Probabilidad y Estadística

Año de cursada 1° AÑO

Clase Nº 6: Distribuciones de Probabilidad.

Contenido: Distribución normal, normal estandarizada. Parámetros.

1. **Presentación:**

Bienvenidos y bienvenidas a una nueva clase de Probabilidad y Estadística. En nuestro encuentro N° 6, avanzaremos con el estudio de las distribuciones especiales de probabilidad.

Anteriormente, habíamos definido herramientas de probabilidad que nos permitían calcular e interpretar probabilidades de algunos fenómenos.

Ahora, en esta clase, avanzaremos analizando cómo, en algunos casos, bajos ciertas variables determinadas, las probabilidades se distribuyen siguiendo una forma a determinarse de acuerdo con las características del conjunto que estemos estudiando.

Este es un encuentro, nos permitirá poner en acción varias herramientas que vinimos desarrollando a lo largo de la cursada.

2. Desarrollo:

La distribución normal

A partir de ejemplos de la vida cotidiana veremos cómo a menudo utilizamos la palabra normal.

Cuando nace un bebé, el médico lo lleva a la balanza y observa que pesa 3,500 Kg. Lo primero que dirá, casi con certeza, es que tiene un peso normal.

Luego, cuando el niño comienza su vida escolar, es probable que alguna psicopedagoga luego de efectuarle un test informe a la maestra que tiene una inteligencia normal.

Cuando ya adolescente asista a un colegio secundario y termine el año con 5 amonestaciones, la madre probablemente tratará de justificarlo diciendo que es un chico normal.

Si al llegar a adulto, se dedica a sembrar papas y dice que ha obtenido un rendimiento de 10.000 Kg. por Ha. seguramente pensará que eso es lo normal.

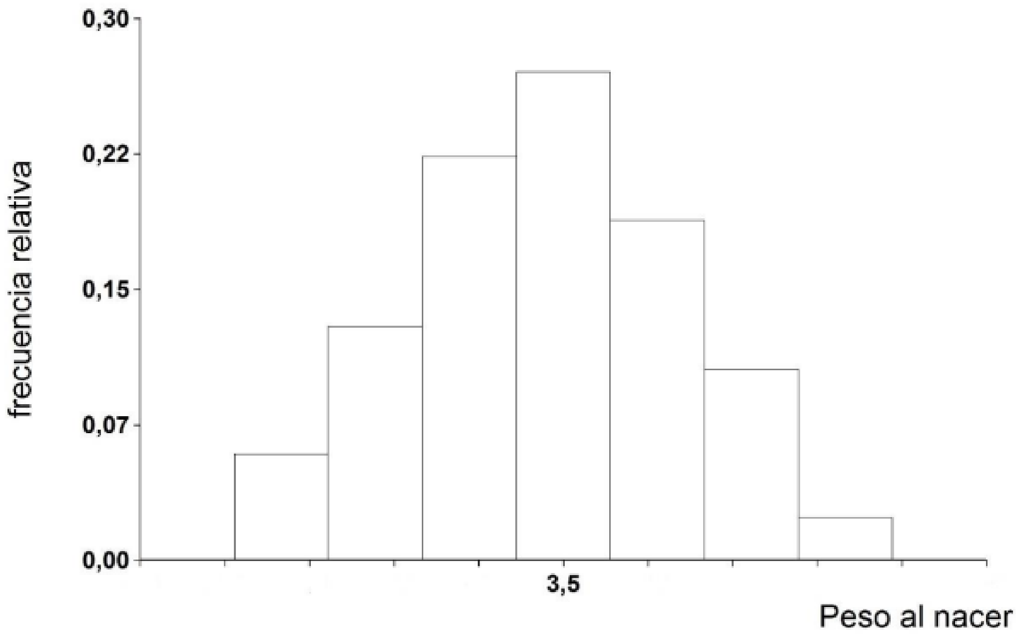
Ahora bien, ¿qué queremos expresar cuando decimos es normal? Lo que queremos decir es que, a la mayoría de las personas, en situaciones similares a las del joven del ejemplo, le pasa lo mismo.

Podemos pensar, por ejemplo, que el 95% de los bebés que nacen pesan entre 3 Kg y 4 Kg, que el mismo porcentaje de niños tienen un coeficiente de inteligencia entre 95 y 105, que el 95% de los adolescentes terminan un año escolar con una cantidad de amonestaciones entre 3 y 7 y que el 95% de las hectáreas sembradas con papas rinden entre 9.000 y 11.000 Kg.

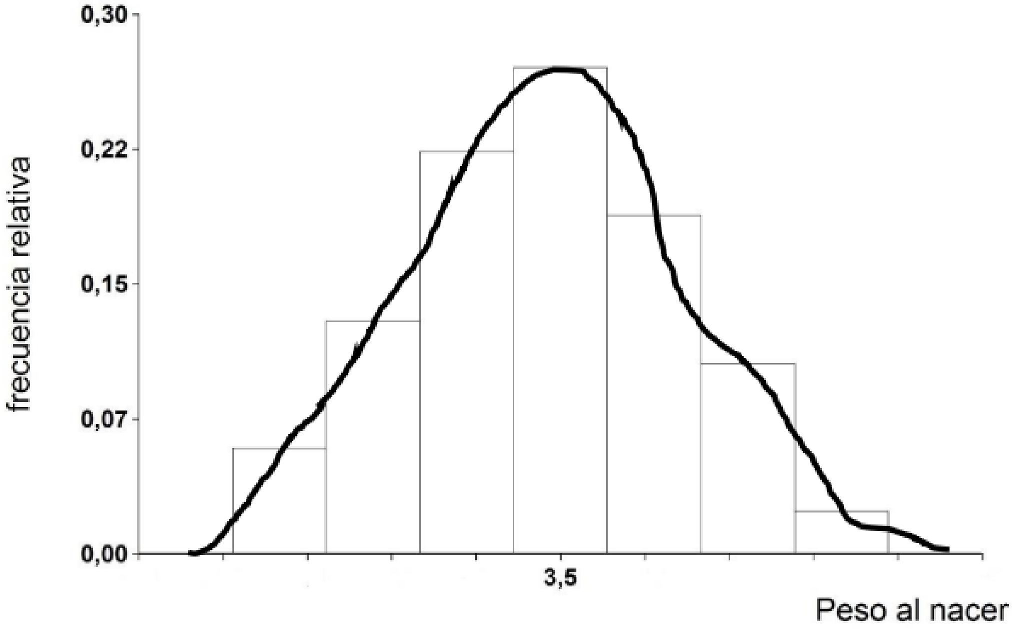
Pero, como toda regla tiene excepciones, siempre habrá un porcentaje mínimo de nacimientos prematuros, de niños superdotados, de adolescentes que siempre se comportan de acuerdo a las normas escolares, o de hectáreas sembradas con papas que sufrieron una determinada plaga y produjeron mucho menos.

Sin embargo, un análisis estadístico lógico nos llevará a concluir que todas estas situaciones tienen una baja probabilidad de ocurrir.

Ahora bien, si un médico se tomara el trabajo de pesar, por ejemplo, a 500 niños recién nacidos, registrar estos valores en una planilla, construir con ellos intervalos de clase, calcular las frecuencias absolutas en dichos intervalos y luego graficar los datos por medio de un histograma, es muy probable que obtenga la siguiente distribución:

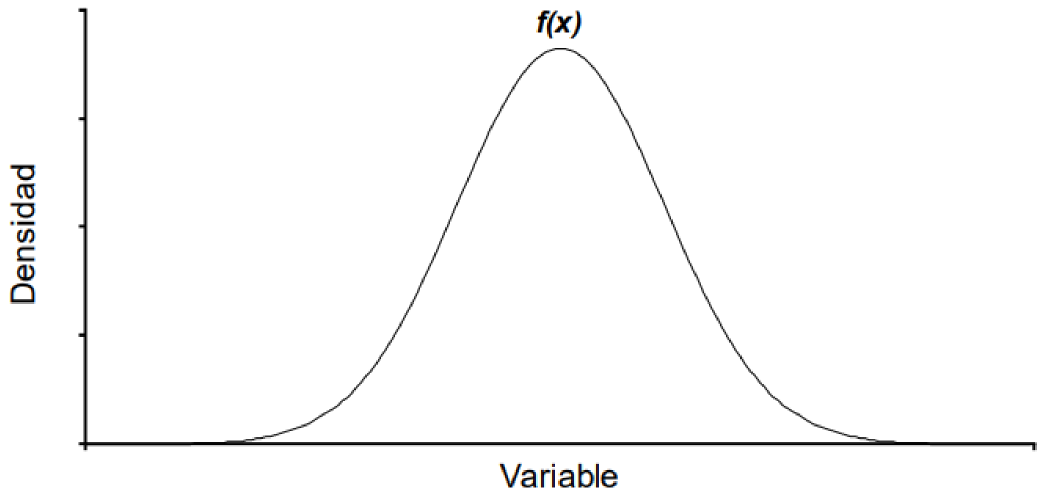


Si trazamos el polígono de frecuencias en forma suavizada, tendremos una curva más o menos simétrica, es decir, así:



Esta curva, obtenida empíricamente, se asemeja a una curva teórica unimodal y perfectamente simétrica que caracteriza a una distribución de probabilidad conocida como **distribución normal**, Gaussiana o de Laplace, o también llamada campana de Gauss. En unidades anteriores estudiamos el concepto de la medida de posición conocida como Moda y vimos que las distribuciones unimodales se caracterizan por presentar un valor bien definido de la variable al que le corresponde la mayor frecuencia absoluta.

La distribución normal presenta un valor de mayor frecuencia y, a partir de él, decae hacia ambos lados con una simetría perfecta. Esta simetría hace que a valores situados a igual distancia del valor modal por izquierda y por derecha de la distribución, les corresponda la misma probabilidad. La representación gráfica de esta distribución es la siguiente:



Esta figura refleja las situaciones expresadas en el ejemplo inicial. Así, en la primera situación planteada, podemos pensar que existe una gran cantidad de bebés que tienen un peso de 3,500 Kg. y a partir de este valor las frecuencias disminuyen a ambos lados respetando la perfecta simetría de la distribución normal.

Para realizar este gráfico se recurre a una expresión matemática genéricamente conocida como función densidad. Cada distribución de probabilidad de una variable aleatoria continua tiene su propia función de densidad, en el caso de la Distribución Normal, su función está dada por la siguiente expresión:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{para } -\infty < x < \infty$$

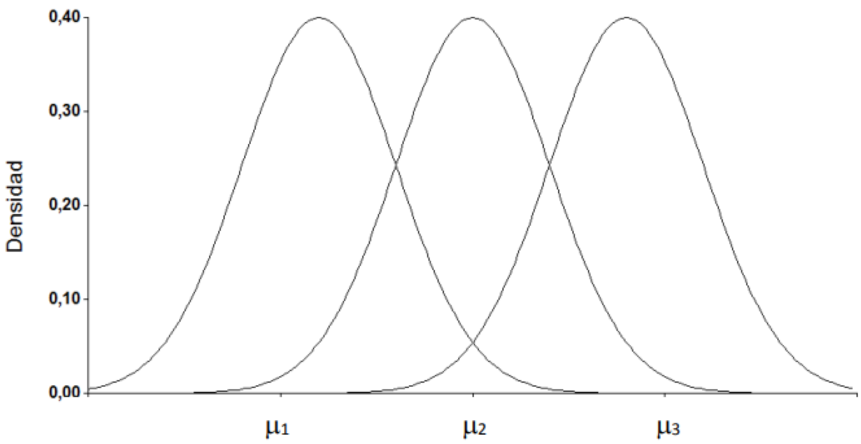
Siendo μ y σ, los dos parámetros que caracterizan a una distribución normal donde: μ=media aritmética o esperanza, y σ=desviación estándar

Por ser la distribución perfectamente simétrica, la media μ coincide con la mediana y la moda. Se encuentra en el punto del eje de abscisas que divide a la distribución en dos partes iguales y, a su vez, registra el valor de la variable de mayor frecuencia.

La desviación estándar a, medida de variabilidad de la distribución, determina la mayor o menor dispersión de los datos alrededor de μ.

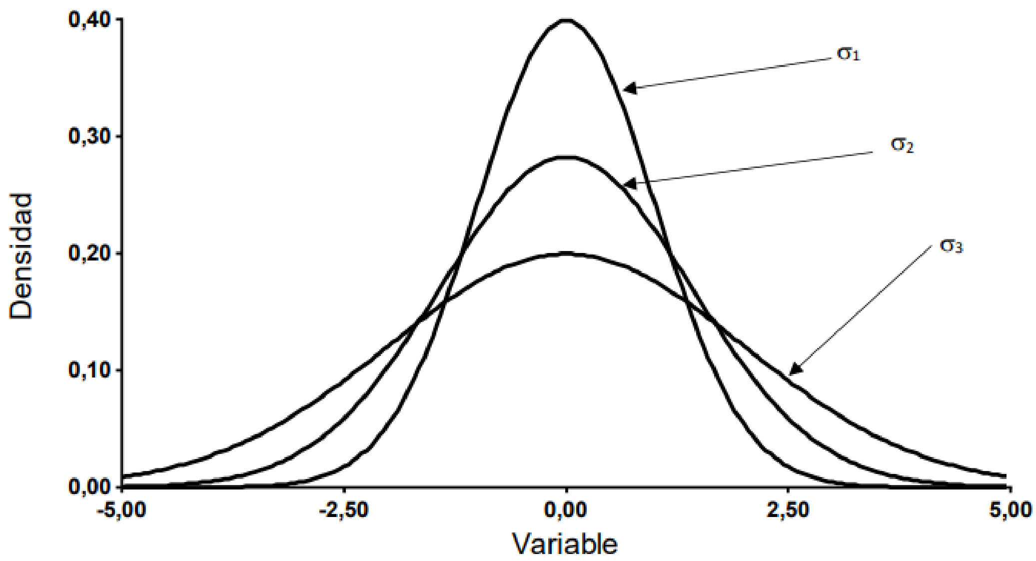
Cuando σ crece, la curva se achata reflejando esta situación. Por ejemplo, si consideramos la variable peso de personas, es razonable pensar que el peso de los niños al nacer presenta una variabilidad mucho menor que el peso de adolescentes de 14 años, ya que a esta edad algunos están desarrollados como un adulto y otros, más retrasados en su crecimiento, tienen todavía un cuerpo infantil.

Si graficamos la distribución de la variable peso al nacer y a los 14 años, la curva correspondiente a esta última situación presentará una forma mucho más achatada, reflejando una mayor dispersión de los pesos alrededor del valor promedio. Los siguientes gráficos ayudarán a comprender mejor estos conceptos:



μ cambia y σ es constante; $\mu 1 < \mu 2 < \mu 3$

Esta situación podría plantearse al estudiar la altura de determinada planta, en tres estadíos de crecimiento: por ejemplo, al mes, a los 3 meses y a los 5 meses. Es probable que la variabilidad sea muy parecida en los tres momentos de la experimentación y, en cambio, que el promedio de altura en el primer mes sea inferior al que registra a los 3 meses y éste, a su vez, Inferior a la altura promedio a los 5 meses.



μ es constante y σ cambia $\sigma_1 < \sigma_2 < \sigma_3$

Un ejemplo de lo expresado en el gráfico anterior podría ser que σ_1 fuera la desviación estándar de los pesos de niños de 10 años en escuelas privadas de clase social alta, con similar estado nutricional, σ_2 la de los pesos de los niños de 10 años de escuelas públicas y privadas, de clase social media y alta, con estado nutricional más variable y σ_3 la medida de variabilidad correspondiente al peso de niños de 10 años de escuelas públicas en general, donde se observa una gran variabilidad respecto del estado nutricional de los alumnos.

En síntesis, en el primer gráfico que muestra las tres curvas normales se observa que un cambio en el valor de μ, desplaza toda la distribución normal pero no altera su forma.

Por el contrario, si analizamos el segundo gráfico vemos que un cambio en el valor de σ no desplaza la curva, sino que altera su forma, haciéndola más estrecha a medida que disminuye su valor.

Características de la distribución normal

Veremos ahora algunas características que presenta la distribución normal.

Como ya hemos expresado anteriormente al hablar de las curvas de distribución de densidad de probabilidad, el área total bajo la curva normal es también igual a 1. $P(-\infty < x < \infty) = 1$

A su vez, al ser la curva simétrica con respecto a μ, el 50% de las observaciones se encuentran por debajo de la media μ y el 50% por encima de ella. O sea: $P(-\infty < x < \mu) = 0{,}50$ y $P(\mu < x < \infty) = 0{,}50$

La distribución normal es una distribución teórica, ideal, que cumple con ciertas propiedades, las que surgen de la aplicación de la fórmula matemática que expresa su función de densidad de probabilidad.

Estas propiedades son:

En el intervalo $(\mu - \sigma; \mu + \sigma)$ se encuentra el 68,2% de las observaciones, o lo que es lo mismo: $P(\mu - \sigma < x < \mu + \sigma) = 0{,}682$

En el intervalo (μ - 2σ; μ + 2σ) se encuentra el 95,4% de las observaciones, o: $P (\mu - 2\sigma < x < \mu + 2\sigma) = 0{,}954$

En el intervalo (μ - 3σ; μ + 3σ) se encuentra el 99,8% de los datos, o:

$P (\mu - 3\sigma < x < \mu + 3\sigma) = 0{,}998$

Veremos la aplicación de estas propiedades por medio de un ejemplo. Supongamos que se está considerando la variable continua "altura de niños" de una determinada edad escolar en la ciudad de Rio Grande. Una vez medida la altura de una muestra de niños y sistematizando la información podríamos haber obtenido los siguientes valores:

μ = 1,30 m. y σ = 0,12 m.

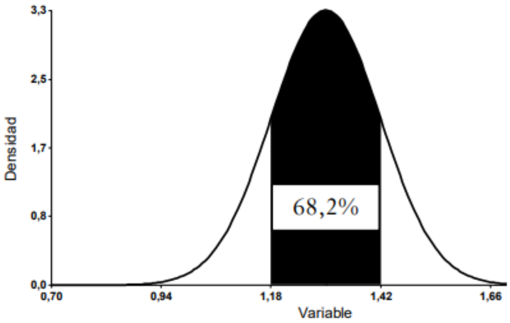
Si la distribución de la variable "altura de niños" es aproximadamente simétrica, se podrían aplicar las propiedades enunciadas precedentemente obteniendo las siguientes conclusiones:

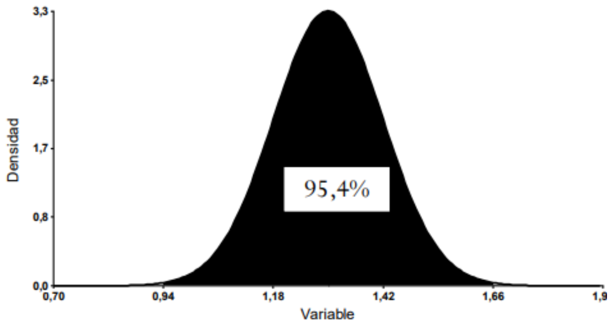
1. El 68,2% de los niños medirá entre (1,30 - 0,12) y (1,30 + 0,12) o, lo que es lo mismo, el 68,2% de estos niños tendrán una altura comprendida en el intervalo (1,18; 1,42)

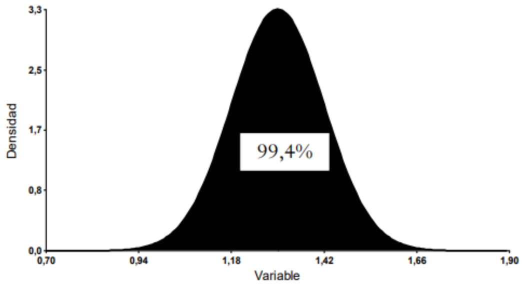
2. El 95,4% de los niños tendrán una altura comprendida en el intervalo (1,06; 1,54)

3. El 99,8% de los niños tendrán una altura comprendida en el intervalo (0,94; 1,66)

Gráficamente, tendremos:







Ahora bien, considerando los distintos valores que pueden tomar los parámetros de la distribución normal, los correspondientes campos de variación son: $-\infty < \mu < \infty$ y $0 < \sigma < \infty$

y ya podemos pensar que existen infinitas curvas normales distintas, que surgen al variar los valores de μ y σ en cada situación experimental particular.

Para calcular probabilidades de la distribución normal deberíamos aplicar una operación matemática llamada integral a la función de densidad, reemplazando los parámetros μ y σ por los valores de alguna combinación en particular. Pero teniendo en cuenta la expresión matemática que caracteriza a la función de densidad de una variable normalmente distribuida, vemos que calcular probabilidades en base a ella, es un procedimiento muy complicado y prácticamente imposible de efectuar manualmente.

Como no existe una sola distribución normal sino toda una familia de infinitas distribuciones normales, cada una correspondiente a una cierta combinación de μ y σ, fue necesario seleccionar una de todas estas combinaciones para que las representara.

Distribución normal estandarizada

Una distribución normal se denomina estandarizada si su media es 0 y su varianza y, por lógica, su desviación estándar, son iguales a 1.

Para obtener una variable normal estandarizada se debe realizar una transformación de la variable normal original X.

Esta transformación da como resultado una nueva variable que llamaremos Z y que surge al restar a la variable X original su media μ y dividirla por su desviación estándar σ.

Entonces: $Z = \frac{X-\mu}{\sigma}$ y su correspondiente función de densidad de probabilidad es:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Como podemos observar, la distribución de probabilidad de una variable estandarizada Z es simétrica con respecto al valor 0 pues ahora μ = 0.

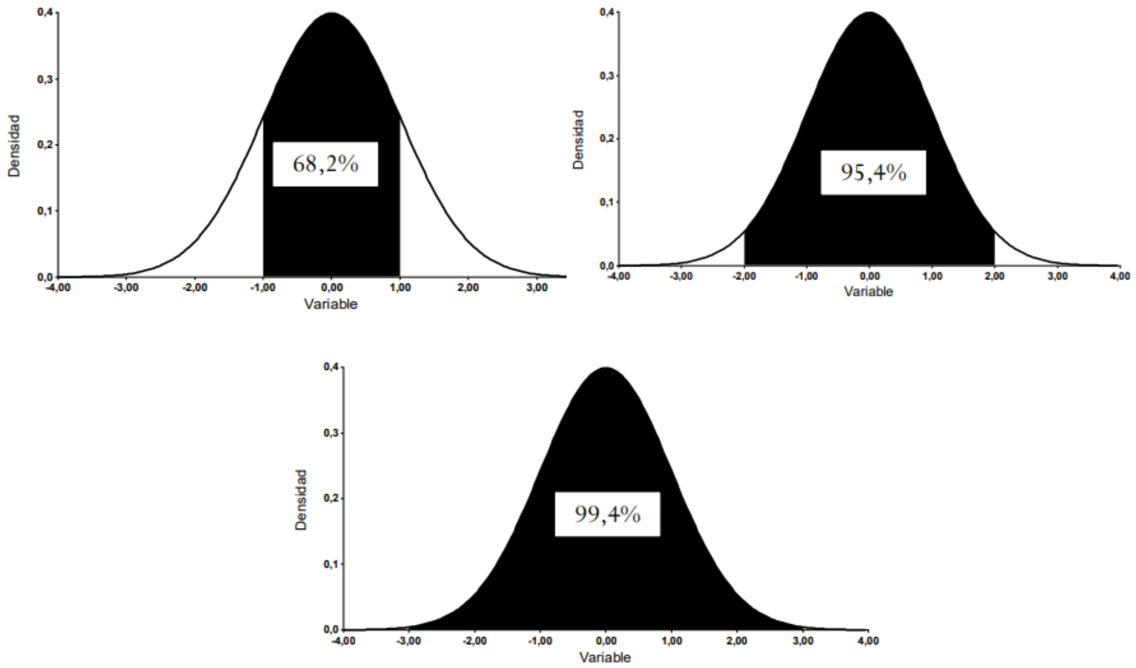
Al trabajar con una variable normal estandarizada también se cumple que:

P(- 1 < z < 1) = 0,684

$P(-2 < z < 2) = 0{,}954$

$P(-3 < z < 3) = 0{,}998$

Gráficamente:



¨Para no extendernos demasiado, les dejo a continuación un link donde podrán ver cómo se usan las tablas de para el cálculo de probabilidades en una distribución normal. Por otro lado, podrán ver cómo realizar los cálculos con el InfoStat en el siguiente link.

Y, por último, como no hemos llegado a trabajar con los demás modelos de distribución de probabilidades le dejo el siguiente link para que lo puedan ver.

3. Actividad Integradora:

1. Los salarios mensuales de los recién graduados que acceden a su primer empleo se distribuyen según una ley normal de media $130000 y desviación típica $60000. Calcular el porcentaje de graduados que cobran:

a) Menos de $60000 al mes

b) Entre $100000 y $150000 al mes

c) Más de $220000 al mes

2. La media de los pesos de 500 estudiantes de un Instituto es 70 kg y la desviación típica 3 kg. Suponiendo que los pesos se distribuyen normalmente, hallar cuántos estudiantes pesan:

a) Entre 60 kg y 65 kg.

b) Más de 90 kg.

c) Menos de 64 kg.

d) 64 kg.

e) 64 kg o menos

4. Cierre:

Llegamos al final de la clase n° 9. Hemos trabajado de sobre un modelo de distribución de probabilidades que resulta muy importante ya que aparece constantemente en el trabajo cuando analizamos determinados fenómenos.

Por otra parte, esta es, además, la última clase del módulo de Probabilidad y Estadísticas. Ha sido un placer haber compartido este espacio de formación con cada uno y cada una de udes. Sin duda hemos realizado un recorrido lleno de muchos aprendizajes.

Les deseo el mayor de lo éxitos en los parciales, y como siempre están disponibles las vías de comunicación por medio del campus para cualquier consulta.

Saludos cordiales.

5. **Bibliografía:**

- ✓ Ruiz Diaz, G. (2019). Estadística y Probabilidad. Notas de cátedra. Universidad Nacional de Formosa.
- ✓ García, J; López, N; Calvo, J. (2011). Estadísticas Básicas para Estudiantes de Ciencias. Facultad de Ciencias Físicas Universidad Complutense de Madrid. España.
- ✓ Wackerly, D; Mendenhall, W; Schaeffer, R. (2010). Estadística Matemática con Aplicaciones. 7ma Ed. Cengage Learning. Santa Fe, México.