Support for Eliminating Tariffs on Electric Vehicles from China (Practice: Comparing Multiple Proportions)

Leonardo L Sanchez

2026-02-23

Goal

This practice file shows **how to do the test manually** (step by step) and then verify in `R`.

Topic covered:

- Comparing multiple proportions (chi-square test of homogeneity)
- Pairwise comparisons using Marascuilo-style simultaneous intervals
- A follow-up two-proportion z-test (West vs East)
- Confidence interval for the overall proportion

Given Data

A random sample of Canadians from four major cities was surveyed.

City	Sample Size (n)	% Yes	Yes	No
Vancouver	400	62%	248	152
Calgary	250	58%	145	105
Toronto	300	52%	156	144
Montreal	250	48%	120	130

```r
# Use the exact counts from the problem (best practice for inference)
city <- c("Vancouver", "Calgary", "Toronto", "Montreal")
n    <- c(400, 250, 300, 250)
yes  <- c(248, 145, 156, 120)
no   <- n - yes
p_hat_city <- yes / n

# Build a clean data frame
df <- data.frame(
  City = city,
  n = n,
  Yes = yes,
  No = no,
  p_hat = p_hat_city
)
df$YesPercent <- 100 * df$p_hat

# Reorder columns for display
```

```r
df <- df[, c("City", "n", "YesPercent", "Yes", "No", "p_hat")]
df
```

```
##        City   n YesPercent Yes  No p_hat
## 1 Vancouver 400         62 248 152  0.62
## 2   Calgary 250         58 145 105  0.58
## 3   Toronto 300         52 156 144  0.52
## 4  Montreal 250         48 120 130  0.48
```

Part A - Test for Differences Among Cities (Chi-square Test of Homogeneity)

1) Hypotheses

Let p_V, p_C, p_T, p_M be the true proportions of "Yes" in Vancouver, Calgary, Toronto, and Montreal.

$$H_0 : p_V = p_C = p_T = p_M$$

$$H_a : \text{At least one population proportion is different}$$

2) Manual idea (what we do by hand)

Under H_0, all cities share **one common proportion**. We estimate it using the pooled sample proportion:

$$\hat{p}_{\text{pooled}} = \frac{\text{total Yes}}{\text{total sample size}}$$

```r
# MANUAL STEP 1: pooled proportion under H0
x_total <- sum(yes)
n_total <- sum(n)
p_pool <- x_total / n_total

x_total
```

```
## [1] 669
```

```r
n_total
```

```
## [1] 1200
```

```r
p_pool
```

```
## [1] 0.5575
```

So,

$$\hat{p}_{\text{pooled}} = \frac{669}{1200} = 0.5575$$

3) Expected counts under H_0

For each city:

- Expected Yes $= n_i \hat{p}_{\text{pooled}}$
- Expected No $= n_i(1 - \hat{p}_{\text{pooled}})$

```r
# MANUAL STEP 2: expected counts under H0
expected_yes <- n * p_pool
expected_no  <- n * (1 - p_pool)

expected_df <- data.frame(
  City = city,
  Obs_Yes = yes,
  Exp_Yes = expected_yes,
  Obs_No = no,
  Exp_No = expected_no
)
expected_df
```

```
##        City Obs_Yes Exp_Yes Obs_No  Exp_No
## 1 Vancouver     248 223.000    152 177.000
## 2   Calgary     145 139.375    105 110.625
## 3   Toronto     156 167.250    144 132.750
## 4  Montreal     120 139.375    130 110.625
```

```r
# Condition check: all expected counts should be >= 5
all(expected_yes >= 5)
```

```
## [1] TRUE
```

```r
all(expected_no >= 5)
```

```
## [1] TRUE
```

4) Chi-square statistic (manual computation)

We compute:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

across all cells in the 4 x 2 table.

```r
# MANUAL STEP 3: contribution from each cell = (O-E)^2/E
contrib_yes <- (yes - expected_yes)^2 / expected_yes
contrib_no  <- (no  - expected_no )^2 / expected_no

contrib_df <- data.frame(
  City = city,
  contrib_yes = contrib_yes,
  contrib_no = contrib_no,
  contrib_total_city = contrib_yes + contrib_no
)
contrib_df
```

```
##        City contrib_yes contrib_no contrib_total_city
## 1 Vancouver   2.8026906  3.5310734          6.3337640
## 2   Calgary   0.2270179  0.2860169          0.5130349
## 3   Toronto   0.7567265  0.9533898          1.7101163
## 4  Montreal   2.6933857  3.3933616          6.0867472
```