ECMT1020: Exam Notes

Chapter 1: Overview

1.1 Analysis of economic data

1.1.1 Types of data

- Value
  - Numerical – naturally recorded and interpreted as numbers (continuous-annual income, or discrete-hrs worked – only integer values)
  - Categorical – belonging to 1+ groups (numbers have no meaning- gender, religion)
- Unit of observation: of actual behaviour in uncontrollable environment
  - Cross-section – different entities collected at common point in time, $x_i, i = 1, \dots n$ (where i is individual observation, n total number/sample size, x is value of variable observed) (e.g. une for state in single year)
  - Time series – same quantity at different time points, $x_i, t = 1, \dots T$ (where t is time period, T total periods, x value of variable observed) (e.g. AU GDP over time)
  - Panel data(x) – different individuals at multiple time points, $x_{i,t}, i = 1, \dots n; t = 1 \dots, T$ (e.g. life expentency by country over time)
- Number of variable
  - Univariate – single data series observing one variable (x_i for CS and x_t for TS data) (e.g. Inflation 1960-95)
  - Bivaritate – 2 related data series ($x_i y_i$ for CS data and $x_t y_t$ for TS data) (e.g. inflation and une over time)
  - Multivariate – 3+ related data series ($x_{1,i} y_{2,i} \dots x_{k,i} y_i$ for CS data and $x_{1,t} y_{2,t} \dots x_{k,t} y_t$ for TS data (e.g. input and output and profit of firm over time)

1.1.2 Steps of data analysis

- Data Summary
  - Visual representation through graphs/charts – single or relationship (+ summary stats)
- Statistical inference
  - Draw conclusions about relationships we can't observe (due to sample not pop.) – probable characteristics and relationship between at pop. level
- Interpretation

If X takes n values, $x_1, x_2 \dots x_{n-1}, x_n$ their sum is

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \cdots + x_{n-1} + x_n$$

- If g(x) is a function of x, then

$$\sum_{i=1}^{n} g(x_i) = g(x_1) + g(x_2) + g(x_3) + \cdots + g(x_n$$

- If "a" and "b" are constant, then
  - $\sum_{i=1}^{n} a = n * a$
  - $\sum_{i=1}^{n} a x_i = a \sum_{i=1}^{n} x_i$
  - $\sum_{i=1}^{n} (a + b x_i) = na + b \sum_{i=1}^{n} x_i$
  - $\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$
  - $\sum_{i=1}^{n} (x_i * y_i) \neq \sum_{i=1}^{n} x_i * \sum_{i=1}^{n} y_i$

Chapter 2: Univariate data

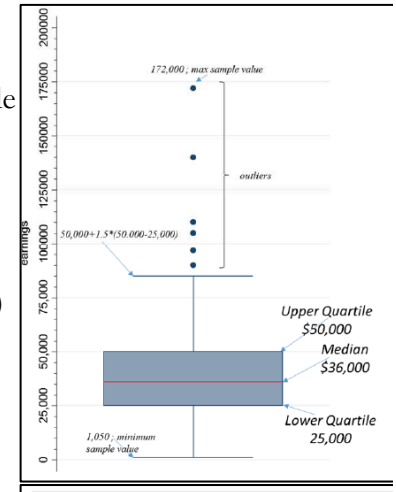
2.1 Univariate Data Summary

- Single series– observations on one variable (numerical - annual earnings/person in sample of women or categorical - expenditure of number of categories)

2.1.2 Summary statistics (numerical) (sample)

- Central Tendency
  - Mean: $\bar{x} = \frac{1}{n}(\Sigma_{i=1}^{n}) x_i$
  - Median: divides ordered data into 2 half's (middle value for odd n, average for even n) – less sensitive to outliers
  - Midrange: average of smallest and largest (outlier sensitive)
  - Mode – most frequent (discrete data where values are meaningful)
- Quartiles, deciles and percentiles
  - Median divides into half, Lower Q ¼ below and Upper Q ¾ below
  - p^{th} percentile – p% observed are equal or less than value
  - Deciles split ordered sample into 10ths
  - Quantile is percentile reported as fraction of one (0.56 quantile = 56th percentile)
- Dispersion

- Variance: $s^2 = \frac{1}{n-1}\Sigma_{i=1}^{n}(x_i - \bar{x})^2$, variance increase = spread wider
- SD (positive): $s = \sqrt{\frac{1}{n-1}\Sigma_{i=1}^{n}(x_i - \bar{x})^2}$, average deviation from mean, larger SD = greater variability
- Coefficient of variation: CV = $\frac{s}{\bar{x}}$, standardised measure, no units, compared across series (sample SD relative to sample mean)
- Range: largest minus smallest value (outlier sensitive)
- Inter-quartile range: variation on sample range (less sensitive) 75th-25th percentile
- Average absolute deviation: $\frac{1}{n}\Sigma_{i=1}^{n}|x_i - \bar{x}|$



Range (or bin)	Frequency	Relative frequency (%)
0-14,999	12	7.0
15,000-29,999	53	31.0
30,000-44,999	52	30.4
45,000-59,999	20	11.7
60,000-74,999	11	6.4
75,000-89,999	16	9.4
90,000-104,999	2	1.2
105,000-119,999	3	1.8
120,000-134,999	0	0.0
135,000-149,999	1	0.6
150,000-164,999	0	0.0
165,000-180,000	1	0.6

- Skewness (asymmetry)
  - Normal distribution: probability within 1SD is 0.684 and 2SD is 0.955
  - Other/Chebychev distribution: probability within 2SD is 0.75
  - Right/positive skewed – long right tail, concentrated left (mean>median & >0)
  - Left/negative skewed – long left tail, concentrated right (mean<median & <0)
  - $\frac{\frac{1}{n}\Sigma_{i=1}^{n}(x_i-\bar{x})^3}{\left[\frac{1}{n}\Sigma_{i=1}^{n}(x_i-\bar{x})^2\right]^{3/2}}$
- Kurtosis (peakedness)
  - Kurt = $\frac{\frac{1}{n}\Sigma_{i=1}^{n}(x_i-\bar{x})^4}{\left[\frac{1}{n}\Sigma_{i=1}^{n}(x_i-\bar{x})^2\right]^2}$
  - Normal distribution K=3, excess Kurt is actual Kurt (K)-3
  - Positive excess – fat tails (greater area) vs. negative excess – skinny tail
- Presented in table or Box and Whisker plot
  - Box – lower, median, upper quartile, Whiskers – min and max, Dots – outliers
  - Upper bar=upper quartile + 1.5x interquartile range, right-skewed

2.1.3 Graphical representations of univariate data

- Histogram
  - Absolute frequency of different values occurring – problem if n is large (hard to read y axis) (vs. relative frequency through % observed not number)
  - Bins = $\sqrt{n}$ (width = largest-smallest / bins)
  - Also Stem and leaf display
  - Smoothed histogram: Continuous (not discrete) probability density (rolling bins/windows and counting fraction within each bin - more weight to observations closer to center of window) – kernel density estimate
- Line chart
  - For natural ordered data (time-series data)
  - Variable value vs. observation number (data vs. index values)
- Pie/bar/column chart
  - Categorical data

Chapter 3: Economic data

3.1 Variables in economic analysis

- GDP
  - Aggregate output, sum of gross value added of all residents and institutional units (+tax – subsidies), geographical location of production, real GDP – control for price inflation
  - GNP - Value of all G&S produced by citizens (GDP + income earned by overseas AU)
- Price indexes
  - $100 \times \frac{Basket\ price\ CY}{Basket\ price\ PY/Base}$ (100 base period)
  - CPI: CPI in 1950 4.6 and 1960 8 – So $1 in 1950 equivalent to $\$1 \times \frac{8}{4.6} = 1.76$ in 1960, to convert from year t dollars to year t+n dollars, divide by CP_t and multiply by CP_{t+n}
  - GDP deflator: chained index controlling for substitution effect (G&S change and become cheaper) in response to change in relative prices
- Labour force statistics
  - Une Rate: 100 x Une / (Une + E) and Participation rate: 100 x labour force / population
- Stock price index
  - $I_t = \Sigma_{j=1}^{J}(w_{jt}\ p_{jt})$ - p_{jt} is price of j stock at t time, and w_{jt} is weight given to j stock