STATISTICAL LABORATORY

Lecture 11: Final Review (Theory and Practice)

Yiu Lim LUI, Allen

Lisbon School of Economics & Management
Universidade de Lisboa

allenlui@iseg.ulisboa.pt

Applied Mathematics for Economics and Management
1st Year / 1st Semester – 2026/2027

Announcements

- **Group Project Report:** Submit as a PDF to `allenlui@iseg.ulisboa.pt` by 20:00 on November 20, 2026.
- **Written Exam:** November 25, 2026, in Amphitheater 24 (Floor 2, Building F1).
- **Exam Instructions:** Available now on Fenix and will be sent via email today.
- **Next Up:** Determining the group project presentation order.

Group Project Presentation Order

Order	Group	Members
____	Group 1	Clara Soares, Jingling Huang, Lara Oliveira, Inês Pereira, Ava Reis, Manuel Alves
____	Group 2	Carina Cruz, Martim Jesus, Beatriz Martins, Lara Ferreira, Rafæl Lourenço, André Cunha
____	Group 3	Maria Rita Vasques, Mateus Fontes, Vasco Menezes, Carlota Loureiro, Afonso Lima
____	Group 4	Maria Medeiros, Madalena Aleluia, Luís Oliveira, Francisco Oliveira, Afonso Andrade
____	Group 5	Maria Marques, António Mamede, Gonçalo Conchinha, Miguel Salgueiro, Luana Borges
____	Group 6	Diogo Brás, Guilherme Santos, Inês Frada, Matilde Nascimento, Margarida Carreira

Presentation Schedule: Will be uploaded to Fenix after class.

Final Review

The curriculum spans six main program modules covered across the semester:

1. **Fundamental Concepts of Statistics** (Data types, scales of measurement, population vs. sample)
2. **Exploratory Data Analysis & Summarizing Data** (Frequency tables, graphs, central tendency, dispersion, skewness, kurtosis)
3. **Association and Relationships Between Variables** (Covariance, correlation, simple linear regression, R^2, contingency tables, association measures)
4. **Index Numbers** (Simple, synthetic, Laspeyres, Paasche, Fisher, chain indices, base shifting, and reconciliation)
5. **Time Series Analysis** (Components, additive/multiplicative models, analytical linear trend, moving averages, seasonality analysis)

Module 1: Fundamental Concepts & Exploratory Data Analysis (Lectures 1–5)

1. Variables and Levels of Measurement

- **Qualitative (Categorical)**:
  - **Nominal**: Categories without inherent order (e.g., gender, nationality, state of residence).
  - **Ordinal**: Ordered categories where relative position matters, but differences are not numerical (e.g., satisfaction scale 1–5, education level).
- **Quantitative (Numerical)**:
  - **Discrete**: Countable values (e.g., number of children, late payments).
  - **Continuous**: Result from measurement processes and take infinite values within a range; decimals possible (e.g., weight, height, price).
  - *Measurement Levels*: **Interval** (meaningful differences, no true zero, such as Celsius temperature) and **Ratio** (true absolute zero exists, such as income or weight).

2. Frequency Distributions & Grouped Data Rules

- **Class Interval Construction (Histograms)**: The set of class intervals must be **inclusive** (collectively exhaustive, covering every observation from min to max) and **non-overlapping** (mutually exclusive, so each value falls into exactly one bin, e.g., $[a, b)$).
- **Sturges' Rule for Number of Classes (k)**: $2^k \geq n$ (or estimated via $k \approx \sqrt{n}$).
- **Class Width (h)**: $h = \dfrac{\max - \min}{k}$ (always round *up* to a convenient number).
- **Class Midpoint (x_j')**: $\dfrac{\text{Lower Limit} + \text{Upper Limit}}{2}$.

3. Measures of Central Tendency & Position

- **Arithmetic Mean**: $\bar{x} = \frac{\sum x_i}{n}$ (Sample) or $\mu = \frac{\sum x_i}{N}$ (Population). Grouped: $\bar{x} = \sum f_j x'_j$.
- **Weighted Mean**: $\bar{x} = \frac{\sum w_i x_i}{\sum w_i}$.
- **Median (Me)**: Middle value at position $\frac{n+1}{2}$. Grouped interpolation formula:

$$Me = l + \frac{0.5 - F^*(l)}{F^*(L) - F^*(l)} \cdot h$$

Module 1: Fundamental Concepts & Exploratory Data Analysis (Lectures 1–5)

- **Quantiles / Percentiles Position**: $(n+1)p$, where r is the integer part and a is the fractional part. If r is not an integer:

$$q_p = (1-a)x_{(r)} + ax_{(r+1)}$$

- **Mode (Mo)**: Most frequent value. Grouped via **King's Formula**:

$$Mo = l + \frac{f^{**}}{f^{*} + f^{**}} \cdot h$$

(where f^{*} is the relative frequency of the class before and f^{**} is the class after the modal class).

- Review exercise 3.1. in lecture 4b.

4. Measures of Dispersion & Concentration

- **Range**: $\max(x_i) - \min(x_i)$
- **Interquartile Range (IQR)**: $IQR = Q_3 - Q_1$. (Outlier inner fences: $Q_1 - 1.5 \cdot IQR$ to $Q_3 + 1.5 \cdot IQR$).
- **Variance (s^2) & Standard Deviation (s)**:

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{\sum x_i^2}{n} - \bar{x}^2 \quad \text{(Ungrouped)}$$

$$s^2 = \sum f_j (x_j' - \bar{x})^2 = \frac{\sum n_j (x_j')^2}{n} - \bar{x}^2 \quad \text{(Grouped)}$$

Module 1: Fundamental Concepts & Exploratory Data Analysis (Lectures 1–5)

- **Coefficient of Variation (CV):** $CV = \frac{s}{\bar{x}} \times 100\%$
- **Gini Index (G) for Concentration:**

$$G = 1 - \frac{\sum_{i=1}^{m-1} Q_i}{\sum_{i=1}^{m-1} P_i}$$

(where P_i is the cumulative population share and Q_i is the cumulative variable/income share).

- Review exercise 3.2 in Lecture 5.

5. Skewness Coefficients

- **Distribution Shape**: Mean $>$ Median $>$ Mo (Right/Positive skew); Mean $<$ Median $<$ Mo (Left/Negative skew); Mean $\approx$ Median $\approx$ Mo (Symmetric).
- **Pearson's Coefficient (g)**: $g = \dfrac{\bar{x} - Mo}{s}$
- **Bowley's Quartile Coefficient (g')**: $g' = \dfrac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)}$
- **Moment Coefficient (g_1)**: $g_1 = \dfrac{m_3}{m_2^{3/2}}$ *(where m_3 is the 3rd central moment and m_2 is the variance)*.
- Review Example I in Lecture 5.

Module 2: Association and Relationships Between Variables (Lectures 6–7)

1. Bivariate Quantitative Analysis (Correlation & Regression)

- **Covariance (s_{xy}):** $s_{xy} = \frac{1}{n}\sum(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n}\sum x_i y_i - \bar{x}\bar{y}$
- **Pearson Correlation Coefficient (r_{xy}):** $r_{xy} = \frac{s_{xy}}{s_x s_y}$ *(Ranges from -1 to $+1$)*
- **Simple Linear Regression Model:** $\hat{y} = \hat{a} + \hat{b}x$

$$\hat{b} = \frac{s_{xy}}{s_x^2} = \frac{s_{yx}}{s_x^2}, \qquad \hat{a} = \bar{y} - \hat{b}\bar{x}$$

- **Goodness of Fit (R^2):** $R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = r^2$ *(for simple linear regression)*.
  - $SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$, $SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$, $SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.
  - *Benchmarks*: > 0.9 (excellent), 0.7–0.9 (good), 0.5–0.7 (moderate), < 0.5 (weak).
  - Review Exercise 4.1 in Lecture 6.

2. Contingency Tables & Association Measures (Qualitative Data)

- **Expected Frequencies**: $n^*_{jk} = \dfrac{n_{j\cdot} \times n_{\cdot k}}{n}$
- **Pearson's Chi-Square (χ^2)**: $\chi^2 = \sum \dfrac{(n_{jk} - n^*_{jk})^2}{n^*_{jk}}$
- **General Tables ($I \times c$)**:

$$\Phi^2 = \frac{\chi^2}{n}, \qquad C = \sqrt{\frac{\chi^2}{\chi^2 + n}} \quad \text{(Contingency Coefficient)}$$

$$V = \sqrt{\frac{\Phi^2}{\min(I-1, c-1)}} \quad \text{(Cramr's } V\text{)},$$

$$T = \sqrt{\frac{\Phi^2}{\sqrt{(I-1)(c-1)}}} \quad \text{(Tschuprow's } T\text{)}$$

- Review exercise 4.4 in Lecture 7.

2×2 **Tables Simplified Formulas**:

$$\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}, \qquad \Phi = \frac{ad-bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

$$Q = \frac{ad-bc}{ad+bc} \quad \text{(Yule's } Q\text{)}, \qquad Y = \frac{\sqrt{ad}-\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}} \quad \text{(Yule's } Y\text{)}$$

Review Example 2 in Lecture 7.

1. Basic & Synthetic Indices

- **Simple Index:** $I_{t|0} = \dfrac{x_t}{x_0} \times 100$
- **Weighted Synthetic Index:** $I_{t|0} = \dfrac{\sum w_k x_{kt}}{\sum w_k x_{k0}} \times 100$

2. Price and Quantity Indices (Laspeyres & Paasche)

- **Laspeyres Price Index (L^P)** *(Fixed base quantities q_0)*:

$$L^P_{t|0} = \frac{\sum q_{k0} p_{kt}}{\sum q_{k0} p_{k0}} \times 100$$

- **Paasche Price Index (P^P)** *(Current period quantities q_t)*:

$$P^P_{t|0} = \frac{\sum q_{kt} p_{kt}}{\sum q_{kt} p_{k0}} \times 100$$

- **Composite / Ideal Indices**:
  - **Fisher Ideal Index (F^P)**: Geometric mean of Laspeyres and Paasche: $F^P = \sqrt{L^P \cdot P^P}$.
  - **Bowley Index (B^P)**: Arithmetic mean: $B^P = \frac{L^P + P^P}{2}$.
  - **Marshall-Edgeworth Index (ME^P)**: Uses average quantities $\frac{q_0 + q_t}{2}$.
- Review worked example in Lecture 8.

3. Chain Indices, Base Shifting, and Reconciliation

- **Link Index (Year-to-Year)**: $I_{t|t-1} = \dfrac{p_t}{p_{t-1}} \times 100$
- **Chain Index**: $I_{t|0}^{\text{chain}} = I_{1|0} \times \dfrac{I_{2|1}}{100} \times \cdots \times \dfrac{I_{t|t-1}}{100}$
- **Changing Base (Shifting)**: $i_{t|b} = \dfrac{i_{t|0}}{i_{b|0}} \times 100$
- **Base Uniformization (Reconciliation)**:
  - To shift from older base 0 to later base b (**Divide**): $I_{t|b}^{C} = \dfrac{I_{t|0}^{A}}{I_{b|0}^{A}} \times 100$
  - To shift from later base b back to older base 0 (**Multiply**):
    $I_{t|0}^{C} = I_{t|b}^{B} \times \dfrac{I_{b|0}^{A}}{100}$
- Review Reconciliation: Situation 1 and 2 in Lecture 9.

1. Decomposition Models

- **Additive Model**: $x_t = T_t + S_t + C_t + e_t$ *(used when seasonal amplitude is constant over time)*.
- **Multiplicative Model**: $x_t = T_t \cdot S_t \cdot C_t \cdot e_t$ *(used when seasonal amplitude grows proportionally with the trend)*.

2. Trend Analysis (Analytical Method)

- **Linear Trend Model**: $T_t = a + bt$
- **Least Squares Parameter Estimation**:

$$\hat{b} = \frac{s_{xt}}{s_t^2}, \qquad \hat{a} = \bar{x} - \hat{b}\bar{t}$$

$$(\textit{For } t = 1, 2, \ldots, n,\ \bar{t} = \frac{n+1}{2} \textit{ and } s_t^2 = \frac{n^2-1}{12}).$$

3. Moving Averages & Seasonality

- **Simple Moving Average (*MA*)**: Smooths short-term fluctuations by averaging k consecutive periods.
- **Centered Moving Average (*CMA*)**: Required for even-period moving averages (such as $MA(4)$) to align values back to proper time indices:

$$CMA_t = \frac{MA(4)_t + MA(4)_{t+1}}{2}$$

- **Seasonal Indices via Moving Averages**:
  - Ratio to moving average: $S_t = \frac{x_t}{CMA_t}$ (Multiplicative).
  - Average ratios by quarter/month and normalize so their average equals 1.
- Review exercise 6.5 in Lecture 10.

THANKS!

Questions?