Sample Final Solutions.

1 (a) Yes. There is an approximate linear relationship. The relationship appears to be positive.

(b) Some evidence of nonconstant variance (variance smaller for larger fitted values). This is not a gross violation of the assumption; I am reasonably comfortable with the linear regression model.

(c) $\widehat{\text{Coronary}} = 29.5 + 0.0557 \text{ Cigarette}$

(d) In the fitted model, a one unit increase in the Cigarette variable is associated with a 0.0557 unit increase in the expected value of the Coronary variable.

(e) The standard deviation of $\hat{\beta}_0$ is approximately equal to $se(\hat{\beta}_0) = 29.5$

$\boxed{2}$ a Yes, it is plausible that $\beta_0 = 0$. If the true intercept were 0, there would be a 33% chance of getting data like observed (this is the interpretation of the p-value for $\hat{\beta}_0$).

Practically, the model says that β_0 is the expected coronary deaths per 100K persons aged 35-64 for countries with no cigarette consumption ("Cigarette" = 0). If β_0 were 0, then contries with no cigarette consumption would have no deaths from coronary heart disease. This is very unlikely.

There is no contradiction; 0 is outside the range of the "Cigarette" variable, so we should not try to interpret β_0 directly.

(b) If the true β_1 were 0, the chance of getting such a large $\hat{\beta}_1$ due to natural variability alone would be extremely small (less than .001 probability). This is indicated by the p-value for $\hat{\beta}_1$.
I do not think that natural variability alone could account for the observed $\hat{\beta}_1$.

(c) Yes, evidence exists. ($p < .01$)
Based on p-value for $\hat{\beta}_1$.

(d) $R^2 = 49.57\%$. weak to moderate relationship

(e) The p-value indicates extremely strong evidence of a relationship ($\beta_1 \neq 0$). The R^2 indicates a weak or moderate relationship (moderate σ^2).
This is not contradictory

3

Population	Sample
weights of <u>all</u> chips in the batch	Weights of $n=10$ chips
mean μ	mean $\bar{x} = 0.8$
	sd $s = 0.03$

(a) 95% CI for μ:

$$\bar{x} \pm t_{.025, n-1} \frac{s}{\sqrt{n}}$$

$$= 0.8 \pm 2.262 \frac{0.03}{\sqrt{10}}$$

$$= 0.8 \pm .021$$

$$= (.779, .821)$$

(b) No. μ is not random, so it doesn't make sense to talk about "chance" or "probability". We would say we are 95% <u>confident</u> that μ is in the interval

4 .83 is not in the 95% CI $\Longrightarrow$ reject H_0 at level 5%

(alternatively, $t = \frac{\bar{x} - .83}{s/\sqrt{n}} = -3.16$, $|t| > t_{.025, n-1}$, reject H_0).

5

Population

increases of all cows that take the drug

mean μ

Sample

increases of the $n=100$ sampled cows

mean $\bar{x} = 11$

sd $s = 50$

ⓐ $H_0: \mu = 0$ (no effect)

$H_a: \mu \neq 0$ (effect)

Note: Ok to do 1-sided H_a instead, but we didn't cover this

ⓑ μ = mean increase in milk production for all cows taking the drug

ⓒ H_0: drug has no effect

H_a: drug has an effect

ⓓ First compute the p-value:

$t = \frac{\bar{x}-0}{s/\sqrt{n}} = \frac{11-0}{50/\sqrt{100}} = 2.2$; $p \approx P(|Z| > 2.2) = .02781$

Reject for $\alpha > .02781$

ⓔ $1000 \cdot p = 27.81$

Questions 6–9 concern the following situation. A random sample of 50 adults were asked how much they spend on lottery tickets, and were interviewed about various socioeconomic variables. The variables are

PercLott = Percentage of total household income spent on the lottery. (This is Y).
YrsEdu = Number of years of education,
Age = The persons Age,
Kids = Number of Children,
Income = Personal income (Thousands of Dollars).

Here is the Minitab regression output:

```
Analysis of Variance

Source      DF   Adj SS   Adj MS  F-Value  P-Value
Regression   4   404.42   101.10    17.72    0.000     (7)
  YrsEdu     1    60.68    60.68    10.63    0.002
  Age        1     0.21     0.21     0.04    0.850
  Kids       1     0.55     0.55     0.10    0.761
  Income     1    23.30    23.30     4.08    0.050
Error       45   256.80     5.71
Total       49   661.22

Model Summary

    S    R-Sq  R-Sq(adj)  R-sq(pred)
2.389  61.16%     57.71%      52.16%

Coefficients

Term           Coef   SE Coef  T-Value  P-Value   VIF
Constant     15.070     2.444     6.17    0.000
YrsEdu      -0.5911    0.1813    -3.26    0.002  1.47     (6)
Age         0.00647   0.03395     0.19    0.850  2.81
Kids         0.0816    0.2665     0.31    0.761  1.93
Income     -0.06663   0.03305    -2.02    0.050  1.58

Regression Equation

PercLott = 15.1 - 0.591 YrsEdu + 0.0065 Age + 0.082 Kids - 0.0666 Income
```

(8)

Problem 6

Based on the output, is there statistical evidence to suggest that relatively educated people spend a different amount on lotteries than relatively uneducated people?

(a) Yes
(b) No

p-value for YrsEdu = .002 < .05;
there is evidence that after adjusting fore Age, Kids, and income, YrsEdu is related to PercLott.

Problem 7

reject $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$
since $p \approx .000$
$(p < .05)$

The results of the F test imply that, beyond a reasonable doubt:

(a) All of the true slope coefficients in the model are nonzero

(b) At least one of the true slope coefficients in the model is nonzero

(c) None of the true slope coefficients in the model is nonzero

(d) All of the estimated slope coefficients are nonzero

(e) At least one of the estimated slope coefficients is nonzero

. . . .

Problem 8

The 95% confidence interval for the true coefficient of YrsEdu is

(a) (-2.12, 3.14)

(b) (-0.5911, 0.5911)

(c) (-1,1)

(d) (-0.956, -0.226)

(e) (-1.06, -0.124).

Approximate 95% CI for β_1:
$\hat{\beta}_1 \pm 2\, se(\hat{\beta}_1) = -.5911 \pm 2(.1813)$
$= (-.954, -.229)$

(Note: exact 95% CI uses $t_{.025, n-k-1} = 2.021$ instead of 2)

. . . .

Problem 9

Performing a two-tailed hypotesis test for the null hypothesis that the true coefficient of YrsEdu is -1, at the 5% level of significance, we:

(a) Reject the null hypothesis

(b) Do not reject the null hypothesis

-1 is not in 95% CI from #8.

. . . .

Alternatively: $t = \dfrac{-.5911 - (-1)}{.1813} = 2.255$,

$|t| > t_{.025, n-k-1}$.

Problem 10

Let's return to the simple regression described in Problem 1. The residual for Greece is:

(a) 1800

(b) 29.45

(c) 31.74

(d) 1768.26

(e) -88.474

For Greece, $x = 1800$ $y = 41.2$

$\hat{y} = 29.5 + (0.0557)(1800)$
$= 129.76$

residual $= y - \hat{y} = 41.2 - 129.76$
$= -88.56$

Note: there is rounding error in the minitab output

Problem 11

A sample of size 100 is going to be taken from a population with mean 3 and variance 25. The probability that the sample mean will exceed 4 is approximately:

(a) .0456

(b) .4207

(c) .0793

(d) .3446

(e) .0228

$P(\bar{X} > 4)$

$= P\left(\frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} > \frac{4 - \mu_{\bar{x}}}{\sigma_{\bar{x}}}\right)$

$= P\left(Z > \frac{4-3}{(\frac{1}{2})}\right)$

$= P(Z > 2) = .02275$

~~= .0228~~

$\mu = 3$ $\sigma^2 = 25$

$n = 100$

$\mu_{\bar{x}} = \mu = 3$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{25}}{\sqrt{100}} = \frac{1}{2}$

Problem 12

Suppose that X and Y are independent random variables with $P(X > 4) = 0.8$ and $P(Y > 5) = 0.6$. The probability that X exceeds 4 and Y exceeds 5 is

let $A = \{X > 4\}$, $B = \{Y > 5\}$.

(a) 1.4

(b) 0.92

(c) 0

(d) 0.48

(e) Not enough information to determine

by independence,

$P(A \cap B) = P(A)\,P(B)$
$= (0.8)(0.6)$
$= 0.48$

....