados de libertad del numerador y denominador, la expresión matemática es:

$$F = \frac{\frac{{S_1}^2}{{\sigma_1}^2}}{\frac{{S_2}^2}{{\sigma_2}^2}} = \frac{{N_1}^2{S_1}^2(N_1-1){\sigma_1}^2}{{N_2}^2{S_2}^2(N_2-1){\sigma_2}^2}$$

$${S_1}^2 = \frac{N_1{S_1}^2}{N_1 - 1}$$

$$S_2{}^2 = \frac{N_2 S_2{}^2}{N_2 - 1}$$

Para el cálculo de los grados de libertad $(n-1)$ es importante tener en cuenta la tabla de distribución F. a partir de la siguiente condición, el numerador es la columna y el denominador es la fila.

Ejemplo: Dos muestras de tamaño 9 y 12 se han tomado en dos poblaciones normalmente distribuidas con varianzas respectivas de 16 y 25. Si las varianzas muéstrales son 20 y 8. Determinar si la primera muestra tiene una varianza significativamente mayor que la segunda al nivel de significación del 5%.
Solución.
Partamos de identificar que existen dos muestras 1 y 2 $N_1 = 9$, $N_2 = 12$, dos distribuciones de varianzas $\sigma_2{}^2 = 16$, $\sigma_2{}^2 = 25$ y la varianzas muéstrales $S_1{}^2 = 20$, $S_2{}^2 = 8$, por ende reemplazamos en la función.

$$F = \frac{\frac{S_1{}^2}{\sigma_1{}^2}}{\frac{S_2{}^2}{\sigma_2{}^2}} = \frac{N_1 \quad S_1{}^2 (N_1-1)\sigma_1{}^2}{N_2{}^2 S_2{}^2 (N_2-1)\sigma_2{}^2} = \frac{(9)(20)(9-1)16}{(12)(8)(12-1)25} = \frac{(9)(20)(8)(16)}{(12)(8)(11)(25)} = \frac{23040}{26400} = 4.03$$

En conclusión los grados de libertad para el numerado y denominador de F son $V_1 = (9-1) = 8$, $V_2 = (12-1) = 11$, se puede indicar que la varianza de la muestra 1, es significativamente mayor que la muestra 2 al nivel de significancia 0.05.

Ejemplo: Se toman dos muestras de tamaño 8 y 10 de dos poblaciones normalmente distribuidas con varianzas respectivas 20 y 36. Hallar la probabilidad de que la varianza de la primera sea doble que la de la segunda.
Solución.
Se organizan los datos de acuerdo al problema;
Muestra $N_1 = 8$
Muestra $N_2 = 10$
Desviaciones de la varianza
$\sigma_1{}^2 = 20$
$\sigma_2{}^2 = 36$
De acuerdo a esta condición se aplica la formula

$$F = \frac{8(S_1^2)(7)(20)}{10(S_1^2)(9)(36)} = 1.85\frac{S_1^2}{S_1^2}$$

$$F = 1.85\frac{S_1^2}{S_1^2} = 1.85(2) = 3.70$$

De esta manera se puede indicar que la probabilidad es menor que 0.05 pero mayor que 0.01.

1.10. Ejercicios de reflexión

- Un contador toma una muestra aleatoria de 60 facturas de venta de una empresa de cosméticos en el aeropuerto El Dorado. Se sabe que en el mes de septiembre de 400 facturas, se encontró un promedio de ventas de 20000 pesos y una desviación estándar de 2000 pesos. ¿Cuál es el intervalo de confianza del 90% para el promedio de las ventas de la empresa?

- Cuando Gregorio Mendel realizó sus famosos experimentos de hibridación con chicharos, uno de sus experimentos produjo vástagos que consistieron en 428 chicharos con vainas verdes y 152 chicharos con vainas verdes y 152 con vainas amarillas. Según la teoría de Mendel ¼ de los chícharos vástagos debían tener vainas amarillas. (Triola, 2013, P.393), Con un nivel de significancia de 0.05, se puede decir que: ¿la proporción de vainas amarillas es igual a 1/4?

- De 1000 casos seleccionados al azar sobre una enfermedad x, en el pulmón, 823 terminaron en muerte. ¿Cuál es el intervalo de confianza para las ambas situaciones del 95% con relación a la tasa de mortalidad?

1.11. Prueba de hipótesis para:

1.11.1. Media con varianza poblacional conocida y desconocida.

Se ha estudiado diferentes aspectos importantes de las pruebas de hipótesis en relación a las medias conocidas y desconocidas de igual manera se ha visto la parte fundamental de las desviaciones estándar de los parámetros poblacionales.

De acuerdo a esta situación cabe resaltar que los métodos para someter a prueba una desviación y estándar o una varianza son los utilizados en una distribución chi-cuadrado vistos en el módulo anterior.

En conclusión, se puede indicar que para calcular la una prueba existe la posibilidad de utilizar varios parámetros poblacionales conociendo o desconociendo las desviaciones o varianzas.

Ejemplo: Un estudio revela que 30 autos que constituyen una muestra aleatoria se condujeron a un promedio de 12500 km durante un año, con una desviación estándar de 2400 km. En relación a lo explicado, indicar la hipótesis en donde los promedios de autos se condujeron a 12000 km durante un año, en cuanto a la alternativa de que el promedio sea mayor emplear el nivel de significancia del 0.05.

Solución

De acuerdo a la información presentada es importante tener en cuenta las siguientes condiciones:

1. Organizar o definir las pruebas de hipótesis.

H_O: $\mu = 12000km$

H_a: $\mu > 12000km$

Ahora bien, si la probabilidad es mayor al promedio entonces se rechaza la hipótesis nula y se acepta la hipótesis alternativa, si la probabilidad es menor que el promedio, entonces se rechaza la nula.

Nos basamos de estas condiciones para calcular z,

$$z = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{12500-12000}{\frac{2400}{\sqrt{30}}} = \frac{500}{\frac{2400}{5.4772}} = \frac{500}{438.180} = 1.14$$

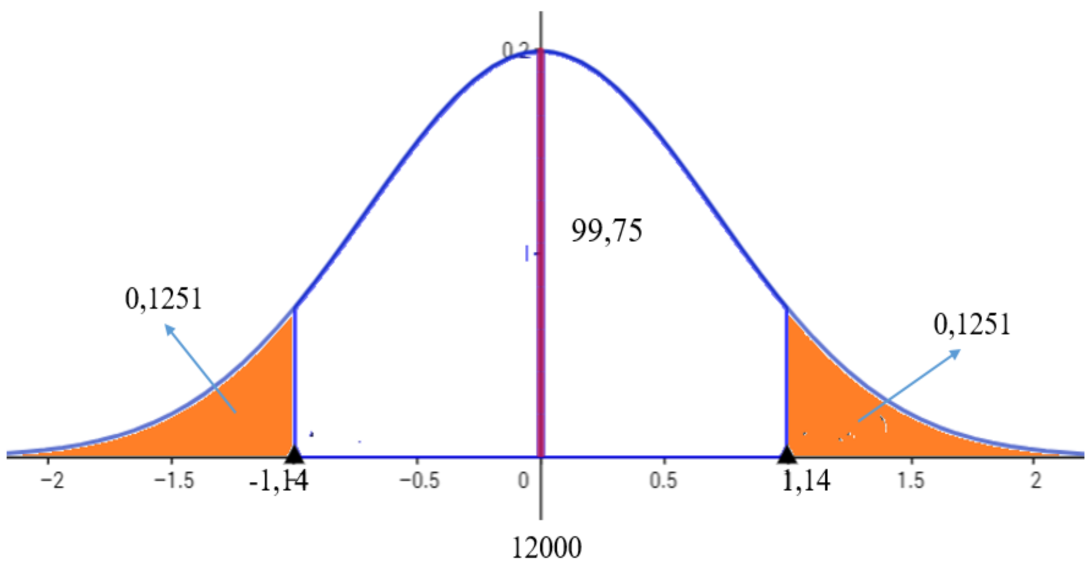
Con este resultado se puede indicar que

La $P(z < -1.14) \rightarrow tabla\ z = 0.1251$

La $P(z < -1.14) \rightarrow tabla\ z = P(Z > 1.14) = 1 - P(Z < 1.14) = 1 - 0.8749 = 0.1251$

En una distribución normal se observa el nivel de significancia del 0.05. ver figura 11

Figura 11: Distribución normal para una prueba de aseveración con desviación estándar.



Fuente: propia del autor

Si se suman los valores de la tabla z, $0.1251 + 0.1251 = 0.25$. por lo tanto, que puede analizar que el nivel de significancia es menor al valor obtenido.

$$0.25 > 0.05$$

En conclusión, los promedios en que condujeron los autos fueron menores a 12000km durante un año.

1.12. Ejercicio de Reflexión.

- Un analista de una empresa de productos de aseo quiere determinar si el lote de una materia prima tiene o no una varianza poblacional mayor a 9 gramos en su cantidad de alcohol. Se realiza un muestreo de 15 elementos y se obtiene una varianza muestral de 11.4 gramos de alcohol. Según la información anterior, ¿Cómo plantearía las hipótesis de la prueba el analista?

- Un analista de una empresa de hidrocarburos afirma que la tasa de toxicidad en los

1.13. Conclusiones

- Para comprender la manera adecuada de interpretar las medidas de variabilidad tales como la varianza y desviación estándar es un buen recurso calcular el coeficiente de variación y tomarlo como un indicador de dispersión.
- Comprender los requisitos teóricos de cada uno de las distribuciones permite diferenciarlas y aplicarlas mejor para dar solución a situaciones problema.
- Identificar los elementos de las distribuciones: Variable aleatoria, función de probabilidad, permite realizar mejor los cálculos de las probabilidades en cada situación problema.
- Se concluye que los intervalos de confianza permiten establecer diferentes postulados y supuestos en estadísticos muestrales grandes y pequeños con la finalidad de organizar decisiones de un parámetro poblacional.
- Los resultados de las pruebas de hipótesis aplicadas permiten comprender la utilización de verificar las hipótesis planteadas a través de los intervalos de confianza y así poder tomar la decisión correcta al finalizar dicho análisis.

1.14. Material de estudio.

Temas que abordan	Referencia bibliográfica (APA)	Ubicación (el link web o la base de datos)
Distribución Binomial	Mario, F. (2013). Distribuciones de probabilidad discreta. (Pearson).(pp. 218 - 223). Ciudad: Ciudad de México	Bases de datos de la UMB, Ebooks

Media, varianza y desviación Estándar para la distribución binomial	Mario, F. (2013). Distribuciones de probabilidad discreta. (Pearson). (pp. 230-231). Ciudad: Ciudad de México	Bases de datos de la UMB, Ebooks
Bases de datos UM	Mario, F. (2013). Probabilidad. Estadística. México: Pearson Warpole, M. (2013). Probabilidad. Estadística y probabilidad para ingeniería y ciencias. San Antonio (Texas): Pearson Newbold, P, Carson W, Thorne, B. (2013). Estadística para administración y economía. México: Pearson	Bases de datos de la UMB, Ebooks
Estimación de proporciones	Johnson, R. A. (2012). Probabilidad y estadística para ingenieros. Pearson Educación.	Bases de datos de la UMB, Ebooks-24
Diferencia entre intervalos de confianza.	Ronald E. Walpole, R. H. (2012). *Probabilidad y estadística para ingeniería y ciencias*. Pearson Educación	Bases de datos de la UMB, Ebooks-24
Intervalos de confianza sobre media, conocida la varianza.	Montgomery, W. W. (2002). Probabilidad y Estadística para ingeniería. Continental.	Biblioteca UMB
Conceptos básicos acerca pruebas de hipótesis (para proporciones)	Triola, M. F. (2013). Estadística. Pearson Educación.	Bases de datos de la UMB, Ebooks-24
Estadístico de prueba para proporciones	Triola, M. F. (2013). Estadística. Pearson Educación.	Bases de datos de la UMB, Ebooks-24
Decisiones y conclusiones de una prueba de	Triola, M. F. (2013). Estadística. Pearson Educación.	Bases de datos de la UMB, Ebooks-24

hipótesis (proporción)		