1)

We need to minimize the error term in $y = ax + b + c$. To do that for all data points, we must minimize $\sum_{i=1}^{m} c_i^2 = \sum_{i=1}^{m} (y_i - (ax_i + b))^2$

To find the minimum, take the derivative with respect to a and b and set them $= 0$

$$f = \sum_{i=1}^{m} (y_i - (ax_i + b))^2$$

$$\frac{df}{da} = \frac{d \sum_{i=1}^{m} (y_i - (ax_i + b))^2}{da} = 0$$

$$\sum_{i=1}^{m} x_i (y_i - (\hat{a}x_i + \hat{b})) = 0$$

$$\sum_{i=1}^{m} x_i (y_i - (\hat{a}x_i + \bar{y} - \hat{a}\bar{x})) = 0$$

$$\sum_{i=1}^{m} x_i (y_i - \bar{y} - \hat{a}(x_i - \bar{x})) = 0$$

$$\sum_{i=1}^{m} x_i (y_i - \bar{y}) - \sum_{i=1}^{m} \hat{a} x_i (x_i - \bar{x}) = 0$$

$$\sum_{i=1}^{m} x_i (y_i - \bar{y}) = \hat{a} \sum_{i=1}^{m} x_i (x_i - \bar{x})$$

$$\hat{a} = \frac{\sum_{i=1}^{m} x_i (y_i - \bar{y})}{\sum_{i=1}^{m} x_i (x_i - \bar{x})}$$

$$\hat{a} = \frac{\sum_{i=1}^{m} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{m} (x_i - \bar{x})^2}$$

$$\frac{df}{db} = \frac{d \sum_{i=1}^{m} (y_i - (ax_i + b))^2}{db} = 0$$

$$-2 \sum_{i=1}^{m} (y_i - (ax_i + b)) = 0$$

$$\sum_{i=1}^{m} (y_i - (ax_i + b)) = 0$$

$$\sum_{i=1}^{m} y_i - \sum_{i=1}^{m} ax_i - \sum_{i=1}^{m} b = 0$$

$$\sum_{i=1}^{m} y_i - \sum_{i=1}^{m} ax_i - mb = 0$$

$$\sum_{i=1}^{m} y_i - \sum_{i=1}^{m} ax_i = mb$$

$$\frac{1}{m} \sum_{i=1}^{m} y_i - \frac{1}{m} a \sum_{i=1}^{m} x_i = b$$

$$\frac{1}{m} \sum_{i=1}^{m} y_i = \bar{y} \text{ and } \frac{1}{m} \sum_{i=1}^{m} x_i = \bar{x}$$

$$\therefore b^* = \bar{y} - a^* \bar{x}$$

3)

a) $P = A(A^TA)^{-1}A^T$

$P^T = (A(A^TA)^{-1}A^T)^T$

$= A^T((A^TA)^{-1})^T(A^T)^T$

$= A^T(A^TA)^{-T}A$

$= A(A^TA)^{-1}A$

$= P$

Since $P = P^T$, P must be symmetric

b) For any nonzero vector v:

$v^TA^TAv = (Av)^TAv = ||Av||^2 > 0$

Since A^TA is symmetric (since the columns of A are linearly independent) and $v^TA^TAv > 0$, A^TA is positive definite and ∴ invertable.

c) Let $n = 1$

Basis: $P^1 = P$

Hypothesis: $P^{n+1} = P$

$P^2 = (A(A^TA)^{-1}A^T)(A(A^TA)^{-1}A^T)$

$= A(A^TA)^{-1}(A^TA)(A^TA)^{-1}A^T$

$= A(I)(A^TA)^{-1}A^T$

$= A(A^TA)^{-1}A^T$

$= P$

∴ by induction $P^n = P$ for any positive integer

d) $Pv = \lambda v \Rightarrow A(A^TA)^{-1}A^Tv = \lambda v$

$A^TA(A^TA)^{-1}A^Tv = \lambda A^Tv$

A^TA is positive definite

A^Tv is nonzero

∴ λ must be 0 or 1

e) $trace(P) = trace(A(A^TA)^{-1}A^T)$

$= trace(A^TA(A^TA)^{-1}) = trace(I) = rank(P)$