Discussion 13

Example 1. Dairy scientists want to examine the effect of two dietary additives (A and B) on milk production. They have four treatment groups: 1 = control, 2 = additive A only; 3 = additive B only, 4 = additive A and B together. The following information is available:

Treatment	Sample Size	Mean yield	Standard deviation
1	5	20	10
2	5	32	10
3	5	36	11
4	5	50	11

Conduct an overall test of the treatment differences. That is, fill in the ANOVA table, state the hypotheses and conduct the test.

Example 2. A designed study was undertaken to relate the rate of oxygen consumption in birds to different temperatures. Ten adult birds (same species) were randomly assigned to some given temperatures (in degrees celsius) and the consumption was determined in ml/g/hr.

temp	X	-6	-3	0	2	4	4	6	6	8	8
rate	Y	4.7	4.0	4.1	3.6	3.7	3.4	3.4	2.8	2.9	3.1

Some partial computations that may be useful are : $\sum(x_i - \bar{x})^2 = 196.90$; $\sum(y_i - \bar{y})^2 = 3.08$; $\sum(x_i - \bar{x})(y_i - \bar{y}) = -22.93$.

(a) Compute the slope of a simple linear regression line for Y on X and interpret it.

(b) Perform a test of the null hypothesis that Y does not depend on X versus the two-sided alternative. Interpret the results.

(c) Consider using the fitted line to estimate the mean rate. Find a 95% confidence interval for the true mean rate of oxygen consumption at a temperature of +5. Interpret the interval. Before the study was conducted an avian physiologist hypothesized that the true rate of oxygen consumption at +5 was 3.0. What can you say about this hypothesis?

Example 3. A landscape ecology study investigated the characteristics of a number o f plots of prairie land. For each plot they measured the "length of edge" and the number of species. The data are presented with some summary statistics.

$$\sum x_i = 162 \quad \sum y_i = 126 \quad \sum x_i y_i = 2648;$$
$$\sum x_i^2 = 3448 \quad \sum y_i^2 = 2200 \quad n = 9.$$

(a) Consider the regression line relating y to x. Calculate the least square estimates of the slope and intercept for the regression line relating number of species to length of edge (i.e. relating y to x).
(b) Construct the ANOVA Table for this regression problem. (Indicate Source, df, SS, MS). You do not need to perform any tests.
(c) Perform a test of the hypothesis $H_0 : b_0 = 10$ versus the two-sided alternative.
(d) The investigators plan to continue to examine additional plots of prairie land. If, tomorrow, they should find a plot whose length of edge is 13, give a prediction of the number of species in that plot, and give a 90% confidence interval for your prediction.

Solutions

Solution 1. $SSError = 4 \times (10^2 \times 2 + 11^2 \times 2) = 1768$ and $SSTreatment = 5 \times [(20 - 34.5)^2 + (32 - 34.5)^2 + (36 - 34.5)^2 + (50 - 34.5)^2] = 2295$. Hence, $SSTotal = 4063$ and the ANOVA table is

source	df	SS	MS
Treatment	3	2295	765
Error	16	1768	110.5
Total	19	4063	

For testing $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$, use $F = MS_{Treatment}/MS_{Error}$ and the observed $F = 6.923$. Compare with $F_{3,16}$, the p-value is less than 0.01. Reject H_0 at 5% and there is strong evidence of treatment differences.

Solution 2. (a) $\hat{b}_1 = -22.93/196.90 = -.116$. This means that for each increase in temperature of one degree C, the rate of oxygen consumption decreases by .116 ml/g/hr.
(b) $SSReg = \hat{b}_1 \times (-22.93) = 2.67$. Thus, the ANOVA table for regression can be written:

source	df	SS	MS
Regression	1	2.67	2.67
Error	8	.41	.051
Total	9	3.08	

The stated test is $H_0 : b_1 = 0$ vs $H_A : b_1 \neq 0$. $F = 2.67/.051 = 52.1$ on 1,8 df. Thus, $p < .001$ and the rate depends strongly on the temperature.
(c) Here we focus on $\hat{Y}$ as an estimator. $se(\hat{Y}) = s\sqrt{(1/n) + (5 - \bar{x})^2/(\sum(x_i - \bar{x})^2)} = 0.079$. Now, $t_{8,.025} = 2.306$ and $\hat{Y} = 3.326$. Thus, a 95% CI for $\hat{Y}$ is $(3.144, 3.508)$. Thus, at 95% confidence, a plausible range for the mean rate of oxygen consumption is from 3.144 to 3.508. Since the value of "3" does not fall within this range, it is not a plausible value with 95% confidence; thus the hypothesis is rejected at $\alpha = 0.05$.

Solution 3. (a) Plugging in the formula for the least square estimates, we get $\hat{b}_1 = 0.714$, and $\hat{b}_0 = \bar{y} - \hat{b}_1\bar{x} = 1.148$.
(b)

source	df	SS	MS
Regression	1	271.32	271.32
Error	7	164.68	23.53
Total	8	436	

(c)

$$T = \frac{1.148 - 10}{\sqrt{23.53}\sqrt{\frac{1}{9} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}}} = -2.15.$$

The p-value is between 0.05 and 0.10.
(d) $\hat{y}_{pred} = 1.148 + 0714 \times 13 = 10.43$. The 90% confidence interval for the prediction is

$$10.43 \pm 1.895\sqrt{23.53}\sqrt{1 + \frac{1}{9} + \frac{(13 - 18)^2}{\sum(x_i - \bar{x})^2}} = 10.43 \pm 9.89.$$