

Probability

Binomial Distribution – Racial Profiling

Prof. Dr. Mostafa Elhosseini

<https://youtube.com/drmelhosseini>

| To conclude...

Summary

If $X \sim \text{Bin}(n, p)$, then the mean and variance of X are given by

$$\mu_X = np \quad (4.5)$$

$$\sigma_X^2 = np(1 - p) \quad (4.6)$$

- The **binomial distribution** has a **bell shape** when **n is large**, so in that case, we can **use the normal distribution** to approximate the binomial distribution and conclude that nearly **all the probability falls between $\mu - 3\sigma$ and $\mu + 3\sigma$**

| Checking for Racial Profiling

In 2006, the New York City Police Department (**NYPD**) confronted approximately **500,000 pedestrians** for suspected criminal violations. Of those confronted, **88.9% were non-white**. Meanwhile, according to the 2006 American Community Survey conducted by the U.S. Bureau of the Census, of the more than **8 million individuals** living in New York City, **44.6% were white**.

- Are the data presented above evidence of racial profiling in police officers' decisions to confront particular individuals?

| Checking for Racial Profiling

- We'll treat the 500,000 confrontations as $n = 500,000$ trials. From the fact that 44.6% of the population was white, we can deduce that the other 55.4% was non-white. Then, if there is no racial profiling, the probability that any given confrontation should involve a non-white suspect is $p = 0.554$ (other things being equal, such as the rate of engaging in criminal activity).
- Suppose also that successive confrontations are independent. (They would not be, for example, if once an individual were stopped, the police followed that individual and repeatedly stopped her or him.)

| Checking for Racial Profiling

- Under these assumptions, for the 500,000 police confrontations, the number of non-whites confronted has a binomial distribution with $n = 500,000$ and $p = 0.554$.
- The binomial distribution with $n = 500,000$ and $p = 0.554$ has $\mu = np = 500,000(0.554) = 277,000$
- And $\sigma = \sqrt{np(1 - p)} = \sqrt{500,000(0.554)(0.446)} = 351$

| Checking for Racial Profiling

- Below we see that this binomial distribution is approximated reasonably well by the **normal distribution** because the number of **trials is so large**.
- Then, the probability of **falling within 3 standard deviations of the mean is close to 1.0**. This is the interval between
- $\mu - 3\sigma = 277,000 - 3(351) = 275,947$
- $\mu + 3\sigma = 277,000 + 3(351) = 278,053$

| Checking for Racial Profiling

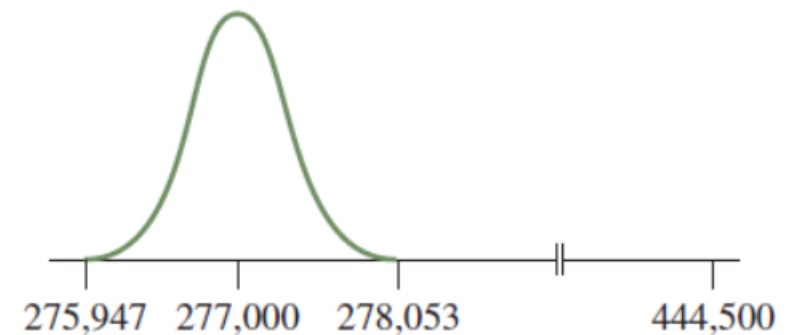
- If no racial profiling is taking place, **we would not be surprised if between about 275,947 and 278,053 of the 500,000 people stopped were non-white**. See the smooth curve approximation for the binomial in the margin. However, 88.9% of all stops, or $500,000 \times 0.8892 = 444,500$, involved nonwhites.
- **This suggests that the number of non-whites stopped is much higher than we would expect if the probability of confronting a pedestrian were the same for each resident, regardless of his or her race**

| Checking for Racial Profiling

- If no racial profiling is taking place, **we would not be surprised if between about 275,947 and 278,053 of the 500,000 people stopped were non-white.** See the smooth curve approximation for the binomial in the margin. However, 88.9% of all stops, or $500,000 \times 0.8892 = 444,500$, involved nonwhites.
- **This suggests that the number of non-whites stopped is much higher than we would expect if the probability of confronting a pedestrian were the same for each resident, regardless of his or her race**

Recall

From Section 6.2, when a distribution has a normal distribution, nearly 100% of the observations fall within 3 standard deviations of the mean. ◀

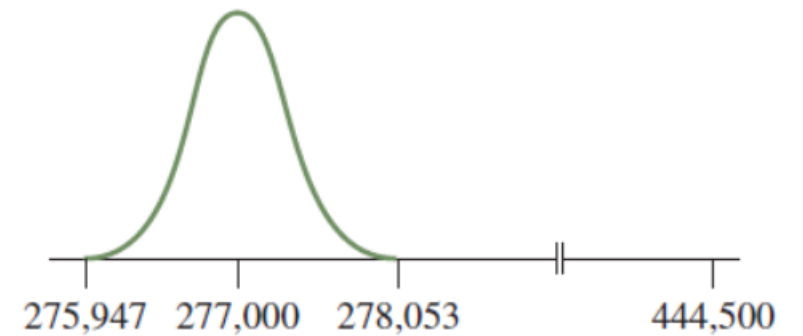


| Checking for Racial Profiling

- The controversial subject of racial profiling continues to receive nationwide attention. A 2013 report by the Bureau of Justice Statistics⁷ based on 2011 survey data found that relatively more black drivers (13%) than white (10%) and Hispanic (10%) drivers were pulled over in a traffic stop and that white drivers were both ticketed and searched at lower rates than black and Hispanic drivers.

Recall

From Section 6.2, when a distribution has a normal distribution, nearly 100% of the observations fall within 3 standard deviations of the mean. ◀



| References

- Statistics for Engineers and Scientists, Fourth Edition, William Navidi
Colorado School of Mines
- Agresti, A., Franklin, C. and Klingenberg, B., 2017. Statistics: The art and science of learning from data (Fourth Edi).
- Khan Academy