4.2

(a) We have $t_x = 69$. $t_y = 83$, $S_x = 4.092676$, $S_y = 5.333333$, $R = 0.8112815$, and $B = 1.202899$.

(d) The mean of the sampling distribution of $\hat{t}_{yr}$ is 83.733; the variance is 208.083 and the bias is 83.733 $-83 = 0.733$. By contrast, the mean of the sampling distribution of $N\bar{y}$ is 83 and its variance is 518.169.

(e) From (4.6),

$$\text{Bias}\,(\hat{\bar{y}}_r) = 0.07073094.$$

4.8 (a) (b) $\hat{B} = \bar{y}/\bar{x} = 297897/647.7467 = 459.8975$. Thus

$$\hat{t}_{yr} = t_x\hat{B} = (2087759)(459.8975) = 960{,}155{,}061.$$

The estimated variance of the residuals about the line $y = \hat{B}x$ is

$$s_e^2 = 149{,}902{,}393{,}481.$$

Using (4.11), then, with farms87 as the auxiliary variable,

$$\text{SE}\,[\hat{t}_{yr}] = 3078\sqrt{1 - \frac{300}{3078}}\sqrt{\frac{s_e^2}{300}} = 65{,}364{,}822.$$

(c) The least squares regression equation is

$$\hat{y} = 267029.8 + 47.65325x$$

Then

$$\hat{\bar{y}}_{\text{reg}} = 267029.8 + 47.65325(647.7467) = 297897.04$$

and

$$\hat{t}_{y\text{reg}} = 3078\hat{\bar{y}}_{\text{reg}} = 916{,}927{,}075.$$

The estimated variance of the residuals from the regression is $s_e^2 = 118{,}293{,}647{,}832$, which implies from (4.19) that

$$\text{SE}\,[\hat{t}_{y\text{reg}}] - 3078\sqrt{1 - \frac{300}{3078}}\sqrt{\frac{s_e^2}{300}} - 58{,}065{,}813.$$

(d) Clearly, for this response, it is better to use acres87 as an auxiliary variable. The correlation of farms87 with acres92 is only 0.06; using farms87 as an auxiliary variable does not improve on the SRS estimate $N\bar{y}$. The correlation of acres92 and acres87, however, exceeds 0.99. Here are the various estimates for the population total of acres92:

Estimate	$\hat{t}$	SE $[\hat{t}]$
SRS, $N\bar{y}$	916,927,110	58,169,381
Ratio, x = acres87	951,513,191	5,344,568
Ratio, x = farms87	960,155,061	65,364,822
Regression, x = farms87	916,927,075	58,065,813

Moral: Ratio estimation can lead to greatly increased precision, but should not be used blindly. In this case, ratio estimation with auxiliary variable farms87 had larger standard error than if no auxiliary information were used at all. The regression estimate of t is similar to $N\bar{y}$, because the regression slope is small relative to the magnitude of the data. The regression slope is not significantly different from 0; as can be seen from the picture in (a), the straight-line regression model does not describe the counties with few but large farms.