Analysis of Continuous Data

Master of Statistical Data Analysis

January 12, 2015

Estimating cholesterol levels through different methods

In Lopes-Virella et al. three different methods are compared for measuring HDL cholesterol levels. In the figure below results from the method based on "ultracentrifuge" are set out against the estimates obtained via a method based on "heparine residu".

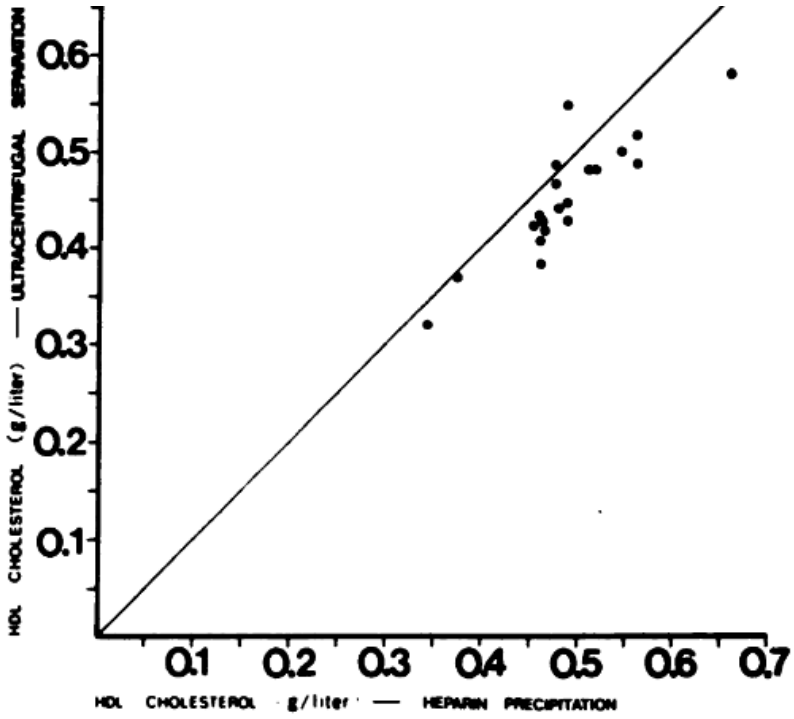

Fig. 3. Cholesterol content in HDL separated by ultracentrifugation (y) and heparin–Mn^{2+} (x) methods
The line was drawn with a slope of 1.00 (slope obtained by linear regression analysis, 0.81; y-intercept, 0.0577; $r = 0.88$)

Table 1. Mean Values for Cholesterol Content of HDL Separated by Different Methods

HDL separation methods	No. subjects	HDL cholesterol (±SD), g/liter
Heparin–Mn^{2+}	20	0.491 ± 0.068
Ultracentrifugation	20	0.455 ± 0.062

1. Write down the theoretical simple linear regression model for the ultracentrifugation estimation in function of the heparine-residu estimation. Clearly define all parameters and symbols.

2. Estimate the MSE of the model using the numbers given and then check this visually on the figure. Give both derivations.

3. Is one of the model assumptions possibly not fulfilled? If so, describe it (no more than one) and explain why based on the figure. Also point to an assumption that cannot be evaluated based on the figure and explain why.

4. Perform the test of prime importance here at the 5% significance level and draw a well motivated conclusion.

5. Give the advantages and disadvantages of this specific test for the underlying research question. Do you see any alternative analysis that the authors could have chosen? Explain and motivate which of the 2 methods seems best and why.

Stock price reaction to management change in distressed companies

Bonnier and Bruner (1989) study the effect of management change on the excess returns to stakeholders (per firm), from a sample of 85 financially under performing firms listed in the New York or American stock exchanges from 1969 to 1983. The outcome variable "excess returns" what has been returned over and above the expected profit to the stakeholders, expressed as a percentage of the expected profit, and this averaged over two days. Specifically, this is averaged over day 0, the day when the announcement of management change was made, and day -1, the preceding day. A negative number means less return than expected. The primary predictor of interest here is the position for which a change is planned: CEO or not, registered in the dummy variable 'title' which indicates 1 for the CEO and 0 otherwise. Two further covariates were used in the analysis presented below: 'relative size of firm' and 'origin of the incoming executive', both are explained in the footnote below table 2.

To investigate the effect of interest the authors selected a hierarchical approach to model fitting: they first entered all main effects, then all two-way interactions and lastly the three-way interaction into the model. Results of the fitted regression models are given in Table 2 of the paper and presented below.

Table 2

Hierarchical regression analysis of firm size, origin, and title effects on announcement-day (−1 and 0) returns.[a]

Variables entered on step 1	Beta	t-test of significance
Size	0.008	$t = 0.07$
Origin	0.056	$t = 0.51$
Title	0.226	$t = 2.02$[b]
Step 1 $R^2 = 0.06$		
Equation $F = 1.74$, df = 3, 81		
Variables entered on step 2	**Beta**	**t-test of significance**
Size × Origin	0.005	$t = 0.03$
Origin × Title	0.546	$t = 2.85$[c]
Size × Title	1.093	$t = 3.35$[c]
Step 2 $R^2 = 0.25$, R^2 increase step 2 = 0.19 ($p \leq 0.001$)		
Equation $F = 4.36$,[c] df = 6, 78		
Change in $F = 6.62$,[c] df = 3, 78		
Variable entered on step 3	**Beta**	**t-test of significance**
Size × Origin × Title	0.262	$t = 0.522$
Step 3 $R^2 = 0.25$ ($p \leq 0.05$) R^2 increase step 3 = 0.003 ($p = ns$)		
Equation $F = 3.74$,[c] df = 7, 77		
Change in $F = 0.274$, df = 1, 77		

[a] Origin = origin of the incoming executive, measured by a dummy variable taking the value 1 if the new appointee came from outside the firm, and 0 otherwise. Title = title of the management position in which change occurred, measured by a dummy variable taking the value 1 if the title included 'Chief Executive Officer', and 0 otherwise. Size = relative size of firm i measured by the ratio of firm i's sales to the sales of the largest firm in the sample.
[b] Significant at the 0.05 level.
[c] Significant at the 0.01 level.

1. For the following questions consider the first part of Table 2.

   (a) Write down the theoretical linear multiple regression model with all main effects for this relation of interest. Use clear notation and explain all symbols.

   (b) Authors state that the standard deviation for the 'excess return' outcome is 0.00565. Derive the SSE and SSR for the model with all main effects using this and the information given in the first part of Table 2.

   (c) Fill in estimated parameters for the regression model in (a) where possible. Give a clear interpretation of each estimate and of the R^2 value.

   (d) Assuming the assumptions hold, calculate a 95% confidence interval for the effect of 'title'. Express in words what this means for the effect of interest in this model.

   (e) Assuming that the linear model assumptions are fulfilled for this model, explain what is being tested with the F statistic in the first part of Table 2. Write down

the null hypothesis and alternative hypothesis for this test, explain how the statistic is computed and check the calculation of the degrees of freedom and of the statistic.

(f) In the paper this paragraph is included: "Reinganum (1985) suggests that small firms may have less complex control structures than large firms and that therefore the effect of a change in one management position will be more meaningful in a small firm than a large firm." To what phenomenon is Reinganum referring here?

(g) How would you judge whether the variables 'size' and 'origin' confound the effect of 'title'? Which additional model would allow you to evaluate this quite directly? Given the available information here, do you think such confounding may be present here? Why or why not? Explain.

(h) If the residuals of the model presented so far were skew to the right. What impact would this have on a prediction interval you would derive from this model for an under performing firm that plans to have a change in senior management? Explain.

2. For the following questions consider the second part of Table 2.

   (a) Write down the theoretical multiple regression model with all main effects and all two-way interactions. Use clear notation and explain all symbols.

   (b) Give a clear interpretation of the interaction effect between size and origin. Explain in words what the value 0.005 represents.

   (c) Interpret the two R^2 values provided for this model and explain how they differ conceptually from each other. Give formulae in your explanations.

   (d) Assuming that the linear model assumptions are fulfilled for this model, explain what is being tested with the F statistic of change in the second part of Table 2. Write down the null hypothesis and alternative hypothesis for this test, explain how the test statistic is computed and check the calculation of the degrees of freedom and of the statistic.

3. Why would the authors use a sample of financially underperforming firms in order to estimate the relation between change of management and excess of returns? Which phenomenon the authors think may be playing here? What are its potential consequences?

4. Suppose that the positive estimate for 'title' (0.226) and its t-statistic (2.02) in the first part of Table 2 were obtained from a simple linear regression with excess returns. Calculate the power to detect the alternative $\beta = 0.3$ when testing $H_0 : \beta = 0$ against $H_a : \beta \neq 0$ at the 5% significance level. To this end, estimate the variance of $\hat{\beta}$ and assume that its population variance is known and equal to this value.

5. How could you adjust the power calculation for a Model that has adjusted for 'size' and 'origin' if we still want to detect an effect of 0.3 for 'title'?

Good luck!