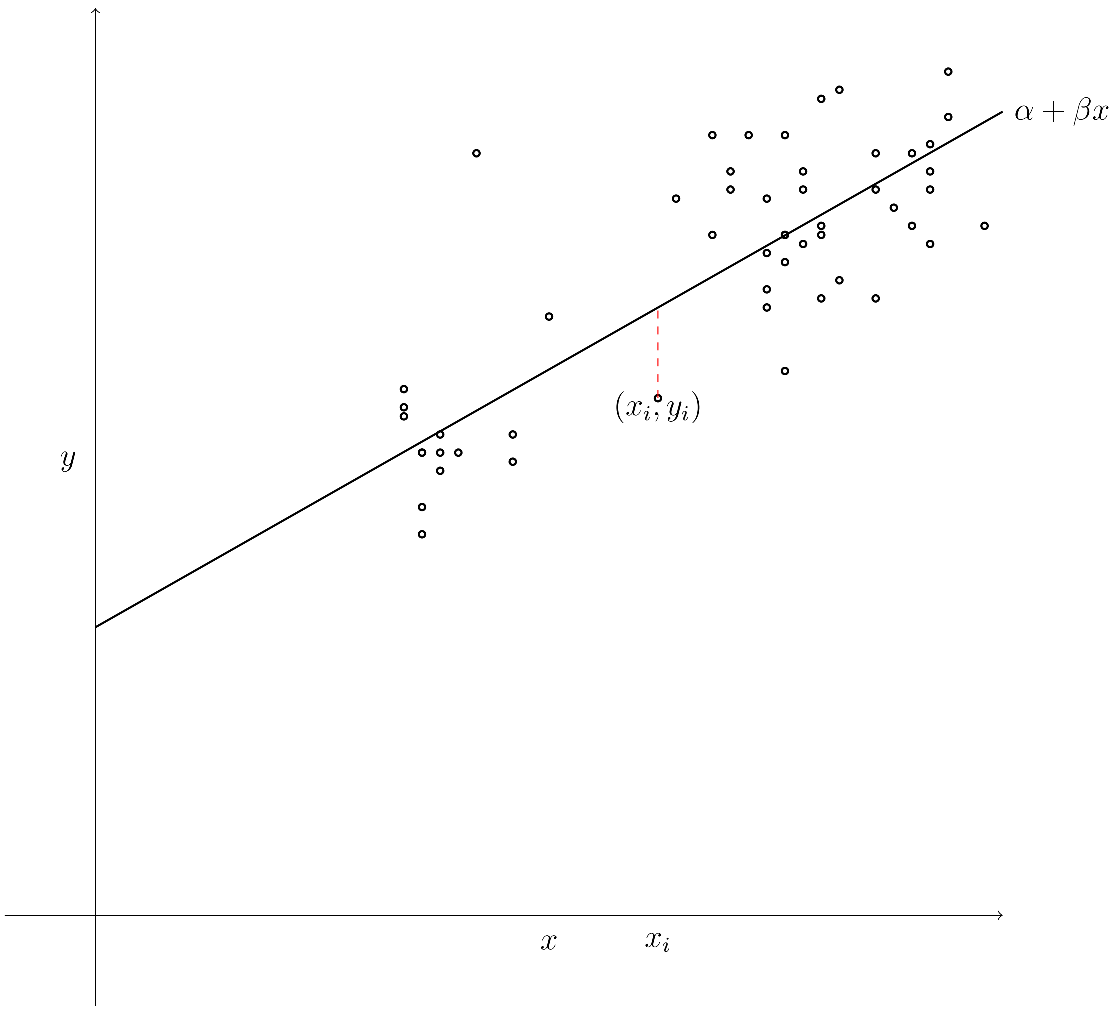
3. Least squares.

Linear models. Response y and factor/predictor x are measured on n subjects: $(x_1, y_1), \ldots, (x_n, y_n)$. Modeling the linear dependence of y on x for estimation and prediction is of interest. With this in view, the following model is explored.
$y_i = \alpha + \beta x_i + \epsilon_i$, $i = 1, \ldots, n$ where ϵ_i are uncorrelated random errors with mean 0 and variance σ^2. Then the *least squares* method is to estimate α and β by:

$$\min_{\alpha,\beta} \sum_{i=1}^{n} (y_i - (\alpha + \beta x_i))^2.$$



Since we need to minimize a quadratic function in α and β, we may simply differentiate it and set the partial derivatives to 0. We then obtain

$$\begin{aligned}
\hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x}, \text{ and} \\
\hat{\beta} &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
&= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \times \frac{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \\
&= r_{xy}\frac{s_y}{s_x},
\end{aligned}$$

where r_{xy} is the correlation coefficient between x and y, and s_x, s_y are the s.d. of x and y, respectively. $\hat{y} = \hat{\alpha} + \hat{\beta}x = \hat{\alpha} + r_{xy}\frac{s_y}{s_x}x$ is the least squares equation to predict y based on x.

Now suppose ϵ_i are i.i.d $N(0, \sigma^2)$. Then y_i are independent and $y_i \sim N(\alpha + \beta x_i, \sigma^2)$. What is the MLE of $(\alpha, \beta, \sigma^2)$? Note that the model is for $y|x$, treating x fixed. Then, we have,

$$\begin{aligned}
f(\mathbf{y}|\alpha, \beta, \sigma^2) &= (2\pi)^{-n/2}\sigma^{-n}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \alpha - \beta x_i)^2\right), \\
\mathcal{L}(\alpha, \beta, \sigma^2) &= -n\log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \alpha - \beta x_i)^2.
\end{aligned}$$

For each fixed σ^2, maximization of $\mathcal{L}(\alpha, \beta, \sigma^2)$ over α, β is the same as minimization of $\sum_{i=1}^n (y_i - \alpha - \beta x_i)^2$. Therefore MLE of (α, β) is the same as the least squares estimate. This shows an optimality property of least squares under normality.

Now we show that the above normal linear model is a 3-parameter exponential family. Note that

$$\begin{aligned}
&\sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 = \sum_{i=1}^n \left(y_i - \hat{\alpha} - \hat{\beta}x_i - (\alpha - \hat{\alpha}) - (\beta - \hat{\beta})x_i\right)^2 \\
&= \sum_{i=1}^n \left(y_i - \hat{\alpha} - \hat{\beta}x_i\right)^2 + \sum_{i=1}^n \left\{-(\alpha - \hat{\alpha}) - (\beta - \hat{\beta})x_i\right\}^2 \\
&\quad -2\sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)\left\{(\alpha - \hat{\alpha}) + (\beta - \hat{\beta})x_i\right\} \\
&= \sum_{i=1}^n \left(y_i - \hat{\alpha} - \hat{\beta}x_i\right)^2 + \sum_{i=1}^n \left\{(\alpha - \hat{\alpha}) + (\beta - \hat{\beta})x_i\right\}^2,
\end{aligned}$$

since

$$\begin{aligned}
&\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}x_i)\left\{(\alpha - \hat{\alpha}) + (\beta - \hat{\beta})x_i\right\} \\
&= \sum_{i=1}^{n}(y_i - \bar{y} - \hat{\beta}(x_i - \bar{x}))\left\{(\alpha - \hat{\alpha}) + (\beta - \hat{\beta})x_i\right\} \\
&= (\alpha - \hat{\alpha})\sum_{i=1}^{n}(y_i - \bar{y}) - (\alpha - \hat{\alpha})\hat{\beta}\sum_{i=1}^{n}(x_i - \bar{x}) + (\beta - \hat{\beta})\sum_{i=1}^{n}(y_i - \bar{y})x_i \\
&\quad -\hat{\beta}(\beta - \hat{\beta})\sum_{i=1}^{n}x_i(x_i - \bar{x}) \\
&= (\beta - \hat{\beta})\sum_{i=1}^{n}x_i\left\{(y_i - \bar{y}) - \hat{\beta}(x_i - \bar{x})\right\} \\
&= (\beta - \hat{\beta})\left\{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) - \hat{\beta}\sum_{i=1}^{n}(x_i - \bar{x})^2\right\} \\
&= 0.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&f(\mathbf{y}|\alpha, \beta, \sigma^2) = (2\pi)^{-n/2}\sigma^{-n} \\
&\times \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}x_i)^2 + \sum_{i=1}^{n}\{(\alpha - \hat{\alpha}) + (\beta - \hat{\beta})x_i\}^2\right]\right),
\end{aligned}$$

and hence $\{\hat{\alpha}, \hat{\beta}, \sum_{i=1}^{n}\left(y_i - \hat{\alpha} - \hat{\beta}x_i\right)^2\}$ is sufficient for $(\alpha, \beta, \sigma^2)$. To show that this is an exponential family, note

$$\begin{aligned}
\sum_{i=1}^{n}\left\{(\alpha - \hat{\alpha}) + (\beta - \hat{\beta})x_i\right\}^2 &= \sum_{i=1}^{n}\left\{(\alpha + \beta x_i) - (\hat{\alpha} + \hat{\beta}x_i)\right\}^2 \\
&= \sum_{i=1}^{n}(\alpha + \beta x_i)^2 + \sum_{i=1}^{n}(\hat{\alpha} + \hat{\beta}x_i)^2 \\
&\quad -2\sum_{i=1}^{n}(\hat{\alpha} + \hat{\beta}x_i)(\alpha + \beta x_i).
\end{aligned}$$

Therefore,

$$
\begin{aligned}
& f(\mathbf{y}|\alpha,\beta,\sigma^2) \\
= \ & \exp\Bigg(\frac{\alpha}{\sigma^2}\sum_{i=1}^n(\hat{\alpha}+\hat{\beta}x_i)+\frac{\beta}{\sigma^2}\sum_{i=1}^n(\hat{\alpha}+\hat{\beta}x_i)x_i \\
& -\frac{1}{2\sigma^2}\left[\sum_{i=1}^n(y_i-\hat{\alpha}-\hat{\beta}x_i)^2+\sum_{i=1}^n(\hat{\alpha}+\hat{\beta}x_i)^2\right] \\
& -\frac{1}{2\sigma^2}\sum_{i=1}^n(\alpha+\beta x_i)^2-n\log(\sigma)-n\log(\sqrt{2\pi})\Bigg),
\end{aligned}
$$

so that we can take $c_1(\alpha,\beta,\sigma^2)=\alpha/\sigma^2$, $T_1(\mathbf{y})=\sum_{i=1}^n(\hat{\alpha}+\hat{\beta}x_i)$, $c_2(\alpha,\beta,\sigma^2)=\beta/\sigma^2$, $T_2(\mathbf{y})=\sum_{i=1}^n(\hat{\alpha}+\hat{\beta}x_i)x_i$, $c_3(\alpha,\beta,\sigma^2)=-1/(2\sigma^2)$,
$T_3(\mathbf{y})=\sum_{i=1}^n(y_i-\hat{\alpha}-\hat{\beta}x_i)^2+\sum_{i=1}^n(\hat{\alpha}+\hat{\beta}x_i)^2$. Then, we can also use exponential family methods to find MLE of (α,β,σ^2). Since we already know the MLE for α and β, and since, for each σ^2,

$$
\max_{\alpha,\beta} L(\alpha,\beta,\sigma^2)=L(\hat{\alpha},\hat{\beta},\sigma^2)=\sigma^{-n}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n(y_i-\hat{\alpha}-\hat{\beta}x_i)^2\right),
$$

we can find the MLE of σ^2 by finding

$$
\max_{\sigma^2} L(\hat{\alpha},\hat{\beta},\sigma^2)=\max_{\sigma^2}(\sigma^2)^{-n/2}\exp\left(-\frac{t^2}{2\sigma^2}\right),
$$

where $t^2=\sum_{i=1}^n(y_i-\hat{\alpha}-\hat{\beta}x_i)^2$. Then,

$$
\begin{aligned}
\mathcal{L}(\hat{\alpha},\hat{\beta},\sigma^2) &= -\frac{n}{2}\log(\sigma^2)-\frac{1}{2\sigma^2}t^2, \\
\frac{\partial\mathcal{L}(\hat{\alpha},\hat{\beta},\sigma^2)}{\partial\sigma^2} &= -\frac{n}{2}\sigma^2+\frac{t^2}{2}\frac{1}{(\sigma^2)^2}=\frac{1}{2\sigma^4}(t^2-n\sigma^2).
\end{aligned}
$$

Check that MLE of σ^2 is

$$
\hat{\sigma}^2=\frac{t^2}{n}=\frac{1}{n}\sum_{i=1}^n(y_i-\hat{\alpha}-\hat{\beta}x_i)^2.
$$