ECON 1310 Mid-term exam revision (key points)

Lecture 1 Descriptive Statistics I

1. Types of data and levels of data

<table>
<tr><td colspan="4">Types of data</td></tr>
<tr><td colspan="2">Quantitative (numerical)</td><td colspan="2">Qualitative (categorical)</td></tr>
<tr><td>Discrete
= variable has a whole counting number
e.g. 1. Number of phone calls
2. number of books
3. number of laptops</td><td>Continuous
= variable that is measureable, can have any real number value, includes numerical values with decimals
e.g. 1, length of time
2. temperature
3. weight</td><td colspan="2">- variables are not able to be measured on a number line since they typically take a “language” value
e.g. 1. type of drink ordered
2. favorite shopping center
3. method of payment</td></tr>
<tr><td colspan="4">Levels of data</td></tr>
<tr><td>Interval
The location of zero is a matter of convenience or convention and not a natural or fixed data point
e.g. 1. Temperature
2. calendar time (dates)</td><td>Ratio
Zero truly represents the value of what is being measured is nothing
e.g. height, weight and volume
Profit and loss, revenue, expenses</td><td>Nominal
Used only to classify or categorize
e.g. 1. types of occupation
2. gender</td><td>Ordinal
Numbers DO indicate rank or order
e.g. student grades</td></tr>
</table>

2. Probability sampling

Approaches	Content	Comparison with SRS
Simple Random Sample (SRS)	All items in the sampling frame have an *equal chance* of being selected	
Systematic Sample	A systematic sample is chosen from an ordered list of the population, with the starting point chosen randomly and each successive sample member selected systematically K = N/n	-Easy to do -Convenient -Evenly distributed across the frame -May reflect some periodicity (hence not be random) -Less statistically efficient
Stratified Sampling	-Population is divided into two or more strata, which are different from each other (mutually exclusive, heterogeneous between strata) -but within each strata the members all have the same characteristic (homogenous within) -a random sample of people is selected from each strata -in the same proportion as they occur in the population -this ensures a representative sample is selected which is more statistically efficient than SRS	-more costly -more closely matches variations in the population -more statistically efficient (i.e. reduction in sampling error) and the *most efficient* of the four methods -requires a smaller sample size than simple random sample for the same level of accuracy
Cluster Sampling	Population is divided into clusters (non-overlapping, mutually exclusive) each of which is representative of the population (has all the	-convenient (and more practical to do)

	characteristics of interest) (homogenous between clusters) -(Heterogeneous within clusters) -clusters often geographic areas -can subdivide into further clusters -a random sample cluster is selected -all items, or a random sample, from the clusters makes up the sample -may be less efficient than SRS	-cost efficient and time efficient to administer if data is widely dispersed -more difficult to analyze data -less statistically efficient (also compared to stratified)

Lecture 2 Descriptive Statistics II

1. **Central Tendency**

Mean	-*Affected* by each value in the data set, including extreme values	Population mean $\mu = \frac{\sum_{i=1}^{N} X_i}{N}$ Sample mean $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}$
Median	-*Not affected* by extremely large or extremely small values i.e. when the data is skewed -Often thought of as the *best measure* of central tendency	In the ordered observation: $median\ position = \frac{n+1}{2}$ If n is *odd*: the value of the median is in the position given by the formula If n is *even*: the value is the average of the two ordered observation values either side of the calculated position value given by the formula