Summary of Hypothesis Testing

Null hypothesis is rejected if p-value is less than significance level. Why?

Confidence interval is determined by z-score corresponding to the significance level.

How is p-value related to probability of null hypothesis to be true?

1.3.7 Chi-squared test for independence

Chi-square distribution

Let a coin tossed n times on an average gives n*p heads where p is the probability of getting a head.

Compute the population variance.

What is the variation in the estimation of p itself?

Partial derivation of Pearson's Chi-square test

$$\hat{p} \sim N\left(\hat{p}; p, \frac{p(1-p)}{n}\right)$$

Then $p' = \frac{n\hat{p} - np}{\sqrt{np}} \sim N(p'; 0, 1-p)$

$$\Rightarrow q = p'^2 = \frac{(n\hat{p} - np)^2}{np} \sim Z(q; 0, 1-p)$$

Here p' is normally distributed with variance $1-p$ and let the distribution of q be Z. We note that $n\hat{p}$ is observed frequency (O) and *np* is the expected frequency (E). Therefore,

$$q = \frac{(O-E)^2}{E} \sim Z(q; 0, 1-p)$$

We can divide the domain of any arbitrary distribution into b different bins. Each bin behaves like tossing a coin because a randomly generated variable from that distribution will either fall in a bin with probability p or not fall in that bin. Adding Z distributions for all the b bins we get **Chi-square distribution** of $b-1$ degrees of freedom (proof available in literature).

$$\sum_{j=1}^{b} \frac{(O_j - E_j)^2}{E_j} \sim \chi^2_{b-1}$$

Note that Z is not χ^2_1, and bin i and bin j are not independent, otherwise the sum would have been χ^2_b with b degrees of freedom.