

Advanced Analysis of Scientific Data (STAT1301)

Assignment 4

The due date/time is given on Blackboard.

1. In this question, you are asked to simulate random data from a linear regression model, and then use the `lm` function to verify how close the parameter estimates are to the true values. Take the true parameters to be $\beta_0 = -1.5$, $\beta_1 = 0.3$, and $\sigma^2 = 0.25$. Let the explanatory variables $\{x_i\}$ lie equally spaced on the interval $[0,1]$; that is, $x_1 = 0$, $x_{i+1} = x_i + 1/100$, $i = 1, \dots, 100$. Next, simulate the error terms ε_i , $i = 1, \dots, 101$ by drawing them independently from a $\mathcal{N}(0, \sigma^2)$ distribution, and set the (outcomes of the) response variables as

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, 101.$$

Now, perform a regression analysis on the $\{(x_i, y_i)\}$.

2. Edwin Hubble discovered that the universe is expanding. If v is a galaxy's recession velocity (relative to any other galaxy) and d is its distance (from that same galaxy), Hubble's law states that

$$v = Hd, \tag{1}$$

where H is known as Hubble's constant. The following are distance (in millions of light-years) and velocity (thousands of miles per second) measurements made on 5 galactic clusters.

distance	68	137	315	405	700
velocity	2.4	4.7	12.0	14.4	26.0

Showing working, derive the least-squares estimate for H .

3. In order to investigate the effectiveness of "walking exercises" for babies, 24 babies (of the same age and sex) were randomly divided into 4 groups. Each group followed a different training program. The table shows the age (in months) when the infants first walked alone.

Group			
A	B	C	D
9	11	11.5	13.25
9.5	10	12	11.5
9.75	10	9	12
10	11.75	11.5	13.5
13	10.5	13.25	11.5
9.5	15	13	11.5

A one-way analysis gives:

One-way ANOVA: A, B, C, D

Analysis of Variance

Source	DF	SS	MS	F	P
Factor	a	14.20	c	e	f
Error	20	b	d		
Total	23	58.49			

A and B underwent the same training program, but the progress of the children in group A was checked every week, whereas the children in group B were checked only once, at the end of the study.

- (a) Evaluate the missing numbers a, b, c, d, e and f in the table.
 - (b) Test whether there is a psychological effect of periodic testing by considering group A and B only.
4. An experiment was performed to determine the effects of four different chemicals on the strength of fabric. Five fabric samples were selected and each chemical type was tested once on each fabric sample. The data are shown below.

Chemical type	Fabric Sample				
	1	2	3	4	5
1	1.3	1.6	0.5	1.2	1.1
2	2.2	2.4	0.4	2.0	1.8
3	1.8	1.7	0.6	1.5	1.3
4	3.9	4.4	2.0	4.1	3.4

- (a) Formulate an appropriate statistical model for the data.
- (b) Perform a 2-way ANOVA analysis on the data.
- (c) Is there an effect of the chemical on the fabric strength? State the appropriate hypotheses, give the test statistic, its degrees of freedom and p-value and state your conclusion.
- (d) Examine the residuals and check for violations of basic assumptions that could invalidate the results.

5. Six different types of paint are tried on five different surfaces. For each combination of paint and surface we measure the quality of coverage. The (partial) results of a statistical analysis are depicted in the table below. There is no replication, thus we have 30 measurements in total. A two-way Analysis of Variance gave the following output:

Two-way ANOVA: coverage versus painttype, surface

Analysis of Variance for coverage

Source	DF	SS	MS	F	P
painttype		519.07			
surface		166.87			
Error		383.93			
Total					

- (a) Complete the ANOVA table
- (b) Test whether there is any difference between the paint types with regard to the coverage quality.
- (c) Is “surface type” an relevant factor in explaining the coverage quality?