# Location Quotient (LQ)

The **location quotient (LQ)** measures the concentration of an industry in a region compared to some benchmark region.

where E is the number of establishments in a given industry (i) and/or region (r). N refers to the benchmark region overall.

# Credibility Adjusted LQ

We can adjust the LQ using the Buhlmann credibility factor (Z):

= 1 since the ratio of to itself will always be 1. The purpose of the credibility adjustment is to pull LQ estimates with less underlying data toward the mean, scaled by the degree of the uncertainty.

# Credibility Factor (Z)

The credibility factor Z is calculated as:

where is a measure of the number of observations in a region (e.g. ); and K is a constant reflect the balance between within-region and between-region variance. It is empirically calculated using the Buhlmann formula:

where EPV is the *Expected Value of Process Variance,* and VHM is the *Variance of Hypothetical Means.*

* **EPV** measures the average variability *within regions* in the proportions of enterprises across industries due to random chance. It quantifies how much industry proportions in a region might fluctuate purely because of random sampling variability. A higher EPV suggests that observed differences within regions are largely due to randomness, not systematic factors.
* **VHM** measures the variability *between regions* in the true proportions of enterprises in each industry. It captures real differences in industry composition across regions. A higher VHM indicates significant disparities in industry concentrations among regions, reflecting actual regional specialization.
* **K** balances that weight between regional data and benchmark data in the credibility adjustment. If K is high (EPV > VHM), random variability within regions dominates, meaning regional data is less credible and low weight is given to traditional LQs. If K is low (VHM > EPV), systematic differences between regions dominate, meaning regional data is more credible and more weight is given to regional LQs. With the formula for credibility factor in mind, as is higher, , meaning regional LQs are trusted more. As K is lower, , meaning regional LQs are trusted less.

# Expected Value of Process Variance (EPV)

EPV measures the average variability within regions, calculated by the following steps.

1. The proportion of each industry in each region is calculated as .
2. The variance for each industry/region is here: , which is derived from the multinomial distribution’s variance for proportions. *Ask GPT to show this if needed.* This says, for each region, what is the *variance* of the proportions of each industry?
3. EPV is then the average variability within regions:

where is the total number of regions and is the total number of industries. *This is the AVERAGE variance of proportions of each industry within each region.*

# Variance of Hypothetical Means (VHM)

VHM measures the variability *between regions*, calculated as:

1. Estimate average proportion for each industry:
2. Compute variance across regions for each industry: .
3. Calculate VHM:

# Calculate the Credibility Factor

Use EPV and VHM to get the constant K: and plug that into the credibility formula for each region:

We can then compute the credibility-adjusted LQ for each industry and region:

# Assumptions

1. **Multinomial Distribution:** each enterprise belongs to exactly one industry. The distribution of enterprises across industries follows a multinomial distribution. For each region , the counts are multinomially distributed with total count and probabilities . The counts of enterprises across industries in a region follow a multinomial distribution. The probability represents the likelihood of an enterprise being in industry , within region .
2. **Independence of Regions:** enterprise distributions are independent across regions.
3. **Homogeneity of Process Variance:** the variance within regions (EPV) is assumed to be similar across all regions and industries.