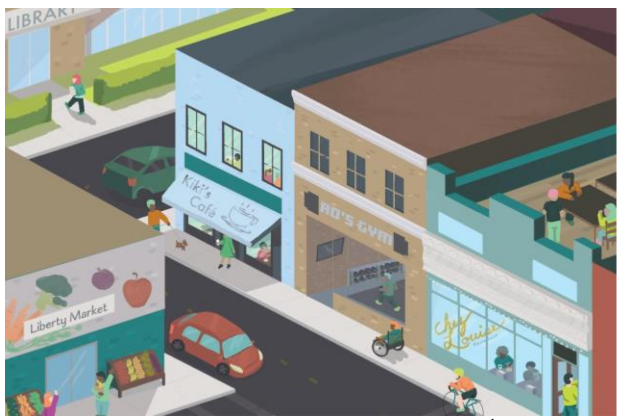
Neighborhood Walk

A citizen science project where undergraduates identify soft targets and crowded venues in their neighborhoods, assess their vulnerabilities, and develop plans to evacuate them in case of natural or manmade threats



Graphic from the American Survey Center[1]

A SENTRY Undergraduate Teaching Module

Prepared by

Donna Beers, Simmons University, donna.beers@simmons.edu
Clifton Morrow, Taylor Business Institute, clifton.morrow@tbiil.edu
Yachi Wanyan, Texas Southern University, yachi.wanyan@tsu.edu

---

[1] Cox, Daniel A. et al. (2021, October 20). *Public Places and Commercial Spaces: How Neighborhood Amenities Foster Trust and Connection in American Communities*. American Survey Center.

https://www.americansurveycenter.org/research/public-places-and-commercial-spaces-how-neighborhood-amenities-foster-trust-and-connection-in-american-communities/

Note to teachers: Teacher notes appear in red in the module, allowing faculty to pull these notes off the teacher version to create a student version of the module.

ACKNOWLEDGEMENT AND DISCLAIMER

This module was developed as a part of the Reconnect 2023 Workshop on Risk Assessment held in Providence, Rhode Island from June 18 – 23, 2023 organized by the SENTRY DHS Center of Excellence, and the Center of Discrete Mathematics and Theoretical Computer Science (DIMACS). The authors would like to thank Dr. Midge Cozzens for her valuable suggestions and continued support.

This module can be used only for educational purposes and cannot be reproduced and sold for commercial purposes without the authors' permission.

Summary of the Module:

The goal of this module is to introduce undergraduates across a variety of disciplines to the vital field of risk assessment. Drawing upon students' life experiences, this module challenges students to identify soft targets in their neighborhood, identify their vulnerabilities, and consider how to protect them from natural and manmade threats. Acting as citizen scientists, students will gather foot traffic data for a facility in their neighborhood. They will analyze and display the foot traffic data they gather in order to determine the time of day and day of the week when their facility has highest occupancy (i.e., greatest vulnerability), and determine an evacuation plan. Overall this module will equip students to advance the safety and security of their communities.

Target Audience: First- year undergraduates.

Prerequisites: Algebra I and elementary descriptive statistics.

Keywords: active learning, hands-on learning, tactile learning.

Learning Goals, Learning Objectives, and Learning Outcomes:

The learning goals, learning objectives, and learning outcomes for the Neighborhood Walk module are specified in Table 1 below. Please note that every exercise in the module is color coded with Bloom's taxonomy index label(s) and cross-referenced to its corresponding learning goal, objective, and outcome.

Anticipated Number of Class Periods: 3

Of special note:

This module includes supplemental online materials freely available in the cloud. Exercises are available in Kahoot-like interactive game style on Quizizz (free, open access, no account needed) and at MyOpenMath (free, secure access via the course ID: 202011 and enrollment key: 423132523). MyOpenMath offers classroom-like options including live polls, assignments and exams.
This module includes several class discussion exercises. Please find a discussion guide, a discussion evaluation form, and discussion rubrics (both classroom and online) made available here by the Searle Center for Advancing Teaching and Learning at Northwestern University.

This module provides optional exit surveys for students and instructors

Table 1. Neighborhood Walk Goals, Objectives, and Learning Outcomes

	Goal	Objective	Learning Outcome	Exercise #	Index (Based on Bloom's Taxonomy)
Module 1	G1 Promote neighborhood awareness, community service, and protecting civilian venues	O1.1 Promote neighborhood awareness. Students conduct a neighborhood walk	Student will be able to find the neighborhood walk score	1.1.1	Remember
		O1.2 Promote community service. Students learn citizen science terminology	Student will be able to define citizen science	1.2.1	Remember, Understand, Apply
			Student will be able to give at least one example of a citizen science project	1.2.2	Remember, Understand
			Student will be able to define disaster citizen science		
			Student will be able to give at least one example of a disaster citizen science project		
			Student will be able to give at least one model for organizing disaster citizen science projects		
			Student will be able to give at least one example of a model for organizing disaster citizen science projects		
		O1.3 Promote the protection of civilian venues. Students learn how to compare and contrast hard with soft targets	Student will be able to differetiate hard targets from soft targets	1.3.1	Remember, Understand
			Student will be able to give at least one example of a soft target in the neighborhood	1.3.2	Understand
		O1.4 Promote the protection of civilian venues. Students learn foot traffic terminology and methodology	Student will be able to define foot traffic	1.4.1	Remember
			Student will be able to give at least one method for measuring foot traffic		
			Student will be able to give at least one method for analyzing foot traffic		
			Student will be able to give at least one reason why foot traffic is important to retail		
			Student will be able to give at least one reason why foot traffic is important to public safety		
			Student will be able to use a bar chart to find the time of peak occupancy	1.4.2	Understand
			Student will be able to compare two bar charts and hypothesize why time of peak occupancy differs by day or date	1.4.3	Apply
			Student will be able to use Excel to compute measures of center (mean, median, mode) and measures of spread (range, variance, standard deviation) of a dataset	1.4.4	Apply
Module 2	G2 Introduce evacuation dynamics and descriptive statistics	O2.1 Introduce key elements in evacuation considerations such as human factors and building factors	Student will be able to discuss building factors and human factors which increase evacuation vulnerability	2.1.1 - 2.1.4	Remember, Understand, Analyze, Evaluate
			Student will be able to discuss the evacuation of a high-rise building	2.1.5	Remember
		O2.2 Introduce and review descriptive statistics	Student will be able to compute measures of central tendency	2.2.1	Apply
			Student will be able to justify the selection of a measure of central tendency to solve a problem	2.2.2	Understand
			Student will be able to discuss conflation	2.2.3	Apply
		O2.3 Introduce basic statistic and graphing functions of Excel as data analysis tools	Student will be able to construct the following types of data visualization graphs: scatter plot, bar chart, pie chart, histogram, box plot	2.2.3 - 2.2.6	Apply
			Student will be able to justify data analysis methods used in decision making senarios.	2.2.8	Analyze
Module 3	G3 Discuss evacuation dynamics, inferential statistics, and future projects	O3.1 Discuss evacuation dynamics through an evacuation drill	Student will be able to conduct an emergency evacuation drill	3.1.1	Remember, Apply, Analyze
			Student will be able to identify factors contributing to the success and/or failure of the emergency evacuation drill		
			Student will be able to state at least one of the FEMA evacuee special needs		
			Student will be able to design a multidimensional observational study of evacuation dynamics	3.1.2	Create
		O3.2 Introduce and review inferential statistics	Student will be able to construct a scatter plot	3.2.1	Apply
			Student will be able to recognize the least order polynomial which fits the plot	3.2.2 and 3.2.6	Remember
			Student will be able to describe a trendline and explain why it fits the scatter plot	3.2.3 and 3.2.7	Remember
			Student will be able to compare a computed R^2 value against the range of possible R^2 values	3.2.4 and 3.2.8	Analyze
			Student will be able to describe a residual plot	3.2.5 and 3.2.9	Remember
		O3.3 Future projects and discussions	Student will be able to point out evacuation challenges posed by venue geometry	3.3.1	Analyze

Index Legend

- Remember
- Understand
- Apply
- Analyze
- Evaluate
- Create

TOPIC ONE – BACKGROUND

This topic will take one class session.

Learning Goals for Topic One

- Students will learn the mission of the Department of Homeland Security (DHS) and the role of university DHS Centers of Excellence (COE) in helping DHS to (1) anticipate threats and challenges to homeland security, and (2) protect soft targets and crowded places (ST-CP).
- Students will learn what citizen science is and how they may participate in their communities to carry out citizen science projects that enhance public health and safety.
- Students will increase their situational awareness of the public venues in their neighborhoods.
- Students will gain understanding of a practical way they may use their knowledge of elementary statistics to assess the vulnerability of a specific, crowded public venue in their neighborhood.

Learning objectives for Topic One

- Students will be able to state the mission of the Department of Homeland Security (DHS) and describe the role of university DHS Centers of Excellence (COE) in helping DHS to (1) anticipate threats and challenges to homeland security, and (2) protect soft targets and crowded places (ST-CP).
- Students will be able to define the following terms: walk score ranking of a city, citizen science, disaster citizen science, model for organizing citizen science projects, soft target, hard target, foot traffic, ways to measure foot traffic.
- Students will be able to give examples of: citizen science projects, disaster citizen science projects, models for organizing citizen science projects, soft targets and hard targets, ways to measure foot traffic; they will also be able to explain how and why foot traffic is important to retail and public safety.
- Students will be able to calculate the walk score ranking for their neighborhood, identify a current or emerging public health or safety issue of concern to their neighborhood, identify what data might help to measure the severity of the issue, and describe how members of their neighborhood might collaborate to gather the needed data, and use tools such as Excel to analyze, represent and communicate statistical findings.

1.1 Introduction

Have you treated yourself lately to a walk around your neighborhood? Not only do neighborhood walks offer numerous health benefits, but they also can help us to discover (or re-discover!) attractions and services nearby that can enrich and enlarge our lives. In popular culture, the long-running children's educational series, *Mister Rogers' Neighborhood*, and the Tom Hanks' movie based on that series, *A Beautiful Day in the Neighborhood* (2019), celebrate the attractions, services, and exceptional people in our neighborhoods.

One of the most walkable neighborhoods in the U.S. is Washington, D.C. The online app, Walk Score®, gives our nation's capital high scores (rankings are out of 100) for walkability, transit, and bike:

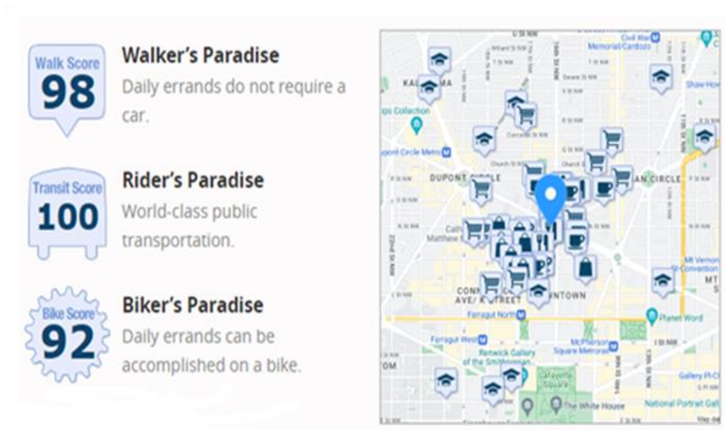


Figure 1. Walk Score of Washington D.C. (Graph accessed Jan 08, 2024)

How is Walk Score computed? The following graphic provides details. (Click on "Methodology" for more details)

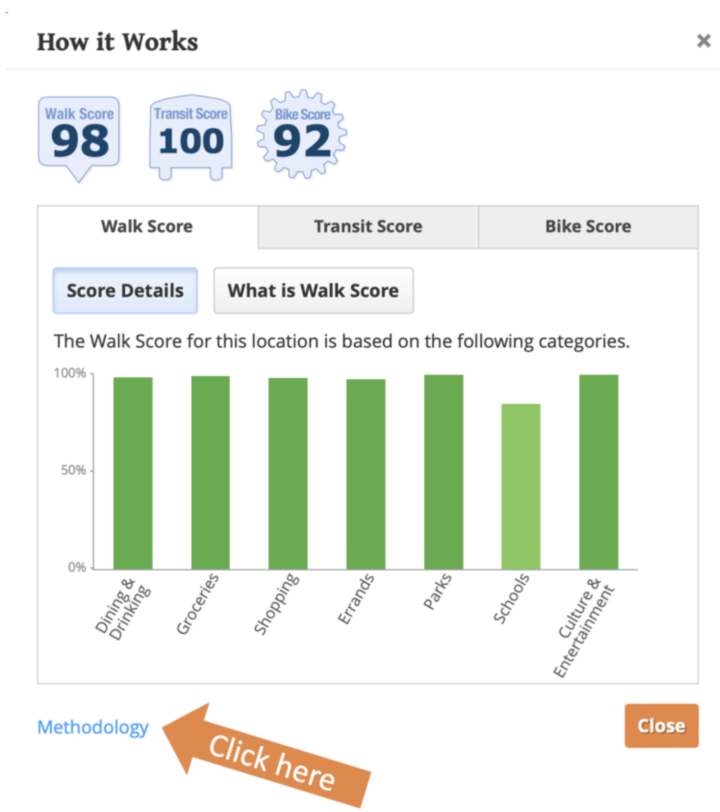


Figure 2. Walk Score Details (Graph accessed Jan 08, 2024)

How walkable is *your* neighborhood? What are its attractions and service?

Exercise 1.1.1 For your neighborhood,

a. Use the Walk Score® widget to find its Walk Score ranking
   - Go to the Walk Score website where you register for the widget, https://www.walkscore.com/professional/walk-score-widget.php
   - Sign up for the widget (with Facebook or Google)
   - Fill in the slot labeled "Get scores for your address"
   - Click on the button About Your Score
b. Give the bar graph from Walk Score that shows the categories for attractions and services in your neighborhood and their rankings on a scale from 0 to 100 percent.

Answer

The answer will vary for each neighborhood. We encourage the teacher to use the Walk Score® widget to prepare an answer key for your neighborhood.

1.2 Community Service: Citizen Science and Disaster Citizen Science

The need to belong is recognized as one of the most basic human needs, and community service can build our sense of belonging to our community and contribute to its vitality. Today students - and people of all ages - volunteer with neighborhood clean-ups, animal shelters, collecting food for food pantries, charity runs, and a host of other activities.

Another way that students are contributing to the well-being of their neighborhoods is through citizen science. What is citizen science?

Exercise 1.2.1 Read the blog, *Let's learn about citizen science*. Based on your reading:

a. What is citizen science?
b. What are examples of citizen science projects?
c. Identify a current or emerging public health or safety issue of concern to your neighborhood, identify what data might help to measure the severity of the issue, and describe how members of your neighborhood might collaborate to gather the needed data.

Answer

a. Citizen science occurs when "Volunteers gather and analyze data for scientific studies…"
b. Examples of citizen science projects include:
   "The Globe at Night project, for instance, crowd-sources data on light pollution."
   "NoiseTube is a similar effort to map noise pollution."
   "Hikers, birders and other nature lovers share wildlife photos on iNaturalist."
c. Answers will vary according to students' locations. The teacher can treat this as an open question, then organize the students to make an opinion poll or survey to gather and tabulate answers.

Citizen science dates back to 1880 when an ornithologist, Wells Woodbridge Cooke, began recording the arrival dates of migratory birds in the Mississippi Valley. Since its inception, citizen science has focused on the collection and analysis of biological and ecological data to advance scientific understanding of the environment, climate, and space.

In recent years, the occurrences of hurricanes, tornados, floods, and public health crises have spurred the growth of another area of citizen science, **disaster citizen science**, defined in the publication, Citizen Science for Disasters - A Guide for Community Groups (Rand Corporation, 2021) as follows**:**

> "Disaster citizen science is the involvement of members of the public in scientific activities relating to disaster preparedness, response, or recovery (e.g., study design or collection or analysis of data). Individuals can pursue citizen science in collaboration with professional scientists or independently. …" [page 9]
>
> "In Flint, Michigan, community members partnered with a university and collected drinking water samples that documented high levels of lead in the water. … " [page 13]
>
> "In Tonawanda, New York, community members organized to collect air samples that demonstrated harmful pollution levels from a nearby factory." [page 13]

Exercise 1.2.2 Read pp. 10-12 of *Citizen Science for Disasters A Guide for Community Groups*.

a. What are three models for organizing disaster citizen science projects
b. Give one example of each of the three models you listed in part (a).

Answer

a. Contributory, Collaborative, and Collegial
b. Contributory: "The U.S. Geological Survey's Did You Feel It? program collects reports on earthquake location, perceived intensity, and damage from people around the world, which are used to create maps of earthquake intensity. Scientists also use the data to study such topics as the extent to which people feel earthquakes."

   Collaborative: "In response to an emerging influenza (H1N109) pandemic, a participatory action research approach, in which communities and scientists collaboratively work together, was carried out in Australia to understand barriers to interventions in Aboriginal and Torres Strait Islander communities. The project also aimed to develop culturally appropriate and effective strategies to reduce influenza risk. To carry out the project, community members worked with researchers to identify the research problems, plan study approaches, prepare data collection instruments, and collect and analyze data. The project identified barriers that would limit the effectiveness of general Australian influenza containment policies but also elicited several strategies for reducing disease spread in the Indigenous communities.

   Collegial: "The Louisiana Bucket Brigade, a nonprofit environmental health and justice organization, addresses petrochemical pollution in the state. One aspect of Louisiana Bucket Brigade's work is to engage residents to collect air samples in their neighborhoods using an inexpensive, Environmental Protection Agency–approved "bucket." The bucket is an easy-to-use air-sampling device. Residents then send these air samples to a lab for analysis. The Louisiana Bucket Brigade also facilitates crowdsourcing of pollution via its online "iWitness Pollution Map," which was set up with the help of academic partners. Although the organization makes use of professional and scientific assistance, its staff and volunteers, who are community members, lead its scientific efforts."

1.3 Disaster Citizen Science: Protecting the Public Venues in our Neighborhoods

Collecting and analyzing samples of drinking water, air, noise, and light are vital activities in disaster citizen science that enhance the safety and well-being of our neighborhoods.

This module will focus on another aspect of disaster citizen science: identifying the soft targets in our neighborhoods, assessing their vulnerabilities, and developing plans to protect them from natural and manmade threats.

"The Department of Homeland Security has a vital mission: to secure the nation from the many threats we face." as stated at https://www.dhs.gov/about-dhs (accessed Jan 08, 2024)

Exercise 1.3.1 Read the web page, https://www.dhs.gov/about-dhs

a. State the DHS mission.
b. When was DHS established.
c. How does DHS fit into the federal government?

Answer

a. "The Department of Homeland Security has a vital mission: to secure the nation from the many threats we face."
b. "The Department of Homeland Security was established in 2002, combining 22 different federal departments and agencies into a unified, integrated Cabinet agency."
c. "The Cabinet is an advisory body made up of the heads of the 15 executive departments. Appointed by the President and confirmed by the Senate, the members of the Cabinet are often the President's closest confidants." https://www.whitehouse.gov/about-the-white-house/our-government/the-executive-branch/

In the next exercise, you will learn how the Department of Homeland Security (DHS) distinguishes between the soft and hard targets among the POI's in our neighborhoods.

Exercise 1.3.2 Read the article, *Northeastern University wins $36M federal contract to build a system to protect schools, churches, and stadiums* (Peter Ramjug, Northeastern Global News, Nov 4, 2021) to answer these questions:

a. What is the difference between a soft target and a hard target?
b. Give six different examples of soft targets in your neighborhood.
c. Among the popular soft targets in your neighborhood, which ones are frequented by college students?

Answer

a. "Soft targets are usually civilian sites such as restaurants and train stations where large numbers of unarmed people gather, and could be vulnerable to an active shooter or some other type of attack. A hard target, by contrast, such as a military base, is usually heavily defended and not open to the public." Here are some differences:
   - soft targets host many unarmed people, hard targets are heavily defended.
   - soft targets are open to the public, hard targets are close to the public.
b. The answer can vary by neighborhood, here are some common soft targets:
   - boutique specialty retail stores. For example, Sephora often has high foot traffic.
   - cell phone stores

- churches
- department stores
- fitness club
- hotels
- pharmacy stores
- restaurants
- schools
- train stations

c. The teacher can treat this as an open question, then organize the students to make an opinion poll or survey to gather and tabulate answers.

Current security and coordination efforts of DHS focus on "Soft Targets and Crowded Places", terminology that is explained as follows:

> "Soft Targets and Crowded Places (ST-CPs) are locations that are easily accessible to large numbers of people and that have limited security or protective measures in place making them vulnerable to attack. ST-CPs can include, but are not limited to, schools, sports venues, transportation systems or hubs, shopping venues, bars and restaurants, hotels, places of worship, tourist attractions, theaters, and civic spaces." (Source: U.S. Department of Homeland Security Soft Targets and Crowded Places Security Plan Overview May 2018)

Considering "crowded places" leads us next to the concept of foot traffic.

1.4 The Basics of Foot Traffic

Whether our goal is to find the most popular restaurant or fitness center or some other Point of Interest (POI) in our neighborhood, it all comes down to counting foot traffic.

Exercise 1.4.1 Read the online Safegraph article, Foot Traffic Data: Calculations, Accurate POIs, & Where to Get It to answer the following questions.

a. What is foot traffic?
b. What are methods for measuring foot traffic? How do we analyze it
c. How is foot traffic important to retail? To public safety?

Answer

a. Foot traffic data associates people's movements with physical places.
b. Methods for measuring foot traffic include:
    - counting WiFi connections
    - manually using a clicker counter every time someone enters a place.
    - mobile device location tracking
    - using tools such as motion sensors, pressure mats or video cameras

    Methods for analyzing foot traffic differ by use case. If you are choosing a new location, you could compare the foot traffic of the new location against an alternative location.
c. Foot traffic is important to retail because it:
    - helps you decide where to open or close a store.
    - helps you learn about your customers.
    - tells you how many people visit your store without buying.

- helps the government decide whether to build a walkway, a sidewalk, or a paved shoulder.

Exercise 1.4.2 Explorations with Google Map's Popular Time function.

Figure 3 shows historical foot traffic graph for the International Spy Museum in Washington, D.C. on Fridays. This graph was obtained by using the Popular Time function of Google Map. Also, Figure 4 shows the Museum's historical foot traffic data for Sundays.

(a) What is the Museum's peak time on Fridays? On Sundays?
(b) What factor(s) may explain the different peak times for Friday and Sunday?

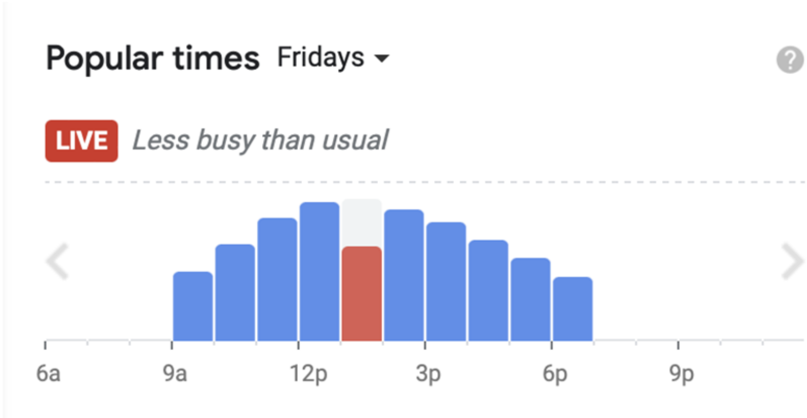


Figure 3. Google Live Data for a venue (Museum)

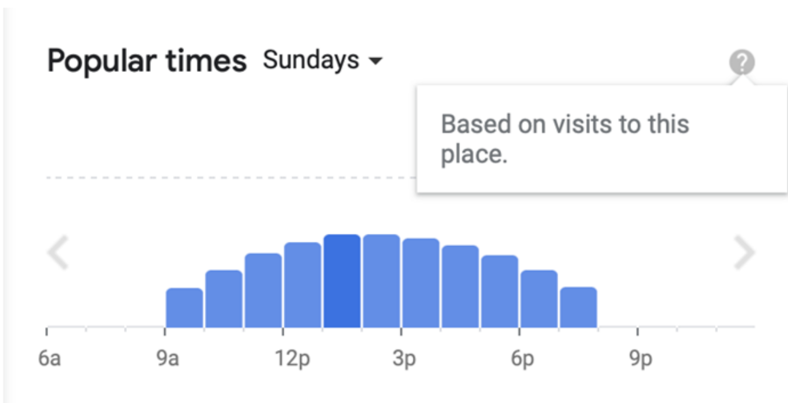


Figure 4. Google Historical Data for the same venue (Museum) (as in Figure 3)

Answer

(a) Peak time on Friday is noon, peak time on Sunday is 2p.
(b) Friday visits might occur during a workday lunch, Sunday visits during an afternoon outing (between meals)

Exercise 1.4.3 Pick a popular restaurant in your neighborhood and use the Popular Times feature of Google to explore how busy it is on a Friday night. What time should you make reservations to avoid a long wait?

Answer

The answer can vary by neighborhood. The teacher should ask the students to recognize times of low occupancy and select one of them.

Exercise 1.4.4 Your instructor will provide you with foot traffic data (*T2_Dataset_Students.xlsx*). For this data, use Excel to compute its measures of center (mean, median, mode) and measures of spread (range, variance, standard deviation).

Calculated results below are given to one decimal place of accuracy. In practice, we round the output to correspond to the accuracy of the input. Please note that we obtain mode and range from direct observations, not calculation. Since we observe people, mode and range are integers.

Statistical Parameters calculated based on combined data from Classes one and two	
Mean	37.4
Median	40.5
Mode	51
Range	69
Variance	345.8
Standard Deviation	18.6

1.5 The Foot Traffic Project

We are concerned with emergency evacuation of a facility. When emergency evacuation is necessary, people in the facility are vulnerable if they fail to evacuate quickly enough. This definition is inspired by the lexicon here.

Over the next week, you will carry out a Citizen Science project. In this project you will discuss factors which affect vulnerability: the geometry of the venue, obstacles in the venue, occupancy, special needs of occupants. One of the easiest factors to measure is occupancy: the more people inside the facility, the more difficult it is to safely evacuate everyone. For a public venue that will be identified by your instructor, you are going to partner with a classmate to:

- Gather foot traffic data (headcount data) for the venue; you and your partner will collect data for a 30-minute period on a day and time that will be assigned by your instructor.
- Consolidate your data with the data collected by the other teams to determine the day of week and time of day of highest occupancy for the venue.
- Identify the specific day and time of day, based on all the team data, to add more customer service to the venue in order to avoid bottlenecks and, equally important, identify when the occupants of the venue are most vulnerable, due to peak crowds.

At our next class, you will

- Learn more about the vulnerability of occupants in the venue here.

- Develop a better understanding of risk reduction planning.

In this project-based assignment, students will have the opportunity to become citizen scientists as they explore the significance of foot traffic data in emergency evacuation planning. After assigning the task, give students 7-10 days (depending on class schedule) for data collection (the neighborhood walk).

Assignment: The instructor chooses one venue, from the range of venues in students' neighborhoods, and assigns student pairs to gather foot traffic data for that venue on a particular day of the week and time of the day, with the goal of identifying the busiest periods and vulnerabilities of the selected venue.

Venue Suggestions:

- School Parking Garage Data. For this option, one student in each pair counts the inflow and the other counts the outflow for a period of at least 30-60 minutes. Students will also need to acquire total parking spaces (number of spots per floor multiply by the number of parking levels). By monitoring the number of vehicles entering and exiting the garage at different times throughout the week, they can gain insights into the peak hours of usage and potential congestion points.
- Movie Theater: In addition to physically counting the foot traffic, students are encouraged to use other available information to validate their data. For example, use an online ticketing platform to collect data on movie ticket purchases. By tracking ticket sales for various showtimes on different days, they can gather valuable foot traffic patterns in the digital realm and identify the most popular screening times.
- Retail stores such as supermarkets. The following session examples are given with a local Apple Store. Most Apple retail stores have a single entrance/exit for customers. By observing the number of customers visiting the store on different days and during various hours, students can gain valuable insights into the store's busiest periods, the popularity of product launches, and the flow of foot traffic in a retail environment. Similarly, Starbucks can be another single entrance/exit venue choice.

In addition to the aforementioned venues, students can explore data collection in other public spaces, such as shopping malls (multiple entrances and exits), stadiums (multiple entrances and exits, but fixed maximum capacity), libraries, or recreational parks (may have very high variabilities due to events, etc). These locations offer valuable foot traffic data. However, it is important to remind students about potential security and privacy concerns when conducting a school project involving data collection, especially associated with certain private establishments.

Analyzing Foot Traffic Data:

Upon gathering the foot traffic data, student teams will use Excel to input and display their dataset, running descriptive statistics to uncover trends and patterns. By creating different types of graphs, they can visualize the busiest times of the venue and determine potential vulnerabilities in emergency situations. Through this exercise, students will learn valuable skills in data analysis, understand the relevance of citizen science in real-world scenarios, and grasp the critical role of foot traffic data in emergency evacuation planning.

Advanced Cohorts: Multivariable from the Start

This module targets beginning statistics, and focuses on foot traffic, with possible extensions to other variables. If you have an advanced cohort, then please operate in a multivariable framework from the start. Encourage your students to walk mindfully, noticing venue geometry, evacuation obstacles, and special needs occupants. An advanced cohort could also gather data on these variables.

Topic Two - How to Interpret and Understand Our Data

This topic will take 90 - 120 minutes.

There are three supplementary materials for this session:

1) T2_Dataset_Teachers.xlsx (Topic 2 teachers' version with worked out demo and graphs)
2) T2_Dataset_Students.xlsx (Topic 2 student version that contains data only)
3) T2_PPT_Teachers.pptx (Supplementary Topic 2 PowerPoint slides for instruction)

We are going to use TWO fictitious foot traffic datasets from a local Apple store in the Houston area for hands-on exercises in this section. Please distribute "*T2_Dataset_Students.xlsx*" to students before class and make this file as well as Excel software available for all students during class. Students should have access to computers with Microsoft Excel installed.

All full-time students have Microsoft Office Package available through school Microsoft 360 subscription. Please contact IT for questions related to how to download and install Excel on a personal computer. The teacher's version (*T2_Dataset_Teachers.xlsx*) has a complete, worked out demo with an optional, easy-to-use data generator to provide more sample files.

Two fictitious groups of students (30 students in each class) during the week of 6/26/23 to 7/2/23 collected the sample data. Store hours are from 9:00 am to 8:00 pm. Once students understand the basics of Excel as a useful tool for data analysis, the neighborhood data collected by students can be compiled and analyzed in a similar fashion.

2.1 Vulnerability and Evacuation

Gathering places such as stadiums, convention centers, train stations, and movie theaters with a large crowds have potential risks of large casualties. Protecting the lives of everyone within a facility is the first priority in the event of an emergency. One common means of protection is evacuation. When we analyze vulnerability and evaluate evacuation plans, many factors come into play. For example, how big (square footage) is the place? How many exits are there? And what is the density of the crowd (count per unit area)? The key to a successful evacuation plan lies in its efficiency (i.e., the time it takes to move the crowd to safety).

Exercise 2.1.1 The graph below (Figure 5) shows two buildings with the same area and built for the same purpose. For the same headcounts, which one has a higher risk for occupants in case of an emergency? Why?

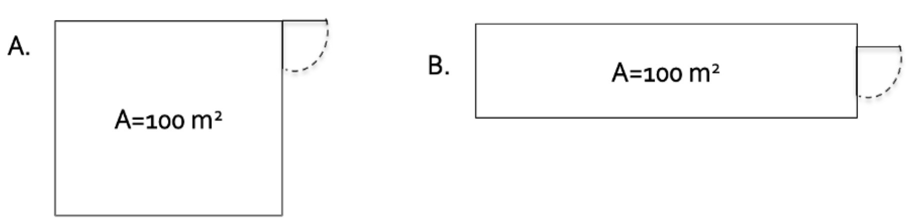


Figure 5. Comparison of Risk (cases A and B)

Answer:

People in the narrower building will have a disadvantage when they evacuate. The following study shows that people prefer wide evacuation corridors.

Exercise 2.1.2 The graph below (Figure 6) shows four buildings with the same area and built for the same purpose. For the same headcounts, which one has the highest risk for occupants in case of an emergency? Why?

Answer:

When A and B are compared, the building in case B with a narrower layout is more vulnerable. However, as you can observe here, when more variables come into play, the comparison becomes a lot more complex.

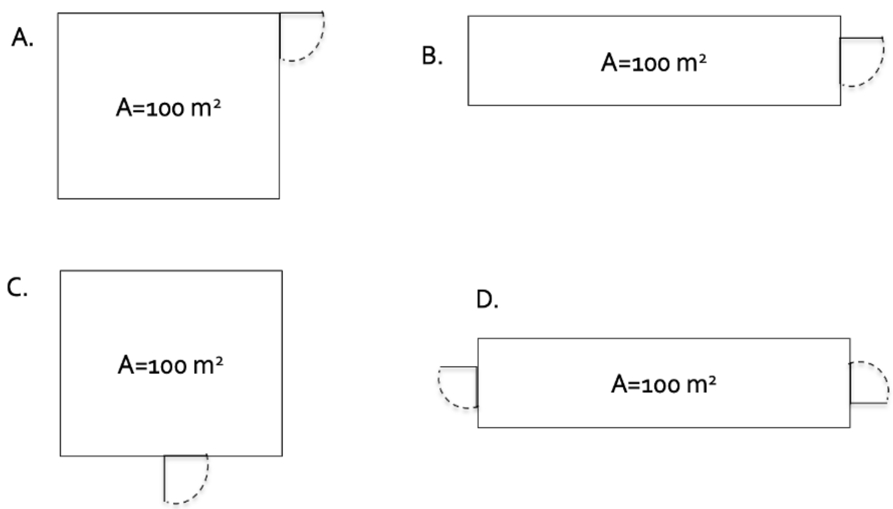


Figure 6. Comparison of Risk (cases A, B, C, D)

Exercise 2.1.3 Discussion: In order for us to evaluate the vulnerability of a building and come up with an efficient evacuation plan, there are many factors that contribute to the risk and evacuation efficiency of a building. Please name at least one factor.

Answer:

Here are a few examples:

- Density of the crowd (headcounts, foot traffic, etc.)
- Number of exits
- The maximum, minimum and average distances to exit(s)
- Response time
- Total Evacuation Time
- Crowd moving behavior
- (more…Refer to ppt page 9 to 11 for supplementary materials)

Exercise 2.1.4 Read the article, [Supporting People with Vulnerabilities Framework](). Based on your reading, answer the following questions:

(a) List at least one factor that enables a person to prepare for and adapt to a disaster.
(b) List at least two characteristics that could reduce a person's ability to prepare for and adapt to a disaster.

Answer:

(a) Strong family connections; Strong social connections.
(b) Diminished sensory awareness; Impaired mobility; Economic limitations; Multiple chronic health conditions; Social limitations.

Exercise 2.1.5 Read the article, [FAQs about Building Evacuations](). Based on your reading, answer the following questions:

(a) List at least one element of emergency preparedness for high-rise building evacuation
(b) In a high-rise building emergency, which occupants should move, and where should they go?
(c) In a high-rise building emergency, should occupants move to the roof and wait for a helicopter rescue?
(d) List at least one action you should take if you are trapped in a high-rise building emergency.

Answer:

(a) Adequate means of egress (exit routes); Early warning (typically through an alarm or voice communication system); Occupant familiarity with the plan through knowledge and practice.
(b) The occupants of the emergency floor and floors immediately above and below it should immediately use the exit stairs to descend to a floor level that is at least a few floors below the emergency floor. The occupants can then reenter the occupied space on those safe floors to await further instructions.
(c) No, a helicopter roof rescue is an extraordinarily dangerous procedure for the occupants, the pilots, and firefighters who may be in the building.
(d) Listed from high preference to low preference:

- Position yourself in an area where you can close the door and seal the cracks to keep smoke out.
- Use a telephone to call the fire department and report your exact location in the building.
- Be patient.
- Signal your position to rescue personnel from a window using a light-colored cloth.
- Open the window slightly to allow fresh air in.
- Do not break a window.

2.2 Using Excel to Interpret Data

The neighborhood walk data that the student teams have collected can be presented in many ways, for example tables, graphs and charts, infographics, maps, dashboards, written reports, and summaries, to name a few. Which format(s) to use depends on the nature of the data, the purpose of the analysis, and the target audience. **The goal, above all, is to effectively communicate statistical information.**

As we go through the following hands-on activities, instruct students to double-click and open the distributed "*T2_Dataset_Students.xlsx*" workbook. Upon launching Excel, students are greeted with pre-entered data as shown in Figure 7.

A dataset can be described by its center, variability, and shape, and/or by its frequency, percentage or proportion in each category. When you open the Excel workbook, you should see something similar to the screenshot below (Figure 7):

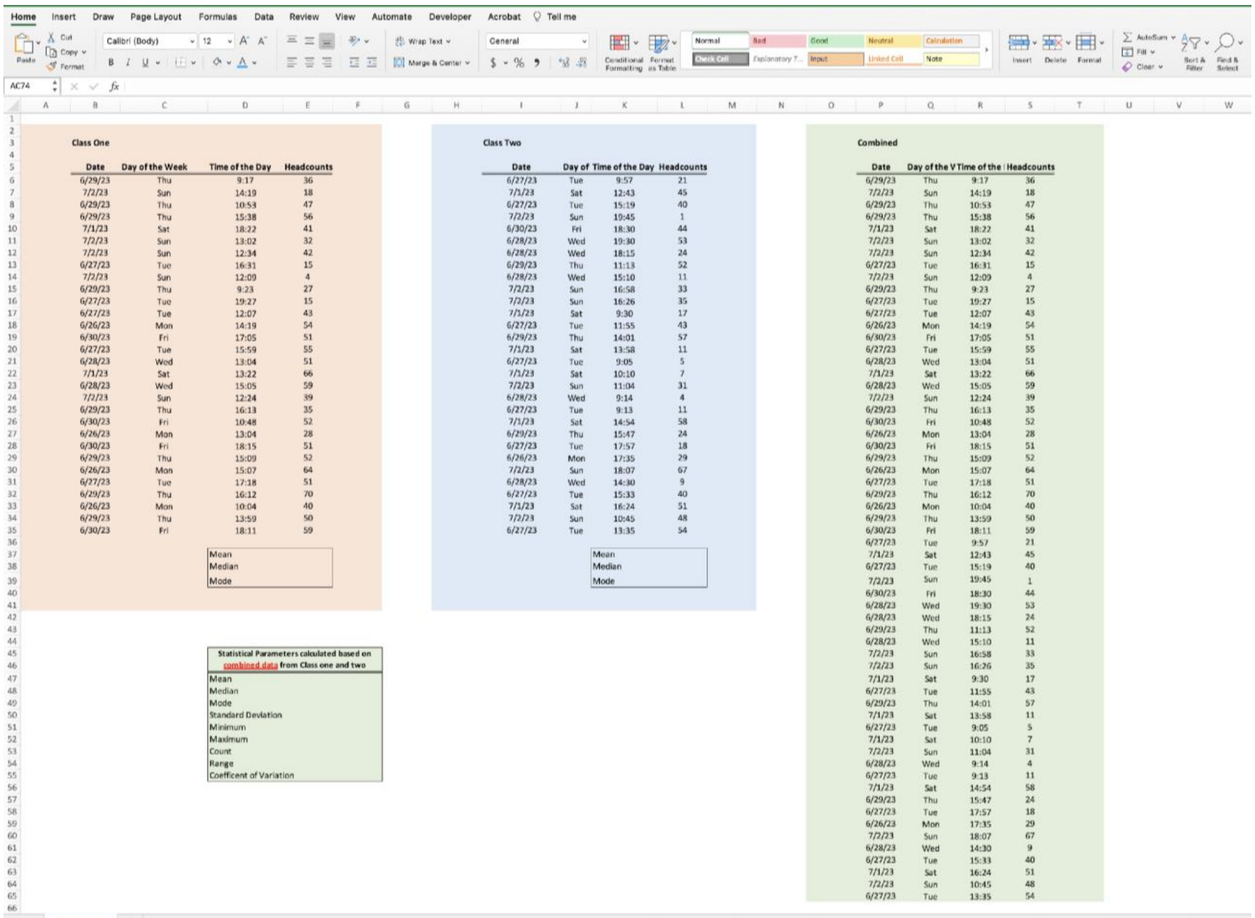


Figure 7. T2_Dataset_Students.xlsx Start Page

Exercise 2.2.1 Using the "*T2_Dataset_Students.xlsx*", carry out statistical calculations (If students are not familiar with Excel, find tutorials on functions AVERAGE, MEDIAN and MODE.). What is the mean, median and mode for Class One (rose pink)? Class Two (baby blue)? And the combined data (light green)?

Answers:

Mean	43.4
Median	48.5
Mode	51

Mean	31.4
Median	32
Mode	11

Mean	37.4
Median	40.5
Mode	51

Exercise 2.2.2 Assume you are the store manager and want to estimate the expected daily number of customers in the store.

(a) Which number would you use? Why?
(b) Discussion: How well do you think a central value characterizes the whole set of data? When do we use mean, median or mode? Why?

Answer:

(a) Mean is the most commonly used measure of central tendency but can be corrupted by outliers. Median is resistant to outlier corruption. Mode occurs the most often. The store manager needs to select the appropriate measure for this location.
(b) Students may notice the values are quite different from different datasets, extend students' thinking by "*T2_PPT_Teachers.pptx*" page 21. The following exercises will guide students to more ways to present and interpret data.

Exercise 2.2.3 For the data, construct scatter plots for Class One, and Class Two respectively.

(a) Try to combine the two scatter plots you have constructed into one graph. Sample graph provided in Figure 8. (Hint: Search how to add data series to an existing graph.)

(b) Look at the scatter plot with two datasets from class one and class two, can we combine data from these two classes? Why or why not?

(a) Construct a scatter plot using the combined data (in light green).

Answer:

(a) Microsoft has this tutorial explaining how to do this.
(b) Yes, we can. They are independent. Once combined, we will have a larger dataset that can improve our statistical analysis. Please enjoy this discussion of conflation here: https://doi.org/10.1063/1.3593373
(c) This is the same plot as in part (a), where we have plotted two separate data series in one graph.

Depending on your specific need, the same dataset can be viewed from many different perspectives. Frequency distribution shows either the actual number of observations falling in each range or the percentage of observations.

Exercise 2.2.4 Let's organize our data chronologically into three ranges: morning (from store opening to noon), afternoon (noon to 6:00 pm) and evening (6:00 pm to store close). Find the following values for morning, afternoon and evening data respectively: Data Count, Mean, Median, Mode, Frequency, Proportion, and Percentage.

Answer:

"*T2_PPT_Teachers.pptx*" pages 23-28 demonstrate how to use Sort in Excel; use page 28 to check answers.

Exercise 2.2.5 Plot pie chart and bar chart using data from Exercise 2.2.3. Compare with scatter plots obtained from Exercise 2.2.2. Discussion: what is each type of visual representation suited for?

Answer:

Please find a sample pie chart and bar chart available at *T2_Dataset_Teachers.xlsx* -> Categorical-Time of Day Demo sheet.

A pie chart provides a clear visual representation of a small number of categories, approximately two to four. A bar chart performs better than a pie chart for five or more categories. Further discussion available at [OpenIntro Statistics] page 66.

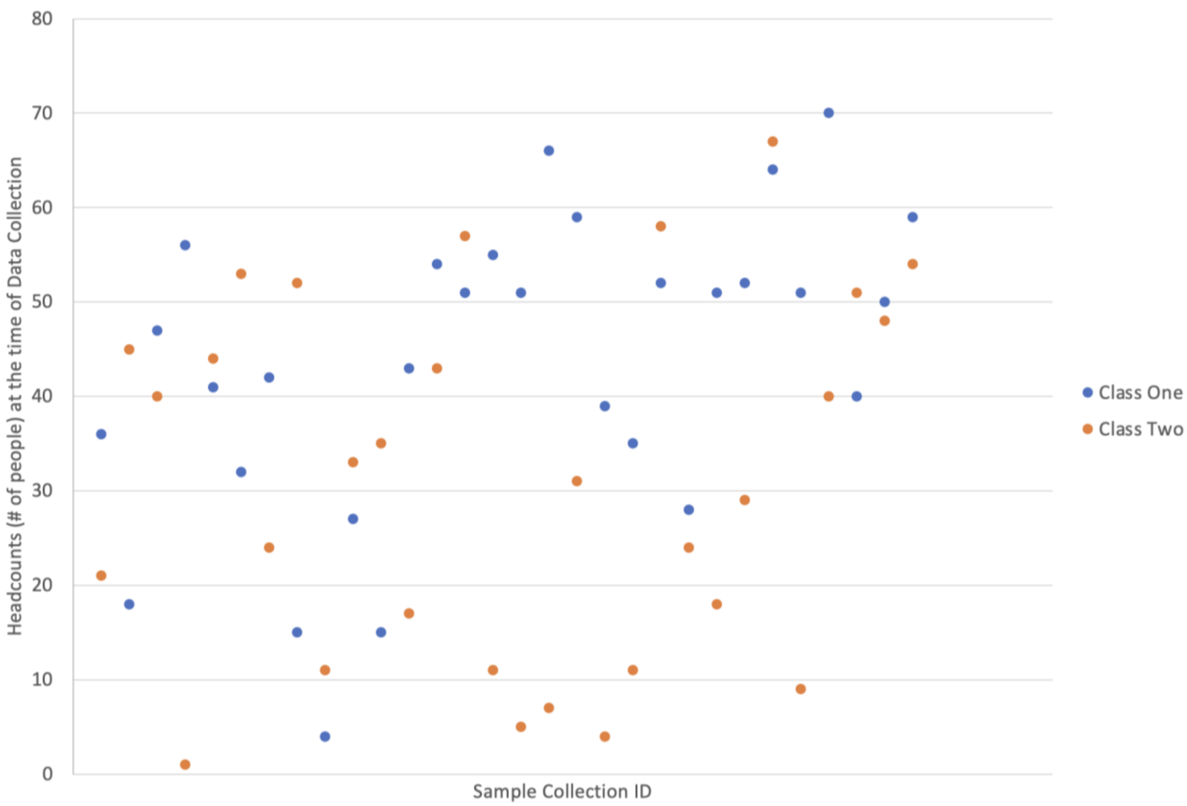


Figure 8. Scatter Plot of Apple Store Occupancy

Exercise 2.2.6 Let's reorganize headcounts into intervals of bin width 10 (0-10, 11-20, 21-30…). Range, variance, and standard deviation measure the amount of variability or dispersion in a dataset. Calculate frequency and cumulative frequency in tables and plot the corresponding histogram. Optionally repeat for different bin widths.

Answer:

Figure 9a shows a sample histogram with bin width 10, and Figure 9b shows a sample plot with a bin width of 5. Use "*T2_PPT_Teachers.pptx*" pages 36 and/or 45 to introduce the formulas for these measures along with simple examples.

Exercise 2.2.7 Use Excel to plot a box and whisker graph for each of the following:

(a) Class one headcount data
(b) Class two headcount data
(c) Combined headcount data

Answer:

Guide students to construct box and whiskers graphs and add them to the same graph for comparison. Refer to "*T2_PPT_Teachers.pptx*" pages 43-44. Optional: extend students' thinking by leading a discussion on good sampling and testing techniques that can minimize data collection variabilities.

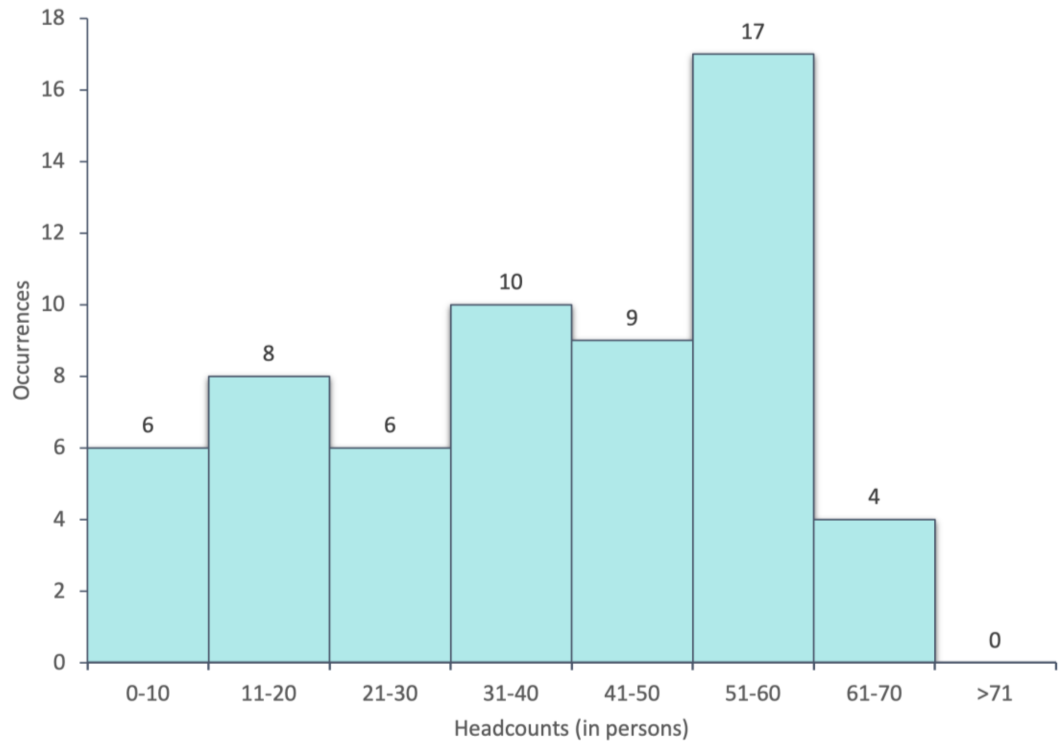


Figure 9a. Histogram of Apple Store Occupancy by intervals of 10

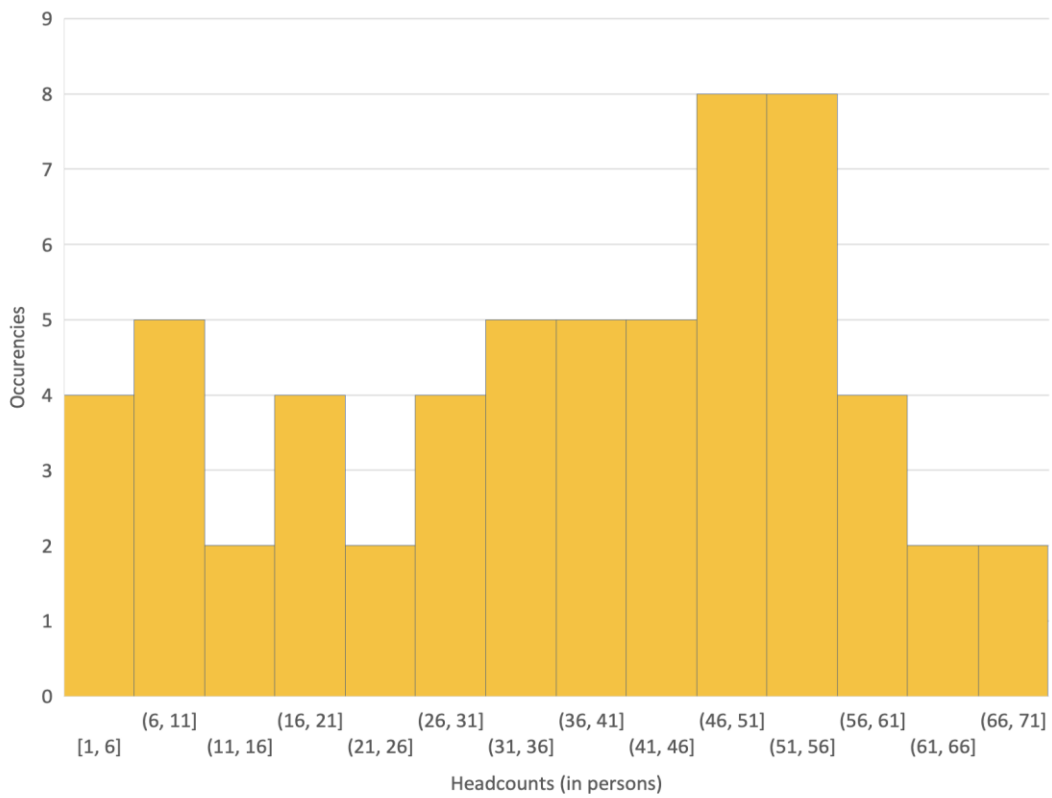


Figure 9b. Histogram of Apple Store Occupancy by interval of 5

Exercise 2.2.8 For the same set of data, there are different ways to present it, numerically or graphically. Let's consider the following scenarios. For each case, please select one numerical method (mean, median, mode) and one graphical method (scatter plot, box plot, histogram) that you think best suits the scenario.

(a) You, as the store manager, want to add manpower for the busiest day.
(b) You, as the chief security officer, want to be prepared for the coming holiday season.
(c) You, as an engineer, want to evaluate the fire safety of the store and evacuation plans.

Answer:

(a) For example, you might want to add one or two more employees for the busiest day of the week. We prefer median due to its resistance to outliers. A histogram can show the busiest day at a glance.
(b) In this case you are more concerned about the ratio of workers to visitors. We prefer mode due to its measure of popularity. A box and whisker plot can show the times of large visitor traffic.
(c) In this case you are more concerned about the projected maximum occupancy. We prefer mean in order to capture the outliers. A scatter plot can show the time of peak traffic.

Topic One discussed citizen science and traffic counts. In Topic Two we used descriptive statistics to summarize our data. In Topic Three we will use inferential statistics to find the time of peak traffic.

TOPIC THREE – EVACUATION AND PREDICTION

Topic Three consists of three independent sections. Please feel free to use one or more of these sections in your class. Each section contains its own references. The exception is the Banaszak study.com reference, used in both 3.2 and 3.3 but stated only in 3.2.

3.1. Classroom Drill

Section 3.1 should take between 35 and 90 minutes, depending on your emergency preparedness (Do you already have a safe destination selected?), evacuation drill mechanics (How many students need to travel how far?), and depth of class discussion.

The goal of this section is to conduct an emergency evacuation drill of your classroom.

Exercise 3.1.1 Class Fire Drill

- Safe Location: Select a destination for the evacuation drill.
- Team Members: Select classroom members to serve as marshal(s), timekeeper(s), and warden. Marshal(s): guide the classroom members to a safe location. Timekeeper: record how long it takes to evacuate the classroom to safety. Warden: verify that all classroom members reach safe location, notifies Timekeeper that drill is over.
- Perform the Drill
- Discuss the Drill: Did any obstacles slow the evacuation time?
    - Facility obstacles: Did the classroom physical plant contain any obstacles which slowed the evacuation time?
    - Organization obstacles: Did poor planning and/or preparation slow the evacuation time?
    - Voluntary obstacles: Did any participants wear restrictive clothing, or send one last text message, or perform other voluntary activities which slowed the evacuation time?
    - Special needs: Did any participants have slow evacuation time due to involuntary conditions? For background, here is the official federal definition of special needs that is stated in the FEMA Evacuee Support Guide:

  "An evacuating population will likely include individuals who need additional or specialized assistance, such as people who: have disabilities; live in institutionalized settings; are fragile elderly; are children; are from diverse cultures and have limited English proficiency; or need transportation assistance. The Federal definition of Special Needs includes five functional needs: communication, medical care, maintaining independence, supervision, and transportation." *Evacuee Support Planning Guide*, FEMA P-760/Catalog No. 09049-2 July 2009, page 26

Answer:
Please employ this fire drill rubric and here is a resource for Exercise 3.1.1. Alert Media:
https://www.alertmedia.com/blog/how-to-conduct-a-fire-drill-at-work/ (Accessed Jan 08, 2024)

Exercise 3.1.2 Use the classroom data and the discussion from Exercise 3.1.1 to modify the external facility foot traffic count exercise in Section 1.5.

Answer:

The point here is for students to consider how the approach they took in Section 1.5 must be modified to reflect the four discussion items.

3.1.1. References

- Disaster Preparedness for Retail Stores (Accessed Jan 08, 2024) is a brief HTML overview with a link to a detailed PDF.
- Emergency Action Plan Check List (OSHA)

3.2. Foot Traffic Count Interpolation

This section (all of 3.2) should take between 70 and 160 minutes, depending on the student background in algebra, Excel, and statistics. This section uses two supplementary excel files named "*T3_Dataset_1.xlsx*" and "*T3_Dataset_2.xlsx*".

The goal of this section is to fit a curve to the foot traffic count data gathered during the previous sections. We can use this curve to predict traffic count values between the data points and find the peak foot traffic.

3.2.1. Model, Data, Population

Section 3.2.1 should take between 3 and 6 minutes, depending on the student background in statistics. Please find resources referred to and/or cited in section 3.2.4. This section discusses the following concepts: curve fitting, data, modeling, noise, Occam's razor, overfitting, parameter, population, sample, signal. Question ID 1229535 at MyOpenMath.com asks the student to match the term with its definition and provides both video and written help for each term. The teacher could provide additional background on these concepts.

To fit a curve to our data, we need to model our data. We should recall that "All models are wrong, but some are useful" as discussed in Box (1976) and remember that our data is a sample from a population. Our goal is to use our data to make a statement about a population parameter.

Our data probably contains both *signal* (information about the population parameter) and *noise* (errors introduced by sampling and/or other sources). We seek a model that will explain the signal without getting distracted by the noise.

We seek the simplest model which explains our data. More complicated models tend to overfit the data and explain the noise better than the signal.

3.2.2. Model Selection for the data in "*T3_Dataset_1.xlsx*" workbook (shortened as T3D1)

Please note that this section uses the fictitious data in the *T3_Dataset_1.xlsx* workbook. You may apply the same methods to your observed traffic count data from Topics One and Two.

Please note that 3.2.2 and 3.2.3 are independent sections. Feel free to cover one or the other or both. Please find resources referred to and/or cited in section 3.2.4

Please note that the use of fictitious data to illustrate a concept has a precedent in Anscombe (1973). This section (3.2.2 only) should take between 30 and 70 minutes, depending on the student background in Excel and statistics.

This section discusses the following concepts: coefficient of determination, data, data exploration, data visualization, interpolation, linear equation, model, observation, overfitting, noise, polynomial equation, prediction, R^2, residuals, rounding, scatter plot, signal, trendline. Question ID 1231017 at MyOpenMath.com asks the student to match the term with its definition and provides both video and written help for several terms. The teacher could provide additional background on these concepts.

We will now explore our data, select a model, apply it to our data, record R^2, plot residuals, interpolate to predict traffic at times without observations, and find the time of peak foot traffic.

Please open *T3_Dataset_1.xlsx*.

Exercise 3.2.1 Exploration of data by data visualization

We explore our data by making a scatter plot of our traffic count data (leftmost sheet named "data" in *T3_Dataset_1.xlsx*. For example, the data set yields the scatter plot in Figure 10 (scatter plot sheet in the workbook).

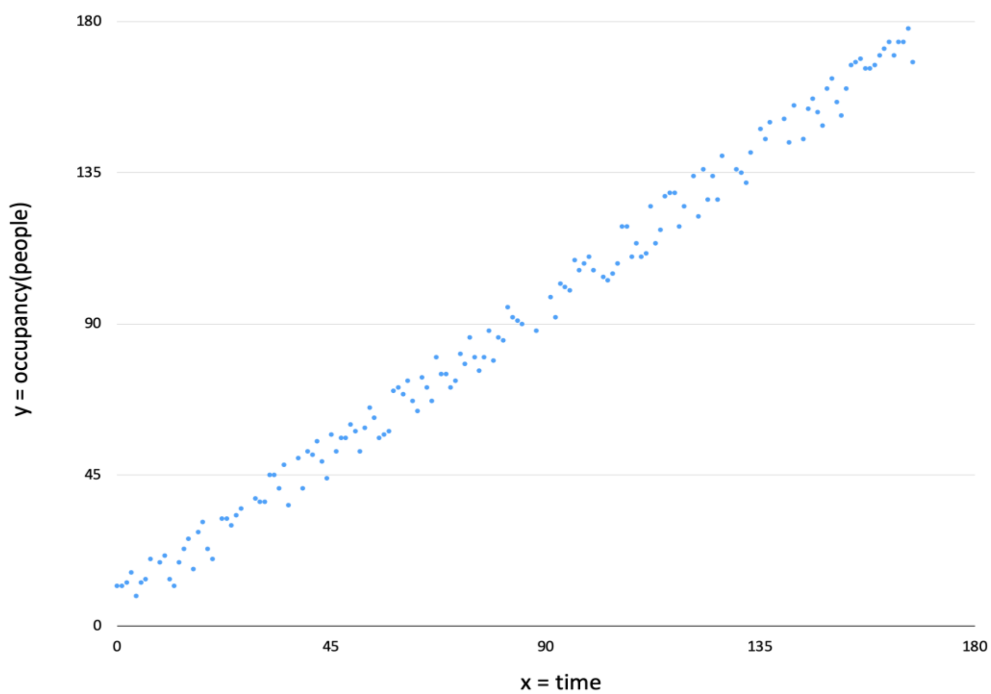


Figure 10. Scatter plot of (x_i, y_i) of Dataset T3D1

Exercise 3.2.2 What does this scatter plot look like? What curve could be used to fit the data?

Answer:
For this T3D1 data set, the signal looks like a line, so we select a linear model.

Use Excel to find the *line of fit* for the data (called *trendline* in Excel): You will find that Excel provides the graph of the line and also computes the *coefficient of determination* statistic, R^2, "... that represents the proportion of the variance for a dependent variable that's explained by an independent variable in a regression model." Fernando (2023).

Another resource for trendline is Banaszak (n.d.). Another resource for R^2 is Diaz & Cetinkaya-Rundel & Barr (2022).

You will find that the dataset *T3_Dataset_1.xlsx* yields the trendline in Figure 11 as $\hat{y} = 0.9994\,x + 7.3588$.

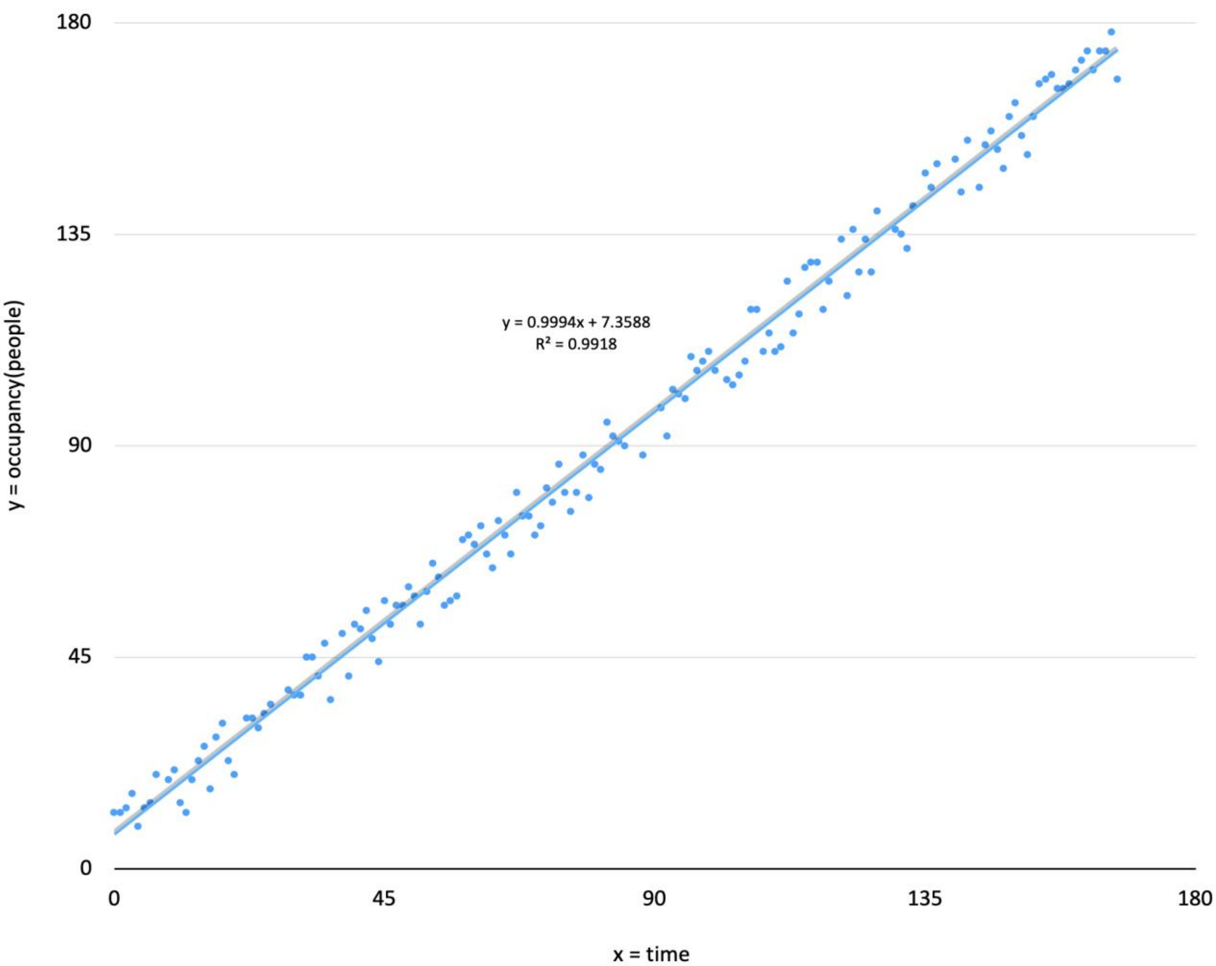


Figure 11. Trendline $\hat{y} = 0.9994\,x + 7.3588$ and $R^2 = 0.9918$ for Dataset T3D1

Exercise 3.2.3 Describe the trendline and discuss how well it fits the signal of the data.

Answer:

The linear trendline looks like a good fit to the data, because it approximately goes through the middle of the data. About half of the data resides above the line and half below.

Exercise 3.2.4 Discuss the R^2 value.

Answer:

We have $R^2 = 0.9918$, very close to $R^2 = 1$. This R^2 value indicates that the linear trendline is good fit to the data.

Plot Residuals: At every data point (x_i, y_i), our trendline gives us the predicted point $(x_i, \widehat{y}_i)$. The *error*, or *residual* is the difference $e_i = y_i - \widehat{y}_i$. We compute these residuals. The simplest residual plot consists of the error between the prediction and the observation at each point. We make this residual plot of the points (x_i, e_i) For example, the data set cpm04 yields the residual plot in Figure 12 (residuals sheet in the T3D1 workbook).

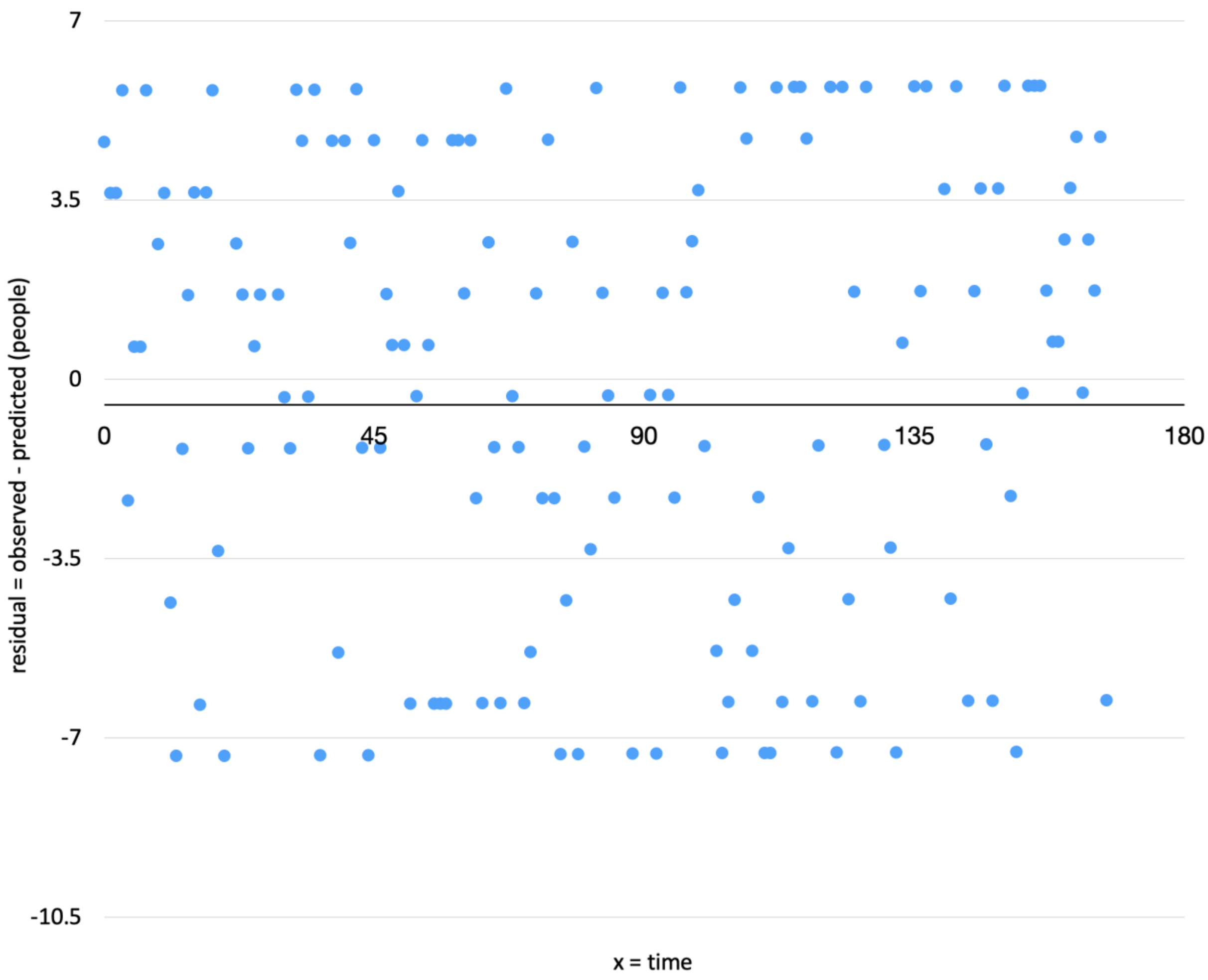


Figure 12. Residuals (x_i, e_i) of the Dataset T3D1

Exercise 3.2.5 Describe the residual plot.

Answer:

We describe the residual plot. The residuals exhibit no clear pattern. This reinforces our notion that the linear trendline provides a good fit to the data.

Interpolation: The scatter plot of the observed data points (x_i, y_i), our selection of the simplest possible model, the trendline $\hat{y} = 0.9994\ x + 7.3588$, $R^2 = 0.9918$ (very close to $R^2 = 1$), and the scatter plot of the residuals (x_i, e_i) combine to give us confidence that our model fits the data well.

Please note that selecting a polynomial trendline often yields even higher R^2 values, and residuals lacking pattern. But these polynomials violate the preference for the simplest model. These polynomials overfit the data. These polynomials model the noise more than the signal.

We can now use our model to interpolate. Our data has no observation for several *x* values. The biggest gap in our *x* values is 3, which occurs between 26-29, 85-88, 88-91, 127-130, and 137-140. We can use our trendline to predict *y* values anywhere in our domain of *x* observations, including *x* values without any *y* observations. For example, at the point x = 87, we can predict

$$\hat{y} \approx\ 0.9994\ x + 7.3588 \approx 0.9994 \times 87 + 7.3588 \approx 94.3066$$

Since we are counting integer people, we predict that 95 people would occupy our facility at the time $x = 87$.

For example, we plot several interpolated values of the data set cpm04 in Figure 13 (Lacey, n.d.).

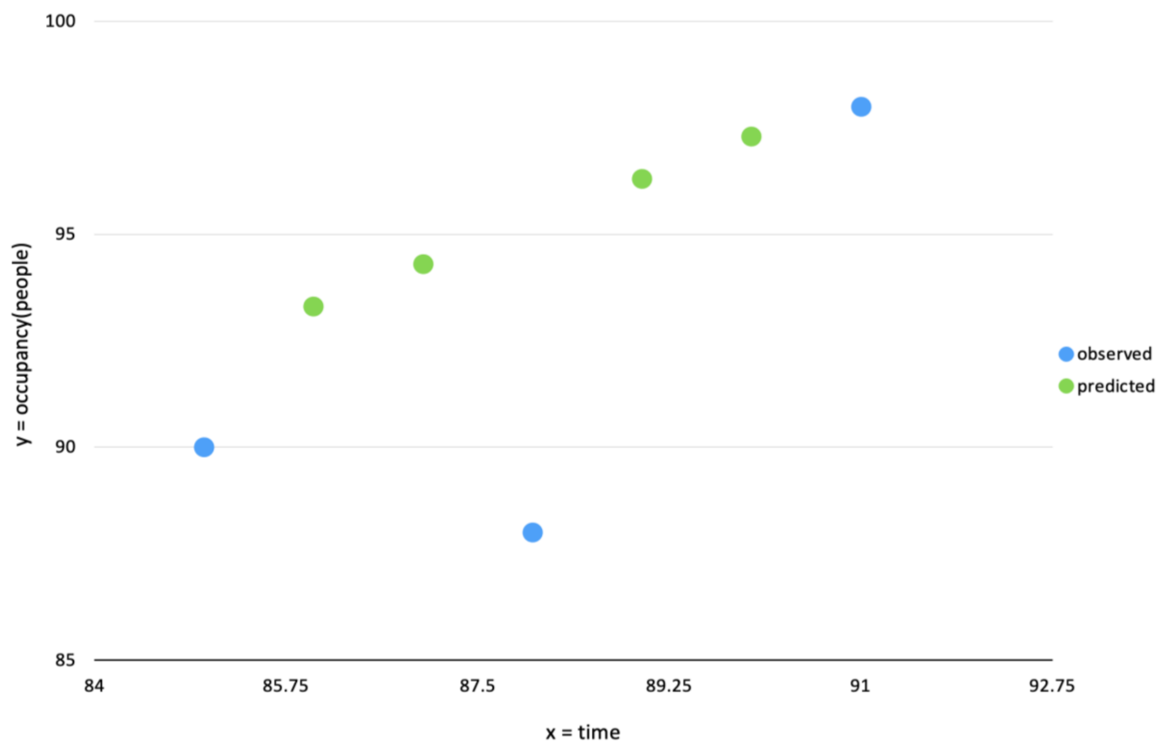


Figure 13. Interpolated values $(x_i, \hat{y}_i)$ added to Dataset T3D1

Find the Peak Traffic: Since we are using a linear model, the trendline $\hat{y} = 0.9994\,x + 7.3588$ predicts the peak traffic at the largest x value, $x = 167$:

$$\hat{y} \approx 0.9994\,x + 7.3588 \approx 0.9994 \times 167 + 7.3588 \approx 174.2586$$

Since we are evacuating people, we do not want to leave anyone behind, we round up, so the trendline predicts peak foot traffic at $(x, \hat{y}) = (167{,}175)$. We observed peak foot traffic at $(x, y) = (166{,}178)$.

3.2.3. Model Selection for the data in "*T3_Dataset_2.xlsx"* workbook (shortened as T3D2)

Please note that this section uses the fictitious data in the *T3_Dataset_2.xlsx* workbook. You may apply the same methods to your observed traffic count data from Topics One and Two. Subsection 3.2.2 (T3D1) and 3.2.3 (T3D2) are independent. Please feel free to cover one or the other or both. Please find resources referred to and/or cited in section 3.2.4. The use of fictitious data to illustrate a concept has a precedent in Anscombe (1973).

This section (3.2.3 only) should take between 35 and 80 minutes, depending on the student background in algebra, Excel, and statistics. (Applying the optional linear model to T3D2 would use another 30 to 70 minutes.)

This section discusses the following concepts: coefficient of determination, curve fitting, data exploration, data visualization, error, equation of parabola in standard form, Excel LINEST function, general form of a quadratic equation, ill-fitted model, interpolation, model, noise, observation, overfitting, parabola, parabola focus, parabola vertex, polynomial equation, polynomial evaluation, prediction, R^2, residual, scatter plot, signal, trendline, underfitting. The teacher could provide additional background on these concepts.

We will now explore our data, select a model, apply it to our data, record R^2, plot residuals, interpolate to predict traffic at times without observations, and find the time of peak foot traffic.

Please open *T3_Dataset_2.xlsx*.

Exploration of data-by-data visualization: We explore our data by making a scatter plot of our traffic count data (leftmost sheet named "*Data*" in the *T3_Dataset_2.xlsx* workbook). For example, the data set yields the scatter plot in Figure 14 (scatter plot sheet in the workbook).

Exercise 3.2.6 What does this scatter plot look like? What curve could be used to fit the data?

Answer:
For this T3D2 dataset, the signal looks like a parabola, so we select a second order polynomial model.

Use Excel to *fit a curve* to the data (called *trendline* in Excel): You will find that Excel provides the graph of the curve and also computes the *coefficient of determination* statistic, R^2, "... that represents the proportion of the variance for a dependent variable that's explained by an independent variable in a regression model." Fernando (2023).

Record R^2: For example, the dataset T3D2 yields the trendline in Figure 15, $\hat{y} = -0.0067x^2 + 1.1243x + 21149$.

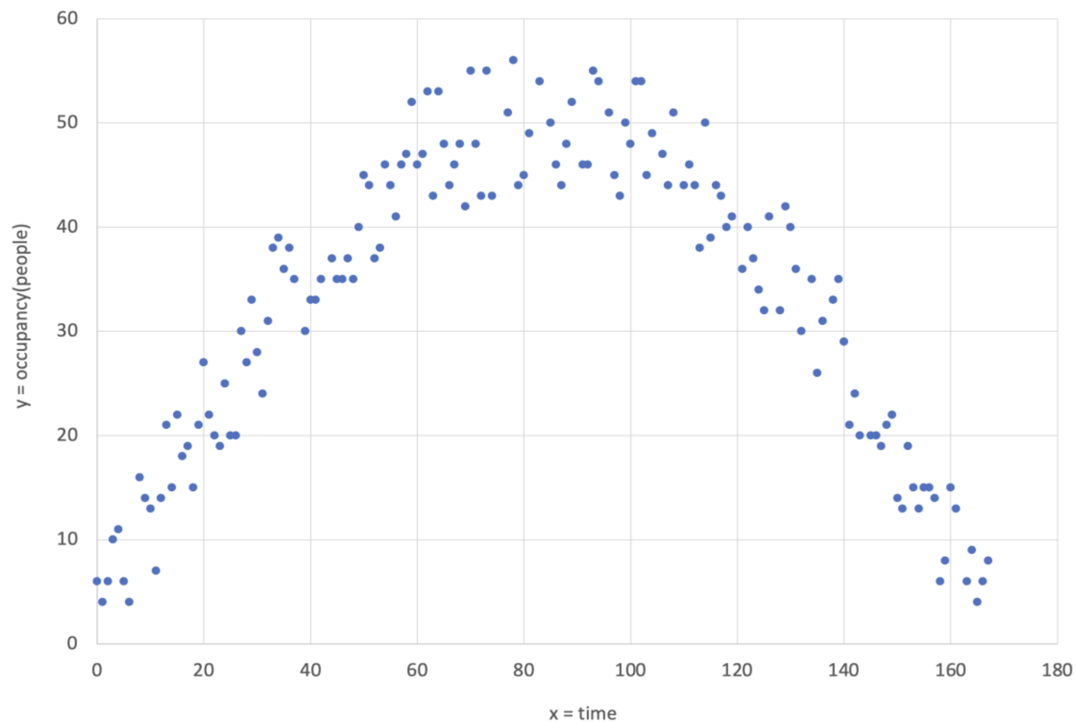


Figure 14. Scatter plot of (x_i, y_i) of T3D2

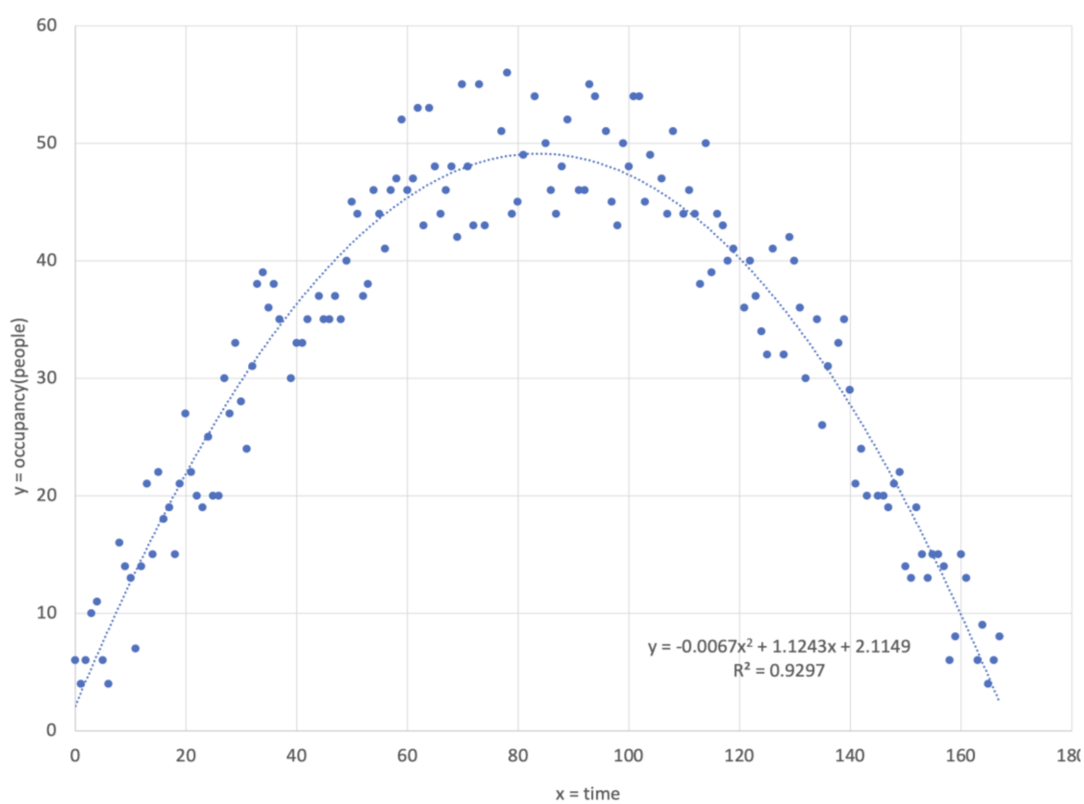


Figure 15. Trendline $\hat{y} = -0.0067x^2 + 1.1243x + 21149$ and $R^2 = 0.9297$ for Dataset T3D2

Exercise 3.2.7 Describe the trendline and discuss how well it fits the data.

Answer:

The second order polynomial trendline looks like a good fit to the data, because it approximately goes through the middle of the data. About half of the data resides above the curve and half below.

Exercise 3.2.8 Discuss the R^2 value.

Answer:

We have $R^2 = 0.9297$, close to $R^2 = 1$. This R^2 value indicates that the second order polynomial trendline is a good fit to the data.

Please note that you could fit a linear trendline to this data, to discuss the example of low R^2 and pattern in the residuals. That is, a linear trendline provides an example of an ill-fitted model, a model that underfits the data, a poor model for the signal.

Plot Residuals: At every data point (x_i, y_i), our trendline gives us the predicted point $(x_i, \hat{y}_i)$. The *error*, or *residual* is the quantity $e_i = y_i - \hat{y}_i$. We compute these residuals. The simplest residual plot consists of the error between the prediction and the observation at each point. We make this residual plot of the points (x_i, e_i). We use the Excel LINEST function to obtain more decimal places of the trendline coefficients as discussed in Zach (2021). For example, the dataset T3D2 yields the residual plot in Figure 16 (residuals sheet in the T3D2 workbook).

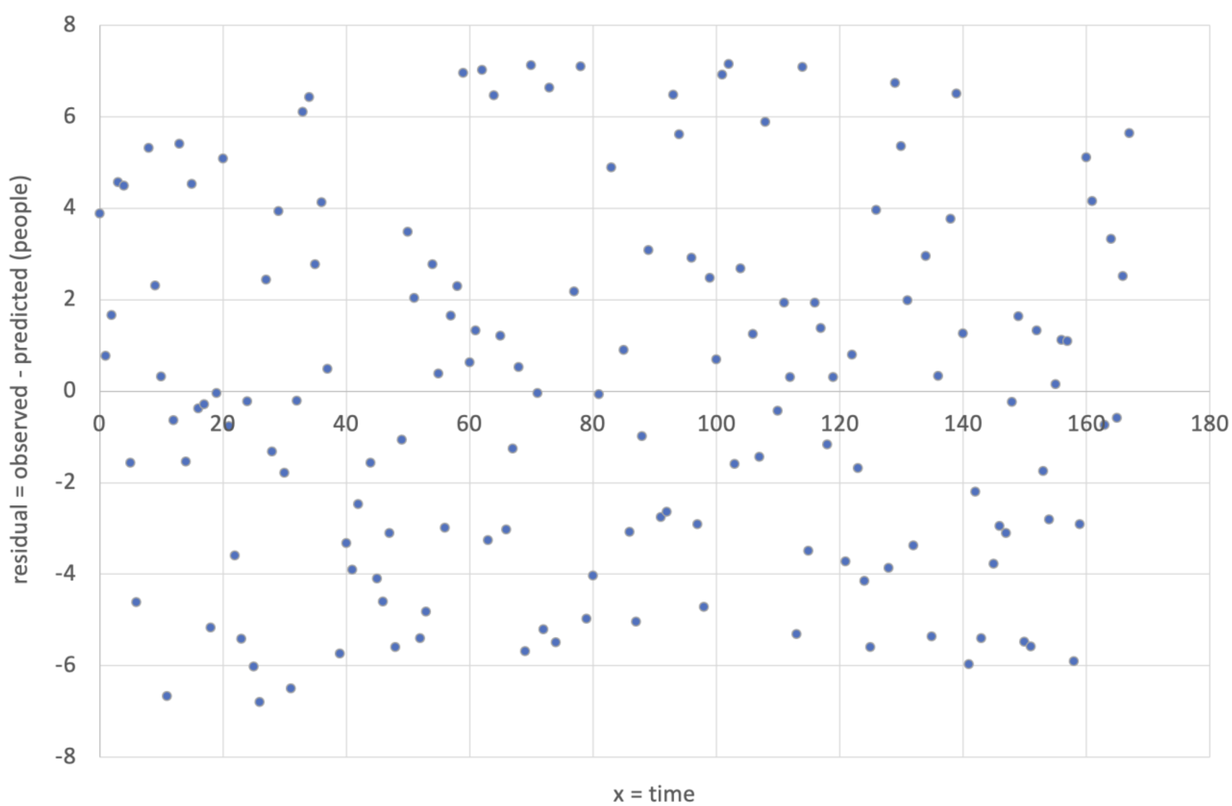


Figure 16. Residuals (x_i, e_i) of the Dataset T3D2

Exercise 3.2.9 Describe the residual plot.

Answer:

We describe the residual plot. The residuals exhibit no clear pattern. This reinforces our notion that the second order polynomial trendline provides a good fit to the data.

Interpolation: The scatter plot of the observed data points (x_i, y_i), our selection of the simplest successful model, the trendline $\hat{y} = -0.0067x^2 + 1.1243x + 21149$, $R^2 = 0.9297$ (close to $R^2 = 1$), and the scatter plot of the residuals (x_i, e_i) combine to give us confidence that our model fits the data well.

Please note that selecting a third order or higher order polynomial trendline often yields even higher R^2 values, and residuals lacking pattern. But these polynomials violate the preference for the simplest successful model. These polynomials overfit the data. These polynomials model the noise more than the signal.

We can now use our model to interpolate. Our data has no observation for several x values. The biggest gap in our x-values is 3, which occurs between 74-77. We can use our trendline to predict $\hat{y}$ values anywhere in our domain of x observations, including x values without any y observations.

For example, at the point $x = 76$, we can predict

$$\hat{y} \approx -0.0067x^2 + 1.1243x + 21149 \approx -0.0067(76)^2 + 1.1243(76) + 2.114 \approx 48.72694$$

Since we are counting people, we state our result as an integer. We predict that 49 people would occupy our facility at the time $x = 76$. For example, we plot several interpolated values of the dataset T3D2 in Figure 17 (Lacey, n.d.).

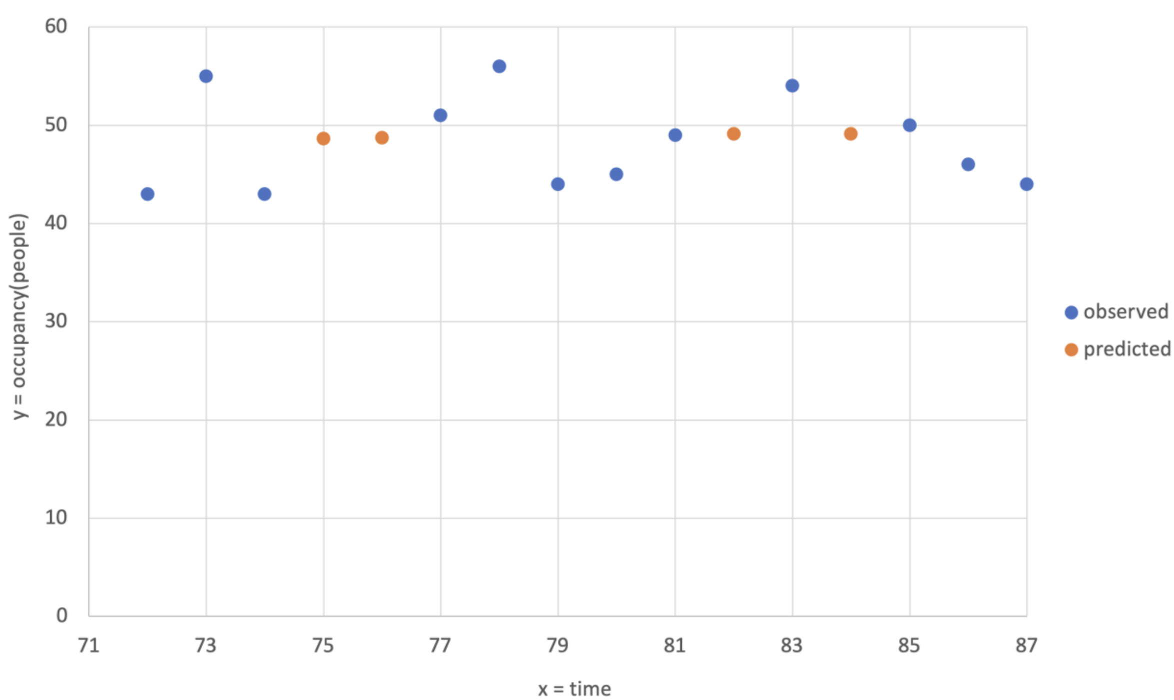


Figure 17. Interpolated Values $(x_i, \hat{y}_i)$ added to the Dataset T3D2

Find the Peak Foot Traffic: Since we are using a second order polynomial model, we can use algebra, as discussed in Abramson & North (2021), to find the time of peak traffic by converting the trendline $\hat{y} = ax^2 + bx + c$ from general form to the standard form $(x - h)^2 = 4p(y - k)$, where (h, k) is the parabola vertex and $(h, k + p)$ is the focus. In general, $(x + \frac{b}{2a})^2 = \frac{1}{a}(\hat{y} - c)$. Our data has $(a, b, c) \approx (-0.0067233, 1.12428672, 2.11492876)$. The parabola vertex occurs at

$$x = -\frac{b}{2a} \approx -\frac{1.12428672}{2 \times (-0.0067233)} \approx -\frac{1.12428672}{-0.01345} \approx 83.61123$$

Alternatively, we can apply *calculus* to the trendline $\hat{y} = -0.0067x^2 + 1.1243x + 21149$ to find the time of peak traffic at the zero of the first derivative. In general form, the trendline is $\hat{y} = ax^2 + bx + c$. The first derivative is

$$\frac{d\hat{y}}{dx} = 2ax + b$$

The zero of the first derivative occurs at $x = -\frac{b}{2a}$. This is the same x value as the coordinate of the parabola vertex found using solely algebra above.

We find the people of peak traffic by evaluating the second order polynomial as

$\hat{y} \approx -0.0067x^2 + 1.1243x + 21149 \approx -0.0067(83.6)^2 + 1.1243(83.6) + 2.11 \approx 49.11642$

Since we are evacuating people, we round up to avoid leaving anyone behind. The trendline predicts peak traffic at $(x, \hat{y}) \approx (84,50)$. We observed peak traffic at $(x, y) = (78,56)$ (CFI Team, n.d.)

3.2.4. Conclusion

In section 3.2.2 (fictitious linear data) and section 3.2.3 (fictitious quadratic data) we explored our data, selected a model, applied it to our data, recorded R^2, plotted residuals, interpolated to predict traffic at times without observations, and found the time of peak traffic.

3.2.5. References

Abramson, J. & North, S. (2021). *College Algebra with Corequisite Support 2e*. OpenStax. https://openstax.org/details/books/college-algebra-corequisite-support-2e Section 8.3 discusses parabolas in general and example 5 on page 1004 converts from general from to standard form.

Anscombe, F. J. (1973). "Graphs in Statistical Analysis", Am. Stat. 27 (1).

Banaszak, B. (n.d.). *What is a Trend Line in Math?*

https://study.com/learn/lesson/trend-line-formula-examples.html

Box, George E. P. (1976). “Science and statistics” (PDF), *Journal of the American Statistical Association*, 71 (356): 791799, doi:10.1080/01621459.1976.10480949.

CFI Team. (n.d.). *Index Match Formula.*

https://corporatefinanceinstitute.com/resources/excel/inde match-formula-excel/

Diaz, D. & Cetinkaya-Rundel, M. & Barr, C.D. (2022). *OpenIntro Statistics*, Fourth Edition. https://www.openintro.org/book/os/ Section 8.2.7 discusses R^2.

Fernando, J. (2023). *R-Squared: Definition, Calculation Formula, Uses, and Limitations*. https://www.investopedia.com/terms/r/r-squared.asp

Lacey, A. (n.d.). *How to Make a Scatter Plot in Excel? 4 Easy Steps.* https://www.simonsezit.com/article/how-to-make-a-scatter-plot-in-excel/

Zach. (2021). *How to Fit a Polynomial Curve in Excel (Step-by-Step).* https://www.statology.org/excel-polynomial-fit/

3.3. Evacuation Simulation

This section (all of 3.3) should take between 5 and 15 minutes, depending on your discussion depth.

In section 3.1 (Classroom Drill) we gathered data by evacuating our own classroom. In section 3.2 (Traffic Count Interpolation) we analyzed our traffic count data to find the time of peak traffic. It might be tempting to use our data to extrapolate the evacuation time. Instead, we should follow the Stoic advice, "Don't extrapolate" as stated in Aurelius (161). A discussion of extrapolation in data science is available at Awati (2022).

To predict the evacuation time for a facility we do not control, we should simulate the facility. Then we can interpolate within our simulated data, to avoid extrapolation error. We leave this simulation as a future project for you to explore.

3.3.1. Available Simulators

Several simulators are available for free for academic use including the following:

- Oak Ridge National Lab IMPACT, as discussed in Koch (2013). (This is freely available.)
- Pathfinder, as discussed in Thunderhead Engineering (2023). (An academic license is available.)

3.3.2. Evacuation Simulation Research

Evacuation simulation research uses custom software to explore problems posed by geometry, psychology, and other factors. A collection of diagrams in Figure 18 from Trivedi and Rao (2018) provide a taste of the challenges.

Exercise 3.3.1 Describe evacuation challenges posed by the rooms in the Trivedi and Rao diagrams.

(a) Physical challenges
(b) Psychological challenges

Answer:
(a) Physical: Many people compete to use one or a small number of exits, sometimes hindered by obstacles.
(b) Psychological: People toward the back of the queue could suffer increased anxiety as they wait to evacuate. In extreme cases, panic can lead to trampling.

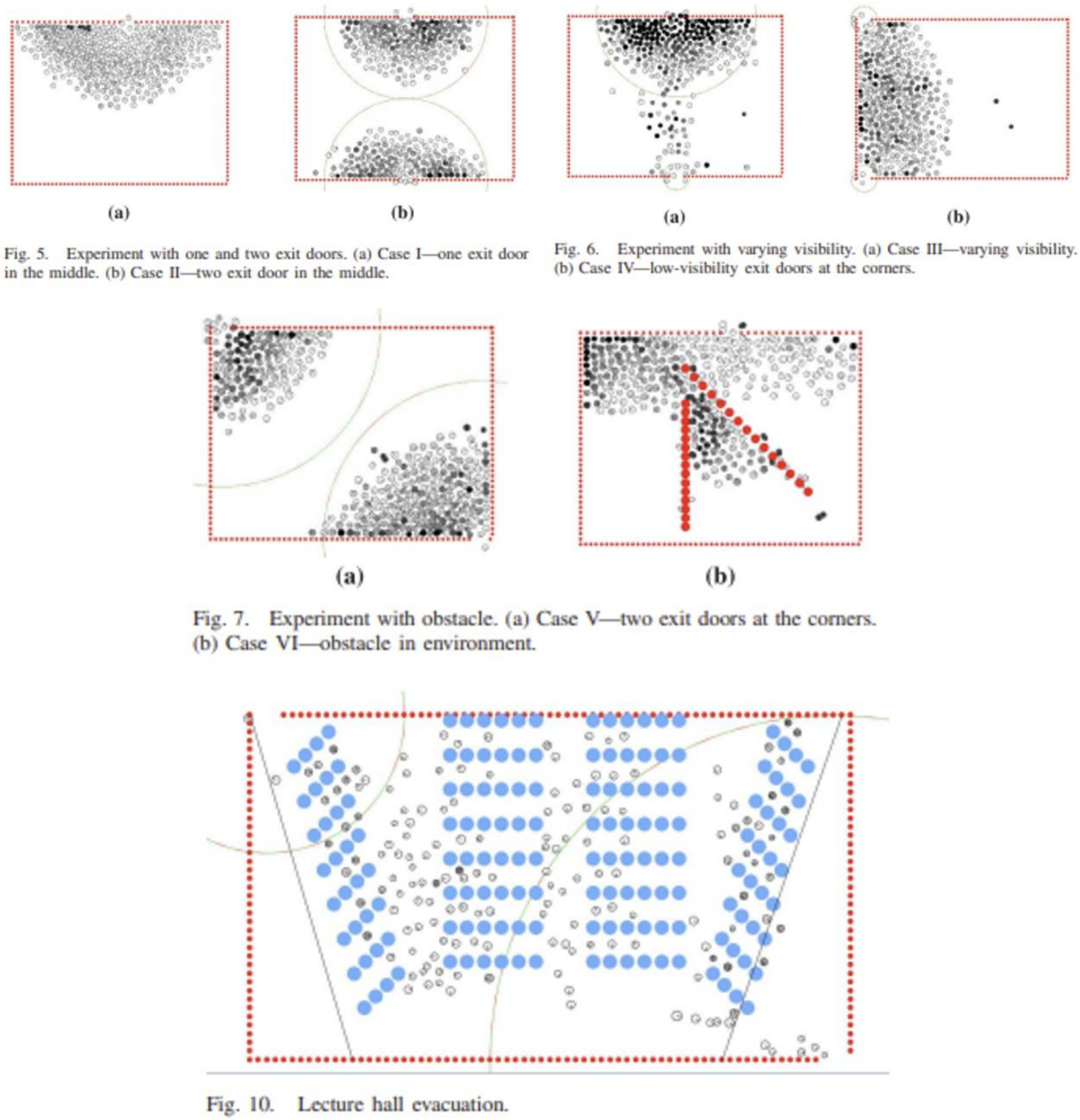


Figure 18. Evacuation Simulation Examples from Recent Study by Trivedi and Rao (2018)

3.3.3. References

Aurelius, M. (161). *Meditations*, Book 8:49.

https://quotecatalog.com/quote/marcus-aurelius-nothing-but-wha-Ra3ZzPp

Awati, R. (2022). *extrapolation and interpolation*,

https://www.techtarget.com/whatis/definition/extrapolation-and-interpolation

Koch, D.B. (2013). *Incident Management Preparedness and Coordination Toolkit*, https://www.osti.gov/biblio/1231647

Thunderhead Engineering. (2023). Pathfinder. https://www.thunderheadeng.com/pathfinder

Video: Agent-Based Modeling and Simulation of Emergency Evacuation Strategies (2018, Trivedi and Rao)

Paper: Agent-Based Modeling and Simulation of Emergency Evacuation Strategies (2018, IEEE) A. Trivedi and S. Rao. "Agent-Based Modeling of Emergency Evacuations Considering Human Panic Behavior," in IEEE Transactions on Computational Social Systems, vol. 5, no. 1, pp. 277-288, March 2018, doi: 10.1109/TCSS.2017.2783332.