ANÁLISIS EN LOS RESIDUOS DE REDES ALTIMÉTRICAS MÍNIMAMENTE CONDICIONADAS

Claudio E. JUSTO y María Beatriz PINTARELLI
Departamento de Agrimensura. Facultad de Ingeniería, UNLP
claudio.justo@ing.unlp.edu.ar

INTRODUCCIÓN

El análisis de residuos es una parte fundamental en el estudio individual de las observaciones en relación al modelo. En el caso de redes altimétricas es sabido que la escasez de grados de libertad es una de las características de su ajuste. Se busca mostrar como con herramientas estadísticas de uso común en el análisis de regresión lineal múltiple es posible realizar una comprobación de residuos a partir de ciertas condiciones. Tales herramientas consisten de estadísticos, histogramas y ploteos de residuos de los que valernos para la toma de decisiones. Para obtener las observaciones a ajustar se determinaron los desniveles entre una serie de marcas físicas de esta Facultad, mediante el método de nivelación compuesta con varios niveles automáticos similares. Con cada relevamiento se realizó un ajuste por Mínimos Cuadrados Ponderados con Condicionamiento Mínimo en el mismo punto fijo. Finalmente también se realizó un ajuste con todas las observaciones con el objeto de verificar la normalidad de las observaciones. Luego del ajuste quedó demostrado que el uso de observaciones duplicadas posibilitó construir histogramas y ploteos más cercanos a las hipótesis previas así como contar con estadísticos capaces de responder estrictamente a una distribución de probabilidades determinada. Además, por medio de la Distancia de Cook se buscó la influencia de las observaciones con residuos atípicos en los parámetros estimados.

PARTE EXPERIMENTAL

Las observaciones (desniveles) fueron tomadas, por estudiantes de la carrera Ingeniero Agrimensor, con niveles automáticos de similares características y durante distintos cursos del dictado de la asignatura Cálculo de Compensación. Se aplicó el modelo topográfico de equipotenciales esféricas concéntricas con estaciones equidistantes entre miras y cada desnivel se midió dos veces.

En la vista aérea se aprecian las ubicaciones de las marcas físicas vinculadas mediante diferencia de alturas. Como datum (cota conocida para evitar singularidad) se adoptó el punto AGRIM_VIEJA (AV).

Se ajustaron las observaciones por Mínimos Cuadrados Ponderados (observaciones con distinta jerarquía) mediante ecuaciones de observación:

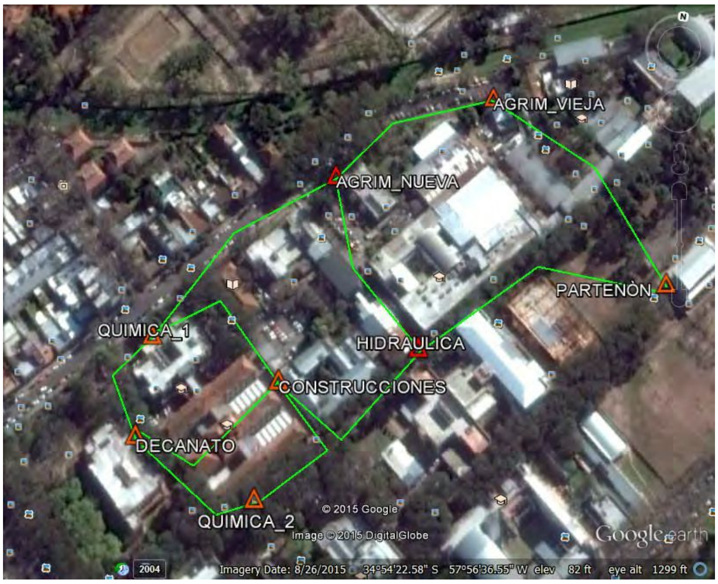
$$\boldsymbol{y} = \boldsymbol{A}.\boldsymbol{X} + \boldsymbol{C} + \boldsymbol{\varepsilon}$$

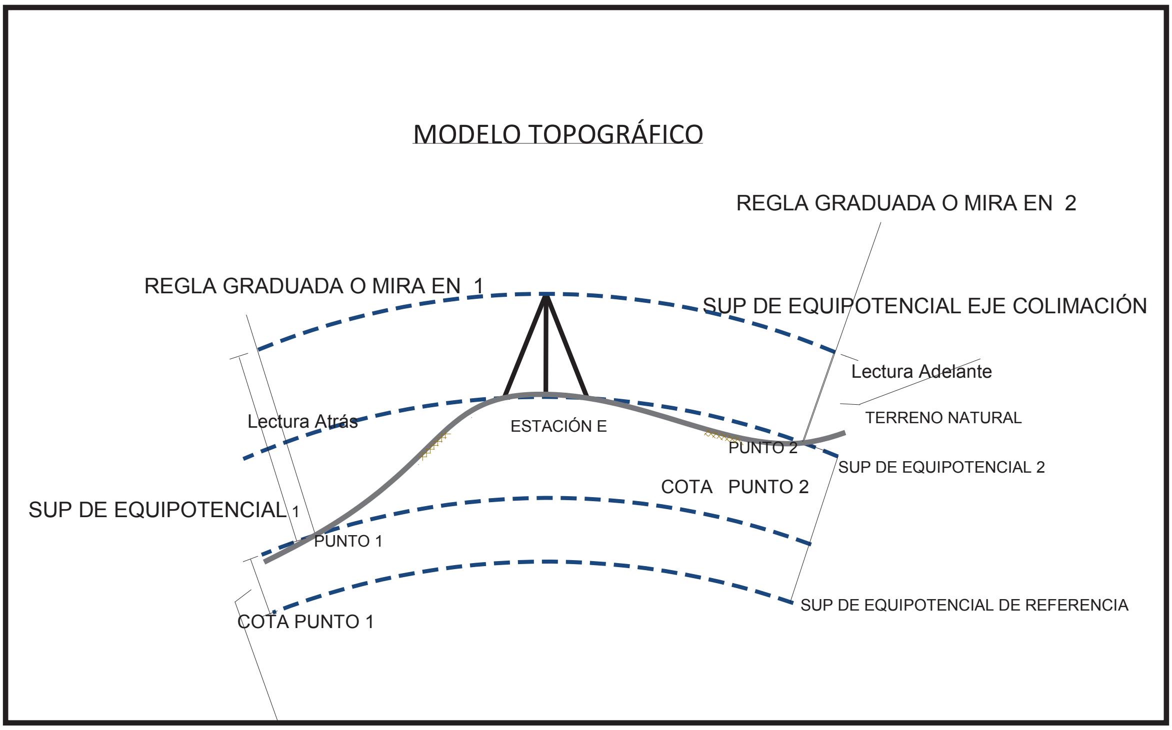
El estudio de este vector $\boldsymbol{\varepsilon}$, luego de ser estimado, es el motivo de este trabajo. Este vector tiene por hipótesis distribución normal con $\boldsymbol{E(\varepsilon)=0}$ mientras que su varianza a priori está expresada por la matriz de varianzas covarianzas $\boldsymbol{V \equiv Var(\varepsilon)}$ y que representa las varianzas de los desniveles medidos ΔH

$$\boldsymbol{Var(\varepsilon) = \sigma^2\ Q}$$

$$Var(\Delta H_{AB}) = \sigma_{km}{}^{2}\, Q_{AB} = V$$

El valor $\sigma_{km}{}^{2}$ es la varianza a priori que resultaría de nivelar un kilómetro en forma simple. La matriz ***Q*** es la matriz de cofactores en cuya diagonal principal se encuentran la cantidad de Km empleados para la determinación de cada desnivel.





Nuestra matriz de pesos ***P*** provendrá de la inversa de ***V***. El vector ***C*** es el que permite introducir el datum mediante un condicionamiento mínimo. Este tipo de condicionamiento implicará que sólo las observaciones determinaran la calidad del ajuste. La resolución del sistema de ecuaciones normales

$$A^t.P.A.\widehat{X} = A'.P.(y + C)$$

Permite obtener una única solución para el vector de cotas $\widehat{X}$. A partir de éste obtenemos el vector de residuos estimados ***e***

$$e = A.\widehat{X} - (y + C)$$

Los residuos podrán estudiarse gráfica y/o analíticamente pero antes debe considerarse su **estandarización o escalamiento** ya que luego del ajuste los residuos tendrán diferente varianza aunque a priori sean las observaciones del mismo peso o precisión [2]. Esto debe hacerse ya sea para estudios gráficos o con estadísticos.

La estudentizaciòn se realiza mediante la expresión,

$$r_i = \frac{e_i}{s_{e_i}} = \frac{e_i}{\sqrt{s^2.q_{e_i}}}$$

Donde s_{e_i} es la varianza de cada residuo estimado para lo cual deberemos contar con la matriz de varianzas covarianzas [1] de los residuos la cual se expresa:

$$S_e^2 = S^2.(P^{-1} - A.(A^t.P.A)^{-1}.A^t) = S^2.Q_e$$

Cuando el modelo es correcto estos residuos r_i tienen varianza constante y $\sigma_{r_i}^2 = 1$.

Los residuos así escalados podrán estudiarse en histogramas o en QQPlots ya que pertenecerán a una misma población. Hay que recordar que se recomienda su empleo cuando el número de observaciones es superior a 20 [2].

Los r_i serán estadísticos con una distribución ***tipo t*** [2] con ***n-r*** grados de libertad. Estrictamente no pertenecen a una distribución ***t*** por no ser independientes el numerador del denominador [2]. Este tipo de escalamiento donde se incluye a la observación sospechada de atípica se denomina **interno**.

Para solucionar la circunstancia de la falta de independencia entre numerador y denominador se puede estimar $S^2{}_{(i)}$ prescindiendo de la observación Δh_i considerada atípica en función de su residuo r_i. Esto independiza numerador y denominador evitando un escalamiento interno [2]

Se obtiene un estadístico con distribución **T de Student** con n-p-1 grados de libertad. La estimación de $S^2{}_{(i)}$ puede ser realizada con la expresión

$$S^2{}_{(i)} = \frac{(n-p).S^2 - \frac{e_i^2}{(1-h_{ii})}}{n-p-1}$$

Recordando que $(1 - h_i)$ es q_{e_i}

Esta expresión evita tener que hacer los ajustes nuevamente. Así el estadístico de estudio será

$$R_i = \frac{e_i}{\sqrt{S^2{}_{(i)} q_{e_i}}}$$

Esto último sólo será posible cuando contemos con observaciones duplicadas.

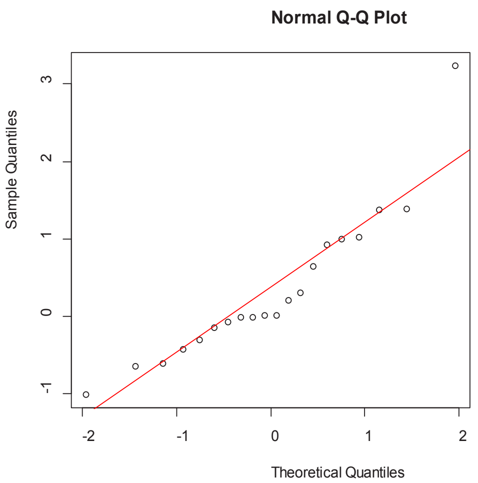
En la tabla se pueden ver los desniveles medidos y sus valores en metros en las primeras dos columnas. En la tercera columna se encuentran los desniveles ajustados al modelo por Mínimos Cuadrados Ponderados y en la cuarta columna los residuos brutos ***e***.

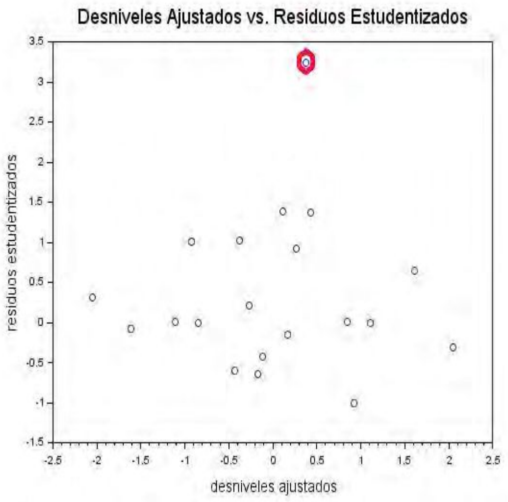
Tramo medido	DH bruto	DH ajustado	*e*	*r*	R	Dist. Cook
AV - AN	+2.046	+2.0460423	0.0004231	-0.30553	-0.294603	0.0085646
Q1 - AN	+1.110	+1.1100164	0.0000164	-0.010755	-0.010333	0.0000084
D – Q1	+0.112	+0.1105502	0.0014498	1.3849766	1.4412095	0.1822203
Q2 – D	+0.381	+0.3764464	0.0045536	3.2334299	7.0212768	0.7538931
P - H	-1.615	-1.6148721	0.0001279	-0.0727518	-0.069912	0.0003866
C – Q2	-0.920	-0.9212172	0.0012172	1.0037532	1.0040679	0.0849196
AN - H	+0.166	+0.1661877	0.0001877	-0.1491127	-0.143386	0.0016492
C-Q1	-0.435	-0.4342206	0.0007794	-0.601502	-0.586118	0.0233012
P - AV	+0.267	+265363	0.0016367	0.9227019	0.9170405	0.0615705
C - H	+0.842	+0.8419835	0.0000165	0.010763	0.0103408	0.0000084
AN - AV	-2.046	-2.0460423	0.0004231	0.30553	0.2946033	0.0085646
AN – Q1	-1.110	-1.1100164	0.0000164	0.010755	0.0103331	0.0000084
Q1 – D	-0.111	-0.1105502	0.0004498	-0.429699	-0.415805	0.0175404
D – Q2	-0.375	-0.3764464	0.0014464	1.0270777	1.0294347	0.0760658
H - P	+1.616	+1.6148721	0.0011279	0.641391	0.6262164	0.0300517
Q2 - C	+0.920	+0.9212172	0.0012172	-1.0037532	-1.004068	0.0849196
H-AN	-0.167	-0.1661877	0.0008123	-0.6451983	-0.630056	0.0308775
Q1-C	+0.436	+0.4342206	0.0017794	1.3732679	1.4269477	0.1214548
AV-P	-0.265	-0.26535	0.0003633	0.2048246	0.1971074	0.003034
H - C	-0.842	-0.8419835	0.0000165	-0.010763	-0.010341	0.0000084

En las últimas tres columnas se ven los residuos estudentizados r_i y los residuos **R de Student** y los valores de la Distancia de Cook.

Esta red resultó con una varianza del ajuste de $S^2 =$ **0.00002304m²** con $n - p = 13$ grados de libertad. La varianza propuesta a priori de $\sigma^2_{Km} =$ 0.000025m² fue satisfactoriamente testeada bilateralmente mediante Chi Cuadrado para $\alpha = 0.05$.

Para estudiar los residuos individualmente primero se determinó la magnitud de los residuos estudentizados ***r*** y posteriormente, sexta columna, la de los residuos R. Los primeros se analizaron para $\alpha = 0.05$ con un $t_{\frac{\alpha}{2}} = 2.16$ [3], y los segundos al tener $n - p - 1 = 12$ grados de libertad, para un $t_{\frac{\alpha}{2}} = 2.18$ [3].Se encontró el valor ***i=4*** alcanzó el valor ***r*** **=** 3.23 superior al $t_{\frac{\alpha}{2}} = 2.16$ [3] considerando a la misma como atípica según este criterio. El mismo residuo cuando se lo escala en forma externa se obtiene un valor de 7.02. Este último residuo tiene distribución estrictamente T de Student ya que numerado y denominador son independientes. Finalmente tenemos el QQPlot de los residuos estudentizados y el gráfico de los residuos estudentizados contra los valores ajustados de los desniveles.





Influencia de las observaciones

Cuando es encontrada una observación atípica puede ser útil obtener una medida de su influencia. La Distancia de Cook [2] mide si la distancia cartesiana, elevada al cuadrado, entre dos soluciones es significativa. Estas soluciones se diferencian en que una de ellas no incluye al dato atípico. Esta distancia puede expresarse en función de los parámetros de la siguiente manera

$$D_i = \frac{(X_{(i)} - X)^t . A^t . P . A . (X_{(i)} - X)}{p . S^2}$$

En la práctica suele emplearse el siguiente estadístico [2].

$$D_i = \frac{r_i^2}{p}\left(\frac{s_{\hat{l}_i}^2}{s_{e_i}^2}\right)$$

Donde r_i es el residuo estudentizado y p la cantidad de parámetros del ajuste. D_i no es un estadístico F sin embargo usar el valor de corte $F_{0.5,p,n-p} \approx 1$ funciona bien en la práctica [2]. Es decir que se mide la significancia de una observación presuntamente influyente. Lo usual es poner en duda la observación cuya $D_i \geq 1$.

CONCLUSIONES

El empleo de técnicas gráficas como QQPlots y ploteo de residuos contra valores ajustados muestra rápidamente el cumplimiento, o no, de hipótesis, como homogeneidad de varianzas y adecuación al modelo, así como también la existencia de valores atípicos. La decisión de calificar las observaciones estudiando el residuo que arrojan luego del ajuste con la construcción de estadísticos con distribución T de Student debe adecuarse a la cantidad de datos. Con redes cuyos desniveles fueron medidos solamente una vez solo pueden estudentizarse en forma interna ya que no puede excluirse ninguna de las observaciones sin perder parte de la red. La estudentización externa, aplicable solo a redes con observaciones dobles mostró cómo los residuos reaccionan en caso de tener una varianza estimada sin el aporte del dato atípico. La Distancia de Cook no mostró ser sensible al valor atípico. Cabe recordar que las circunstancias que generen el apartamiento del valor esperado se distribuyen a toda la red. La Distancia de Cook mide la influencia en todos los parámetros estimados. Existen estadísticos que miden la influencia sobre cada uno de los parámetros estimados siendo posiblemente más pertinente su empleo.

BIBLIOGRAFÍA

[1] Wolf, Ghilani . 2006 ”*Adjustment Computations*”. 4 ta Edition. Wiley & Sons

[2] Montgomery, Peck, Vining.2007 “*Introducción al Análisis de Regresión*” LIneal. 3ª. Edición. Ed. Patria.

[3] Walpole, Myers, Myers “*Probabilidad y Estadística para Ingenieros*”. Sexta Edición. Pearson Educación.