**Exercise 2.1 Based on the results in this chapter, suppose you were asked to summarize what you learned about whether first babies arrive late.**

**Which summary statistics would you use if you wanted to get a story on the evening news?**

To create a story on evening news on to find whether first babies arrive late, I would use the most common summary statistic ‘mean’, which is meant to describe the central tendency of distribution. In this scenario, we can determine the answer by using the average pregnancy length of n women. Additionally, we could use the standard deviation to show the expected deviations.

**Which ones would you use if you wanted to reassure an anxious patient?**

To reassure an anxious patient, we could use the ‘effect size’ summary statistic or by calculating Cohen’s d, by calculating the difference in the means of two groups i.e., firsts and others. Though the difference is of small fraction, this estimate could be accurate and could be used to assure an anxious patient. However, it is unlikely that anyone would notice this fractional difference at all.

**Finally, imagine that you are Cecil Adams, author of The Straight Dope (http://straightdope.com), and your job is to answer the question, “Do first babies arrive late?” Write a paragraph that uses the results in this chapter to answer the question clearly, precisely, and honestly.**

According to histogram plot data, first babies’ pregnancy length is less than other babies’ pregnancy length which means first babies do not arrive late.

Chart, histogram

Description automatically generated

Histograms are useful because they make the most frequent values immediately apparent. But they are not the best choice for comparing two distributions. In this example, there are fewer “first babies” than “others,” so some of the apparent differences in the histograms are due to sample sizes.

To answer this precisely, we can use summary statistics to calculate mean, standard deviation and variance. For all live births, the mean pregnancy length is 38.6 weeks, the standard deviation is 2.7 weeks, which means we should expect deviations of 2-3 weeks to be common. Variance of pregnancy length is 7.3, which is hard to interpret, especially since the units are weeks2 , or “square weeks.” Variance is useful in some calculations, but it is not a good summary statistic.

We don’t need Variance because it gives a minute difference between 2 groups.

Another statistic effect size is intended to describe the size of an effect. For example, to describe the difference between two groups, one obvious choice is the difference in the means. Mean pregnancy length for first babies is 38.601; for other babies it is 38.523. The difference is 0.078 weeks, which works out to 13 hours. As a fraction of the typical pregnancy length, this difference is about 0.2% which is too small, so we cannot use this statistic.

Overall, we can use the Histogram and Mean to answer the question, “Do first babies arrive late?”. The answer it seems to be, No – the first babes arrive little early compared to the other babies that are not first for their moms.

**Exercise 2.4 Using the variable totalwgt\_lb, investigate whether first babies are lighter or heavier than others. Compute Cohen’s d to quantify the difference between the groups. How does it compare to the difference in pregnancy length?**

Chart, histogram

Description automatically generated

From the above histogram plot, we can see that first babies are lighter than the other babies. Based on the other summary statistics we have seen that there is not much difference in value for Mean, Variance and Standard deviation. So, we can conclude this question based on the histogram plot that first babies are relatively lighter than the other babies.

Cohen’s difference between two groups for total weight and pregnancy length has been calculated above. Cohen’s difference between two groups for total weight is -0.089 which is negative that means that there is no improvement/significant difference between both groups. Also, Cohen’s difference between two groups for pregnancy length is 0.029 which is small too. So these two values doesn’t add any weight to our investigation.