(1) Note:

\[
y = \beta_0 + \beta_1 x + u \\
= \beta_0 + \beta_1 x + \alpha_0 + (u - \alpha_0) \\
= (\beta_0 + \alpha_0) + \beta_1 x + \epsilon \\
= \delta_0 + \beta_1 x + \epsilon
\]

where \( \delta_0 = \beta_0 + \alpha_0 \) and \( \epsilon = u - \alpha_0 \). Note \( E(\epsilon) = E(u - \alpha_0) = E(u) - \alpha_0 = \alpha_0 - \alpha_0 = 0 \). Thus, the regression model can be written as one with a new intercept and error term, where the latter is mean-zero.

(2) Note: Your answers may differ a little from mine due to rounding error.

(i) Let \( x = \text{ACT} \) and \( y = \text{GPA} \). Note that, from the raw data \( \bar{x} = 25.875, \bar{y} = 3.21 \).

Further, we calculate

\[
\sum_i (x_i - \bar{x})^2 = 56.875, \quad \text{and} \quad \sum_i (x_i - \bar{x})(y_i - \bar{y}) = 5.812.
\]

Thus,

\[
\hat{\beta}_1 = .102.
\]

As for the intercept,

\[
\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 3.21 - .102(25.875) = .571.
\]

Thus, our estimated regression equation is

\[
\hat{\text{GPA}} = .571 + .102\text{ACT}.
\]

The intercept here does not have a particularly useful interpretation, since we don’t observe anyone in the sample with ACT scores near zero. Finally, if someone increases their ACT score by 5 points, the expected effect on GPA is \( .102(5) = .51 \).

In Figure 1, we plot the fitted regression line and scatterplot of this data.

(ii) The following table presents the fitted values and estimated residuals for each observation. Note that the fitted values are obtained as \( \hat{\text{GPA}}_i = .571 + .102\text{ACT}_i \), and the estimated residuals are \( \hat{u}_i = \text{GPA}_i - \hat{\text{GPA}}_i \).
Figure 1: Fitted Regression line and Scatterplot using GPA/ACT data

<table>
<thead>
<tr>
<th>Observation Number</th>
<th>Fitted Value</th>
<th>Estimated Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.71</td>
<td>0.086</td>
</tr>
<tr>
<td>2</td>
<td>3.02</td>
<td>0.379</td>
</tr>
<tr>
<td>3</td>
<td>3.23</td>
<td>-0.225</td>
</tr>
<tr>
<td>4</td>
<td>3.33</td>
<td>0.173</td>
</tr>
<tr>
<td>5</td>
<td>3.53</td>
<td>0.068</td>
</tr>
<tr>
<td>6</td>
<td>3.12</td>
<td>-0.123</td>
</tr>
<tr>
<td>7</td>
<td>3.12</td>
<td>-0.423</td>
</tr>
<tr>
<td>8</td>
<td>3.63</td>
<td>0.066</td>
</tr>
</tbody>
</table>

The sum of the final column of the table above is indeed close to zero. (Note that this comes from the first-order condition for the intercept).
(iii) 

\[ E(GPA|ACT = 20) = 0.571 + 0.102(20) = 2.61. \]

(3), part(i): The intercept tells you the expected birthweight of a child whose mother never smoked while pregnant. Thus,

\[ E(BirthWeight|Cigs = 0) = 119.77. \]

Since there 16 ounces per pound, this suggests that the average birthweight for mothers who do not smoke is about 7.5 pounds. (This seems sensible). Also note that there are likely to be lots of women in the sample who never smoke, and thus we would expect to get a good estimate of the intercept in this application.

The slope on Cigs is negative, suggesting cigarette consumption while pregnant lowers birthweight. Specifically, for every one cigarette consumed per day, birth weight is expected to be lowered by .05 ounces.

\[ E(Birthweight|Cigs = 20) = 119.77 - 0.514(20) \approx 109.5. \]

Thus, we expect that birthweight will fall by about 10.3 ounces, on average, for women who smoke one pack of cigarettes per day.

(ii) It is difficult to argue that this picks up a causal relationship. Women who smoke while pregnant probably possess lots of other characteristics that are different from women who do not smoke. For example, their access to prenatal care, eating habits, exercise habits, alcohol consumption patterns, etc. may be different from women who do not smoke. In addition, these omitted factors probably influence birthweight. So, before arguing a causal link here, one should probably at least attempt to control for these other factors.

(4, part (i). The intercept here is not reliable. It implies that individuals with no income have negative amounts of consumption. While we might interpret this as “borrowing,” this is a stretch, since no data points in the sample have negative values of consumption! The problem here is that no individuals in the data have income values near zero, and so it is difficult to estimate the intercept of this relationship.

(ii) 

\[ Consumption = -124.84 + 0.853(30,000) = 25,465. \]

Thus, we would expect that an individual earning $30,000 per year will consume $25,465 of that income.
First, note
\[ \bar{y} = \frac{1}{n} \left[ \sum_{i: D_i = 1} y_i + \sum_{i: D_i = 0} y_i \right], \]
which we can write as
\[ \frac{n_1 \bar{y}_1 + (n - n_1) \bar{y}_0}{n}, \]
using the given definitions.

First, let’s start with the slope:
\[ \hat{\beta}_1 = \frac{\sum_i (D_i - \bar{D})(y_i - \bar{y})}{\sum_i (D_i - \bar{D})^2} = \frac{\sum_i (D_i - \bar{D})y_i}{\sum_i (D_i - \bar{D})^2}. \]

For the numerator of this expression, note:
\[ \sum_i (D_i - \bar{D})y_i = \sum_{i: D_i = 1} y_i - n_1/n \sum_i y_i \]
\[ = n_1 \bar{y}_1 - n_1 \bar{y} \]
\[ = n_1(\bar{y}_1 - \bar{y}) \]

As for the denominator:
\[ \sum_i (D_i - \bar{D})^2 = \sum_i D_i^2 - 2D_i \bar{D} + \bar{D}^2 \]
\[ = n_1 - 2n \bar{D}^2 + n \bar{D}^2 \]
\[ = n_1 - n \bar{D}^2 \]
\[ = n_1(1 - n_1/n). \]

Thus, putting these together, we obtain:
\[ \hat{\beta}_1 = n(\bar{y}_1 - \bar{y})/(n - n_1). \]

Subbing in our identity for \( \bar{y} \) in the first part of this question gives
\[ \hat{\beta}_1 = \bar{y}_1 - \bar{y}_0. \]

As for the intercept, note
\[ \hat{\beta}_0 = \bar{y} - [\bar{y}_1 - \bar{y}_0](n_1/n) \]
\[ = (n_1/n) \bar{y}_1 + [(n - n_1)/n] \bar{y}_0 - [\bar{y}_1 - \bar{y}_0](n_1/n) \]
\[ = \bar{y}_0, \]
as claimed.