LM09 · QUANTITATIVE METHODS · CFA LEVEL I

Parametric and Non-Parametric Tests of Independence

Tests of correlation, Spearman rank, contingency tables, chi-square

What We Test

Two related inference problems: (1) is the **correlation** between two variables zero (or not)? and (2) are two **categorical variables** independent? Parametric and non-parametric tools address both.

Learning Outcome Statements

LOS a Explain parametric and non-parametric tests of the hypothesis that the population correlation coefficient equals zero, and determine whether the hypothesis is rejected at a given level of significance

LOS b Explain tests of independence based on contingency table data

LOS A · PARAMETRIC TEST OF CORRELATION

Test of Pearson Correlation

Hypotheses

- Two-sided: H_0: ρ = 0; H_a: ρ ≠ 0
- Left-tailed: H_0: ρ ≥ 0; H_a: ρ < 0
- Right-tailed: H_0: ρ ≤ 0; H_a: ρ > 0

$$r_{xy} = \mathrm{Cov}(x, y)/(S_x \ \cdot \ S_y) \quad (Pearson/bivariatecorrelation)$$

Test Statistic

$$t_{n-2} = r \quad (n-2)/ \quad (1-r^2)$$

Large-Sample Warning

As n ↑, H_0 is rejected for even **very small correlations**. In big datasets, almost any r will be statistically significant.

Example: r = 0.02, n = 10,000

t = 0.02√9,998 / √(1 − 0.02^2) ≈ **2**, vs critical t = 1.96 ⇒ reject H_0!

WORKED EXAMPLE

Pearson Correlation Significance Test

WORKED EXAMPLE

Oil Prices vs Energy Stocks

GIVEN

$r = 0.7986$ between oil prices and monthly energy-stock returns in Country A, January 2014 through December 2018 ($n = 60$ months). Test at $\alpha = 0.05$.

HYPOTHESES

H_0: $\rho = 0$ (true population correlation is zero)

H_a: $\rho \neq 0$

CALCULATION

$t = 0.7986 \cdot \sqrt{(60 - 2)} / \sqrt{(1 - 0.7986^2)}$

$= 0.7986 \cdot \sqrt{58} / \sqrt{(1 - 0.6378)}$

$= 6.0820 / 0.6019 = \mathbf{10.1052}$

Critical value: $t_{58,\ 0.025} \approx \mathbf{2.00}$

DECISION

Since $|10.1052| > 2.00$, **reject H_0**. The population correlation is significantly different from zero.

Exam Tip: The magnitude of r needed to reject H_0: $\rho = 0$ decreases as n increases because (1) critical t falls and (2) the t-statistic's magnitude rises for any given r.

LOS A · NON-PARAMETRIC TEST

Spearman Rank Correlation Coefficient

Used when the **normality assumption** for X or Y is violated, or **outliers** are present. It is a correlation calculated on *rank values* rather than raw observations.

Procedure

1. Rank all X values from largest (rank 1) to smallest (rank n). For a tie, assign the **average rank** (e.g., 3 tied for 6th → each gets (6+7+8)/3 = 7).
2. Repeat for Y.
3. On the original paired data, compute $d_i^2 = (\text{rank } X_i - \text{rank } Y_i)^2$.
4. Compute $r_S = 1 - [\, 6 \cdot \Sigma_{i=1}^{n} d_i^2 \,] / [\, n(n^2 - 1) \,]$

Testing r_S (n > 30)

$$t_{n-2} = r_S \cdot \quad (n-2)/ \quad (1 - r_S^2)$$

Exam Tip: Use Spearman when: normality violated, outliers present, or data are ordinal (ranks).

CHECK YOUR UNDERSTANDING · QUIZ 1

Quiz 1

Select the best answer. Feedback appears after each click.

Q1. The parametric test statistic for H_0: ρ = 0 is distributed with:

Q2. r = 0.60, n = 30. The t-statistic is closest to:

Q3. Spearman rank correlation is preferred over Pearson when:

Q4. For r = 0.02, n = 10,000, the Pearson t-statistic is approximately 2. This shows:

LOS B · TESTS OF INDEPENDENCE

Contingency Tables & Chi-Square Test

A **contingency table** is a two-way array. Rows = attributes of one variable, columns = attributes of the other. Cells hold observed frequencies. We test whether the two characteristics are **independent**.

Test Statistic (non-parametric, right-tailed)

$$\chi^2 = \Sigma_{i=1}^{m}(O_{ij} - E_{ij})^2/E_{ij}$$

$$df = (r-1)(c-1)$$

where m = total cells, O_{ij} = observed in cell, E_{ij} = expected if independent.

$$E_{ij} = (row_i total) \times (column_j total)/overalltotal$$

Exam Tip: Chi-square is always **right-tailed** because squared deviations can't be negative. Larger χ^2 → observed further from expected → reject independence.

WORKED EXAMPLE

Size vs Type Independence Test

WORKED EXAMPLE

Stock Size and Type

GIVEN

3 × 3 contingency table of stocks classified by size (S/M/L) and type (V/B/G). Overall total = 1,594. df = (3−1)(3−1) = **4**.

E_{SV} = (503 × 148)/1594 = **46.703**

E_{MG} = (366 × 381)/1594 = **87.482**

O_{SV} = 50, O_{MG} = 122.

HYPOTHESES

H_0: size and type are independent

H_a: size and type are NOT independent

CELL CONTRIBUTIONS (EXAMPLES)

SV cell: $(O-E)^2/E = (50 - 46.703)^2 / 46.703 =$ **0.2328**

Summing all nine cells: χ^2 = **32.08025**

Critical value: CHISQ.INV(0.95, 4) = **9.4877**

DECISION

32.08 > 9.49 ⇒ **Reject H_0**. Size and type are NOT independent.

Standardized Residuals

$$SR_{ij} = (O_{ij} - E_{ij})/\quad E_{ij}$$

- SR > 0: more observations than expected if independent
- SR < 0: fewer observations than expected if independent

Example: $SR_{SV} = (50 - 46.703)/\sqrt{46.703} =$ **0.48**

$SR_{MG} = (122 - 87.482)/\sqrt{87.482} =$ **3.69** (significantly more than expected).

CHECK YOUR UNDERSTANDING · QUIZ 2

Quiz 2

Select the best answer. Feedback appears after each click.

Q1. For a 3 × 4 contingency table, degrees of freedom for the chi-square test of independence are:

Q2. Expected frequency in cell (i, j) of a contingency table equals:

Q3. The chi-square test of independence is:

Q4. A standardized residual of −2.5 indicates the cell has:

Q5. Which of the following tests is parametric?

CHECK YOUR UNDERSTANDING · QUIZ 3

Quiz 3

Select the best answer. Feedback appears after each click.

Q1. r = 0.7986, n = 60. The t-statistic is closest to:

Q2. Spearman correlation for a sample of n = 10 (X, Y) paired observations with $\Sigma d_i^2 = 12$ is:

Q3. A researcher uses a chi-square test of independence on a 2 × 3 contingency table with $\chi^2 = 7.8$. Critical value $\chi^2_{0.05,\,2} = 5.99$. The researcher should:

MODULE COMPLETE

LM09 · Parametric and Non-Parametric Tests of Independence Summary

Final Score

0 / 0

Cheat Sheet

$$Pearson : t = r \quad (n-2)/ \quad (1-r^2), df = n-2$$

$$Spearman : r_S = 1 - 6\Sigma d_i^2/[n(n^2-1)]$$

$$\chi^2 = \Sigma(O_{ij} - E_{ij})^2/E_{ij}, df = (r-1)(c-1)$$

$$E_{ij} = (rowtotal)(coltotal)/grandtotal$$

$$SR_{ij} = (O_{ij} - E_{ij})/ \quad E_{ij}$$

$$Chi - squaretestisalwaysright - tailed$$

Last-Minute Exam Checklist

Oil/energy example: r=0.7986, n=60, t=10.1052, reject ρ=0

Size/type example: χ^2=32.08 > 9.49, reject independence

Expected: E_{SV}=(503×148)/1594=46.703

Std residual: SR_{MG}=3.69 = more than expected

Spearman: rank-based, robust to outliers

Small r, large n: can still be significant (r=0.02, n=10,000)

Step 1 of 10