

Probabilidad y Estadística:

Trabajo Práctico Integrador I

Nombre: Farias, Gustavo

Comisión: M2025-13

Matrícula: 101662

Repositorio GitHub:

<https://github.com/Lucenear/UTN-TUPaD-TPs/tree/main/Probabilidad%20y%20Estadistica>

Consignas:

1) Haciendo uso del vocabulario técnico, identificar con relación al caso propuesto:

- a. Población y Muestra
- b. Unidad Elemental
- c. Variables en estudio, detallando su tipo

2. Construir la/s Tabla/s de Frecuencias y calcular todas las frecuencias de las siguientes variables:

- a. Tiempo en horas semanales dedicadas al estudio. (Determinar la cantidad óptima de intervalos a utilizar)
- b. Nivel de satisfacción con la Carrera.
- c. A partir de la tabla obtenida en el punto a. realizar la interpretación de todas las frecuencias correspondientes al cuarto intervalo en el contexto del caso planteado.
- d. A partir de la tabla obtenida en el punto b. realizar la interpretación de todas las frecuencias correspondientes a la categoría “Satisficho”

3. Calcular, para las variables definidas en el punto 2, las medidas descriptivas de tendencia central, posición y dispersión, interpretando sus resultados en términos del problema planteado. (Para la variable categórica solo calcular Moda, Mediana y Cuartiles).

4. Representar gráficamente las variables definidas en el punto 2 y realizar el correspondiente análisis.

- a. Elegir una frecuencia (Frecuencia Absoluta o Frecuencia Relativa) para la variable Tiempo en horas semanales dedicadas al estudio y construir un Histograma.
- b. Construir un Diagrama Circular que represente porcentualmente el Nivel de satisfacción con la carrera.
- c. Realizar el análisis de los gráficos obtenidos en el contexto del caso planteado.

5. Suponiendo que los datos corresponden toda la población y son seleccionados 16 estudiantes, calcular las siguientes probabilidades:

- a. Más de 9 estudiantes estén muy satisfechos con la carrera.
- b. Entre 4 y 8 estudiantes estén satisfechos con la carrera.
- c. Menos de 5 estudiantes estén insatisfechos con la carrera.
- d. Exactamente 10 estudiantes estén muy insatisfechos con la carrera.

6. En el horario consultas de cierta materia se reciben en promedio 15 consultas en media hora. Calcular las siguientes probabilidades:

- a. Que lleguen por lo menos 6 consultas en 20 minutos.
- b. Que lleguen a lo sumo 12 consultas en 40 minutos.
- c. Más de 7 y menos de 10 consultas en 30 minutos.

7. Utilizando el modelo normal:

- a. Calcular la probabilidad de que un estudiante seleccionado aleatoriamente tenga una estatura mayor o igual que 179 cm.
- b. Calcular la probabilidad de que un estudiante seleccionado aleatoriamente tenga una estatura comprendida entre 147 cm. y 172 cm.
- c. Hallar el valor que excede al 97,5% de las estaturas.

Respuestas:

1) Identificación:

a.

Población: Son todos los estudiantes de segundo año de la Tecnicatura Universitaria en Programación de la Universidad INNOVA XXII con sede en Ushuaia.

Muestra: Es el grupo de alumnos que fueron encuestados por la consultora STUDIO X el 11 de Marzo de 2025.

b.

Unidad elemental: Corresponde a cada estudiante que fue encuestado.

c.

- Género: Cualitativa nominal
- Fuma: Cualitativa nominal

- Nro de hermanos: Cuantitativa discreta
- Peso: Cuantitativa continua
- Estatura: Cuantitativa continua
- Edad: Cuantitativa discreta
- Gastos semanales Aliment.: Cuantitativa continua
- Cant. de materias aprob.: Cuantitativa discreta
- Satisfacción con la carrera: Cualitativa ordinal
- Tiempo semanal en HS dedic. Est.: Cuantitativo continua
- Trabaja: Cualitativa ordinal

2)

a. Tiempo en horas semanales dedicadas al estudio. Utilizamos la función de Sturges

Intervalo	Frecuencia	Frecuencia_Relativa	Frecuencia_Acumulada	Frecuencia_Relativa_Acumulada
[0,2.78)	1	0.0040	1	0.0040
[2.78,5.56)	3	0.0121	4	0.0161
[5.56,8.33)	16	0.0645	20	0.0806
[8.33,11.1)	43	0.1734	63	0.2540
[11.1,13.9)	44	0.1774	107	0.4314
[13.9,16.7)	69	0.2782	176	0.7096
[16.7,19.4)	51	0.2056	227	0.9152
[19.4,22.2)	17	0.0685	244	0.9837
[22.2,25)	4	0.0161	248	0.9998

b. Nivel de satisfacción con la Carrera.

SATISFACCIÓN CON LA CARRERA	Frecuencia	Frecuencia_Relativa	Frecuencia_Acumulada	Frecuencia_Relativa_Acumulada
Muy satisfecho	138	0.552	138	0.552
Satisfecho	77	0.308	215	0.860
Insatisfecho	16	0.064	231	0.924
Muy insatisfecho	19	0.076	250	1.000

c. Interpretación del cuarto intervalo de tiempo:

- Frecuencia absoluta: Hay 43 estudiantes que dedican entre 8,33 y 11,1 horas semanales al estudio.
- Frecuencia relativa: Este intervalo representa aproximadamente el 17,34% del total de estudiantes.
- Frecuencia acumulada: Sumando desde el primer intervalo hasta el cuarto, 63 estudiantes dedican hasta 11,1 horas semanales al estudio.
- Frecuencia relativa acumulada: Indica que el 25,4% de los estudiantes dedica hasta 11,1 horas semanales al estudio.

d. Realizar la interpretación de todas las frecuencias correspondientes a la categoría “Satisfecho”.

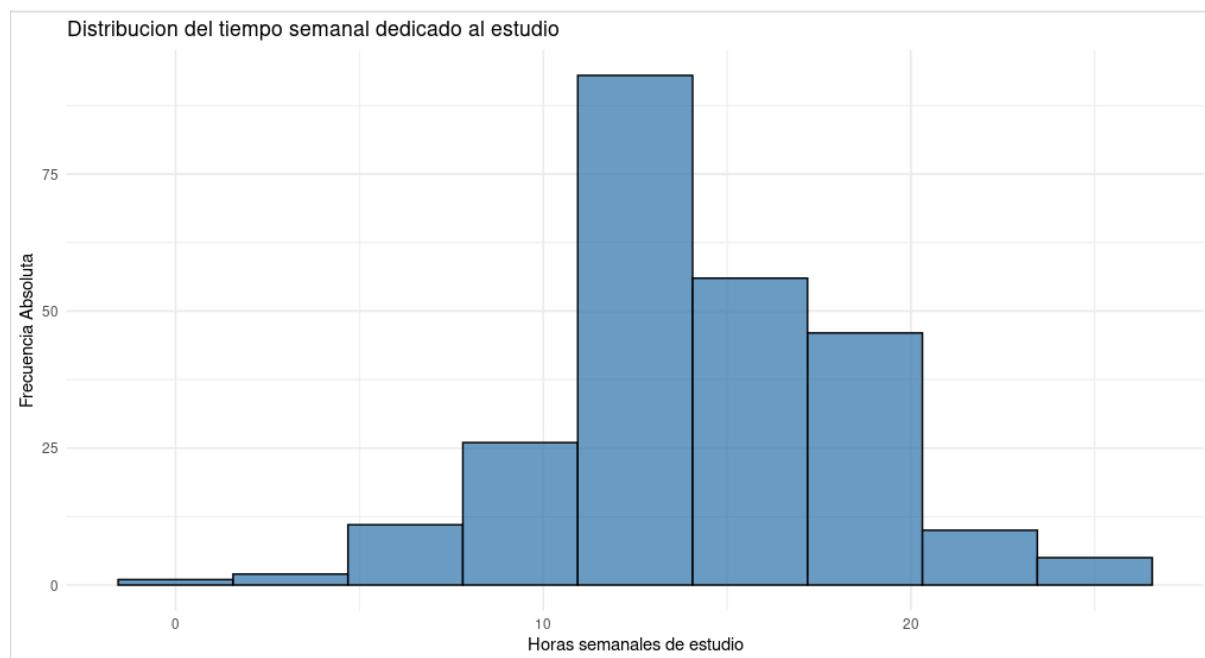
- Frecuencia absoluta: Muestra que 77 estudiantes se sienten satisfechos con la carrera.

- Frecuencia relativa: Los 77 estudiantes representan un 30,8% del total de estudiantes.
- Frecuencia acumulada: Sumando desde “Muy satisfecho” hasta “Satisficho”, 215 estudiantes muestran un nivel de satisfacción positivo.
- Frecuencia relativa acumulada: Esto indica que el 86% de los estudiantes está satisfecho o muy satisfecho con la carrera.

3. Cálculo de medidas descriptivas de variables del punto anterior:

Medidas descriptivas - Tiempo semanal de estudio:

- Media: 14.3 horas
- Mediana: 14 horas
- Moda: 16 horas
- Cuartil 1 (Q1): 11.25 horas
- Cuartil 2 (Mediana): 14 horas
- Cuartil 3 (Q3): 17 horas
- Desviación estandar: 4.17 horas
- Varianza: 17.38 horas²
- Rango: 25 horas
- Coeficiente de variación: 29.16 %

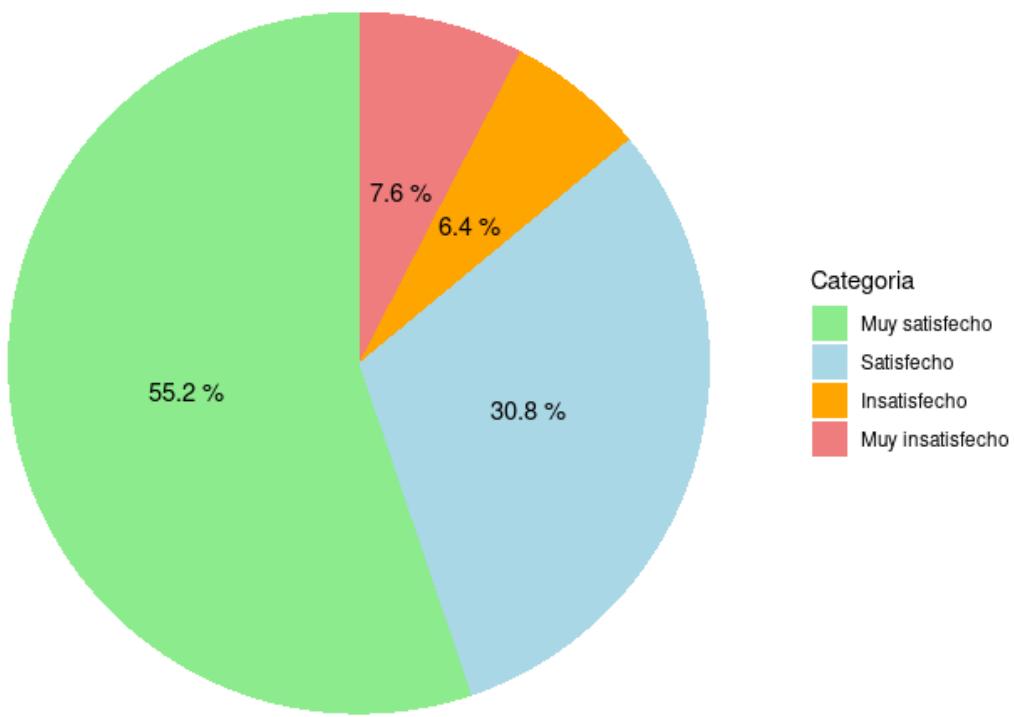


Medidas descriptivas - Satisfacción con la carrera

- Moda: Muy satisfecho

- Mediana: Muy satisfecho
- Cuartil 1 (Q1): Satisfecho
- Cuartil 3 (Q3): Muy satisfecho

Nivel de satisfacción con la carrera



4. Análisis de los resultados obtenidos:

Tiempo semanal dedicado al estudio

- Los estudiantes dedican en promedio 14,3 horas semanales al estudio.
- La mediana se ubica en 14 horas, lo que significa que la mitad estudia menos de 14 horas y la otra mitad más.
- La mayoría concentra su tiempo entre 11 y 17 horas semanales (cuartiles).
- El rango va de 0 a 25 horas, mostrando cierta dispersión: algunos estudian muy poco y otros bastante más.
- La moda es 16 horas, lo más frecuente en la muestra.
- El coeficiente de variación (29,2%) indica una dispersión moderada.
- En el histograma se observa que el mayor grupo de estudiantes estudia entre 10 y 15 horas semanales, con una caída gradual hacia

valores más altos.

Nivel de satisfacción con la carrera

- La gran mayoría se muestra positiva respecto a la carrera:
 - 55,2% satisfecho
 - 30,8% muy satisfecho
- En contraste, solo un 6,4% insatisfecho y un 7,6% muy insatisfecho.
- La mediana y la moda caen en la categoría “Muy satisfecho”, mostrando un sesgo hacia la valoración positiva.
- En los cuartiles, el 75% de los estudiantes se encuentra entre “Satisfecho” y “Muy satisfecho”.

5. Distribución Binomial:

- a. La probabilidad de que más de 9 de los 16 estudiantes seleccionados estén muy satisfechos con la carrera es del 37.22%. Esto indica que, aunque el nivel de satisfacción general es alto, no es altamente probable obtener una muestra tan extremadamente positiva.

```
> # 5a. Mas de 9 muy satisfechos  
> prob_5a <- 1 - pbinom(9, n, p_muy_satisfecho)  
> cat("5a. P(X > 9) =", round(prob_5a, 4), "\n")  
5a. P(X > 9) = 0.3722
```

- b. La probabilidad de que entre 4 y 8 estudiantes de los 16 seleccionados sientan satisfechos con la carrera es del 74.49%. Este valor elevado refleja que esta categoría es frecuente en la población, por lo que es esperable que aparezca dentro de este rango en una muestra aleatoria

```
> # 5b. Entre 4 y 8 satisfechos  
> prob_5b <- pbinom(8, n, p_satisfecho) - pbinom(3, n, p_satisfecho)  
> cat("5b. P(4 <= Y <= 8) =", round(prob_5b, 4), "\n")  
5b. P(4 <= Y <= 8) = 0.7449
```

- c. La probabilidad de que menos de 5 estudiantes de los 16 seleccionados estén insatisfechos con la carrera es del 99.74%. Dado que la proporción de estudiantes insatisfechos en la población es baja (6.4%), es casi seguro que en una muestra de este tamaño no se superen los 5 casos.

```

> # 5c. Menos de 5 insatisfechos
> prob_5c <- pbinom(4, n, p_insatisficho)
> cat("5c. P(Z < 5) =", round(prob_5c, 4), "\n")
5c. P(Z < 5) = 0.9974

```

- d. La probabilidad de que exactamente 10 de los 16 estudiantes seleccionados estén muy insatisfechos con la carrera es prácticamente nula (0%). Esto se debe a que solo el 7.6% de la población se encuentra en esta categoría, lo cual hace extremadamente improbable observar tantos casos en una muestra tan pequeña.

```

> # 5d. Exactamente 10 muy insatisfechos
> prob_5d <- dbinom(10, n, p_muy_insatisficho)
> cat("5d. P(W = 10) =", round(prob_5d, 4), "\n")
5d. P(W = 10) = 0

```

6. Distribucion de Poisson

- a. La probabilidad de que lleguen al menos 6 consultas en un período de 20 minutos es del 93.29%. Considerando que la tasa promedio en ese tiempo es de 10 consultas, es muy probable superar o igualar este umbral bajo condiciones normales

```

> # 6a. Al menos 6 consultas en 20 minutos
> lambda_20 <- lambda_30min * (20/30) # 10
> prob_6a <- 1 - ppois(5, lambda_20)
> cat("6a. P(X >= 6 en 20 min) =", round(prob_6a, 4), "\n")
6a. P(X >= 6 en 20 min) = 0.9329

```

- b. La probabilidad de que lleguen como máximo 12 consultas en 40 minutos es del 3.90%. Dado que la tasa esperada en ese intervalo es de 20 consultas, un número tan bajo como 12 representa un evento poco común, posiblemente indicativo de baja demanda o error en el registro.

```

> # 6b. A lo sumo 12 consultas en 40 minutos
> lambda_40 <- lambda_30min * (40/30) # 20
> prob_6b <- ppois(12, lambda_40)
> cat("6b. P(X <= 12 en 40 min) =", round(prob_6b, 4), "\n")
6b. P(X <= 12 en 40 min) = 0.039

```

- c. La probabilidad de que lleguen exactamente 8 o 9 consultas en 30 minutos es del 5.19%. Aunque estos valores están cercanos a la media (15), su probabilidad individual es baja, ya que la distribución de Poisson asigna mayor masa probabilística a valores más altos cuando la media es grande.

```

> # 6c. Mas de 7 y menos de 10 en 30 minutos -> P(X=8) + P(X=9)
> prob_6c <- dpois(8, lambda_30min) + dpois(9, lambda_30min)
> cat("6c. P(8 o 9 en 30 min) =", round(prob_6c, 4), "\n")
6c. P(8 o 9 en 30 min) = 0.0519

```

7. Modelo Normal - Estatura

- a. La probabilidad de que un estudiante seleccionado al azar mida 179 cm o más es del 5.26%. Esto indica que aproximadamente 1 de cada 19 estudiantes supera esta altura, lo cual se considera un valor relativamente alto dentro de la distribución.

```

> # 7a. Estatura >= 179 cm
> prob_7a <- 1 - pnorm(179, mu, sigma)
> cat("7a. P(X >= 179 cm) =", round(prob_7a, 4), "\n")
7a. P(X >= 179 cm) = 0.0526

```

- b. La probabilidad de que un estudiante tenga una estatura comprendida entre 147 cm y 172 cm es del 76.17%. Esta franja cubre la mayoría de la población, representando el rango de estaturas más comunes entre los estudiantes de la tecnicatura.

```

> # 7b. Entre 147 y 172 cm
> prob_7b <- pnorm(172, mu, sigma) - pnorm(147, mu, sigma)
> cat("7b. P(147 < X < 172) =", round(prob_7b, 4), "\n")
7b. P(147 < X < 172) = 0.7617

```

- c. El valor de estatura que excede al 97.5% de los estudiantes (es decir, solo el 2.5% mide menos) es de 150.6 cm. Este percentil puede utilizarse como umbral mínimo en estudios descriptivos o diseño ergonómico.

```

> # 7c. Percentil 2.5 (excede al 97.5%)
> valor_7c <- qnorm(0.025, mu, sigma)
> cat("7c. Percentil 2.5 =", round(valor_7c, 1), "cm\n")
7c. Percentil 2.5 = 150.6 cm

```

```

--- PARAMETROS USADOS EN MODELO NORMAL ---
> cat("Media de estatura:", round(mu, 1), "cm\n")
Media de estatura: 166.2 cm
> cat("Desvio estandar:", round(sigma, 1), "cm\n")
Desvio estandar: 7.9 cm

```

8. Suponiendo que los datos corresponden a una población, a través de un Muestreo Aleatorio Simple, seleccione 6 muestras de 20 estudiantes y calcule para cada una de ellas el peso promedio.

- ¿Coinciden los promedios de las muestras con el parámetro?
- ¿Cómo son los promedios muestrales entre sí? Obtenga conclusiones en el contexto del problema planteado.

Media poblacional (peso):

```
> # Media poblacional del peso
> mu_peso <- mean(datos$`PESO KG.`, na.rm = TRUE)
> round(mu_peso, 2)
[1] 63.92
```

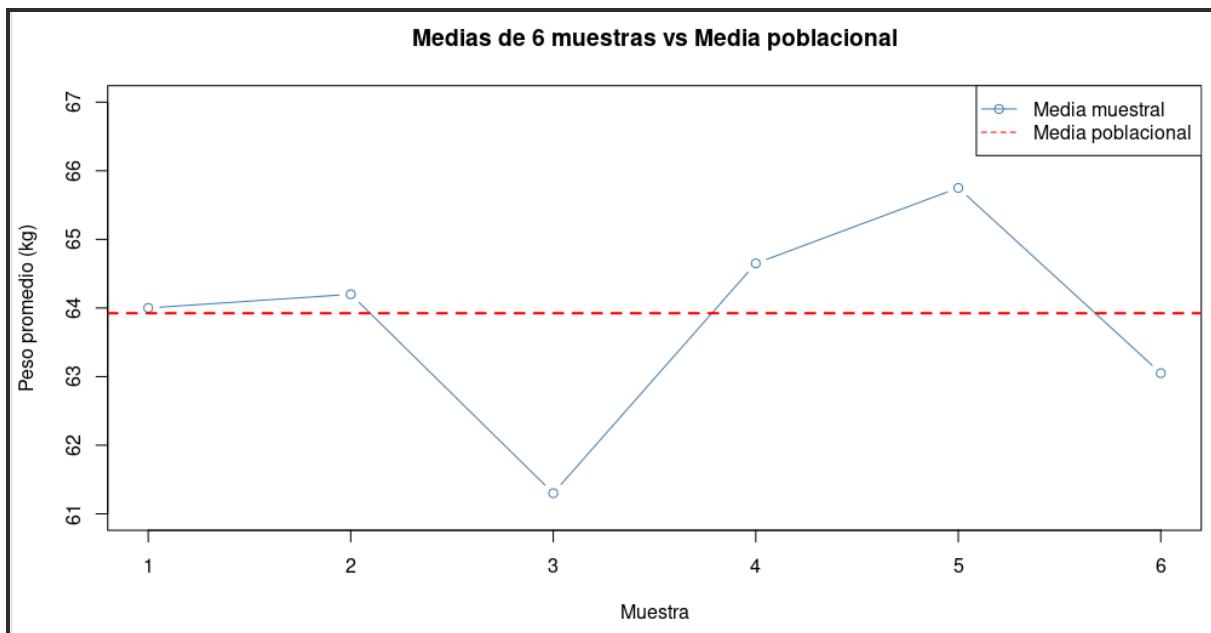
Muestras aleatorias (peso):

```
Muestra 1 : Promedio de peso = 64 kg
Muestra 2 : Promedio de peso = 64.2 kg
Muestra 3 : Promedio de peso = 61.3 kg
Muestra 4 : Promedio de peso = 64.65 kg
Muestra 5 : Promedio de peso = 65.75 kg
Muestra 6 : Promedio de peso = 63.05 kg
```

Media promedio de muestras:

```
> cat("Media promedio de las 6 muestras:", round(mean(medias_muestrales), 2), "kg\n")
Media promedio de las 6 muestras: 63.83 kg
```

Gráfico de medias muestrales (peso):



En esta consigna se realizó un muestreo aleatorio simple de 6 muestras de tamaño 20 estudiantes cada una, con el objetivo de estimar el peso promedio de la población.

La media poblacional del peso es de 63.92 kg. Las medias obtenidas en las 6 muestras fueron: 64.00, 64.20, 61.30, 64.65, 65.75 y 63.05 kg.

Se observa que ninguna media muestral coincide exactamente con el parámetro, lo cual es esperable debido al error de muestreo. Sin embargo, todas están próximas al valor real, y su media promedio (63.83 kg) es prácticamente idéntica a la poblacional.

Además, el desvío estándar de las medias (1.52 kg) indica que las estimaciones son estables y consistentes.

En conclusión, el muestreo aleatorio simple permite obtener estimaciones confiables del parámetro poblacional, incluso con muestras relativamente pequeñas.