Example on Biotechnology: Using the water fern Azolla to produce Hydrogen Fuel [Example Combining All Concepts]

(c) Calculate the linear correlation coefficient for the relationship between atmospheric N2 and hydrogen production. What is the exact P-value?

$$SS_{Total} = SS_{REGR} + SS_{Error} = 493414.347 + 19405.153 = 512819.5$$
$$R^2 = \frac{SS_{REGR}}{SS_{Total}} = \frac{493414.347}{512819.5} = 0.96216$$
$$r = -\sqrt{R^2} = -\sqrt{0.96216} = -0.981$$

P-value (r) = P-value (t-test for slope) = $1.72 \times 10{-}5$

Note: if asked for whethere there's negative correlation between N_2 and hydrogen (left-tailed), the P-value would be $1.72 \times 10{-}5/2 = 0.0000086$

d.
$$\hat{\sigma} = \sqrt{MS_{Error}} = \sqrt{\frac{SS_{Error}}{n-2}} = \sqrt{\frac{19405.153}{8-2}} = \sqrt{3234.1922} = 56.8996$$

e.
$R^2 = 0.96216$
Therefore, 96.22% of the variability in hydorgen production is explained by the level of atmotspheric nitrogen / regression model.

f. At the 1% significance level, test the hypothesis that there is a negative relationship between atmospheric nitrogen and hydrogen production. Carry out the most appropriate, showing all steps (give both the exact P-value and the value from the table).

Note: One-tailed hypothesis, so use t-test, not F-test.

$H_0 : \beta_1 = 0$ (There is no relationship)
$H_a : \beta_1 < 0$(There is a negative relationship)

$$t = \frac{\beta_1}{\hat{\sigma}/\sqrt{S_{xx}}} = \frac{\beta_1}{SE(\hat{\beta}_1)} = \frac{-9.946}{0.805} = -12.355$$

$df = n - 2 = 8 - 2 = 6$
P-value: $P < 0.0005$
Exact P-value: $P = (1.72 \times 10^{-5})/2 = 8.6 \times 10^{-6}$. There is extremely strong evidence against H_0.
$P < \alpha$ (0.01), therefore reject H_0.

At the 1% significance level, the data provide sufficient evidence to conclude that there is a negative relationship between atmospheric nitrogen and hydrogen production.

g. What is the F-statistic and the P-value (both the exact value and the value from the table)?

$F = t^2 = (-12.355)^2 = 152.562$
$df = (1,\ n-2) = (1,\ 6)$
$P < 0.001$
Exact P-value = P (two-tailed t-test) = 1.72×10^{-5}

h. Find the margin of error for a 99% confidence interval for the expected value of hydrogen production at 20% atmospheric nitrogen.

At $df = n - 2 - 6,\ t_{\alpha/2} = t_{0.001/2} = t_{0.005} = 3.707$
$$ME = t_{\alpha/2} \times \hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}} = 3.707 \times 56.86996\sqrt{\frac{1}{8} + \frac{(20-36.25)^2}{4987.5}} = 88.93$$

The margin error for a 99% confidence interval for the expected value of hydrogen production at 20% atmospheric nitrogen is 88.93

4.9 More on Assumptions and Transformations of Data

> Checking Assumptions with Scatterplots, Residual Plots and Normal Probability Plots
> Example 1: No assumptions violated
> Example 2: Equal standard deviations violated
> Example 3: Linearity and standard deviations and outliers violated
> Example 4: Normality plot shows normality assumption not violated, but cannot infer any other violations

Example on Log Transformed Data:
a.
$k = final - initial = 50 - 40 = 10$
$e^{k\beta_1} = e^{10 \times 0.0912} = 2.489$

Interpretation: The median rate of hip fractures at 50 years will be 2.489 times the median rate at 40 years.
OR:
An additive change of 10 years in age is associated with a multiplicative change of 2.489 in the median of the annual rate of hip fractures.

d.
$(e^{5.029}, e^{5.384}) = (152.780, 217.892)$
It is estimated with 95% confidence that the rate of hip fractures for a person that is 80 years old is between 152.780 and 217.892 fractures per 100000 people.