Punto 1.

Suponga que $A$ es una matriz de tamaño $mxn$. Para que el método de mínimos cuadrados funcione, necesitamos que la ecuación $A^TAx = b$ tenga solución única, esto es, que la matriz $A^TA$ sea no singular. Probaremos que esto ocurre si y sólo si las columnas de $A$ son linealmente independientes.

(Necesidad) Sean $v_1, v_2, ..v_n$ los vectores columna de $A$. Si dichos vectores son linealmente dependientes, entonces uno de ellos, supongamos sin pérdida de generalidad que es $v_n$, se puede escribir como combinación lineal del resto. Es decir, existen constantes $c_1, c_2, ..c_{n-1}$ tales que $v_n = c_1v_1 + c_2v_2 + \cdots c_{n-1}v_{n-1}$. Entonces ,si $A = [v_1, v_2, ...v_n]$, tenemos que $A^TA = [A^Tv_1, A^Tv_2, ...A^Tv_n] = [A^Tv_1, A^Tv_2, ...c_1A^Tv_1 + c_2A^Tv_2 + \cdots c_{n-1}A^Tv_{n-1}]$, luego el último vector columna de $A^TA$ puede escribirse como combinación lineal del resto, pero al ser $A^TA$ una matriz cuadrada esto implica que su determinante es cero y es singular. Por lo tanto, si $A^TA$ es no singular,$v_1, v_2, ..v_n$ deben ser linealmente independientes.

(Suficiencia) Suponga que las columnas de $A$ son linealmente independientes. Para ver que $A^TA$ es no singular, sea $x$ una solución al sistema $A^TAx = 0$, y mostraremos que $x = 0$. Observe que $x^TA^TAx = (Ax)^T(Ax) = ||Ax||^2 = 0$, de modo que también se debe tener que $Ax = 0$. Pero como las columnas de $A$ son linealmente independientes, el rango de $A$ es cero y la única solución de este sistema es la trivial, de modo que la única solución del sistema $A^TAx = 0$ también es la trivial. Concluimos que $A^TA$ debe ser no singular.

Para dar un ejemplo particular de cómo se podría aplicar este criterio, suponga que queremos usar el método de mínimos cuadrados para aproximar un conjunto de datos de tamaño $m$ con un modelo que tiene $n$ grados de libertad( por ejemplo, un polinomio de grado $n$ tiene $n+1$ grados de libertad). El criterio anterior nos dice que si $m < n$, entonces el método de mínimos cuadrados no es aplicable. ¿Por qué?. Pues el número de grados de libertad viene a ser el número de columnas de $A$ y los datos el número de filas, pero no es posible que $n$ vectores de tamaño $m$ sean linealmente independientes si $n > m$, de modo que el método de mínimos cuadrados por el criterio anterior no funcionará en este caso. Para dar un ejemplo pequeño, no podemos modelar un conjunto de 2 datos con un polinomio de grado 2, pues existen infinitos polinomios de este tipo que se ajustan perfectamente a estos datos.