

Trabajo RLS
Fecha de entrega: marzo 22 de 2021 hasta mediodía

Los ejercicios 1 a 3 deben ser realizados por todos los grupos. De los ejercicios restantes, a cada grupo han asignados dos (ver Tabla 7). Para la entrega de la solución debe anexar las gráficas pedidas y el programa R usado en un apéndice o anexo.

Se debe entregar un archivo pdf cuyo nombre sea el apellido de los integrantes. Por ejemplo para el grupo uno el nombre del archivo pdf será: RiosVelasquezYepes.pdf

1. Considere el modelo de regresión lineal simple $Y = 200 + 5X + E$, $E \sim N(0, 16)$.
 - a) Halle la distribución de Y para $X = 10, 20, 40$.
 - b) Explique el significado de β_0 y β_1 en este caso, asuma que la cobertura del modelo incluye a $X = 0$.
2. Suponga un modelo de regresión lineal simple $Y = 100 + 20X + E$, con $E \sim N(0, 25)$. Si X es observado en un valor de 5, ¿cuál es la probabilidad de que Y tome un valor entre 195 y 205?
3. La función de regresión que relaciona el volumen de producción (Y) de un operario después de haber tomado un programa de entrenamiento, con el volumen de producción que tenía antes del programa de entrenamiento (X), es $E[Y|X] = 20 + 0,95X$, con X variando entre 40 y 100. Un analista afirma que el programa de entrenamiento no conduce a un incremento en el volumen de producción promedio debido a que $\beta_1 < 1$. Comente acerca de tal afirmación.
4. En un estudio con personas mayores se quería ver la relación entre la actividad física y la frecuencia de resfriados, se pidió a los participantes que supervisaran el tiempo semanal que dedicaban al ejercicio durante un período de cinco años y la frecuencia de los resfriados. El estudio demostró que existe una relación estadística negativa entre el tiempo dedicado al ejercicio y la frecuencia de los resfriados. El investigador concluyó que en personas mayores, aumentar el tiempo dedicado al ejercicio es una estrategia eficaz para reducir la frecuencia de los resfriados.
 - a) ¿Los datos obtenidos en el estudio fueron datos de observacionales o experimentales?
 - b) Comente sobre la validez de las conclusiones de investigador.
 - c) Identifique otras dos o tres variables explicativas que puedan afectar simultáneamente el tiempo invertido en ejercicio y la frecuencia de resfriados en personas mayores.
 - d) ¿Cómo podría cambiarse el estudio para obtener una conclusión válida sobre la relación causal entre la cantidad de ejercicio y la frecuencia de los resfriados en personas mayores?
5. Se solicitó a los programadores informáticos empleados por un desarrollador de software que participaran en un seminario de formación de un mes de duración. Durante el seminario, se pidió a cada empleado que registrara el número de horas dedicadas a la preparación de clases cada semana. Después de completar el seminario, se midió el nivel de productividad de cada participante. Se encontró una relación estadística lineal positiva entre los niveles de productividad y tiempo dedicado a la preparación de las clases de los participantes. El líder del seminario concluyó que los aumentos en la productividad de los empleados son causados por un mayor tiempo de preparación de clases.
 - a) ¿Los datos obtenidos en el estudio fueron datos de observacionales o experimentales?
 - b) Comente sobre la validez de las conclusiones del líder del seminario.
 - c) Identifique otras dos o tres variables explicativas que puedan afectar simultáneamente los niveles de productividad y el tiempo de preparación de clases de los participantes.
 - d) ¿Cómo podría cambiarse el estudio para obtener una conclusión válida sobre la relación causal entre el tiempo de preparación de la clase y el nivel de productividad de los empleados?

ID sujeto	Nota final 1 ^{er} año (GPA)	Puntaje de ingreso
1	3.10	5.50
2	2.30	4.80
3	3.00	4.70
4	1.90	3.90
5	2.50	4.50
6	3.70	6.20
7	3.40	6.00
8	2.60	5.20
9	2.80	4.70
10	1.60	4.30
11	2.00	4.90
12	2.90	5.40
13	2.30	5.00
14	3.20	6.30
15	1.80	4.60
16	1.40	4.30
17	2.00	5.00
18	3.80	5.90
19	2.20	4.10
20	1.50	4.70

Tabla 1: Datos Problema 6

6. El Director de admisiones de una universidad aplicó un nuevo examen de admisión a 20 estudiantes de primer año seleccionados aleatoriamente, con el fin de determinar si el promedio de notas al final del primer año (GPA) puede ser predicho a partir del puntaje de la prueba de ingreso. Los resultados se presentan en la Tabla 1.
 - a) Identifique la variable respuesta y la variable predictora. Haga un gráfico de dispersión con la curva de regresión loess ¿Puede ser el MRLS apropiado?
 - b) Encuentre e interprete $\hat{\beta}_0$ y $\hat{\beta}_1$.
 - c) Obtenga la estimación de la desviación estándar de los parámetros estimados. Indique claramente cuánto vale cada término involucrado en los cálculos. Construya un I.C del 95 % para los parámetros del modelo de regresión asumiendo que son válidos los supuestos. Pruebe la significancia del modelo de regresión e interprete a la luz del problema.
 - d) Obtenga una estimación puntual, un intervalo de confianza y un intervalo de predicción para la media de GPA para estudiantes con un puntaje en la prueba de ingreso de 5.0.
 - e) A partir de los datos puede decirse que por cada incremento unitario en el puntaje de ingreso, en promedio la nota final del primer año se incrementa en más de una unidad?
7. Una empresa distribuye cierto computador de escritorio y proporciona servicio de reparación y mantenimiento preventivo de tales equipos. Los datos en la Tabla 2 fueron tomados de 18 solicitudes recibidas de mantenimiento preventivo. Sea X: el número de equipos servidos y Y: el tiempo en minutos dedicado por el técnico que atiende el servicio.
 - a) Haga un gráfico de dispersión con la curva de regresión loess ¿Puede ser el MRLS apropiado?
 - b) Encuentre e interprete $\hat{\beta}_0$ y $\hat{\beta}_1$.
 - c) Obtenga la estimación de la desviación estándar de los parámetros estimados. Indique claramente cuánto vale cada término involucrado en los cálculos. Construya un I.C del 95 % para los parámetros del modelo de regresión asumiendo que son válidos los supuestos. Pruebe la significancia del modelo de regresión e interprete a la luz del problema.

ID servicio	Tiempo en minutos	Número de equipos
1	97.00	7
2	86.00	6
3	78.00	5
4	10.00	1
5	75.00	5
6	62.00	4
7	101.00	7
8	39.00	3
9	53.00	4
10	33.00	2
11	118.00	8
12	65.00	5
13	25.00	2
14	71.00	5
15	105.00	7
16	17.00	1
17	49.00	4
18	68.00	5

Tabla 2: Datos Problema 7

- d) Obtenga una estimación puntual, un intervalo de confianza y un intervalo de predicción para la media del tiempo de atención de un servicio cuando el número de equipos atendidos es 5.
- e) A partir de los datos puede decirse que por cada incremento unitario en el número de equipos atendidos, en promedio el tiempo en minutos del servicio aumenta en más de 15 minutos?
8. Una sustancia empleada en investigación médica y biológica es transportada por carga aérea en cajas de cartón conteniendo 1000 ampollas de la sustancia. En la Tabla 3 se presentan los datos obtenidos para 10 embarques y corresponden al X: número de veces que las cajas son transferidas de un avión a otro en la ruta de embarque y el Y: número de ampollas que fueron halladas quebradas a la llegada.
- a) Haga un gráfico de dispersión con la curva de regresión loess ¿Puede ser el MRLS apropiado?
- b) Encuentre e interprete $\hat{\beta}_0$ y $\hat{\beta}_1$.
- c) Obtenga la estimación de la desviación estándar de los parámetros estimados. Indique claramente cuánto vale cada término involucrado en los cálculos. Construya un I.C del 95 % para los parámetros del modelo de regresión asumiendo que son válidos los supuestos. Pruebe la significancia del modelo de regresión e interprete a la luz del problema.
- d) Obtenga una estimación puntual, un intervalo de confianza y un intervalo de predicción para la media del número de ampollas que fueron halladas quebradas cuando la caja es transferida una única vez.
- e) Pruebe estadísticamente la veracidad de la siguiente afirmación: Por cada unidad en que se aumenta el número de transferencias entre aviones realizadas durante el transporte del producto, el número promedio de ampollas que llegan quebradas a su destino se incrementa en más de 4 unidades
9. Se espera que la masa muscular de una persona disminuya con la edad. Para explorar esta relación en las mujeres, un nutricionista selecciona aleatoriamente 4 mujeres de cada grupo de edad en rangos de 10 años comenzando con 40 años y terminando en 79 años. Sea X: edad y Y: medida de la masa muscular. Los datos obtenidos se presentan en la Tabla 4.

ID embarque	Número de ampollas quebradas	Número de transferen- cias
1	16	1
2	9	0
3	17	2
4	12	0
5	22	3
6	13	1
7	8	0
8	15	1
9	19	2
10	11	0

Tabla 3: Datos Problema 8

ID mujer	Masa muscular	Edad
1	82.00	71.00
2	91.00	64.00
3	100.00	43.00
4	68.00	67.00
5	87.00	56.00
6	73.00	73.00
7	78.00	68.00
8	80.00	56.00
9	65.00	76.00
10	84.00	65.00
11	116.00	45.00
12	76.00	58.00
13	97.00	45.00
14	100.00	53.00
15	105.00	49.00
16	77.00	78.00

Tabla 4: Datos Problema 9

- a) Haga un gráfico de dispersión con la curva de regresión loess ¿Puede ser el MRLS apropiado?
 - b) Encuentre e interprete $\hat{\beta}_0$ y $\hat{\beta}_1$.
 - c) Obtenga la estimación de la desviación estándar de los parámetros estimados. Indique claramente cuánto vale cada término involucrado en los cálculos. Construya un I.C del 95 % para los parámetros del modelo de regresión asumiendo que son válidos los supuestos. Pruebe la significancia del modelo de regresión e interprete a la luz del problema.
 - d) Obtenga una estimación puntual, un intervalo de confianza y un intervalo de predicción para la media de la masa muscular para una mujer con 65 años.
 - e) Pruebe estadísticamente la veracidad de la siguiente afirmación: Por cada unidad en que se aumenta la edad de la mujer, la masa muscular en promedio disminuye en más de una unidad.
10. Un criminólogo estudia la relación entre la densidad poblacional y la tasa de robos para ciudades de tamaño poblacional mediano. Para una muestra aleatoria simple de 16 ciudades se recolectaron los siguientes datos, X: densidad poblacional (No. Habitantes/Ud. de área), Y: tasa de robos en el último año (No. de robos por cada 100000 habitantes). Los datos se muestran en la Tabla 5.

ID ciudad	Tasa de robos	Densidad poblacional
1	209.00	59.00
2	180.00	49.00
3	195.00	75.00
4	192.00	54.00
5	215.00	78.00
6	197.00	56.00
7	208.00	60.00
8	189.00	82.00
9	213.00	69.00
10	201.00	83.00
11	214.00	88.00
12	212.00	94.00
13	205.00	47.00
14	186.00	65.00
15	200.00	89.00
16	204.00	70.00

Tabla 5: Datos Problema 10

- a) Haga un gráfico de dispersión con la curva de regresión loess ¿Puede ser el MRLS apropiado?
 - b) Encuentre e interprete $\hat{\beta}_0$ y $\hat{\beta}_1$.
 - c) Obtenga la estimación de la desviación estándar de los parámetros estimados. Indique claramente cuánto vale cada término involucrado en los cálculos. Construya un I.C del 95 % para los parámetros del modelo de regresión asumiendo que son válidos los supuestos. Pruebe la significancia del modelo de regresión e interprete a la luz del problema.
 - d) Obtenga una estimación puntual, un intervalo de confianza y un intervalo de predicción para la media de la tasa de robos en el último año cuando la densidad poblacional es 78.
 - e) Pruebe estadísticamente la veracidad de la siguiente afirmación: Por cada unidad en que se aumenta la densidad poblacional, la tasa de robos del último año aumenta en promedio en más de una unidad.
11. Una empresa productora de piezas de repuesto para equipos de refrigeración realiza un estudio para establecer el tamaño óptimo de los lotes que produce. Uno de sus objetivos es determinar la relación entre el tamaño de los lotes y las horas de trabajo requeridas para producir el lote, para ello tomó datos sobre tamaño de lote y horas de trabajo para 25 órdenes de producción recientes. Las condiciones de producción fueron estables durante el período de seis meses en los cuales fueron hechas las 25 corridas de producción y se espera que tales condiciones continúen en los próximos tres años. La Tabla 6 contiene la información recolectada.
- a) Identifique la variable respuesta y la variable predictora. Haga un gráfico de dispersión con la curva de regresión loess ¿Puede ser el MRLS apropiado?
 - b) Encuentre e interprete $\hat{\beta}_0$ y $\hat{\beta}_1$.
 - c) Obtenga la estimación de la desviación estándar de los parámetros estimados. Indique claramente cuánto vale cada término involucrado en los cálculos. Construya un I.C del 95 % para los parámetros del modelo de regresión asumiendo que son válidos los supuestos. Pruebe la significancia del modelo de regresión e interprete a la luz del problema.
 - d) Obtenga una estimación puntual, un intervalo de confianza y un intervalo de predicción para la media de las horas de trabajo de un lote de tamaño 80.
 - e) Pruebe estadísticamente la veracidad de la siguiente afirmación: Por cada unidad en que se aumenta el tamaño del lote, las horas de trabajo aumentan en promedio en más de 4 unidades.

ID lote	Tamaño lote	Horas de trabajo
1	80	399.00
2	30	121.00
3	50	221.00
4	90	376.00
5	70	361.00
6	60	224.00
7	120	546.00
8	80	352.00
9	100	353.00
10	50	157.00
11	40	160.00
12	70	252.00
13	90	389.00
14	20	113.00
15	110	435.00
16	100	420
17	30	212
18	50	268
19	90	377
20	110	421
21	30	273
22	90	468
23	40	244
24	80	342
25	70	323

Tabla 6: Datos Problema 11

Grupo	Integrantes	Problemas
1	Rios Castro, Kaline Andrea	4 y 6
	Velásquez Hurtado, Jorge Alexis	
	Yepes Pareja, Maria Fernanda	
2	Gallego Villa, Julián Alejandro	5 y 7
	Garcia Muñoz, Jhonatan Smith	
	Pico Arredondo, Daniela	
3	Franco Valencia, Santiago	4 y 8
	Martínez Echavarría, Juan Pablo	
	Salazar Mejía, Alejandro	
4	Aristizabal Echeverri, Genaro Alfonso	5 y 9
	Rios Garcia, Jhon Alexander	
	Zabaleta Cardeño, Carmen Daniela	
5	Carvajal Torres, Santiago	4 y 10
	Nieto Morales, Alexis Andrés	
	Pabón Palacio, Antonio	
6	Bula Isaza, Juan Daniel	5 y 11
	Giraldo Hernandez, John Carlos	
	Velez Rivera, Vanessa	
7	Duque Calle, David	4 y 7
	Hoyos Arias, John Daniel	
	Naranjo Garcia, Yised Katerine	
8	Granada Alvarez, Santiago	5 y 6
	Tous Diaz, Cleidy Jimena	
	Vásquez Gómez, Kleider Stiven	
9	Cordoba Espinosa, Catherine Andrea	4 y 9
	Julio Alvarez, Ashly Marcela	
	Posada Mazo, Estefania	
10	Alvarez Morales, Guillermo	5 y 8
	Arzuaga Gonzalez, Maria Isabel	
	Murillo Anzola, Christian Camilo	
11	Gómez Valencia, Beatriz Valentina	4 y 11
	Vanegas Castaño, Valentina	

Tabla 7: Grupos y asignación de problemas