Homework 1

Kevin Linares

2025-01-13

The following data are from Google Trends show the number of times that the term "film noir" was searched using Google.

week	film_noir
2022-10-02	68
2022-10-09	73
2022-10-16	58
2022-10-23	59
2022-10-30	72
2022-11-06	70
2022-11-13	77
2022-11-20	57
2022-11-27	56
2022-12-04	76
2022-12-11	63
2022-12-18	52

1. a) [5 points] Calculate maximum likelihood estimate of p (i.e. the proportion of all 781 searches that occurred in each week). Graph these 12 proportions.

    1. The maximum likelihood estimation of a proportion p is given as, which is the success over trials:

$$\theta = g(p), \text{ is } \widehat{\theta} = g(\widehat{p})$$

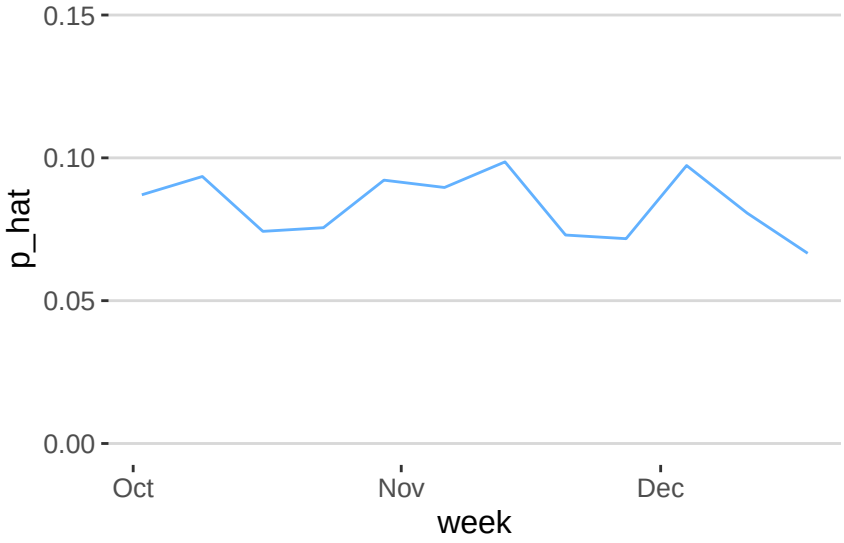
```r
gtrends <- gtrends |>
  mutate(p_hat = film_noir / sum(film_noir))

kable(gtrends)
```

week	film_noir	p_hat
2022-10-02	68	0.0870679
2022-10-09	73	0.0934699
2022-10-16	58	0.0742638
2022-10-23	59	0.0755442
2022-10-30	72	0.0921895
2022-11-06	70	0.0896287
2022-11-13	77	0.0985915
2022-11-20	57	0.0729834
2022-11-27	56	0.0717029
2022-12-04	76	0.0973111
2022-12-11	63	0.0806658
2022-12-18	52	0.0665813

2. We visualize the proportions by week date.

```r
gtrends |>
  ggplot(aes(x=week, y=p_hat)) +
  geom_line(color="dodgerblue", alpha=.7) +
  ylim(0, .15) +
  theme_hc()
```



1. b) [5 points] Write the null hypothesis that the proportion of searches for "film noir" is the same each week. Also, write the alternative hypothesis (i.e., that there has been a change in the proportion of searches each week).

    1. The null hypothesis, $H_0 : p_1 = p_2 = p_3 ... = p_{12}$, is that the weekly proportion of web searches for "film_noir" is the same across weeks. The alternative hypothesis, $H_a : p_1 \neq p_2 \neq p_3 ... \neq p_{12}$, is that there is at least one proportion not equal to the other weekly proportions of web searches, in other words, there has been a change in the proportion of searches each week.

2. c) [15 points] Compute the X2 and G2 statistics. What do these tell us?

    - The Pearson Chi-square test

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

- where $O_i = n_i$ is the observed count in the i^{th} category, and $E_i = np_{0i}$ i
- The likelihood ratio statistic G^2 is a statistical method used to compare the goodness

$$G^2 = 2 \sum_{i=1}^{k} O_i ln(\frac{O_i}{E_i})$$