UDACITY

9 — Chi-Squared tests

9.1 Scales of measurement

> **Definition 9.1 — Ordinal Data.** There is a clear order in the data set but the distance between data points is unimportant.

> **Definition 9.2 — Interval Data.** Similar to an ordinal set of data in that there is a clear ranking, but each group is divided into equal intervals

> **Definition 9.3 — Ratio Data.** Similar to interval data except there exists an absolute zero.

> **Definition 9.4 — Nominal Data.** This is the same as qualitative data, where we differentiates between items or subjects based only on their names and/or categories and other qualitative classifications they belong to.

Type of Data	Example	Data
Ordinal	Ranks in a race	1st, 2nd, 3rd
Interval	Temperature in Celsius	$-10° - 0°, 1° - 10°, 11° - 20°$
Ratio	Percentage correct on test	$0 - 10\%, 11 - 20\%, 21 - 30\%$
Nominal	Shirt Colors	Red, Blue, Yellow, White

Table 9.1: Examples of different scales of measurement

■ **Example 9.1**   ■

9.2 Chi-Square GOF test

The Chi-Square GOF test allows us to see how well observed values match expected values for a certain variable. In particular we compare the frequencies of our data sets.

9.2.1 Chi-Square test of independence

This variation of the Chi-Square test is used to determine if 2 nominal variables are independent. In particular we use the marginal totals.

9.3 Practice Problem

Problem 9.1 A poker-dealing machine is supposed to deal cards at random, as if from an infinite deck. In a test, you counted 1600 cards, and observed the following: table[h]

Suit	Count	
Spades	404	
Hearts	420	Card counts
Diamonds	400	
Clubs	376	

Could it be that the suits are equally likely? Or are these discrepancies too much to be random?