

PROBABILIDAD Y ESTADÍSTICA

GUÍA DE EJERCICIOS

Unidad 1: ESTADISTICA DESCRIPTIVA.

Cuestionario Orientador para la Investigación Bibliográfica.

- 1) En estadística se manejan los conceptos de población y de muestra. ¿Qué se entiende por población? ¿Qué elementos pueden formar parte de una población?
- 2) ¿Qué es una muestra? ¿Cómo se pueden armar muestras representativas de la población en estudio? ¿Qué propiedades debe reunir una muestra?
- 3) Los datos que pueden adquirirse mediante un muestreo, pueden ser denominados como cualitativos o cuantitativos; a su vez, los cuantitativos pueden trabajarse como continuos o discretos. Elaborar definiciones para cada uno de estos tipos de datos. Especificar cuál es la diferencia entre ellos
- 4) ¿Cómo se pueden organizar los datos cualitativos? ¿Qué tipo de análisis estadístico se puede realizar con este tipo de datos? ¿Cuáles son las limitaciones en el tratamiento de este tipo de datos?
- 5) ¿Cómo se pueden organizar los datos cuantitativos? ¿Cuál es el procesamiento posible para este tipo de datos? ¿Qué diferencias se pueden encontrar entre el tratamiento de variables discretas y continuas? ¿Qué ocurre con el procesamiento cuando hay pocos o muchos datos?
- 6) Elaboremos una clasificación de todos los gráficos que se estudian en la unidad uno y especifiquemos para qué tipo de variable se utiliza cada uno.
- 7) En el procesamiento de datos cuantitativos se pueden utilizar medidas de ubicación, de dispersión o de forma. Explicar qué tipo de información brinda cada una.
- 8) Realizar un resumen de las principales medidas estadísticas y definir el modo en que se determina cada una. Explicar además, cuáles son las ventajas y desventajas de cada una de esas cantidades como medidas descriptivas
- 9) Analizar al menos un par de libros de Estadística y determinar qué propiedades debe poseer la media aritmética o promedio.
- 10) Los datos de tipo cuantitativo continuo se pueden agrupar en intervalos. Explicar cuáles son los requisitos que debe tener la muestra para que este agrupamiento sea factible. Explicar además qué información relevante se puede obtener a partir del agrupamiento.
- 11) En un proceso productivo se fabrican ruedas de automóviles, después del balanceo se mide la diferencia de carga en gramos, entre las dos caras de la cubierta. Los valores obtenidos son los siguientes:

2 PROBABILIDAD Y ESTADÍSTICA

23	22	21	34	31	10	13
----	----	----	----	----	----	----

11-a) Calcular el promedio y el desvío de esta muestra.

11-b) Clasificar los datos en intervalos, de acuerdo con la tabla siguiente:

Intervalo	Frecuencia
10 - 19,9	
20 - 29,9	
30 - 40	

Una vez completa la tabla anterior, determinar el promedio y el desvío con los datos agrupados.

11-c) Comparar los resultados obtenidos en los dos ítems anteriores. Explicar por qué motivo son diferentes los resultados. Elaborar una opinión sobre la conveniencia de calcular las medidas a partir de datos agrupados.

12) ¿Qué tipo de conclusiones se pueden elaborar a partir del análisis descriptivo de datos? ¿Cuál es su utilidad?

13) Una de las propiedades que se analizan al estudiar una muestra de datos cuantitativos continuos, es la asimetría. Expliquemos qué es lo que representa la asimetría. Identifiquemos medidas que permiten valorarla.

Ejercicios Propuestos.

1. Una cooperativa telefónica de cierta ciudad, que tiene varios miles de habitantes, decide realizar una encuesta telefónica entre sus abonados, indagando sobre diversos aspectos con el objetivo de explorar la posible ampliación de servicios. Se consideran en el estudio únicamente las casas de familia.

- Defina claramente cuál es la población correspondiente a este estudio.
- Explique de que manera seleccionaría una muestra representativa de esa población para implementar la encuesta.
- A continuación se listan los datos que se solicitan a cada uno de los encuestados. Indicar el tipo de cada uno:

- número promedio de pulsos por bimestre
 - duración de la llamada más larga hecha en el mes
 - color del aparato telefónico
 - facturación del último bimestre
 - ocupación del jefe de familia
 - número de integrantes del grupo familiar
 - barrio en que está ubicada la vivienda
 - si tiene un módem conectado a una computadora
-

2. Un relevamiento realizado sobre los alumnos de una carrera universitaria, arrojó como resultado que, entre los que trabajan, 45 lo hacen en relación de dependencia, 57 son trabajadores independientes, 12 son comerciantes o industriales, mientras que 6 ya realizan tareas profesionales. Organizar la información para facilitar su presentación e interpretación.
3. Se desea estudiar en determinada empresa la insatisfacción en el trabajo, se realiza una encuesta y se registran las causas de insatisfacción.

<i>Causa de Insatisfacción</i>	Infraestructura	Horario	Capacitación	Sueldo	Otros
<i>Frecuencia</i>	18	4	2	30	6

- a) Ordenar los datos por frecuencias decrecientes, considerando el ítem “Otros” (que agrupa varias causas menores de defectos) al final de la escala.
 - b) Calcular las frecuencias relativas, y dibujar el diagrama de barras con esas frecuencias.
 - c) Analizar el diagrama obtenido. ¿Qué interpretación se puede realizar?.
 - d) ¿Cuáles serían las conclusiones o recomendaciones de ese análisis?.
4. Durante un período de 30 días se realiza un estudio sobre el número de reclamos diarios por problemas en los sistemas de computación desarrollados por una cierta Compañía. Los valores registrados durante ese lapso fueron:

3 2 1 2 1 3 0 2 1 4 4 0 1 2 0
2 5 1 4 3 2 1 3 4 3 1 2 5 0 2

- a) Elaborar la tabla de frecuencias simples y acumuladas.
 - b) Graficar las frecuencias relativas.
 - c) Analizar la distribución de frecuencias de la variable.
 - d) Calcular el porcentaje de días en que se producen hasta 2 reclamos.
 - e) Si se producen más de 3 reclamos, hay que subcontratar el servicio de mantenimiento. Calcular el porcentaje de veces que eso ocurre.
5. En el proceso de control final de tanques de combustible para automóviles, se realiza una prueba para verificar la estanqueidad del recipiente. Suponga que estas observaciones se organizan del siguiente modo: se observan cien tanques consecutivos y se registra la cantidad de depósitos que pierden; luego se observan otras cien unidades y nuevamente se registra el número de tanques con pérdidas. De este modo, se inspeccionan sesenta muestras y se obtienen los siguientes valores.

4	2	1	7	2	0	5	4	3	5
2	4	4	1	1	2	4	1	5	4
3	2	3	4	4	6	5	5	2	2
3	4	4	4	2	6	2	5	6	4
1	2	0	4	5	3	1	3	5	4
2	3	3	5	5	9	4	5	4	3

4 PROBABILIDAD Y ESTADÍSTICA

Analizar qué tipos de datos son estos. Construir una Tabla de Frecuencias y elaborar gráficos adecuados. Calcular las principales medidas. La Dirección de la empresa había adoptado como objetivo que la cantidad de tanque con problemas debía ser menor o igual que cinco; analizar si dicho objetivo puede considerarse satisfecho.

6. Se desea comparar el rendimiento alcanzado en un curso de capacitación por los operarios de los turnos Mañana y Tarde de una cierta fábrica. Para ello dispone de los puntajes [de 0 a 100] obtenidos en la evaluación final:

Turno Mañana							Turno Tarde						
63	55	68	83	71	65	56	76	79	54	91	72	48	47
54	47	55	75	81	73	69	53	79	67	56	62	57	49
68	69	73	34	76	61	69	54	48	58	59	59	56	47
59	59	66	46	43	72	64	51	52	69	62	69	52	71
79	54	92	69	57	75	73	38	59	66	60	71	71	73
72	82	48	67	71	64	65	64	69	60	67	57	95	53
36	49	89	59	66	62	49	81	79	64	62	83	64	59
55							45						

- Elaborar la tabla de frecuencias correspondiente a los resultados de cada turno. Utilice intervalos de longitud 10, comenzando desde el cero.
- Dibujar los histogramas de frecuencias relativas.
- Dibujar los polígonos de frecuencias.
- Calcular el porcentaje de operarios del Turno Mañana que sacaron entre 50 y 80 puntos, y la cantidad de operarios del Turno Tarde que sacaron más de 70 puntos.
- Comparar las dos distribuciones de frecuencias. Explicar en qué se parecen, y en qué se diferencian.
- Formular una opinión sobre el rendimiento evidenciado en el curso por los dos turnos de trabajo.

7. Analizar las siguientes muestras:

Variable analizada	Datos de la muestra						
Duración de una tarea	10	7	13	9	11	19	1
Duración de una lamparita	6	28	2	11	18	5	7
Duración del cartucho de tinta de una impresora	20	17	10	6	16	1	14

En todos los casos realizar las siguientes tareas:

- Dibujar un diagrama de puntos.
- Dibujar el diagrama de caja.
- Calcular las medidas de posición y de dispersión.
- Interpretar los resultados obtenidos.

8. Usando datos de los ejercicios anteriores:

- a) Indicar para el Ejercicio N°3, si pueden calcular medidas analíticas. Fundamentar la respuesta.
- b) Calcular para el Ejercicio N°4, las medidas de posición. Comparar la media, mediana y moda, y comentar conclusiones. Calcular las medidas de dispersión.
- c) Calcular para el Ejercicio N°5, las medidas analíticas para los puntajes correspondientes a los dos turnos de trabajo. Comparar los resultados obtenidos, y elaborar conclusiones al respecto. Si el cálculo de las medidas analíticas se realiza en base a los datos agrupados ¿los resultados van a ser los mismos? Fundamentar la respuesta.

9. Los siguientes datos muestran 80 mediciones de la emisión diaria (en toneladas) de óxido de azufre de una planta industrial:

15.8 - 26.4 - 17.3 - 11.2 - 23.9 - 24.8 - 18.7 - 13.9 - 9.0 - 13.2
 22.7 - 9.8 - 6.2 - 14.7 - 17.5 - 26.1 - 12.8 - 28.6 - 17.6 - 23.7
 26.8 - 22.7 - 18.0 - 20.5 - 11.0 - 20.9 - 15.5 - 19.4 - 16.7 - 10.7
 19.1 - 15.2 - 22.9 - 26.6 - 20.4 - 21.4 - 19.2 - 21.6 - 16.9 - 19.0
 18.5 - 23.0 - 24.6 - 20.1 - 16.2 - 18.0 - 7.7 - 13.5 - 23.5 - 14.5
 14.4 - 29.6 - 19.4 - 17.0 - 20.8 - 24.3 - 22.5 - 24.6 - 18.4 - 18.1
 8.3 - 21.9 - 12.3 - 22.3 - 13.3 - 11.8 - 19.3 - 20.0 - 25.7 - 31.8
 25.9 - 10.5 - 15.9 - 27.5 - 18.1 - 17.9 - 9.4 - 24.1 - 20.1 - 28.5

Describir la muestra utilizando convenientemente los conceptos estudiados.

10. El Departamento de Control de Calidad de una fábrica seleccionó una muestra de remaches para evaluar su resistencia. Estos remaches se utilizan para la fijación de una pieza que, de desprenderse, pone en riesgo la vida del usuario de la maquinaria que se produce. Los valores obtenidos (expresados en Newton) son los siguientes:

18,56 18,57 18,55 18,58 18,60 18,59

Sabemos que si los procesos se encuentran bajo control, la dispersión debería ser reducida y el comportamiento de la variable bajo estudio debería poder describirse mediante un modelo normal. Para verificarlo:

- a) Realice una descripción adecuada de la muestra.
- b) Con base en el estudio realizado, ¿puede pensarse que la muestra proviene de una población con distribución Normal? Justifique utilizando todos los argumentos que le brinda la estadística descriptiva.

Ejercicio Adicional Integrador de la Unidad

Caso de estudio: la empresa de transporte

Una empresa de transporte de cargas desea investigar las características del servicio que brinda. Interesa en particular uno de dichos servicios. Se trata del viaje entre dos ciudades, a las cuales puede denominarse como A y B. Las variables que deben estudiarse son las siguientes:

Tipo de producto transportado: el mismo reconoce los siguientes valores:

Tipo de producto	Código
Papelería	1

Insumos industriales	2
Maquinarias	3
Herramientas	4
Residuos no peligrosos	5
Insumos agrícolas	6
Otros	7

En la tabla anterior se utilizan códigos para facilitar el manejo de la información. Los mismos no tienen entidad numérica. Simplemente, para no escribir muchas veces insumos industriales, por ejemplo, se utiliza simplemente el código 2.

Controles camineros: la variable representa el hecho de que el camión haya o no recibido controles camineros en su recorrido. Los valores y códigos son los siguientes:

Controles camineros	Código
Tuvo	1
No tuvo	2

Medidas de seguridad: previo a cada viaje se verifican diversos aspectos relacionados con la seguridad de las unidades. El criterio es que el camión no puede partir si tiene un problema de seguridad serio. En cambio se toleran problemas leves. Los valores y códigos son los siguientes:

Medidas de seguridad	Código
En perfectas condiciones	1
Algún problema leve	2
Más de un problema leve	3

Problemas de mecánica: al finalizar cada viaje se revisan las unidades con el auxilio de una lista de chequeo y se determina la cantidad de problemas que se encuentran. Esta variable representa la cantidad de problemas encontrados. Por ejemplo, si el camión tiene problemas para el arranque y una de las mangueras del motor parece a punto de romperse, la variable toma el valor dos.

Tiempo de viaje: mide el tiempo transcurrido desde que la unidad parte, hasta que arriba a destino. Se mide en minutos.

Tiempo de reparación posterior: mide el tiempo que se destina a la reparación, cuando se encuentran problemas mecánicos. Se mide en horas. Debe notarse que en caso de no presentarse problemas, la variable no se observa.

A fin de realizar el estudio, la empresa registra las variables mencionadas para sesenta viajes consecutivos. Los resultados obtenidos se presentan en una tabla al final de este enunciado.

Actividad 1: Población y muestra

Trabaje sobre el caso de la empresa de transporte que se presenta en TABLA 1, conforme a las siguientes consignas:

- Identifique las diferentes variables que se consideran en dicho caso.

- ❑ Especifique cuál es la población y cuál la muestra, para dicho caso.
- ❑ Identifique a qué tipo de variable corresponde: cualitativa, discreta o continua.

Actividad 2: Tratamiento de variables cualitativas

Las variables cualitativas deben ser tratadas mediante la construcción de tablas de frecuencia y gráficos adecuados. Por otra parte, es importante realizar una interpretación adecuada de los resultados obtenidos.

La propuesta es que usted trabaje las variables cualitativas del caso de la empresa de transporte. Para facilitar su avance, tenga presente el material bibliográfico y los datos de la tabla 1.

Actividad 3: Tratamiento de variables discretas

Los camiones utilizados para los servicios de transporte sufren desgaste y averías, lo que obliga a su reparación. Sucede que una reparación de urgencia siempre tiene un costo mayor que una intervención que se ha programado previamente.

Por ese motivo, nuestra empresa de transporte se encuentra interesada en desarrollar un buen sistema de mantenimiento preventivo de los camiones. La idea es que las actividades de reparación sean programadas y se reduzcan al mínimo las situaciones de urgencia.

Una de las consecuencias de esta iniciativa es que en la revisión que se hace rutinariamente al final de cada viaje, la variable **problemas de mecánica** tienda a cero. Si el preventivo es absolutamente eficaz, la cantidad de problemas encontrados al finalizar los viajes debe ser mínima.

Por ese motivo, la empresa estableció una meta para la variable **problemas de mecánica**. Se desea que la cantidad de problemas encontrados sea menor o igual que dos.

Efectúe el análisis descriptivo de esa variable y elabore una opinión acerca de si se logró o no alcanzar la meta establecida.

Actividad 4: Tratamiento de variables continuas

En el caso de la empresa de transporte, las variables cuantitativas analizadas son el Tiempo de viaje y el Tiempo de reparación posterior, cuando se producen desperfectos. Para estos datos, debemos aplicar toda la batería de herramientas estadísticas que conocemos.

Pero además, consideremos que hemos realizado esfuerzos para que tanto el viaje como las reparaciones constituyan procesos ordenados, controlados y previsibles. Esto es, hemos entrenado a la gente que hace el trabajo, compramos herramental y equipos, efectuamos un importante esfuerzo para garantizar el correcto mantenimiento de las unidades.

Si estas tareas hubieran dado el resultado esperado, entonces las dos variables deberían variar muy poco y tener una distribución de frecuencias compatible con las propiedades del modelo Normal de probabilidades.

Con esta información, debemos:

- Analizar la variable Tiempo de viajes, para determinar si se satisface el criterio adoptado para valorar el grado de control sobre las causas que afectan la duración de esta tarea.
 - Analizar la variable tiempos de reparación, para determinar si se satisface el criterio adoptado para valorar el grado de control sobre las causas que afectan la duración de esta tarea.
-

8 PROBABILIDAD Y ESTADÍSTICA

- Separar los tiempos de viajes de aquellos recorridos que tuvieron a lo sumo un problema mecánico, de los que tuvieron dos o más problemas. Analizar si puede suponerse que las propiedades de la variable cambian cuando es elevado el número de problemas.

TABLA 1: Datos obtenidos del muestreo CASO EMPRESA DE TRANSPORTE

Viaje nro.	Tipo de Producto	Controles	Medidas Seguridad	Problemas de Mecánica	Tiempo viaje (En minutos)	Tiempo reparación posterior (En horas)
1	2	2	2	1	714	12,87
2	2	2	1	3	924	12,01
3	4	2	1	0	775	
4	6	2	1	2	836	0,47
5	6	2	1	2	662	5,58
6	7	1	3	0	791	
7	1	2	2	4	641	1,14
8	2	1	2	3	734	3,20
9	6	2	1	3	686	4,95
10	2	2	1	0	720	
11	2	2	2	1	793	5,47
12	2	2	1	1	677	0,28
13	1	2	2	1	718	11,28
14	2	1	1	0	789	
15	2	2	2	2	689	7,44
16	1	2	1	2	628	0,86
17	2	2	2	1	707	1,73
18	2	1	2	2	735	14,59
19	4	2	1	1	721	7,91
20	2	1	2	0	714	
21	2	1	3	0	716	
22	2	1	2	1	783	2,94
23	6	1	1	0	640	
24	4	2	2	2	766	2,49
25	3	1	1	3	796	2,01
26	2	1	1	1	636	1,94
27	7	2	2	1	756	0,32
28	5	1	3	2	805	6,17
29	7	2	2	0	830	
30	2	2	1	1	783	9,08
31	7	1	1	2	856	1,99
32	2	2	1	1	798	0,37
33	4	2	1	1	684	6,00
34	5	2	1	2	571	12,45
35	7	2	1	1	646	8,19
36	4	1	2	0	600	

37	2	2	1	0	727	
Viaje nro.	Tipo de Producto	Controles	Medidas Seguridad	Problemas de Mecánica	Tiempo viaje (En minutos)	Tiempo reparación posterior (En horas)
38	5	2	1	2	820	14,06
39	2	2	2	0	732	
40	5	2	2	1	767	3,28
41	2	2	2	2	596	3,31
42	2	1	1	2	646	0,64
43	2	2	1	2	604	2,68
44	2	2	1	0	652	
45	4	2	1	1	663	0,64
46	4	2	1	0	797	
47	2	2	3	2	729	1,46
48	7	2	1	0	717	
49	2	1	2	3	656	5,24
50	2	1	1	0	731	
51	1	2	1	0	920	
52	7	2	1	3	772	17,06
53	2	2	1	0	693	
54	2	2	3	1	631	1,68
55	5	2	2	2	754	0,26
56	4	1	1	1	795	7,77
57	5	2	2	1	792	7,51
58	5	2	1	0	750	
59	2	2	1	2	748	15,46
60	2	1	1	1	627	1,65

Unidad 2: PROBABILIDAD y VARIABLES ALEATORIAS.

Cuestionario Orientador para la Investigación Bibliográfica.

- 1) ¿Qué se entiende por experimento aleatorio? Explique en qué consiste el espacio muestral de un experimento aleatorio. ¿Qué es un suceso (o evento)?
- 2) ¿Cómo se puede definir el concepto de probabilidad? ¿Cuáles son sus propiedades?
- 3) Explique de qué manera se pueden asignar probabilidades, y cuál es la regla general de asignación.
- 4) Explique qué significa afirmar que dos sucesos sean: excluyentes, exhaustivos, independientes. ¿Qué se entiende por complemento de un suceso?
- 5) ¿En que consisten las reglas de la suma y del producto de probabilidades? ¿Para qué se utilizan?

10 PROBABILIDAD Y ESTADÍSTICA

- 6) ¿Cómo afecta la regla de la suma el hecho de que los eventos sean excluyentes o no excluyentes?
- 7) ¿Cómo afecta la regla de la multiplicación el hecho de que los eventos sean dependientes o independientes?
- 8) Elabore una definición de probabilidad condicional, y explique cómo se calcula.
- 9) ¿Qué es una variable aleatoria? ¿Qué representa la distribución de probabilidades de una variable aleatoria?
- 10) ¿Cuál es la diferencia entre las variables aleatorias discretas y continuas? ¿Cómo es la representación gráfica de c/u?
- 11) ¿Cuáles son las propiedades de una función de probabilidad? ¿Y de una función de densidad de probabilidad?
- 12) ¿Qué es la función de distribución acumulada? ¿Para qué se utiliza? ¿Cómo se la calcula?
- 13) ¿De qué manera se puede caracterizar el comportamiento de una variable aleatoria (o de su distribución de probabilidades)?
- 14) Explique cómo se calcula el valor esperado y la varianza de una variable aleatoria. Indique qué significado tiene cada una de esas cantidades estadísticas.

Ejercicios Propuestos.

1. Proponer ejemplos de Experimentos Aleatorios. Para cada experimento determinar si el espacio muestral es de tamaño finito o infinito. Dentro de cada espacio definir eventos excluyentes y eventos no excluyentes. Acordar con el grupo el significado de la afirmación “un evento ha ocurrido”.
2. El texto propone determinar la probabilidad de ocurrencia de un evento contando la cantidad de puntos favorables y la cantidad de posibles en el espacio muestral, y luego hacer:

$$P(A) = n_A / n \quad \text{en cambio, en la introducción se planteó:}$$

$$P(A) = \lim_{n \rightarrow \infty} (n_A / n)$$

Elaborar una opinión sobre las situaciones en las que se aplica cada regla. Decidir cuál de ellas es la general.

3. Un edificio tiene dos ascensores. La probabilidad que en un cierto momento funcione el ascensor A es de 0,90; por su parte, la probabilidad de que funcione B es de 0,80; la probabilidad que funcionen los dos de manera independiente es 0,72. Si en un día de cuarenta grados llegamos al edificio para visitar a un amigo que vive en el décimo piso, ¿cuál es la probabilidad de que no se deba subir por la escalera?
 4. De un mazo de naipes españoles (40 naipes) se extrae una carta al azar. Determinar la probabilidad de que:
 - a) la carta sea un REY
 - b) la carta sea una COPA
 - c) la carta sea un REY o una COPA.
 - d) la carta no sea ni REY ni COPA.
 5. Proponer ejemplos en que se apliquen los conceptos de probabilidad condicional y de independencia.
 6. En términos de probabilidades: ¿es lo mismo decir: $P(A \cap B)$ que $P(B \cap A)$?
 7. En una caja hay diez piezas, de las cuales siete son Buenas (B) y las restantes son Defectuosas (D). Se extraen dos piezas, sucesivamente y sin reposición.
 - a) Determinar la Probabilidad de que:
 - las dos sean buenas.
 - la 1° sea buena y la 2° defectuosa.
 - la 1° sea defectuosa y la 2° buena.
 - las dos sean defectuosas.
 - b) Listar todos los **eventos compuestos** del punto anterior y verificar que la suma de sus probabilidades sea uno.
 - c) Calcular la Probabilidad de que la primera pieza sea Buena y comparar con la Probabilidad Marginal del resultado Buena.
 - d) Calcular la Probabilidad de que la segunda pieza sea Buena y comparar con la Probabilidad Marginal del resultado Buena.
 8. Los eventos pueden ser clasificados en
Excluyentes o No excluyentes
Dependientes o Independientes
Establecer relaciones entre las dos clasificaciones posibles, desarrollando proposiciones del tipo: “Si dos eventos son excluyentes entonces son”.
 9. Una planta extractora de aceite recibe de tres fincas cargamento con olivares. En porcentaje la cantidad que recibe de la Finca “Del Sol” es 15%, Finca “Las Tres Maria” 27% y de la Finca “Agrocatamarca” 58%, siendo la proporción de cultivares en mal estado 5%, 7% y 17% respectivamente. Las distintas fincas tienen un porcentaje de materia prima en mal estado debido al sistema de recolección. Este
-

sistema consiste en recoger los cultivos que están en la tierra de los cuales una parte considerable presentan un estado de putrefacción. El Gerente de Calidad necesita saber: ¿cuál es la probabilidad de que la planta extractora de aceite reciba un cargamento con olivares en mal estado? Si esa probabilidad sobrepasa el 10% les exigirá a cada una de las fincas optimizar la recolección para de esta manera obtener menor porcentaje de semillas en mal estado.

10. En la producción de la corona de una caja de cambio para automóvil, las operaciones fundamentales son:

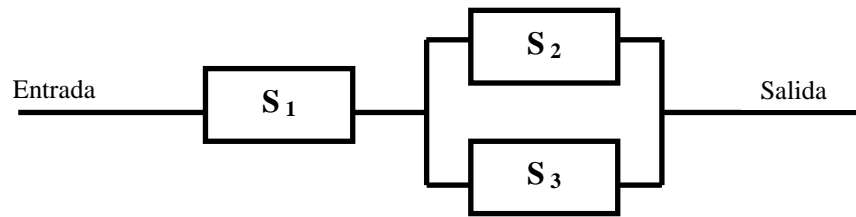
Dentado: en la cual se crean los dientes cortando el acero;

Afeitado: donde se logra la terminación de la pieza, asegurando características como diámetro, hélice, envolvente, separación entre dientes, excentricidad, etc.

La probabilidad de encontrar una pieza que no cumpla con las especificaciones técnicas (No conformidad (NC) como consecuencia de problemas en el dentado es 0,05, en tanto que por fallas de afeitado asciende a 0,12. Por otra parte, la probabilidad que la pieza presente defectos en las dos operaciones es de 0,03.

- a) Determinar la probabilidad que una pieza se considere No Conforme por alguno de esos dos motivos.
 - b) Determinar la probabilidad que falle el afeitado, dado que ha fallado el dentado.
 - c) ¿Son independientes los dos tipos de no conformidad?
11. Una fábrica se abastece de dos fuentes de energía, denominadas A y B. La probabilidad de que la fuente A funcione es de 0.97, que funcione B es 0.95, y que funcionen ambas simultáneamente es de 0.965. Responder:
- a) ¿Son independientes ambas fuentes de abastecimiento? Justificar.
 - b) ¿Cuál es la probabilidad de que funcione la fuente B sabiendo que está funcionando A?
12. Una fábrica distribuye el modelo M24 de cortadora de césped en el interior de la provincia, haciendo la entrega en tiempo y forma el 85% de las veces. De las entregas que realiza en tiempo y forma, sólo el 2% presenta defectos en su embalaje; en tanto que de las entregas que no realizan en tiempo y forma, el 3% presenta defectos en su embalaje.
- a) Calcule la probabilidad de que un pedido presente defectos en su embalaje y no llegue en tiempo y forma.
 - b) Si una entrega no presenta defectos en su embalaje ¿Cuál es la probabilidad de que haya llegado en tiempo y forma
13. En un proceso productivo un obrero atiende 3 máquinas M1, M2 y M3 que funcionan independientemente. En un día cualquiera la probabilidad de falla de M1 es 0,1, mientras que para las otras máquinas es de 0,05. ¿Cuál es la probabilidad de que en un día determinado fallen dos máquinas?
-

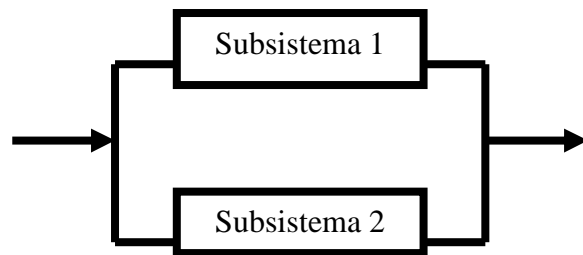
14. En una fábrica hay dos máquinas M1 y M2 para producir los componentes de un aparato. M1 produce el 40% de las componentes y M2 el 60%. Se sabe que el 5% de los componentes producidos por M1 y que el 2% de los producidos por M2, son defectuosos. Si un elemento defectuoso es extraído al azar. ¿Cuál es la probabilidad de que haya sido producido por M1?
15. La confiabilidad o fiabilidad (R) se define como la probabilidad de que un sistema funcione adecuadamente durante un período dado en su aplicación prevista. La siguiente figura esquematiza un sistema constituido por tres componentes tales que, el sistema cumple adecuadamente con su objetivo si funciona el dispositivo S_1 y alguno de los dispositivos S_2 o S_3 . Con base en lo anterior y los valores de fiabilidad de cada dispositivo, calcule la fiabilidad del sistema de la figura sabiendo que: $R(S_1)=0.8$ $R(S_2)=0.6$ $R(S_3)=0.9$



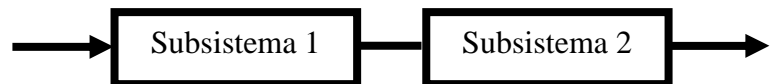
16. Los siguientes esquemas representan sendos procesos que se desarrollan en cierta fábrica. Con el objetivo de calcular la fiabilidad de cada sistema se estudia la confiabilidad de cada subsistema. Considere los valores de confiabilidad indicados más abajo para cada subsistema.

$R(S_1) = 0.02$ y $R(S_2) = 0.06$ en ambos casos. Suponiendo que los subsistemas funcionan independientemente, calcule la fiabilidad de cada sistema.

a) Sistema 1



b) Sistema 2



17. Se considera una célula en el instante $t=0$. En el instante $t=1$ la célula puede: o bien reproducirse, dividiéndose en dos con probabilidad $3/4$, o bien morir con probabilidad $1/4$. Si la célula se divide, entonces en el tiempo $t=2$ cada uno de sus dos descendientes puede también subdividirse o morir, con las mismas probabilidades que antes,

independientemente uno de otro. En el tiempo $t=2$, ¿cuál es la cantidad máxima posible de células y cuál es la probabilidad de que se dé dicha cantidad?

18. Sean A y B dos sucesos asociados con un determinado experimento. Se sabe que $P(A) = 0.4$ y que $P(A \cup B) = 0.7$. Calcular la probabilidad del suceso B, bajo las siguientes condiciones:

- Los sucesos son excluyentes.
- Los sucesos son independientes.

19. En una encuesta de mercado para un supermercado se clasificó a 1000 clientes del negocio según el sexo y su lugar de residencia (ya sea en el barrio donde se sitúa el negocio -Local- o en barrios vecinos -Alrededores-). La proporción de los clientes que responden a cada una de las 4 categorías son:

	Hombres	
Mujeres		
Local	0,17	0,67
Alrededores	0,04	0,12

Suponiendo que se selecciona al azar de este grupo una persona cualquiera. Calcular:

- b) La probabilidad de que el cliente resida en los alrededores.
 - c) La probabilidad de que el cliente sea hombre.
 - d) La probabilidad de que el cliente sea mujer y viva en el barrio.
 - e) Determinar la cantidad de clientes mujeres que tiene el negocio.
 - f) ¿Cuántos clientes son del propio barrio?
- ¿Qué proporción de los clientes femeninos del negocio son de barrios vecinos?
20. Una empresa de transporte de cargas desea investigar las características del servicio que brinda. Interesa en particular uno de dichos servicios. Se trata del viaje entre dos ciudades, a las cuales puede denominarse como A y B.
- En este momento se investiga si existe asociación entre el tiempo que insumen los camiones en realizar dicho viaje y el hecho de que hayan sido controlados por la caminera. A los fines de organizar la información, se categoriza el tiempo (medido en minutos) en tres categorías a saber:
- Normal (Más de 689 minutos y menos de 806 minutos)
 - Menos de lo normal (Entre 571 minutos y 689 minutos)
 - Más de lo normal (Entre 806 minutos y 904 minutos)

Con base en la información obtenida a partir de un gran número de viajes, se obtienen dos tablas de contingencia. La primera de las tablas se muestra a continuación:

		Controles		
		Tuvo	No tuvo	
Tiempo de viaje (en minutos)	Más de lo normal	0,1	0,22	
	Normal	0,18	0,40	0,58
	Menos de lo normal			0,10
		0,30		

Tabla 1

Como notará la tabla está incompleta, será necesario completar dicha información y justificar adecuadamente los pasos instrumentados para tal fin.

- Identifique el tipo de probabilidades que utiliza para completar la tabla.
- Explicite el evento cuya probabilidad de ocurrencia es 0,18. Indique el tipo de evento.
- Calcule la probabilidad de que un camión elegido al azar haya tenido control caminero o el tiempo de viaje haya sido el esperado.
- Se sabe de antemano que el tiempo que ha insumido un viaje es más de lo esperado. Indique la probabilidad de que no haya sido controlado por la caminera. ¿Qué puede decir acerca de la independencia de estas situaciones?
- La situación que dio motivo a la confección de la tabla precedente (Tabla 1) era determinar si existe asociación entre el tiempo que insumen los camiones en realizar el viaje y el hecho de que hayan sido controlados por la caminera. A partir de la información presentada en la correspondiente tabla, ¿es posible resolver dicho interrogante? En caso de ser posible realice el procedimiento correspondiente y justifíquelo.
- Otra de las tablas confeccionada es la siguiente:

	Más de lo normal	Normal	Menos de lo normal
Tuvo	0,33	0,60	0,07
No tuvo	0,31	0,57	0,11

Tabla 2

Caracterice la información presentada en la Tabla 2. Establezca las comparaciones pertinentes con la información presentada en la Tabla 1.

IMPORTANTE: Recomendamos realizar una síntesis de lo abordado hasta el momento al respecto de las operaciones que pueden realizarse con las probabilidades, como también las propiedades que permitan optimizar las estrategias de cálculo

- Las variables aleatorias tienen siempre una población y una función de probabilidad. Proponer ejemplos y construir los correspondientes gráficos.
- Sea el experimento consistente en arrojar dos monedas y ver la cara expuesta. Especifique el espacio muestral de este experimento. Se define la variable aleatoria X : *Número de caras en dos lanzamientos*. Determinar la distribución de probabilidades de X y confeccionar el gráfico de la función de probabilidad.

23. Calcular la Función de Distribución Acumulada para la variable aleatoria del ejercicio anterior. Calcular la probabilidad de que se obtengan menos de dos caras en los dos lanzamientos de la moneda, y que se obtenga al menos 1 cara.
24. Un distribuidor comercializa un artículo que tiene un precio unitario de \$ 20. Sin embargo sus clientes le reconocen los siguientes adicionales:
- \$ 18 si es embarcado de la fábrica en buenas condiciones y llega a tiempo;
 - \$ 8 si es embarcado en la fábrica en buenas condiciones, pero llega fuera de plazo.

Además se dispone de la siguiente información:

- El 50% de los artículos llegan a tiempo y son embarcados en buenas condiciones.
- El 30% tienen buenas condiciones pero no llegan a tiempo.
- El 20% no son embarcados en buenas condiciones

¿Cuál es el ingreso medio que puede esperar el distribuidor por artículo?

25. Un empleado de un concesionario de ventas de gaseosas en un estadio de fútbol debe elegir entre trabajar en el mostrador recibiendo una suma fija de \$ 50 por jornada, o recorrer las tribunas trabajando a comisión. Si elige esta última alternativa, el empleado puede ganar \$ 90 durante una noche calurosa, \$ 70 en una noche templada, \$ 45 en una noche fresca y \$ 15 en una noche fría. En la época del año considerada las probabilidades de noche calurosa, templada, fresca y fría son respectivamente 0,1; 0,3; 0,4 y 0,2. Determinar que forma de trabajo le conviene elegir a este empleado, justificando la respuesta.
26. Una A.F.J.P. tiene una cierta cantidad de dinero para invertir, y dispone de tres alternativas de inversión. Los resultados -Ganancia o Pérdida- de cada alternativa dependen de la evolución de la economía, de acuerdo a lo que se indica en la siguiente tabla:

Evolución de la Economía	Alternativa A	Alternativa B	Alternativa C
Declina	\$ 500	-\$ 2000	- \$ 7000
Se mantiene estacionaria	\$ 1000	\$ 2000	- \$ 1000
Se expande	\$ 2000	\$ 5000	\$ 20000

En base a su experiencia la A.F.J.P. sabe que existe una probabilidad de 0,30 de que la economía decline, de 0,50 de que se mantenga estacionaria, y de 0,20 de que se expanda. Analizar estos datos e identificar la mejor alternativa.

27. Tres máquinas funcionan en paralelo. La probabilidad de que una máquina cualquiera no funcione es de 0,30. Se define la variable aleatoria X: *Nº de máquinas paradas en un instante dado*.
-

- a. Determinar la distribución de probabilidades de dicha variable y calcular las probabilidades de que en un instante cualquiera haya paradas menos de dos máquinas, que haya paradas dos ó más máquinas; que haya paradas una ó dos máquinas.
 - b. Calcular el valor esperado y la varianza de la variable. Interpretar resultados.
28. Se desea usar la función $f(x)=a \cdot x^2$ definida para $0 \leq x \leq 1$ como función de densidad de probabilidad. Determinar el valor de la constante a y calcular la función de distribución acumulada.
29. Una compañía que realiza perforaciones para agua, evalúa los resultados obtenidos para los pozos realizados con una calificación que va entre cero (fracaso total) y uno (éxito total).
En base a experiencias previas se ha determinado que la v.a. “X: calificación asignada a un pozo” se puede modelar mediante la siguiente f.d.p.: $f(x) = 2x$ para $0 \leq x \leq 1$.
- c. Determinar si $f(x)$ cumple con las condiciones de una función de densidad de probabilidad.
 - d. En caso afirmativo, dibujar e interpretar la forma de dicha función.
 - e. Determinar la función de distribución acumulada.
 - f. Calcular la probabilidad de que una calificación sea menor a 0,9, que sea mayor a 0,9, y que esté entre 0,6 y 0,9.
 - g. Calcular el valor esperado y la varianza de la variable. Interpretar.

Unidad 3: MODELOS de PROBABILIDADES.

Cuestionario Orientador para la Investigación Bibliográfica.

- 1) ¿Qué se entiende por modelo o distribución de probabilidad de una variable aleatoria? Elaborar una opinión sobre los motivos por los cuáles se trabaja con modelos y no con las verdaderas distribuciones de las variables.
- 2) Resumir las principales propiedades del modelo Binomial y del modelo de Poisson. Destacar especialmente cuáles son:
 - Los principales supuestos de este modelo.
 - Los Parámetros.
 - La Función de Probabilidad.
 - Las expresiones que permiten determinar el valor esperado y la varianza de la variable.
 - Las posibles asimetrías que puede representar el modelo.
- 3) Cuando una variable aleatoria tiene distribución aproximadamente Normal:
 - ¿Cuál es la forma de la distribución y qué parámetros tiene? ¿Qué indican esos parámetros?

- ¿Cuál es su función de densidad?
- ¿Qué sucede si 2 distribuciones Normales tienen medias iguales y distintas desviaciones estándar? y ¿Qué sucede si tienen medias distintas y desviaciones iguales?
- ¿Cómo se calcula la probabilidad debajo de una curva Normal?
- ¿Cómo se define la variable Normal Estandarizada? ¿Cuáles son sus parámetros?
- Para una distribución Normal Estándar encontrar el área:
 - entre $Z = -1,97$ y $Z = 0,86$
 - para $Z > 1,84$
 - entre $-1 < Z < 1$; $-2 < Z < 2$ y $-3 < Z < 3$.

Representar gráficamente los resultados y elaborar una conclusión al respecto.

4) Distintas variables aleatorias, como por ejemplo la **vida útil de una herramienta de corte** o el **tiempo entre llegadas de dos vehículos sucesivos a una cabina de peaje**, tienen comportamientos que pueden ser descriptos con el **Modelo Exponencial**.

- ¿Cómo es la gráfica de esta distribución?
- ¿Cuál es la función de densidad de probabilidad? ¿Cuáles son sus parámetros?
- Determinar la función de distribución acumulada para este modelo.

5) Investigar sobre el Teorema del Límite Central y determinar cuáles son las propiedades que se demuestran con el mismo. Determinar por qué razón el teorema explica, que distribuciones Discretas como Binomial y Poisson, tiendan a comportarse como una Normal bajo ciertas condiciones.

Ejercicios Propuestos.

1. El 10% de las llantas que produce una fábrica resultan con defectos. Si de la producción se forma un lote de diez unidades seleccionadas al azar. Encontrar respuestas a las siguientes cuestiones:
 - a) ¿Cuál es la variable de interés?
 - b) ¿Puede ser representada mediante el modelo Binomial? Justifique su respuesta.
 - c) Indicar a que se considera "éxito" y cuál es su probabilidad.
 - d) Calcular es la probabilidad que:
 - Haya exactamente dos llantas defectuosas.
 - Ninguna esté defectuosa.
 - Haya más de 2 defectuosas.
 - e) Determinar la media y el desvío estándar de la variable.
 - f) Analizar si la asimetría es positiva o negativa.
 2. Un examen del tipo "*múltiple choice*" consta de 10 preguntas y para aprobar, son necesarias como mínimo 6 respuestas correctas.
 - a) Si cada una de las preguntas tiene 2 respuestas posibles, averiguar la probabilidad de aprobar el examen sólo por azar.
 - b) Si el número de respuestas posibles para cada pregunta se eleva a 4, ¿cuál sería ahora la probabilidad de aprobar por azar?
-

3. En un precinto policial, se estudian las llamadas recibidas por hora y se tienen los siguientes datos promedios :
 - Promedio por hora : 180 llamadas
 - Promedio por minuto : 3 llamadas
 - Promedio por segundo : $1/20$ llamadas
 - a) Establecer qué tipo de variable se estudia, qué mide y cuál es su recorrido.
 - b) Proponer un modelo de probabilidades que permita representar esta variable. Justificar detalladamente la elección.
 - c) Proponer valores adecuados para los parámetros del modelo.
 - d) Determinar la probabilidad que se reciban exactamente 2 llamadas en 1 minuto.
 - e) Determinar la probabilidad que se reciban más de 2 llamadas en un minuto.
 - f) Hallar $E(X)$ y $V(X)$.
 - g) Determinar la probabilidad que se reciban exactamente 2 llamadas en 4 minutos.

 4. Una máquina sufre, en promedio, 3 desperfectos por mes que obligan a su detención para realizar reparaciones.
 - a) Especificar cuál es la variable en estudio y proponer un modelo de probabilidades que permita representarla. Justifique claramente su elección.
 Calcular:
 - b) La probabilidad de que durante un mes cualquiera no sufra ninguna detención.
 - c) La probabilidad de que durante un mes cualquiera sufra al menos una detención.
 - d) La probabilidad de que durante un mes cualquiera sufra entre 1 y 3 detenciones (ambos valores incluidos).
 - e) La probabilidad de que en dos meses sufra 5 detenciones.

 5. En un Laboratorio se controla un espesante celulósico para ello, se prepara una solución al 1% del espesante en agua y se le agregan unas gotas de amoníaco ya que su pH óptimo es mayor a 8. Luego se procede a medir su viscosidad en un roto-viscosímetro, la cual tiene distribución Normal con una media de 20 poises y un desvío de 2 poises. No se aprueba el espesante celulósico si su viscosidad es inferior a 18 poise o superior a 22 poise ya que no cumpliría con los parámetros considerados aceptables. Determine la probabilidad de aceptar el espesante.
 Nota: La viscosidad en poises se define como la magnitud de la fuerza (medida en dinas por centímetro cuadrado de superficie) necesaria para mantener —en situación de equilibrio— una diferencia de velocidad de 1 cm por segundo entre capas separadas por 1 cm.

 6. Se regula una máquina despachadora de refrescos para que el contenido de cada vaso tenga distribución Normal con media 200 ml. y con un desvío de 15 ml. Determinar:
 - a) ¿Qué proporción de vasos tendrán más de 224 ml.?
-

- b) ¿Qué proporción de vasos tendrán más de 195 ml.?
 - c) ¿Cuál es la probabilidad de que un vaso contenga entre 191 y 215 ml ?
 - d) ¿Cuál es la probabilidad de que un vaso contenga entre 180 y 195 ml ?
 - e) Si se utilizan vasos de 230 ml. ¿Cuántos vasos se derramarán en los próximos 1000 despachos?
 - f) Determinar un contenido x_1 tal que el 80% de los vasos contengan más de esa cantidad.
 - g) Determinar un contenido x_2 tal que el 99% de los vasos contengan menos de esa cantidad.
 - h) Determinemos dos valores de contenido x_1 y x_2 (simétricos respecto a la media) tal que el 95% de los vasos están comprendidos entre esos extremos.
7. Se estudia una máquina que llena bolsas de cemento y para ello se define la variable aleatoria X: *contenido de una bolsa*. Sabiendo que X tiene distribución Normal, con media 50 Kg. y desvío 0.5 Kg.
Determinar:
- a) La probabilidad de que una bolsa contenga menos de 51 Kg.
 - b) La probabilidad de que una bolsa contenga más de 51 Kg.
 - c) La probabilidad de que una bolsa contenga entre 49.25 y 49.75 Kg.
 - d) La probabilidad de que una bolsa contenga entre 49 y 51 Kg.
 - e) Si un 20 % de las bolsas contienen más de x_1 Kg. Hallar el valor de x_1 .
 - f) Calcular los contenidos x_2 y x_3 , valores simétricos con respecto a la media que encierran un 50 % de la distribución.
 - g) Al controlar el llenado, se planea desechar todas las bolsas que contengan más de 51 Kg. o menos de 49 Kg. Calcular
 - g 1). La probabilidad de que 3 bolsas fabricadas sucesivamente (de manera independiente) deban ser desechadas en el control de llenado.
 - g 2). La probabilidad de que en un lote de 15 bolsas a las que se les va a controlar el llenado, 3 deban ser desechadas.
8. La variable *Distancia entre dos baches en una ruta*, tiene Distribución Exponencial con media 2 Km. Determinar:
- a) La probabilidad que entre dos baches haya menos de 1200 metros.
 - b) La probabilidad que entre dos baches haya más de 1500 metros.
 - c) La distancia tal que el 80% de las separaciones son mayores que la misma.
9. El *Tiempo entre dos paradas consecutivas de una máquina* tiene un promedio de 25 hrs. Averiguar la probabilidad de que entre dos paradas consecutivas pasen más de 32 hrs.
10. Una máquina consume aproximadamente 15 KW por hora de energía, con un desvío de 1 KW. Si se tienen 30 máquinas de ese tipo operando:
- a) ¿Cuál será la distribución del consumo total de las 30 máquinas? ¿Cuál será la media y el desvío del consumo total?
-

- b) ¿Cuál es la probabilidad de que el consumo total supere los 460 KW?
 - c) ¿Cuál es el consumo total que es superado sólo el 10 % de las veces?
11. Se conoce que entre los alumnos del nivel secundario aproximadamente un 25 % son lectores habituales. Calcule la probabilidad de que en un grupo de 50 alumnos haya al menos 20 que sean lectores habituales.
12. A una central telefónica llegan en promedio 3 llamadas por minuto. Determinar la probabilidad que en dos horas lleguen al menos 350 llamadas.
13. El tiempo de atención de un trámite, en una oficina pública, tiene distribución Exponencial con media 20 minutos. Se entregan 25 turnos a primera hora. Calcular la probabilidad que en las 8 horas de trabajo se pueda atender a todos los interesados.

Unidad 4: INFERENCIA ESTADÍSTICA.

a) Estimación de Parámetros

Cuestionario Orientador para la Investigación Bibliográfica.

- 1) ¿Qué es un parámetro y qué es un estimador? ¿Cuál es la diferencia entre ambos conceptos? Explicar cuál es la razón por la que se dice que los estimadores son variables aleatorias.
 - 2) Suponga que estudiamos la variable continua: “Tiempo de reparación de una máquina, ante un cierto tipo de fallas (en minutos)”. Suponga además, que la verdadera media poblacional es $\mu = 135'$. Se toma una muestra de veinte observaciones de esa variable. Analice si es posible que el promedio de la muestra sea exactamente igual a $135'$. Justifique conceptualmente la respuesta.
 - 3) ¿Cómo es la distribución de la media aritmética de muestras extraídas de una población Normal? ¿Qué pasa si la variable no tiene distribución Normal, o su distribución es desconocida?
 - 4) Representar gráficamente una variable con distribución Normal con media 200 y con desvío 20. Superpuesta con la figura anterior, graficar la distribución de la media aritmética de muestras de tamaño $n=4, 25$ y 100 , de esa población.
 - 5) El promedio de una muestra se considera un buen estimador de la media de la población. Explicar los motivos de la afirmación anterior.
 - 6) Analizando el comportamiento de los estimadores, se acostumbra calificarlos como: Insesgados; Consistentes; Eficientes y Suficientes. Explicar bajo qué condiciones se otorga cada una de esas calificaciones. Indicar en particular cuáles son las medidas
-
-

muestrales que pueden utilizarse como estimadores puntuales insesgados de los parámetros μ y σ^2 .

- 7) Construya una lista de estimadores convenientes para los parámetros de los siguientes modelos: Binomial; Poisson; Normal; Exponencial.
- 8) Explique las principales características de la distribución Chi-cuadrado. Deduzca una cantidad estadística que posea dicha distribución de probabilidad. Mencione mediante un ejemplo cómo se utiliza la tabla correspondiente. Grafique una Chi-cuadrado con 12 grados de libertad.
- 9) Explique las principales características de la distribución t de Student. Ejemplifique el modo en que se utiliza la tabla correspondiente. Grafique en forma superpuesta una t de Student con 6, 20 y 120 grados de libertad.
- 10) Los parámetros de una población pueden estimarse en forma puntual y por intervalo. Explicar en qué consiste cada una y cuáles son las ventajas de la estimación por intervalos de confianza.
- 11) Realice la deducción de los siguientes intervalos:
 - Intervalos de confianza para la media de una Normal con varianza conocida.
 - Intervalos de confianza para la media de una Normal con varianza no conocida.
 - Intervalos de confianza para la varianza de una Normal.

Aplice para ello el método del pivote.

- 12) Cuando se hace una estimación por intervalo, la aproximación es más precisa cuanto menor es la longitud. ¿De qué depende la longitud de un intervalo de confianza? ¿Cómo se puede hacer para disminuir dicha longitud, conservando el nivel de confianza?
- 13) El Nivel de Confianza es una decisión importante a la hora de realizar una estimación por intervalos. Explique qué es lo que representa el Nivel de Confianza.
- 14) Suponga que estudiamos la variable: “Tiempo de reparación de una máquina, ante un cierto tipo de fallas (en minutos)”. Como parte del estudio estimamos la media poblacional con un intervalo de 90% de confianza y obtenemos los siguientes extremos:

$$E_{\text{inferior}} = 120'; E_{\text{superior}} = 150'.$$

Al respecto, analicemos las siguientes cuestiones:

- a) ¿Podemos estar seguros de que el intervalo comprende a la verdadera media?.
- b) ¿Es correcta la siguiente expresión?: $P(120' < \mu < 150') = 0,90$.

Ejercicios Propuestos.

1. Una compañía de electrónica fabrica resistores que tienen una resistencia promedio de 100Ω y una desviación estándar de 10Ω . Si se sabe que la distribución de la resistencia es Normal:
-

- a) ¿Cuál es la probabilidad de que al tomar una muestra de 49 resistores, la resistencia promedio sea menor de 97Ω ?
 - b) Si en lugar de tomar una muestra 49 se considera una de 10 resistores ¿cuál será el valor de la probabilidad antes mencionada?
 - c) ¿Es importante la hipótesis de normalidad de la variable en ambos casos? ¿por qué?
2. Una máquina llena bolsas de cemento. Se estudia la variable X: *contenido de una bolsa*. Si todo funciona normalmente, esta variable debe tener distribución Normal, con una media de 50 Kg. y un desvío de 0.5 Kg. Se controla diariamente el funcionamiento de la máquina. Este control se puede realizar midiendo el contenido de una o varias bolsas de cemento.
- a) Si se extrae sólo una bolsa, calcular los valores x_1 y x_2 , simétricos con respecto a la media especificada, que encierran una probabilidad del 95 %.
 - b) Si se extraen 5 bolsas al azar y se calcula la media aritmética de sus contenidos, determinar los valores x_1 y x_2 , simétricos con respecto a la media, que encierran nuevamente una probabilidad del 95 %.
 - c) Calcular lo mismo que en el punto anterior, para el caso en que se extraiga una muestra de 100 bolsas.
 - d) Comparar los resultados obtenidos en a), b) y c) y elaborar una opinión al respecto. Representar gráficamente las distribuciones obtenidas.
3. Se considera que el tiempo de duración de cierta batería se distribuye con media μ y varianza 1. Se toma una muestra aleatoria de tamaño 7 con cuya información se desea realizar una estimación puntual de la duración promedio poblacional. Con ese fin se sugieren los siguientes dos estimadores:

$$\mu_1 = \frac{X_1 + X_2 + \dots + X_7}{7} \quad \mu_2 = \frac{2X_1 - X_6 + X_4}{2}$$

- a) ¿Son insesgados esos estimadores?
 - b) ¿Cuál estimador es el “mejor”? ¿En qué sentido es mejor?
4. Con relación a la máquina que embolsa cemento del ítem 2, se extrae una muestra de 10 bolsas, que tienen los siguientes pesos (en Kg.):

49,67 49,79 49,79 51,25 49,18 50,36 48,75 50,12 50,24 50,40

- a) Calcular estimaciones insesgadas de la media y de la varianza poblacional.
- b) Teniendo en cuenta las condiciones en las que debe trabajar la máquina (en cuanto al nivel medio y a la dispersión del contenido de cada bolsa si todo funciona normalmente) ¿qué opinión puede formular sobre el funcionamiento de la máquina, empleando las estimaciones puntuales obtenidas en el punto a)?

5. Especialistas en Geología han determinado, gracias a muchos estudios, que la densidad del oro puede ser modelada con una distribución Normal, con media 19,3 gramos por centímetro cúbico y con desvío 0,3 gramos por centímetro cúbico. Por otra parte, para evaluar la densidad de cualquier material, se realizan varias mediciones con un equipo denominado espectrómetro y se determina el promedio de las mismas. Suponiendo que se coloca en el espectrómetro un trozo de mineral y después de realizar 20 mediciones se determina que el promedio de las mismas es 19,05 gr./cm³. Asumiendo que la dispersión no varía, estimar un intervalo del 90% de confianza para la verdadera densidad media poblacional del metal analizado. En base al resultado obtenido, ¿es posible suponer que la muestra es de oro puro?
6. Con relación al ejercicio anterior, determinar el tamaño que debería tener la muestra para duplicar la precisión de intervalo obtenido (es decir que la longitud del intervalo de confianza se reduzca a la mitad).
7. Retomando el ejercicio 4, empleando los resultados del muestreo:
 - a) Suponiendo que el contenido de las bolsas tiene distribución Normal, calcular un intervalo de confianza del 95% para el contenido medio y la dispersión de las bolsas.
 - b) Elaborar una interpretación sobre los intervalos obtenidos.
 - c) ¿Coincide la opinión en este caso con lo concluido en el Ejercicio 4, respecto del funcionamiento de la máquina en cuanto al centrado y la dispersión del proceso?
8. Siguiendo con el análisis de las bolsas de cemento, suponga ahora que un comprador de cemento exige un control periódico del contenido de las bolsas en base al uso de muestras de 50 seleccionadas al azar de la producción. Después de procesar la primera muestra, se obtuvo un contenido medio de 49,91 Kg., con un desvío de 0,65 Kg.
 - a) Asumiendo que la variable analizada tiene distribución Normal, calcular un intervalo de confianza del 90 % del contenido medio verdadero de las bolsas de cemento. Interpretar el resultado obtenido.
 - b) ¿Quedará conforme el cliente con los resultados de este control? ¿Por qué?
 - c) Días más tarde, un nuevo control arroja el siguiente intervalo para el contenido medio de cemento, con una confianza del 99 %: [49.75;49.95] ¿Qué actitud tendrá el cliente en base a este nuevo resultado?
 - d) Calcular un intervalo de confianza para la varianza. Determinar si la dispersión se mantiene cercana al valor normal del proceso ($\sigma^2 = 0,25$).
9. Al Gerente del Departamento de Servicios al Cliente de una compañía de gas le gustaría estimar el tiempo promedio que transcurre entre la solicitud de servicio y su conexión. Se seleccionó una muestra aleatoria de 15 casas de los registros disponibles del año anterior. Los resultados obtenidos (en número de días) se presentan de la manera siguiente:

114 78 96 137 78 103 117 126 86 99 114 72 104 73 86

- a) Identificar con claridad la variable aleatoria estudiada.
 - b) Obtener estimaciones puntuales del promedio y desvío de dicha variable.
 - c) Calcular una estimación de intervalo de confianza de 95% del tiempo de espera promedio de la población durante el año anterior.
 - d) Establecer una estimación de intervalo de confianza de 90% para la varianza del tiempo de espera de la población durante el año anterior.
 - e) La dirección de la compañía ha fijado como meta no superar los 80 días para realizar la conexión. En base a los resultados de este análisis ¿se ha logrado alcanzar ese objetivo?
10. La resistencia a la rotura de los cables producidos por una fabricante tienen una media de 180 Kg. y un desvío estándar de 10 Kg. Mediante una nueva técnica en el proceso de fabricación se aspira a que esta resistencia pueda ser incrementada. Se ensaya una muestra de 20 unidades producidas con el nuevo método y se encuentra que su resistencia media es de 185 Kg. y el desvío 12 Kg.
- a) ¿Puede sostenerse que, en efecto, existe un aumento significativo de la resistencia de los cables? ¿La varianza se mantiene en los mismos valores? Trabajar con un nivel de confianza del 90 %.
 - b) ¿Qué valor usó para la varianza en los cálculos anteriores? ¿Por qué?
11. El encargado del mantenimiento de una máquina compara el tiempo de reparación (en hrs.) de la misma mediante el procedimiento convencional de trabajo, con un nuevo método que resulta más costoso pero que tiende a reducir los tiempos de reparación. Para ello saca una muestra de 10 registros con el procedimiento actual, y otra de 15 valores con el nuevo método:
- | | | | | | | | | | | | | | | | |
|--------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Viejo | 28 | 25 | 25 | 29 | 24 | 29 | 31 | 28 | 31 | 27 | | | | | |
| Nuevo | 26 | 25 | 24 | 24 | 25 | 22 | 27 | 26 | 27 | 27 | 21 | 23 | 30 | 26 | 25 |
- a) Suponiendo que en ambos casos la variable es Normal, hallar los correspondientes intervalos de confianza para la media y varianza poblacionales. Seleccionar y justificar el nivel de confianza empleado.
 - b) Determinar si puede suponerse que ambos procedimientos poseen la misma dispersión.
 - c) Determinar si puede suponerse que el nuevo procedimiento reduce significativamente el tiempo medio de reparación de la máquina.
- Sugerencia: en b) y c) dibujar los intervalos de confianza obtenidos para permitir su comparación en forma gráfica.

c) Pruebas de Hipótesis

Cuestionario Orientador para la Investigación Bibliográfica.

- 1) Imagine a las pruebas de hipótesis como un proceso donde se desarrolla una secuencia de actividades. Elabore un diagrama de flujo que represente estas actividades y prevea las posibles respuestas.

- 2) Todas las pruebas de hipótesis son muy similares en su estructura. Sin embargo se diferencian entre sí en una cuestión fundamental. Especifique cuál es esa cuestión. Elabore ejemplos donde esa diferencia se ponga de manifiesto.
- 3) Elabore un diagrama o representación gráfica que represente todas las pruebas de hipótesis que se analizan en la esta unidad temática.
- 4) Las pruebas de hipótesis no proveen un resultado certero, sino que por el contrario pueden presentar errores. Especifique cuáles son los posibles errores y determine las probabilidades de cada uno. Además analice cuáles son las condiciones que determinan dichas probabilidades y el modo en que se pueden calcular.
- 5) Suponga que una empresa lo contrata con una excelente remuneración, para que usted se encargue de hacer el control de recepción de bolsas de semillas. La idea es que la media del contenido de las bolsas sea 40 kg o más. Todos los días a las seis de la mañana, llegan los camiones de los proveedores trayendo una gran cantidad de bolsas cada uno. Usted puede tomar una pequeña muestra de proveedor, pesar las bolsas y decidir si acepta o no la partida. Por supuesto, si recibe bolsas que no están en condiciones, su contrato corre serio riesgo de corte.
Especifique como plantearía una prueba genérica que le permita decidir sobre cada partida. Además determine como le conviene adoptar el nivel de significancia: ¿grande o pequeño?.
Nota: suponga que todas las bolsas que un proveedor trae en un mismo día, tienen un comportamiento estadístico homogéneo.
- 6) En las pruebas sobre parámetros que se analizan en la presente unidad temática, se contrasta el valor del estadístico contra un valor de referencia que corresponde a las siguientes distribuciones: Normal estándar, t de Student o Chi cuadrado. Por ejemplo, la Normal Estándar tiene media cero, por lo que para decidir en las pruebas que emplean esta distribución, en definitiva se determina si el resultado se encuentra cerca o lejos del cero, esto es, el valor de referencia es cero. Grafique cada una de las restantes distribuciones y especifique cuál es el valor medio o valor de referencia de las mismas.

Ejercicios Propuestos.

1. Una máquina tiene una cierta producción media horaria, digamos μ_0 . Se está pensando en remplazar la máquina por una nueva. Explicar cómo formularía la Hipótesis Alternativa en cada uno de los siguientes casos:
 - a) Se hará el reemplazo siempre y cuando la nueva máquina no tenga una producción significativamente diferente de la actual.
 - b) Se incorporará la nueva máquina solo si su producción horaria es significativamente mayor que la actual.
 - c) Se hará el cambio siempre y cuando la nueva máquina no tenga una producción significativamente menor que la actual.
 2. Explicar qué criterio se emplearía para seleccionar el nivel de significación de la prueba, en los siguientes casos:
-

- a) Un cliente debe decidir la compra de baterías. El negocio es conveniente sólo si la duración media de las mismas es superior a las 1500 horas de uso continuo. En caso contrario, conviene usar otras alternativas, que tienen igual o menor costo.
 - b) Un fabricante de cemento debe comprobar si la máquina llena 50 Kg. como mínimo. Si rechaza que es cierto cuando en realidad lo es, debe parar el proceso, llamar a mantenimiento, revisar la máquina y efectuar pruebas de ajuste que llevan mucho tiempo. Si no rechaza y en realidad la máquina llena menos de 50 Kg., se cambia la bolsa con problemas por una nueva.
3. Considerando nuevamente el ejemplo de la máquina que llena bolsas de cemento de 50 Kg. Se asume que la variable “contenido de una bolsa” tiene distribución Normal y que la dispersión del proceso se mantiene aproximadamente constante, con un desvío $\sigma = 0,5$ Kg. Se presentan las siguientes situaciones:
- a) Se saca una muestra de 30 bolsas, que arroja una media de 50,21 Kg. Comprobar si la media del proceso es significativamente diferente a los 50 Kg., para un nivel de significación del 1 %.
 - b) Un cliente está insatisfecho con una partida de 1000 bolsas, sosteniendo que el contenido de las mismas es sensiblemente menor a los 50 Kg. Como sería muy difícil pesar las 1000 bolsas, se acuerda tomar una muestra de 25 bolsas seleccionadas al azar, que da una media de 49,89 Kg. Comprobar en base a la muestra si el cliente efectivamente tiene razón, a un nivel de significación del 10 %.
 - c) Un tiempo después el responsable del proceso sospecha que por problemas de ajuste, la máquina está embolsando más de 50 Kg. Para confirmar si esto es cierto, se saca una muestra de 40 bolsas, que produce una media de 50,18 Kg. Decidir, para un nivel de significación del 5 %, si la sospecha es razonable o no.
4. Volviendo al ejemplo planteado en Estimación sobre la densidad del oro, que según estudios realizados puede ser modelada con una distribución Normal, con media 19,3 gramos por centímetro cúbico y con desvío 0,3 gramos por centímetro cúbico. El análisis de un trozo de mineral con espectrómetro había arrojado después de realizar 20 mediciones un promedio de 19,05 g/cm³.
- a) Suponiendo que la dispersión no cambia, comprobar considerando un nivel de significación del 5%, si el material analizado puede considerarse de oro puro.
 - b) ¿Coincide la conclusión de esta prueba con la obtenida mediante intervalo de confianza calculado en la Unidad anterior?
5. Volviendo al Ejercicio 2 de Estimación, relacionado con el análisis del contenido de cemento, un muestreo de 10 bolsas había arrojado los siguientes pesos (en Kg.):
- 49,67 49,79 49,79 51,25 49,18 50,36 48,75 50,12 50,24 50,40
- a) Comprobar si la media se mantiene en el valor nominal (50 kg.).
 - b) Comprobar si la dispersión no varía significativamente ($\sigma^2 = 0,25$).
 - c) En ambos casos seleccionar el nivel de significación apropiado, justificando en término de los errores que se pueden cometer al realizar las pruebas.

- d) ¿Qué valor usa para la varianza en este caso? ¿Por qué?
6. La Jojoba (*Simmondsia Chinensis*) es una especie endémica de las regiones áridas. En la Argentina, se cultiva en la región Noroeste, presentando muy poca variabilidad genética, lo que determina un producto uniforme con buena cotización comercial. Aproximadamente el 50% de las semillas de jojoba es “aceite”. En los Estados Unidos mediante transformaciones genéticas han desarrollado una nueva variedad la cual supone contendrá un porcentaje mayor de aceite. Para verificar la nueva variedad se siembran 20 parcelas y se obtiene que el rendimiento promedio de las semillas es de 53,65% de aceite con un desvío estándar del 2%. Verificar, para un nivel de significación del 5 %, si el porcentaje de aceite se mantiene o cambia con la transformación genética.
7. Una empresa automotriz se dedica a producir unidades de gran confort y prestaciones. Uno de los aspectos que diferencian sus productos respecto a la competencia es la seguridad. Para verificar su sistema de frenos, se realiza el siguiente ensayo: con el vehículo a 60 Km./h., se clavan los frenos y se mide la distancia hasta que el vehículo se detiene. Con el sistema actual la variable distancia de frenado tiene distribución Normal con media 16,5 m. y desvío 0,5 m. Para mejorar esta prestación, se prueba un nuevo sistema con veinticinco autos, obteniéndose un promedio de 15,8 m. con un desvío de 0,55 m.. Dado que el cambio de sistema requiere una fuerte inversión para el desarrollo de proveedores, se desea cambiar el sistema sólo si se encuentran evidencias de que la mejora es significativa. Para resolver el problema:
- Analizar si la dispersión realmente se incrementó, utilizando una prueba de hipótesis con un nivel de significancia de 0,10.
 - Decidir si puede suponerse que la distancia media de frenado se ha reducido, utilizando nuevamente un nivel de significancia de 0,10.
 - Analizar si el nivel de significación adoptado puede considerarse adecuado para la decisión que se desea tomar. Para fundamentar la opinión considerar cuál de los dos tipos de errores es más preocupante en cada una de las pruebas realizadas.
8. Las autoridades de una región han establecido que el vuelco de agua con temperaturas no superiores a los 55°C no afecta al ecosistema fluvial. Una planta industrial que está al lado del río posee un sistema de enfriamiento por agua y la descarga se realiza mediante una cañería de corto recorrido que desemboca en el río. Para verificar si el volcamiento de agua cumple con lo requerido, se mide la temperatura de 25 muestras de agua a la salida de la descarga en el río, obteniéndose los siguientes valores muestrales: Promedio = 57,5°C y Desvío = 2,5°C. Suponiendo que la temperatura del agua en el lugar de vuelco tiene distribución Normal, verificar si se cumple o no con lo requerido por las reglamentaciones. Analizar en este contexto, cuál de los errores (de tipo I y de tipo II) tendría consecuencias más serias.
9. Recordemos que en Estimación la Gerencia de una compañía de gas quería estimar el tiempo promedio que transcurre entre la solicitud de servicio y su conexión (en días), para lo cual se había extraído la siguiente muestra:
-

78 96 137 78 103 117 126 86 99 114 72 104 73
86

- a) La dirección de la compañía había fijado como meta no superar los 80 días para realizar la conexión. Compruebe mediante una prueba de hipótesis apropiada si se cumple con ese objetivo.
- b) Explique cuál sería el nivel de significación más apropiado para realiza esta prueba.

10. Retomando el estudio planteado en estimación sobre la resistencia a la rotura de los cables producidos por una fabricante. Ésta tenía una media de 180 Kg. y un desvío estándar de 10Kg. Ahora, mediante una nueva técnica en el proceso de fabricación se aspira a que dicha resistencia pueda ser incrementada, por ello se ensayó una muestra de 20 unidades producidas con el nuevo método y se encontró que su resistencia media era de 185 Kg. y el desvío 12 Kg.

- a) Empleando ahora la prueba de hipótesis correspondiente, comprobar si efectivamente existe un aumento significativo en la resistencia de los cables producidos con el nuevo método, para un nivel de significación del 5 %.
- b) Comprobar si la varianza del nuevo método no ha cambiado significativamente, para el mismo nivel de significación.

Unidad 5: ANÁLISIS DE RELACIONES ENTRE VARIABLES.

Cuestionario Orientador para la Investigación Bibliográfica.

- 1) Cuando analizamos dos variables X e Y, que suponemos relacionadas, podemos aproximar dicha relación con una recta que tiene la siguiente forma: $y = a + b x$. Para obtener los valores de los coeficientes utilizamos las expresiones de Mínimos Cuadrados:

$$b = S_{xy} / S_x^2 \qquad a = \bar{y} - b \bar{x}$$

A fin de verificar la validez de estas expresiones podemos realizar un experimento. En efecto, si partimos de la recta: $y = 10 - 2 x$, y le asignamos algunos valores a la variable independiente, podremos obtener un juego de pares de datos, como el siguiente:

X	1	2	3
Y	8	6	4

Ahora apliquemos sobre estos datos las expresiones para **a** y **b**, y observemos el resultado.

- 2) Como todos sabemos, algunas empresas tienen estrategias orientadas directamente a la venta, donde más que el producto en si, interesa la carga publicitaria que tenga el mismo. Claros ejemplos de estos comportamientos son las empresas productoras de gaseosas o las de zapatillas.

Supongamos que una de estas empresas registra sus valores de **Inversión en Publicidad** y de **Ventas Brutas**, obteniendo los siguientes valores:

Publicidad (en millones)	2	6	8	5	1	7	4
Ventas (en millones)	6	11	13	7	4	10	7

- a) Representemos los datos en un par de ejes coordenados, colocando en abscisas la inversión publicitaria y en ordenadas las ventas. Vale la pena destacar que este tipo de gráficos se denomina habitualmente **Diagrama de Dispersión**.
 - b) Estimemos una recta de regresión de las Ventas en función de la Publicidad.
 - c) Con dicha recta realicemos una determinación de cuál puede ser la Venta si se invierten tres millones en Publicidad. Reflexionemos sobre las siguientes cuestiones: ¿esta determinación será perfecta o tendrá un cierto error?; ¿a cuánto puede ascender dicho error?
- 3) Para representar formalmente este tipo de problemas los matemáticos han propuesto el siguiente modelo lineal:

$$Y = \alpha + \beta X + \varepsilon$$

donde α y β son constantes, en tanto que Y varía de acuerdo a como varíe X y donde ε introduce el efecto aleatorio, es decir que genera variaciones no controladas en Y . Esta componente aleatoria se supone con distribución Normal, con media cero y un desvío propia al que se identifica como σ_ε .

Para comprender el sentido de ese modelo realizaremos un experimento con alguna Planilla de Cálculo:

- a) seleccionemos la expresión de una recta cualquiera con la forma $Y = a + bX$
 - b) asignemos a X unos veinte valores diferentes en una columna y calculemos los correspondientes valores de Y .
 - c) generemos aleatoriamente la misma cantidad de valores de una Distribución Normal con media cero y varianza igual a nueve
 - d) sumemos a los valores anteriores de Y , los números aleatorios generados, obteniendo así los Y definitivos
 - e) solicitemos al soft el cálculo de la regresión, explorando las distintas opciones disponibles. Luego busquemos en la salida las aproximaciones obtenidas para los coeficientes de la recta y para el desvío de la componente aleatoria: σ_ε .
- 4) ¿Qué información brindan los restantes elementos de esta salida?. En particular investigue las siguientes cuestiones:
- a) ¿Qué representa el coeficiente de determinación?; ¿cómo se interpreta?.
 - b) ¿Cómo puede utilizarse la información contenida en la tabla denominada “Análisis de la varianza”?; ¿cómo se construye dicha tabla?.
 - c) ¿De qué modo se estructura la tabla de coeficientes?; ¿cuáles son las pruebas de hipótesis relacionadas con esa tabla?.
-

- d) ¿Cómo se interpreta el gráfico de los residuos con los valores de X ?
- e) ¿Cómo se interpreta el gráfico de probabilidades normales?
- 5) Para deducir las expresiones de cálculo de los coeficientes a y b , se utiliza el Método de los Mínimos Cuadrados. Analicemos y discutamos toda la deducción de dichas expresiones. Si se intenta minimizar los errores en un sentido que no sea el de las Y , ¿serían diferentes las fórmulas?. Si intercambiamos los roles entre las variables y determinamos una regresión de X contra Y ¿obtendremos los mismos coeficientes?.
- 6) Si la recta no resulta una buena representación de la relación entre las variables deberemos utilizar algún otro tipo de función, por ejemplo una Exponencial o una Potencial. Para estimar los coeficientes aplicaremos las fórmulas anteriores para a y b , transformando previamente los datos.
- Para probar esta posibilidad, ajustemos una Exponencial de la forma: $Y = a b^x$, a los datos del problema Publicidad Ventas, calculando además los errores y el Coeficiente de Determinación.
- Comparemos los resultados con los de la recta. ¿Cuál de los modelos es preferible?.
- 7) Desarrollemos paso a paso, las Pruebas de Hipótesis que permiten verificar los valores de α y β . Apliquemos dichas pruebas para determinar si es posible suponer que esos coeficientes son iguales a cero en el ejemplo Publicidad - Ventas. ¿Qué significa la suposición de que $\beta=0$?

Ejercicios Propuestos.

1. Se estudia la relación entre la superficie de una vivienda y el consumo mensual de kerosén empleado para calefaccionar la misma, en un determinado barrio de la ciudad. Una muestra arrojó los siguientes resultados:

Superficie(m ² .)	Consumo (lts.)
69	20
135	40
121	35
103	35
97	26
78	24

- a) Identificar cuál es la variable dependiente en este estudio.
- b) Dibujar el diagrama de dispersión y especificar si se observa algún tipo de relación entre las variables.
- c) Calcular la covarianza y el coeficiente de correlación. Interpretar los valores obtenidos.
- d) Ajustar un modelo lineal simple por mínimos cuadrados y dibujar el modelo obtenido en el diagrama de dispersión.

- e) Calcular el coeficiente de determinación e interpretar el resultado obtenido.
 - f) En base a los resultados del análisis explicar si es razonable tratar de explicar el consumo de kerosén en ese barrio a partir de la superficie de las viviendas.
2. Se realiza un estudio similar al del caso anterior en otro barrio de la ciudad, de características socio-económicas diferentes. Una muestra de 10 viviendas produjo los siguientes resultados:

Superficie (m².)	Consumo (lts.)
177	49
187	52
198	50
170	49
166	50
190	49
170	50
183	48
192	51
198	49

- a) Realizar el mismo análisis estadístico del caso anterior.
 - b) Comprobar si el coeficiente de correlación es significativo.
 - c) Comprobar si la pendiente del modelo es significativa.
 - d) ¿Resulta útil en este caso, el análisis conjunto de las variables?
3. Un laboratorio está interesado en analizar los compuestos antioxidantes en semillas de jojoba, para lo cual se deben determinar los contenidos de fenoles totales en las mismas.
- El ensayo consiste en realizar una extracción total de los fenoles de las semillas mediante una extracción sólido-líquida con metanol durante 8 horas. El extracto metanólico se seca completamente en evaporador rotatorio, luego se realizan diluciones conocidas del extracto seco, y sobre éstas se determinan los fenoles totales.
- La curva de calibración, que relaciona concentración de fenoles con la absorbancia¹ de luz, se obtuvo con soluciones de fenol patrón en metanol. La determinación de fenoles totales se realizó por medio de una reacción con el reactivo Folin-Ciocalteu, siguiendo el procedimiento que se detalla a continuación:

- 1º. Se colocó en un tubo de ensayo 8,4ml de agua destilada y 0,1ml de la solución preparada del extracto en metanol.

¹ absorbancia. f. Fís. Medida de la atenuación de una radiación al atravesar una sustancia, que se expresa como el logaritmo de la relación entre la intensidad saliente y la entrante. (Diccionario de Real Academia Española)

- 2°. Se agregaron 0,5ml del reactivo Folin-Ciocalteu y 1ml de solución saturada de carbonato de sodio.
- 3°. Se midió la absorción de la solución (azul-celeste) en espectrofotómetro UV-visible a 760nm.

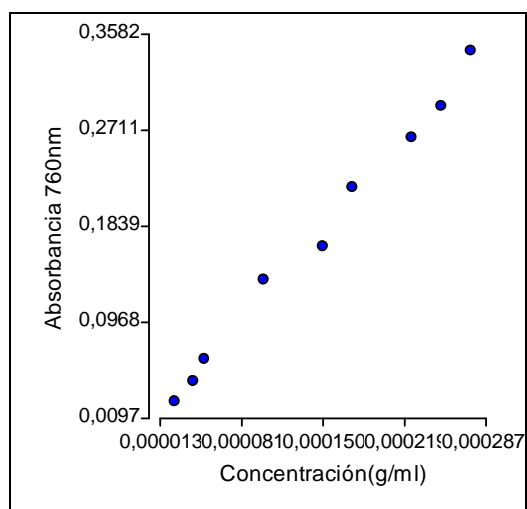
Por interpolación en la curva de calibración con fenol patrón se determina el contenido de fenoles totales de la muestra, en g/ml de solución del extracto, y luego por cálculos de conversión (según la dilución del extracto metanólico) se obtiene el valor en porcentaje de fenoles totales de las semillas de jojoba (1 g/ml de fenol en las diluciones de las muestras equivale a 1g/ml de fenol patrón).

Utilizando el procedimiento descrito anteriormente, se obtuvieron los datos, necesarios para elaborar la curva de calibración. Las diferentes concentraciones corresponden a distintas soluciones de la muestra patrón.

Concentración(g/ml)	Absorbancia 760nm
0,000025	0,0255
0,000040	0,0437
0,000050	0,0628
0,000100	0,1352
0,000150	0,1655
0,000175	0,2186
0,000225	0,2636
0,000250	0,2933
0,000275	0,3424

Basándose en la salida de software que se presenta, realice las siguientes actividades:

- Analice el diagrama de dispersión.
- Construya un modelo adecuado (curva de calibración). Verifique la adecuación del modelo propuesto y el cumplimiento de los supuestos teóricos. **Justifique todos los pasos.**
- Analice además si son significativas la ordenada al origen y la pendiente. **Justifique sus afirmaciones.**
- Elabore una opinión sobre la calidad de la curva de calibración obtenida. **Justifique claramente.**
- ¿Puede estimar la concentración que cabría esperar para una absorbancia de 0,20? Si es así calcúlela, caso contrario **justifique** su respuesta.
- ¿Puede estimar la absorbancia que cabría esperar para una concentración de 0,0002? Si es así calcúlela, caso contrario **justifique** su respuesta.



Coeficientes de correlación

Correlacion de Pearson: coeficientes\probabilidades

	Concentración(g/ml)	Absorbancia 760nm
Concentración(g/ml)	1,00	1,1E-08
Absorbancia 760nm	1,00	1,00

Análisis de regresión lineal

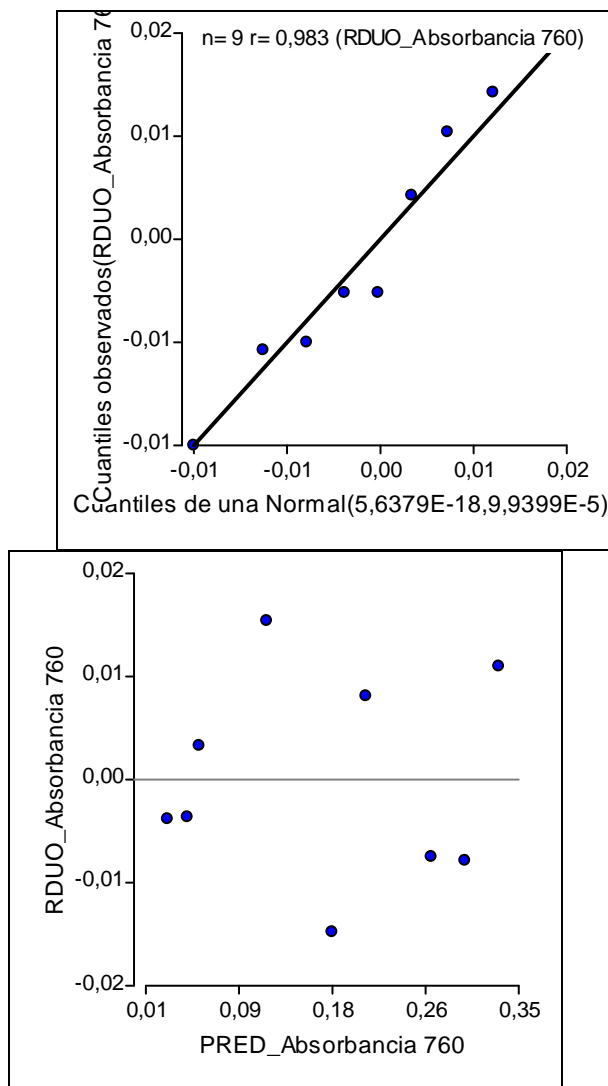
Variable	N	R ²	R ² Aj
Absorbancia 760nm	9	0,99	0,99

Coeficientes de regresión y estadísticos asociados

Coef.	Est.	E.E.	LI(95%)	LS(95%)	T
Valor p					
const	-9,7E-04	0,01	-0,02	0,01	-0,14
0,8897					
Concentración(g/ml)	1208,76	39,80	1114,63	1302,88	30,37
<0,0001					

Tabla de análisis de la varianza SC Tipo III

	FV	SC	gl	CM	F	Valor p
Modelo		0,10	1	0,10	922,20	
<0,0001						
Concentración(g/ml)		0,10	1	0,10	922,20	
<0,0001						
Error		8,0E-04	7	1,1E-04		
Total		0,11	8			



4. Se almacenó maní tostado a temperatura ambiente durante 120 días. Cada 28 días se tomaron muestras y se determinaron dos variables químicas: Índice de Peróxido e Índice de p-Anisidina, para determinar su deterioro por oxidación. Los datos obtenidos fueron:

Tiempo (días)	Índice de Peróxido	Índice de p-Anisidina
0	1.30	0.10
28	11.27	1.01
56	24.08	2.54
84	36.74	3.26
112	50.30	4.15

- Identifique las variables dependientes e independientes.
- Grafique los diagramas de dispersión para cada variable dependiente.

- c) Determine si las variables dependientes se correlacionan significativamente. Interprete el resultado.
- d) Ajuste un modelo lineal simple por mínimos cuadrados para cada variable dependiente en función de la variable independiente, calcule los respectivos coeficientes de determinación, e interprete los resultados obtenidos.
- e) Considerando que el Código Alimentario Argentino fija para el Índice de Peroxido un valor máximo de 10, ¿Cuál sería la vida útil estimada (en días) de estas muestras de maní tostado, en estas condiciones de almacenaje?
5. En los últimos meses la empresa de transporte interurbano “Caminos Cordobeses” ha recibido un número considerable de reclamos por parte de los pasajeros al respecto de incumplimientos en los horarios pautados de arribo a las distintas paradas uno de los tramos que realiza. Los pasajeros expresan que a veces se espera poco y otras veces mucho.
- Se decide entonces realizar una investigación que permita determinar si la percepción de los pasajeros es acertada y se enfoca la misma en el estudio de la posible relación de la variable Distancia (Km) entre paradas con el Tiempo de retraso (Min).
- Se obtiene la siguiente muestra aleatoria:

<i>Distancia entre paradas (Km.)</i>	10	15	20	25	30	35	40	45	50	55
<i>Tiempo (min.)</i>	2	3	6	8	9	14	17	19	20	22

A partir de la muestra indicada se obtiene la siguiente información del software Infostat. Observe que hay celdas incompletas.

Variable	N	R ²
Tiempo (min)	10	¿?

Coeficientes de regresión y estadísticos asociados						
Coef	Est.	E.E.	LI(95%)	LS(95%)	T	p-valor
Const	-3,6	0,82	-5,49	-1,71	¿?	0,0023
Distancia (Km)	0,48	0,02	0,43	0,53	¿?	

Cuadro de Análisis de la Varianza					
F.V.	SC	gl	CM	F	p-valor
Distancia (Km)	475,2	¿?	475,2	¿?	<0,0001
Error	¿?	8	¿?		
Total	484	9			

- a) Complete los datos faltantes en la salida de Infostat expuesta. Justifique.
- b) Interprete los valores de los coeficientes de correlación y de determinación, obtenidos en este ajuste. Explícite el concepto de los mismos.
- c) Realice el planteo de una prueba de hipótesis sobre la pendiente del modelo, Justifique el planteo y el resultado de la misma.

- d) Realice el planteo de una prueba de hipótesis, que permita realizar una valoración global del modelo, a partir de la tabla de análisis de varianza y justifique con claridad el resultado de la misma
- e) Explique la evaluación que debe realizarse sobre los residuos para que el modelo sea aceptado y efectúe dicha evaluación.
- f) Elabore una conclusión respecto del análisis realizado.

IMPORTANTE: Recomendamos revisar exhaustivamente los conceptos implicados en el análisis de regresión lineal simple y consecuentemente las pruebas estadísticas y análisis gráficos constitutivos del mismo.

6. Para la fabricación de panes nutraceuticos se utiliza harina de maní y soja. Es de vital importancia que las semillas estén sanas para la elaboración de la harina ya que de otra manera se favorecerían los procesos de oxidación de ácidos grasos lo que provocaría el aumento del Índice de Acidez y el Índice de Peroxidación dando como resultado el enranciamiento del aceite (olor y sabor desagradable). El proceso de oxidación se inicia con una enzima llamada lipooxigenasa que actúa sobre el ácido linoléico el cual tiene un doble enlace seguido de dos simples enlaces y así sucesivamente. La enzima actúa cambiando la posición de los dobles enlaces formando dienos conjugados. Se necesita analizar dos cultivares que son genéticamente distintos; cada uno cuenta con semillas enfermas y sanas. Las semillas enfermas contienen en su cubierta hongos. La determinación de la actividad de la enzima lipooxigenasa se realiza midiendo en UV la absorbancia de los dienos conjugados. La enzima se alimenta de las proteínas de las semillas al igual que los hongos. Usando los datos de la tabla, analice de cada cultivar las semillas sanas y enfermas. Verificar si existe una relación lineal entre la variable aleatoria Absorbancia con respecto al tiempo tanto para las semillas sanas como enfermas. Justificar.

Tiempo (seg.)	S1 (Absorb.)	S2 (Absorb.)	E1 (Absorb.)	E2 (Absorb.)
0	0,52	0,40	0,29	0,36
30	1,01	1,03	0,48	0,41
60	1,70	1,48	0,52	0,54
90	2,12	1,78	0,63	0,62
120	2,28	1,96	0,71	0,70
150	2,50	1,97	0,78	0,77
180	2,75	2,12	0,86	0,86

S1 y S2= semillas sanas; E1 y E2= semillas enfermas del cultivar 1 y 2.