1. From $Y_i = \beta_0 + \beta_1 X_i + u_i$, we see that $u_i = Y_i - \beta_0 - \beta_1 X_i$. Then

$$
E(u_i \mid X_i) = E((Y_i - \beta_0 - \beta_1 X_i) \mid X_i)
= E(Y_i \mid X_i) - \beta_0 - \beta_1 X_i
= \beta_0 + \beta_1 X_i - \beta_0 - \beta_1 X_i
$$

where the last equality uses the stated assumption that $E(Y_i \mid X_i) = \beta_0 + \beta_1 X_i$.

**Note:** Some argued: From $Y_i = \beta_0 + \beta_1 X_i + u_i$ we have $E(Y_i \mid X_i) = \beta_0 + \beta_1 X_i + E(u_i \mid X_i)$. Since it is known that $E(u_i \mid X_i) = 0$, then $E(Y_i \mid X_i) = \beta_0 + \beta_1 X_i$. This argument is backwards to what was supposed to be shown. The stated assumption was $E(Y_i \mid X_i) = \beta_0 + \beta_1 X_i$, and the question was to show that this implies that $E(u_i \mid X_i) = 0$.

2. In the model $Y_i = \beta_0 + u_i$, the SSR is

$$
SSR(\beta_0) = \sum_{i=1}^{n} (Y_i - \beta_0)^2 = \sum_{i=1}^{n} Y_i^2 - 2\beta_0 \sum_{i=1}^{n} Y_i + n\beta_0^2
$$

The derivative with respect to $\beta_0$ is

$$
\frac{\partial}{\partial \beta_0} SSR(\beta_0) = -2 \sum_{i=1}^{n} Y_i + n2\beta_0
$$

This is zero at $\beta_0 = \bar{Y}$. This is the value which minimizes the SSR.

**Note:** Some argued: In the model $Y_i = \beta_0 + \beta_1 X_i + u_i$, the OLS estimator of $\beta_0$ is $\hat{\beta}_0 = \bar{Y} - \bar{X}\hat{\beta}_1$. Since there is no $X$ in the present model, then $\hat{\beta}_0 = \bar{Y}$. This argument is NOT correct. Since there is no $X$ in the model, the derivation of $\hat{\beta}_0 = \bar{Y} - \bar{X}\hat{\beta}_1$ does not apply.

3. This looks like the model we studied in class, except the dependent variable is $\log(Y_i)$, rather than $Y_i$. This has no effect on the unbiasedness of the OLS estimates, since we can rewrite the model as

$$
Z_i = \beta_0 + \beta_1 X_i + u_i
$$

$$
E(u_i \mid X_i) = 0.
$$

where $Z_i = \log(Y_i)$. This answer is sufficient for complete credit.
If you write out the math explicitly, note that

\[ \hat{\beta}_1 = \frac{\sum_{i=1}^{n} (X_i - \bar{X}) \log(Y_i)}{\sum_{i=1}^{n} (X_i - \bar{X})^2} = \beta_1 + \frac{\sum_{i=1}^{n} (X_i - \bar{X}) u_i}{\sum_{i=1}^{n} (X_i - \bar{X})^2}, \]

and the expectation of this is \( \beta_1 \), as shown in class and the text.

4. By the OLS formula and the fact that \( \bar{X} = 0 \) and \( \bar{Y} = 0 \),

\[ \hat{\beta}_1 = \frac{\sum_{i=1}^{n} (X_i - \bar{X}) Y_i}{\sum_{i=1}^{n} (X_i - \bar{X})^2} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2} = \frac{50}{100} = \frac{1}{2}, \]

and

\[ \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 0 \]

5. First, we need to find the marginal distribution of \( Y \), conditional on \( X = 10 \). Note that \( P(X = 10) = (20 + 5 + 5)/100 = 30/100 \). Hence

\[ P(Y = 5 \mid X = 10) = \frac{20}{100} = \frac{2}{3}, \]

\[ P(Y = 8 \mid X = 10) = \frac{5}{100} = \frac{1}{6}, \]

\[ P(Y = 14 \mid X = 10) = \frac{5}{100} = \frac{1}{6}. \]

Hence

\[ E(Y \mid X = 10) = 5 \cdot P(Y = 5 \mid X = 10) + 8 \cdot P(Y = 8 \mid X = 10) + 14 \cdot P(Y = 14 \mid X = 10) = 5 \cdot \frac{2}{3} + 8 \cdot \frac{1}{6} + 14 \cdot \frac{1}{6} = 7 \]

6. \( \log(\text{Wage}_i) = 4.822 + 0.257 \log(\text{Sales}_i) \)

   (a) The coefficient on \( \log(\text{Sales}_i) \) is the elasticity of a CEO’s salary with respect to firm sales. The estimate is that this elasticity is 0.257. This means that for an increase in Sales by 1%, the CEO’s salary increases by an expected 0.257%.

   (b) An elasticity is independent of units of measurement, so we should expect that \( \beta_1 \) will not change if \( \text{Wage} \) is measured in dollars rather than thousands of dollars. Indeed, if we define \( W_i^* = 1000W_i \), so that \( W_i^* \) is wage in dollars, then

\[ \log(W_i^*) = \log(1000W_i) = \log(1000) + \log(W_i) \]
so the impact of rescaling \( W \) is the same as adding a constant to the dependent variable. This impacts the estimated intercept \( \hat{\beta}_0 \), but not the estimated slope \( \hat{\beta}_1 \). The OLS intercept increases from 4.822 to 4.822 + log(1000) = 11.7 (remember, by \( \log(x) = \log_e(x) \), not \( \log_{10}(x) \)), and the estimated slope remains at 0.257.

7.

(a) Since \( \beta_1 = 1 \), then we can write the model as \( Z_i = \beta_0 + u_i \) where \( Z_i = Y_i - X_i \). This is an intercept-only model. The OLS estimate of the intercept is \( \hat{\beta}_0 = \bar{Z} = \bar{Y} - \bar{X} \)

(b) First, observe that since \( Y_i = \beta_0 + X_i + u_i \), it follows that \( \bar{Y} = \beta_0 + \bar{X} + \bar{u} \). Hence

\[
\hat{\beta}_0 = \bar{Y} - \bar{X} = \beta_0 + \bar{u}
\]

and then

\[
E (\hat{\beta}_0) = E (\beta_0 + \bar{u}) = \beta_0 + E (\bar{u}) = \beta_0 + \frac{1}{n} \sum_{i=1}^{n} E (u_i) = \beta_0
\]

(c) From (1), we have \( \hat{\beta}_0 = \beta_0 + \bar{u} \). Then

\[
\text{Var} (\hat{\beta}_0) = E (\hat{\beta}_0 - E (\hat{\beta}_0))^2 = E (\hat{\beta}_0 - \beta_0)^2 = E (\bar{u})^2 = \frac{\sigma^2}{n}
\]

where \( \sigma^2 = E (u_i^2) \). The last step should be known from your statistics class. If you want to see the calculation in detail,

\[
E (\bar{u})^2 = E \left( \frac{1}{n} \sum_{i=1}^{n} u_i \right)^2 = \frac{1}{n^2} \sum_{j=1}^{n} \sum_{i=1}^{n} E (u_i u_j) = \frac{1}{n^2} \sum_{i=1}^{n} E (u_i^2) = \frac{\sigma^2}{n}
\]