Excercise 3

a)

$$\mathbf{y} = f(\mathbf{x}) + \boldsymbol{\epsilon} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$y_i = \sum_{j=0}^{p-1} X_{ij}\beta_j + \mathcal{N}\big(0, \sigma^2\big)$$

$$\langle y_i \rangle = \left\langle \sum_{j=0}^{p-1} X_{ij}\beta_j \right\rangle + \big\langle \mathcal{N}\big(0, \sigma^2\big)\big\rangle$$

$\sum_{j=0}^{p-1} X_{ij}\beta_j$ contains only the input values $\mathbf{x}$ stored in the design matrix $\mathbf{X}$ and some parameters $\boldsymbol{\beta}$, i.e it has nothing to do with probability. So the expectation value is just itself, $\left\langle \sum_{j=0}^{p-1} X_{ij}\beta_j \right\rangle = \sum_{j=0}^{p-1} X_{ij}\beta_j$.

The expectation value of $\mathcal{N}\big(0, \sigma^2\big)$ is zero by definition, so $\big\langle \mathcal{N}\big(0, \sigma^2\big)\big\rangle = 0$.

$$\langle y_i \rangle = \sum_{j=0}^{p-1} X_{ij}\beta_j$$

(or $\langle y_i \rangle = \mathbf{X}_{i,*}\beta$ in Morten's notation).

The variance of some linear combination of independent stochastic variables $x = \sum_{i=0}^{n-1} \alpha_i x_i$ is $\mathrm{Var}(x) = \sum_{i=0}^{n-1} \alpha_i^2 \mathrm{Var}(x_i)$.

$$\mathrm{Var}(y_i) = \mathrm{Var}\left(\sum_{j=0}^{p-1} X_{ij}\beta_j \right) + \mathrm{Var}\big(\mathcal{N}\big(0, \sigma^2\big)\big)$$

$\sum_{j=0}^{p-1} X_{ij}\beta_j$ can be viewed as a "stochastic variable" that only gives one value, so the variance is zero. The variance of $\mathcal{N}\big(0, \sigma^2\big)$ is σ^2 by definition.

$$\mathrm{Var}(y_i) = \sigma^2$$

Adding a constant (in this case $\sum_{j=0}^{p-1} X_{ij}\beta_j$) to a normal distribution gives another normal distribution that is shifted. For example $\mathcal{N}\left(\mu, \sigma^2\right) + a = \mathcal{N}\left(\mu + a, \sigma^2\right)$.

Since μ in this case is $\mu = 0$ the values y_i are normally distributed with variance σ^2 and mean $\sum_{j=0}^{p-1} X_{ij}\beta_j$.

b)

The Ordinary Least Squares expression for $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

In index notation this is

$$\beta_i = \sum_j \left(\mathbf{X}^T\mathbf{X}\right)_{ij}^{-1} \sum_k \left(\mathbf{X}^T\right)_{jk} y_k = \sum_j \sum_k \left(\mathbf{X}^T\mathbf{X}\right)_{ij}^{-1} \left(\mathbf{X}^T\right)_{jk} y_k$$

$\left(\mathbf{X}^T\mathbf{X}\right)_{ij}^{-1} \left(\mathbf{X}^T\right)_{jk}$ can be regarded as constants since they stem from the input data $\mathbf{x}$ which has nothing to do with probability.

$$\langle \beta_i \rangle = \sum_j \sum_k \left(\mathbf{X}^T\mathbf{X}\right)_{ij}^{-1} \left(\mathbf{X}^T\right)_{jk} \langle y_k \rangle$$

Having already established from 3a) that $\langle y_k \rangle = \sum_{l=0}^{p-1} (\mathbf{X})_{kl}\beta_l$ we have

$$\langle \beta_i \rangle = \sum_j \sum_k \left(\mathbf{X}^T\mathbf{X}\right)_{ij}^{-1} \left(\mathbf{X}^T\right)_{jk} \sum_{l=0}^{p-1} (\mathbf{X})_{kl}\beta_l$$

$$= \sum_{j} \left(\mathbf{X}^T\mathbf{X}\right)^{-1}_{ij} \sum_{k} \left(\mathbf{X}^T\right)_{jk} \sum_{l=0}^{p-1} (\mathbf{X})_{kl}\beta_l$$

Let $\langle\beta_i\rangle$ be the elements of a vector $\mathbf{b}$. Then we can write the above as the matrix equation

$$\mathbf{b} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}$$

$$= \boldsymbol{\beta}$$

since $\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\mathbf{X}^T\mathbf{X}\right) = \mathbf{I}$ is the identity matrix. Thus the components of $\mathbf{b}$ are the components of $\boldsymbol{\beta}$, i.e $\langle\beta_i\rangle = \beta_i$.