**HDA Workshop**

Lab #2 – Regression & Forecasting

**Exercise 1 (Forecasting industrial production) – 5 mins**

Here is the index of industrial production in the UK from 1990 to 2017.

|  |  |
| --- | --- |
| 1990 | 100.4 |
| 1991 | 97.0 |
| 1992 | 97.3 |
| 1993 | 99.5 |
| 1994 | 104.8 |
| 1995 | 106.7 |
| 1996 | 108.1 |
| 1997 | 109.6 |
| 1998 | 110.9 |
| 1999 | 112.2 |
| 2000 | 114.2 |
| 2001 | 112.4 |
| 2002 | 110.8 |
| 2003 | 110.1 |
| 2004 | 110.9 |
| 2005 | 110.0 |
| 2006 | 110.3 |
| 2007 | 110.7 |
| 2008 | 107.5 |
| 2009 | 97.2 |
| 2010 | 100.0 |
| 2011 | 98.8 |
| 2012 | 96.3 |
| 2013 | 97.3 |
| 2014 | 98.8 |
| 2015 | 100.0 |
| 2016 | 101.3 |
| 2017 | 103.5 |

Forecast the price of the index for the period 2018-2022 using simple time series forecasting/regression.

**Exercise 2 (It’s show time!) – 20 mins**

The owner of Showtime Movie Theatres, Inc. would like to estimate weekly gross revenue as a function of advertising expenditures. Historical data for a sample of 8 weeks is provided below.

|  |  |  |
| --- | --- | --- |
| Weekly Gross Revenue (£1000s) | Television Advertising (£1000s) | Newspaper Advertising (£1000s) |
| 96 | 5.0 | 1.5 |
| 90 | 2.0 | 2.0 |
| 95 | 4.0 | 1.5 |
| 92 | 2.5 | 2.5 |
| 95 | 3.0 | 3.3 |
| 94 | 3.5 | 2.3 |
| 94 | 2.5 | 4.2 |
| 94 | 3.0 | 2.5 |

1. Construct a scatter chart to determine whether any linear regression appear to exist between (1) the weekly gross revenue and television advertising expenditures; and (2) the weekly gross revenue and newspaper expenditures.
2. If suitable, find the best-fitting linear regression lines using the Excel *Trendline* tool for each of the chart constructed in part (a).
3. Use the Excel functions *SLOPE* and *INTERCEPT* to confirm the slope and intercept of the best-fitting linear regression lines identified in part (b).
4. Use the Excel function TREND to estimate the weekly gross revenue for a television advertising expenditure of £2,800.
5. Apply the Excel regression tool using the amount of television advertising as the independent variable and interpret all key regression results in the output.
6. Apply the Excel regression tool using both television advertising and newspaper advertising as the independent variables.
7. How does the R2 differ between the two models? Which one is better?
8. What is the estimate of weekly gross revenue for a week when £3500 is spent on television advertising and £1800 is spent on newspaper advertising?

**Exercise 3 (A model with categorical variables) – 20 mins**

The file *Employee Salaries.xlsx* provides salary and age data for 35 employees, along with an indicator of whether or not the employees have an MBA (Yes or No).

1. Apply the Excel regression tool to predict salary as a function of the other variables and interpret all key regression results in the output.
2. What is the estimated salary of a 30-year old employee with an MBA?
3. By how much having an MBA increases the employee salary?

Hint: Note that the MBA indicator variable is categorical. Regression analysis requires numerical data so you need to code the categorical variables (i.e. replace “No” by 0 and “Yes” by 1).