Inference Guide – Categorical Data Distributions (χ^2)

One Variable

One Sample [df = *# of cells/categories* – 1]
└▷ compared with population model

H_0: *distribution = specified model*
H_A: *distribution ≠ specified model (right sided)*

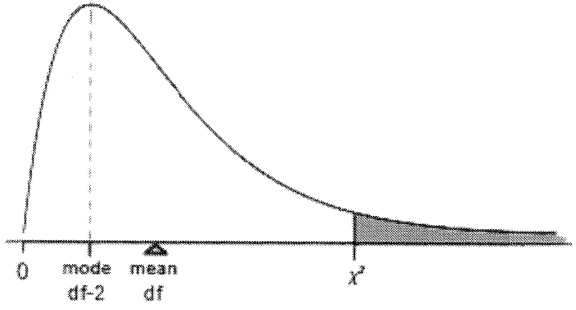
A0 Data are counts.
C0 (Are they?)
A1 Individuals/data independent.
C1 SRS and *n* < 10% population.

A2 Sample large enough
C2 All expected counts ≥ 5.

χ^2 test for Goodness-of-Fit [df = *# of cells* – 1]

$$\chi^2 = \sum_{all\ cells} \frac{(Obs - Exp)^2}{Exp}$$

One Variable ↓	Obs Counts	Exp Value (Counts)	Residuals (Obs-Exp)	$(Resid)^2$	Component $\frac{(Obs-Exp)^2}{Exp}$
Cat. 1		∑Obs*hyp			
Cat. 2		∑Obs*hyp			
Cat. 3		∑Obs*hyp			
Cat. 4		∑Obs*hyp			
df = #cat- 1	∑Obs				$\chi^2 = \sum$



P-value = χ^2cdf (χ^2, 999, df)

Or use: χ^2GOF-Test (L_{Obs}, L_{Exp}, df)
If reject H_0, then ☆

Many Groups

Many Groups [df = (r – 1)(c – 1)]
└▷ compared on one variable

H_0: *distributions = for each group*
H_A: *distributions ≠ for each group (right sided)*

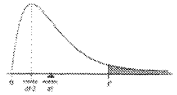
A0 Data are counts.
C0 (Are they?)
A1 Individuals/data in each group independent.
C1 SRSs and *n* < 10% populations
OR random allocation.
A2 Groups large enough
C2 All expected counts ≥ 5.

χ^2 test for Homogeneity [df = (*r* – 1)(*c* – 1)]

$$\chi^2 = \sum_{all\ cells} \frac{(Obs - Exp)^2}{Exp}$$

$$Exp_{cell} = \frac{(row\ total)(column\ total)}{grand\ total}$$

One Variable ↓	Group 1 Obs Counts	Group 2 Obs Counts	Group 3 Obs Counts	Total
Cat. 1	Obs / Exp	Obs / Exp	Obs / Exp	
Cat. 2	Obs / Exp	Obs / Exp	Obs / Exp	
Cat. 3	Obs / Exp	Obs / Exp	Obs / Exp	
Total				

Draw curve → 

P-value = χ^2cdf (χ^2, 999, df)

Or use: MATRIX, STAT TESTS, χ^2-Test.
If reject H_0, then ☆

Two Variables

One Sample [df = (r – 1)(c – 1)]
└▷ classified on two variable

H_0: *Variable 1 and Variable 2 = independent.*
H_A: *Variable 1 and Variable 2 ≠ independent.*

A0 Data are counts.
C0 (Are they?)
A1 Individuals/data independent.
C1 SRS and *n* < 10% population.

A2 Sample large enough
C2 All expected counts ≥ 5.

χ^2 test for Independence [df = (*r* – 1)(*c* – 1)]

Var. 2 → Var. 1 ↓	V2-Cat. 1 Obs Counts	V2-Cat. 2 Obs Counts	V2-Cat. 3 Obs Counts	Total
V1-Cat. 1	Obs / Exp	Obs / Exp	Obs / Exp	
V1-Cat. 2	Obs / Exp	Obs / Exp	Obs / Exp	
V1-Cat. 3	Obs / Exp	Obs / Exp	Obs / Exp	
Total				

Draw curve and calculate P-Value ←

☆ examine the standardized residuals, $\frac{(Obs - Exp}{\sqrt{Exp}}$
to reveal how the data deviate from H_0.
Think of each component as a z-score^2, so