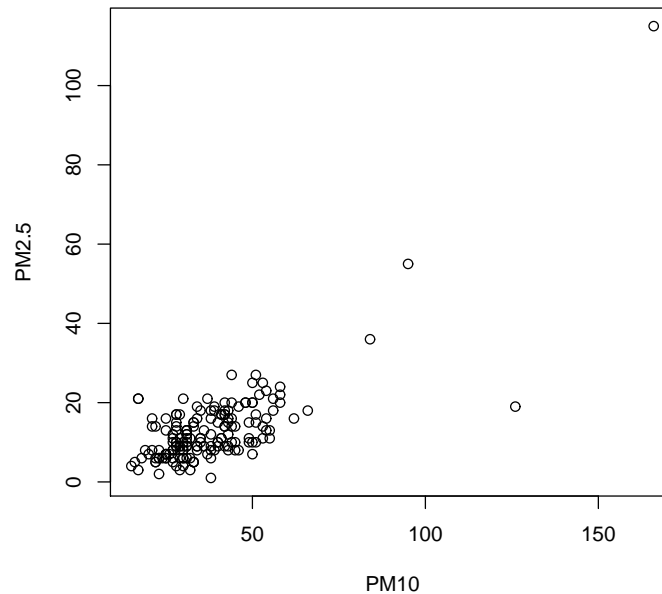


TAREA 5, 2018
Diagnósticos en el Análisis de Regresión,
Regresión a Través del Origen,
Regresión Robusta
PROFESOR JAVIER OLAYA

1. El problema

La figura 1 muestra el diagrama de dispersión de mediciones a lo largo de una semana de las variables $PM_{2.5}$ y PM_{10} en la estación Compartir del sistema de vigilancia de la calidad del aire de Cali. Los datos están contenidos en el archivo “pm2016.csv” (ver Campus Virtual). Una medida de sumo interés

Figura 1: $PM_{2.5}$ vs PM_{10}



en estudios ambientales es la fracción $PM_{2.5}/PM_{10}$, que ha sido estimada usando modelos de regresión lineal simple. Se deduce de la figura 1, que algunos puntos podrían tener altos residuales o tener alta influencia o ser atípicos (o todas las anteriores). El propósito de esta tarea es ajustar un modelo lineal simple a estos datos, con el propósito de estimar la fracción $PM_{2.5}/PM_{10}$ y decidir sobre si existen puntos con altos residuales, puntos con alto leverage, puntos influyentes o puntos atípicos. Una solución alterna, si se identifican atípicos, es construir un modelo de regresión robusta.

2. Procedimiento

Para dar respuesta a estas preguntas, cada grupo deberá proceder de la siguiente manera

1. Ajuste un modelo lineal simple con estas dos variables. Para estimar la fracción $PM_{2.5}/PM_{10}$, elija una de las dos variables como variable de respuesta y justifique su elección.
2. Juzgue su modelo y concluya. ¿Es posible asumir que el intercepto de su modelo es igual a cero?
3. Conduzca pruebas de diagnóstico para identificar posibles puntos con altos residuales, puntos con alto leverage, puntos influyentes o puntos atípicos. Si los encuentra, ¿qué haría?
4. En seguida, aún si se concluye que NO es posible asumir que el intercepto de su modelo sea igual a cero, ajuste de nuevo el modelo, pero asuma ahora que el intercepto es igual a 0. ¿Es mejor este modelo que el anterior? ¿Por qué?
5. Conduzca de nuevo las pruebas de diagnóstico y formule sus conclusiones.
6. Elija uno de los dos modelos y sustente su elección.
7. En seguida, verifique si los supuestos del modelo se cumplen para el modelo que eligió.
8. Finalmente ajuste un modelo de regresión robusta a sus datos. ¿Es este nuevo modelo mejor que los anteriores? ¿Por qué sí o por qué no?

Fecha de entrega

Miércoles 30 de mayo de 2018 en el Campus Virtual, a más tardar a la media noche.

Las presentaciones se entregan el día martes 22 de mayo de 2018, a más tardar a las 10 de la noche. Por favor tener presente que quienes no suben su presentación al campus virtual, están informando de esa manera que no harán su exposición oral y en consecuencia perderán el derecho a presentar el informe escrito.

Recuerden que según dice la gente, los ratones dejaron de comerse el queso el día que se inventaron las excusas.