One way to solve this is to appeal to facts about the multivariate normal distribution that say that if

$$\left(\begin{array}{c} y \\ \boldsymbol{x} \end{array}\right) \sim \mathcal{N}\left(\left(\begin{array}{c} \mu_y \\ \boldsymbol{\mu}_x \end{array}\right), \left(\begin{array}{cc} \sigma_y^2 & \Sigma_{y\boldsymbol{x}} \\ \Sigma_{\boldsymbol{x}y} & \Sigma_{\boldsymbol{x}\boldsymbol{x}} \end{array}\right)\right),$$

where $\boldsymbol{x}$ is potentially a vector of variables, e.g. $\boldsymbol{x} = \left(\begin{array}{c} x \\ z \end{array}\right)$ in the context of this problem, then $y|\boldsymbol{x}$ has a normal distribution with mean

$$E\left[y|x\right] = \mu_y + \Sigma_{y\boldsymbol{x}}\Sigma_{\boldsymbol{x}\boldsymbol{x}}^{-1}\left(\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{x}}\right)$$

and variance

$$\sigma^2 - \Sigma_{y\boldsymbol{x}}\Sigma_{\boldsymbol{x}\boldsymbol{x}}^{-1}\Sigma_{\boldsymbol{x}y}.$$

See for instance the Wikipedia page on the multivariate normal distribution. From here, the problem is pretty mechanical.

For the first part, we have

$$E\left(y|x\right) = 1 + \frac{3}{5}\left(x-2\right) = -\frac{1}{5} + \frac{3}{5}x.$$

For the second part, we have

$$E\left(y|x,z\right) = 1 + \left(\begin{array}{cc} 1 & 3 \end{array}\right)\left(\begin{array}{cc} 5 & 2 \\ 2 & 6 \end{array}\right)^{-1}\left(\begin{array}{c} x-2 \\ z-4 \end{array}\right) = 1 + \left(\begin{array}{cc} \frac{8}{13} & -\frac{1}{26} \end{array}\right)\left(\begin{array}{c} x-2 \\ z-4 \end{array}\right) = \frac{1}{13} + \frac{8}{13}x - \frac{1}{26}z.$$

However, this solution is not very instructive – it starts out with a fact about the normal distribution that mostly solves the problem for us. Furthermore, to derive the fact ourself (i.e., deriving the expressions for conditional distribution from the multivariate normal distribution) would be a pain and not necessarily illuminating.

Another way to proceed is to define

$$\varepsilon_i \equiv y_i - E\left[y_i|x_i\right],$$

and write

$$y_i = E\left[y_i|x_i\right] + \varepsilon_i. \tag{1}$$

Now, start with two facts that are less strong that the expression for $y|x$ above:

1. $y|x$ is normally distributed
2. $E\left[y|x\right]$ linear in x

Now, the idea is to consider the properties of equation (1) and argue that we can apply our linear regression results to understand what its coefficients will be.

Given the second fact, we can write

$$y_i = \alpha + \beta x_i + \varepsilon_i, \tag{2}$$

where

$$E[y|x] = \alpha + \beta x.$$

Furthermore, by the definition of ε_i , we can use the law of iterated expectations to argue that

$$E(\varepsilon_i|x_i) = E(y_i - E[y_i|x_i]|x_i) = E[y_i|x_i] - E[y_i|x_i] = 0.$$

Furthermore, if we consider independently simulated data on x, we would have $E(\boldsymbol{\varepsilon}|\boldsymbol{X}) = 0$, the strict exogeneity condition.

Finally, notice that fact 1 implies that ε_i is normally distributed.

Now, equation (2) satisfies all the assumptions for the linear regression model with normally distributed error terms.

Following the slides, we know that regressing y on x (and a constant) with data simulated from the assumed distribution will yield the following coefficient estimates:

$$\widehat{\beta} = \frac{s_{xy}}{s_x^2} \qquad\qquad \widehat{\alpha} = \overline{y} - \widehat{\beta}\overline{x}$$

where s_{xy} refers to the sample covariance, s_x^2 refers to the sample variance, and $\overline{y}$, $\overline{x}$ refer to sample means.

It's intuitive and easy to verify that the sample moments correspond to the population moments: $E[s_{xy}] = \sigma_{xy}$ and $E(s_x^2) = \sigma^2$. Note: for the sample moments to be unbiased, we have to divide by $n-1$ instead of n (see Bessel's correction).

Informally, this seems to suggest that

$$E\left[\widehat{\beta}\right] \approx \frac{\sigma_{xy}}{\sigma_x^2},$$

which would then imply

$$E[\widehat{\alpha}] \approx \mu_y - \frac{\sigma_{xy}}{\sigma_x^2}\mu_x.$$

If these results hold with equality, we can plug in the parameters from the population distribution, and we arrived at the same solution as above. (A similar argument can be applied to $E[y|x,z]$.) For large sample sizes, it's intuitive that the sample moments s_{xy} and s_x^2 will be very close to their population analogs, so it makes sense that $\widehat{\beta}$ should be close to $\frac{\sigma_{xy}}{\sigma_x^2}$ for a large sample.

To make this proof strategy formally complete, we need to argue that equality does hold

$$E\left[\widehat{\beta}\right] = E\left[\frac{s_{xy}}{s_x^2}\right] = \frac{E[s_{xy}]}{E[s_x^2]} = \frac{\sigma_{xy}}{\sigma_x^2}.$$

The non-obvious part is

$$E\left[\frac{s_{xy}}{s_x^2}\right] = \frac{E[s_{xy}]}{E[s_x^2]},$$

which is true because s_{xy} and s_x^2 are independent, but it's not immediately obvious that they are independent because they both rely on the same data. The independence argument can be established by starting with $n = 2$ and extending the argument by induction, or it can be done more elegantly with a change in variables. See this discussion for an example of the latter proof strategy.