Natalie Belmonte

30 April 2024

CSCI 3327- Probability and Applied Statistics

Dataset Analysis: Atlantic City Temperatures

Introduction- why temperature?

When choosing a dataset to analyze, I selected something that I had some amount of prior experience with: the daily high and low temperatures in a certain region for a given amount of time. A few years ago, I worked on a personal project involving the daily temperature for an entire year in my hometown. Someone had created a cross stitch pattern from daily temperatures, and I was fascinated by it. Though I have not worked on the project in some time, I still have a interest in how daily temperatures change over time.

This report is using data from the National Weather Service (1), sourced from the Atlantic City Airport for the month of March 2024. I have attached the full generated PDF of data to this report for reference. Though thirty-one days does not provide enough data to completely analyze a change in temperature over time, it gives enough information to complete a statistical analysis on specific data points.

Date	Temperature			
	Maximum	Minimum	Average	Departure
2024-03-01	45	21	33.0	-5.7
2024-03-02	53	43	48.0	9.0
2024-03-03	69	42	55.5	16.3
2024-03-04	60	37	48.5	9.1
2024-03-05	53	43	48.0	8.3
2024-03-06	53	46	49.5	9.6
2024-03-07	56	48	52.0	11.9
2024-03-08	53	38	45.5	5.1
2024-03-09	55	40	47.5	6.9
2024-03-10	54	35	44.5	3.6

These are the first ten entries in the chart I am using for my dataset. Though I am focusing on maximum/minimum values, average and departure (the deviation from average) may be referenced. All numbers are in Fahrenheit.

Chapter 2- Probability

As an introduction to general probability terms, this chapter focuses on topics such as set theory, countable probability, conditional probability, and random variables. The first three sections have been omitted for relevancy.

1. (Example 2.1, a, b, e): A journalist picks two days from the first five days of March and wants to find a day where the minimum temperature was above 40 degrees.
    a. List the sample space. There are three days above 40 (A1, A2, A3) and two days below (B1, B2).
    E1 = {A1, A2} E2 = {A2, A3} E3 = {A1, A3} E4 = {B1, B2} E5 = {B1, A1}
    E6 = {B1, A2} E7 = {B1, A3} E8 = {B2, A1} E9 = {B2, A2} E10 = {B2, A3}
    b. Let A be the event that both days are above 40. Sample points of A: {E1, E2, E3}.

    e. Find the probability of event A. P(A) = P(E1) + P(E2) + P(E3) = 3/10 (chance that both days are above 40).

2. (Exercise 2.27, b, c) Assuming the same situation as above.

b. What is the probability of at least one day being above 40? P(A) = {E1, E2, E3, E5, E6, E7, E8, E9, E10} = 9/10

c. What is the probability that at most one day is above 40? P(A) = {E5, E6, E7, E8, E9, E10} = 6/10

3. (Exercise 2.36) There are seven days in a week. How many ways can these days be listed?

7! = 7*6*5*4*3*2*1 = 5040 possible combinations of seven days.

4. (Example 2.16) Three different days are chosen at random (X, Y, and Z). The following events are defined:

A: X has the highest maximum.
B: X has the second highest maximum.
C: Y has the highest maximum.

If these numbers are found randomly, is event A independent of events B and C?

Define the events such that E1 = {XYZ} E2 = {YXZ} E3 = {ZXY} E4 = {XZY} E5 = {YZX} E6 = {ZYX} in order from highest to lowest temperature.

A = {E1, E4}, B = {E2, E3}, C = {E2, E5}.

P(A) = 1/3, P(B) = 1/3. $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{3}}{\frac{1}{3}} = 1.$ $P(A|C) = 0.$

So, A and C are independent events, but A and B are dependent events.

5. (Exercise 2.110) March is divided into two groups of 10/21 days. In the first part, there is a 10% chance that the maximum is above 60. In the second part, there's a 43% chance. If a day is picked at random, what's the probability its max will be above 60?

E1 = in group 1, above: $\frac{10}{31} * \frac{1}{10} = \frac{1}{31}$ = 3.2%

E2 = in group 1, below: $\frac{9}{31}$ = 29%

E3 = in group 2, above: $\frac{21}{31} * \frac{43}{100} = 29.1\%$

E4 = in group 2, below: 38.6%

$A = E1 \cup E3 = 32.3\%$ (chance of the selected day having a max temperature above 60)

6. (Exercise 2.139) Refer to above. Let Y represent the number of days with a max above 60 after three days have been selected. Compute the possibilities for each Y.

P(0) = .677*.677*.677 = 31% (Chance that no days have a max temperature above 60)
P(1) = .677*.677*.323 = 14.8% (Chance that one day has a max temperature above 60)
P(2) = .677*.323*.323 = 7.1% (Chance that two days…)
P(3) = .323*.323*.323 = 3.4% (Chance that three days…)

Chapter 3- Discrete Random Variables and Their Probability Distributions

This chapter analyzes how discrete random variables can be utilized in statistical situations and introduces several probability distributions. This chapter represents the vast majority of the first half of this course.

1. (Example 3.1) Two days are picked at random from a week in March. There is a 2/7 probability of the day's minimum temperature being below 40. Let Y denote the number of days below 40, and find probability distribution

   Days can be selected $\binom{7}{2}$ = 21 ways.

   P(0) = $\frac{\binom{2}{0}\binom{5}{2}}{21} = \frac{10}{21}$ (chance of every day having a min temperature above 40)

   P(1) = $\frac{\binom{2}{1}\binom{5}{1}}{21} = \frac{10}{21}$ (chance of one day having a min temperature below 40)

   P(2) = $\frac{\binom{2}{2}\binom{5}{0}}{21} = \frac{1}{21}$ (chance of both days having a min temperature below 40)

2. (Example 3.2) Find the mean, variance, and the standard deviation of Y as shown above.
   a. $\mu = (0)\left(\frac{10}{21}\right) + (1)\left(\frac{10}{21}\right) + 2\left(\frac{1}{21}\right) = .571$
   b. $\sigma^2 = E((Y-\mu)^2) = (0-.571)^2\left(\frac{10}{21}\right) + (1-.571)^2\left(\frac{10}{21}\right) + (2-.571)^2\left(\frac{1}{21}\right) = .03$
   c. $\sigma = .173$

3. (Example 3.7) Out of 31 days, 3 have a maximum temperature above 70. If a sample of seven days are taken, find the probability that at least one has a temperature above 70.

$$P(at\ least\ one\ day\ above) = 1 - p(0) = 1 - \binom{7}{0}p^0q^7 = 1 - \frac{28^7}{31} = .510$$

   So, there is a 51% chance that all seven days have a max temperature below 70.

4. (Example 3.11) From above, $\frac{3}{31}$ days have a maximum temperature above 70. Find the probability that three days, selected at random, will not have a maximum temperature above 70.

$$P(\textit{three days below } 70) = P(Y \geq 3) = 1 - \left(\frac{3}{31}\right) - \left(\frac{28}{31}\right)\left(\frac{3}{31}\right) - \left(\frac{28}{31}\right)^2\left(\frac{3}{31}\right) = .737$$

There is about a 74% chance of selecting three days in a row that have max temperatures below 70.

5. (Example 3.17) From a month with 31 days, 7 are selected at random. If there are ten days with a max temperature above 60, what is the probability that less than three days are above 60?

   $N = 31, n = 7, r = 10. Y = 0, 1, 2.$

   $P(Y \leq 2) = p(0) + p(1) + p(2) = \frac{\binom{10}{0}\binom{21}{7}}{\binom{31}{7}} + \frac{\binom{10}{1}\binom{21}{6}}{\binom{31}{7}} + \frac{\binom{10}{2}\binom{21}{5}}{\binom{31}{7}} = .599$

   There is about a 60% chance of less than three days having temperatures above 60.

6. (Class notes) 3 times in 31 days, the highest temperature goes above 70 degrees. What is the probability that, in 31 days, there are 0 that are above 70?

   $\lambda = 3, y = 0, p(0) = \dfrac{3^0 e^{-3}}{0!} = e^{-3} = .050$

   There is a 5% chance of no days being hotter than 70 over the course of 31 days.

7. (Class notes) The average lowest temperature in March is between 23 and 47, with an average of around 35 degrees. What percentage of low temperatures fall between 23 and 47?

   Within number = 12, with a standard deviation of 6. So, k = 2. How many values are 2 units from 35?

   $1 - \dfrac{1}{k^2} = 1 - \dfrac{1}{4} = \dfrac{3}{4}$

   75% of low temperatures are between 23 and 47.

Works Citied

US Department of Commerce, NOAA. *Climate*, NOAA's National Weather Service, 3 Mar. 2022, www.weather.gov/wrh/Climate?wfo=phi.

Data sourced from chart generated on this website.

Waskerly, Dennis D., et al. *Mathematical Statistics with Applications*. Seventh ed., Thomson Brooks/Cole.

Problems adapted from example problems in chapters 2-5.