Slide 1

Regression with Dummy Variables

The Data - Productivity Improvement (revisited)

An economist compiled data on productivity improvements last year for a sample of firms producing electronic computing equipment. The firms were classified according to the level of their average expenditures for research and development in the past three years (low, moderate, high). The results of the study follow (productivity improvement is measured on a scale from 0 to 100).

- Y – Productivity improvement (measured on a scale from 0 to 100), which is continuous.
- X – level of their average expenditures for research and development in the past three years (low, moderate, high), which is categorical and has three levels.

Slide 2

	R&D Expenditure Level		
	low	**moderate**	**high**
	7.6	6.7	8.5
	8.2	8.1	9.7
	6.8	9.4	10.1
	5.8	8.6	7.8
	6.9	7.8	9.6
	6.6	7.7	9.5
	6.3	8.9	
	7.7	7.9	
	6.0	8.5	
		8.7	
		7.1	
		8.4	
size ($\mathbf{n_k}$)	9	12	6
mean ($\bar{\mathbf{y}}_{\mathbf{k}\cdot}$)	6.878	8.133	9.200
variance ($\mathbf{s_k^2}$)	0.662	0.573	0.752

Note: the grand mean is $\bar{y}_{\cdot\cdot} = 7.952$ and variance $s_y^2 = 1.365$.

Slide 3

Regression Approach with Dummy Variables

To approach by multiple regression, define dummy variables (Z_1 and Z_2) such that

$$Z_{i1} = \begin{cases} 1 & \text{if the } i\text{-th firm is in the Moderate category} \\ 0 & \text{otherwise} \end{cases}$$

$$Z_{i2} = \begin{cases} 1 & \text{if the } i\text{-th firm is in the High category} \\ 0 & \text{otherwise} \end{cases}$$

Then, consider the following multiple linear regression model

$$y_i = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \varepsilon_i,$$

where $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$.

Slide 4

Fit Performance

Parameter Estimates

	Estimate	S.E.	t value	$2\,\Pr(>|t|)$
Intercept	(1)	0.267	25.790	0.000
factor(moderate)	(2)	0.353	3.559	0.002
factor(high)	(3)	0.422	5.507	0.000

Analysis of Variance Table

	df	SS	MS	F	$\Pr(>F)$
Model	(1)	(4)	(7)	(9)	0.000043307
Residuals	(2)	(5)	(8)		
Total	(3)	(6)			

Related Questions

1. Complete the ANOVA Table.
2. Find an unbiased estimate for σ^2.
3. Compute and interpret the coefficient of determination.
4. Give the LS estimates of $(\beta_0, \beta_1, \beta_2)$.
5. Test H_0: $\mu_{\text{low}} = \mu_{\text{moderate}} = \mu_{\text{high}}$ at $\alpha = .05$.
6. Test H_0: $\mu_{\text{low}} \geq 8$ at $\alpha = .01$
7. Test H_0: $\mu_{\text{low}} \geq \mu_{\text{high}}$ at $\alpha = .05$.

Slide 5

Answers

Parameter Estimates

	Estimate	S.E.	t value	$2\,\Pr(> \lvert t \rvert)$
Intercept	6.878	0.267	25.790	0.000
factor(moderate)	1.256	0.353	3.559	0.002
factor(high)	2.322	0.422	5.507	0.000

Analysis of Variance Table

Source	df	SS	MS	F	$\Pr(> F)$
Model	2	20.125	10.063	15.721	0.000043307
Error	24	15.362	0.640		
Total	26	35.487			

Slide 6

Answers (continued)

For question (6), since $\beta_0 = \mu_{\text{low}}$, it is equivalent to test H_0: $\beta_0 \leq 8$ at $\alpha = .01$. Therefore, a t test is applicable and the t test statistic is

$$t = \frac{\hat{\beta}_0 - 8}{s.e.(\hat{\beta}_0)} = \frac{6.878 - 8}{.267} = -4.2022.$$

Decision Rule: Reject H_0 if $t < t_{.01}^{(24)} = -2.492$. Hence the null is rejected.

Slide 7