2020-21 Example Scenario

Flooding & Preventative Measures

Instructor Guide

A program of The Actuarial Foundation

Modeling The Future Challenge

Introduction

Each year, billions of dollars are lost due to flooding across the United States. In 2019 alone, $3.75 billion was lost; however, this is far from the largest annual loss the US has suffered from flooding. That title goes to 2017, when severe storms, excess precipitation, and hurricanes cost over $60 billion in damage!

These losses come in many forms, most notably perhaps are properly losses. Floods, like most natural disasters, do not discriminate on the type of buildings they damage; from farms to strip-malls, to family homes and everything in-between, floods can cause untold damage to anyone. For individual homeowners, flooding can be one of the most destructive risks of owning a home. However, there are many factors that can lower the risk of loss for homeowners.

Information on the elevation of the property, whether the property has a basement, or other under-ground assets, the size of the property, and whether it has preventative measures to help protect against flooding are all factors that can be used to model the risk of loss for each property in the event that a flood occurs.

In this scenario, you have been provided data about 7,000 households across the Midwest with 1-year term policies that cover water damage to the structure. The data is all from one insurance company, Sphinx Insurance, which is reinsured by the National Flood Insurance Program (NFIP) under the Federal Emergency Management Agency. Sphinx is interested in understanding how to better price their insurance products and help protect families in the flood-prone states of the Midwest. Sphinx currently insures home-owners in the states of Iowa, Illinois, Missouri, Kansas, and Nebraska.

Data summary:

Sphinx has provided you with data on 7,000 of their losses from written policies in the last year, cleaning the data from any policyholders who did not have a claim. All households are located within Iowa, Illinois, Missouri, Kansas, or Nebraska, and policies were written uniformly across each state (1400 each). Households are classified by their policy ID number, the elevation of their home (in feet), as well as by indication if they have preventative measures against flooding (i.e. sump pumps, or drainage ditches).

Use the data in the attached spreadsheet to answer the questions provided and help the Sphinx Flood Insurance CEO make decisions about how to update or add to their insurance policies.

Part 1:
Problem Definition

Questions from this part of the scenario build upon Part 1 of the Actuarial Process. It may be valuable to review this section of the Actuarial Process Guide before answering the questions below.

1. Identify at least two factors other than those mentioned in the spreadsheet provided (i.e. elevation and preventative countermeasures) might be valuable in helping to separate the level of risk into more specific possible likelihoods and severities?

Students could include many things here. We are looking for any additional data that might be used to separate the probability of a loss occurring, or how large of a loss it is. Some examples include:

- The value of the home itself
- The value of any materials in the home
- Whether the home has a basement or other underground assets
- Whether the home has good drainage ditches, or other water removal systems

2. Describe in no more than a few sentences how insurance helps homeowners mitigate the risks of flooding.

Insurance covers the costs of a very large loss, and lowers the variability in how much a policy holder needs to pay out in any year. So although a policy holder pays a little bit of money every year, they will not have a risk of having to pay a very large amount of money in any year.

Students can describe this in many ways, but the key point that should be included is that insurance helps the policy holder NOT have any one-time, large losses.

3. Which type of risk mitigation strategy would the installation of a sump-pump, or other type of preventative counter-measure be (hint: reference the three types of risk mitigation strategies from the Actuarial Process Guide, Section 1.3)?

These are examples of modifying outcomes. The inclusion of preventative measures in a home will (if they work properly) lower either the severity or the frequency of the risk.

4. Provide examples of two other groups (beyond homeowners) who could be at risk due to flooding of communities in the midwestern states discussed in this scenario. Identify who they are and describe the risk in written terms.

- If a city floods, the local businesses could lose their customers and ultimately their business if no one can buy anything from their store.
- Local governments could be at risk of excessive expenditures to repair damaged infrastructure
- Hospitals and EMS services could be at risk of dealing with injuries and healthcare related to the flood
- Utilities and power companies could have service interruptions that they need to fix.

Part 2:
Data Identification & Analysis

Questions from this part of the scenario build upon Part 2 of the Actuarial Process. It may be valuable to review this section of the Actuarial Process Guide before answering the questions below.

5. What critical series of information is NOT included in your spreadsheet that would be required to determine the frequency of the losses in this scenario?

Students should be able to identify that the dataset provided does NOT include policies that didn't have a loss. So it will be impossible to understand the frequency of the loss without having this information that lets you know the total number of policies written by the insurance company during this timeframe.

6. What is one way that the dataset provided needs to be "cleaned?"

There are two examples of clearly bad data in the spreadsheet. Students should be able to identify either of the following:

- Line 15 "Kansas" and "no" (columns B and D) are switched.
- Lines 102-104 are bad data and should be removed

7. How does having more information such as the elevation of the property and whether it includes preventative counter measures or not, help in characterizing the potential risk for loss?

This kind of information is valuable in helping to separate the possible outcomes and making the risk characterization more fine-tuned and specific.

8. Beyond refining the possible severity or frequency of potential losses, what other information could be valuable to this scenario that is not provided? (Hint: see Actuarial Process Guide Section 2.1 for general types of valuable data).

In this scenario we are not provided with any time series information to determine if there are historical trends. This may be valuable. Students could also describe additional data in any of these five categories from the Actuarial Process Guide, but should be able to justify how the data they identify would help in that category.

Part 3:
Mathematical Modeling

Questions from this part of the scenario build upon Part 2 of the Actuarial Process. It may be valuable to review this section of the Actuarial Process Guide before answering the questions below.

8. Of those with a claim, what is the average claim amount for policyholders in each state?

	Total loss	AVG Claim
Kansas	18649193.99	13320.8529
Illinois	14130930.22	10093.5216
Iowa	17914041.56	12795.744
Missouri	16974175.77	12124.4113
Nebraska	15363917.58	10974.2268

We find the average claim amount in each state by summing the total claims and dividing by the number of claims (1400 in each state).

9. If a Nebraska policyholder did not have preventative measures installed in their home, what is the probability that they had a claim over $14,000 last year, given they had a loss?

$P(A|B) = \frac{P(A \cap B)}{P(B)}$, so then:

$$P(claim > \$14{,}000) = \frac{\sum_{i=1}^{7000} y_i * I(state = Nebraska, claim > \$14{,}000, prev.measures = False)}{\sum_{i=1}^{7000} y_i * I(state = Nebraska, prev.measures = False)}$$

I(.) is an indicator function which returns 1 when every condition in its interior is true, and 0 when its interior is false.

P(claim > $14,000 | state = Nebraska, prev. measures = False) = 0.2376761

10. In Illinois, what is the difference between the expected loss of a policyholder who has preventative protection installed and the expected loss of a policyholder who does not have any preventative measures installed?

The expected loss of an Illinois policy holder who has countermeasures installed is $9774, the expected loss of policy holders who do not have counter measures is $10,282. So the difference is $508.

11. What is the variance of elevation in each state? (Hint: You have all of the claims data, so use the population variance)

Iowa: 15158.77108 feet

Illinois: 10,343.55813 feet

Missouri: 30,467.01778 feet

Kansas: 90,683.90593 feet

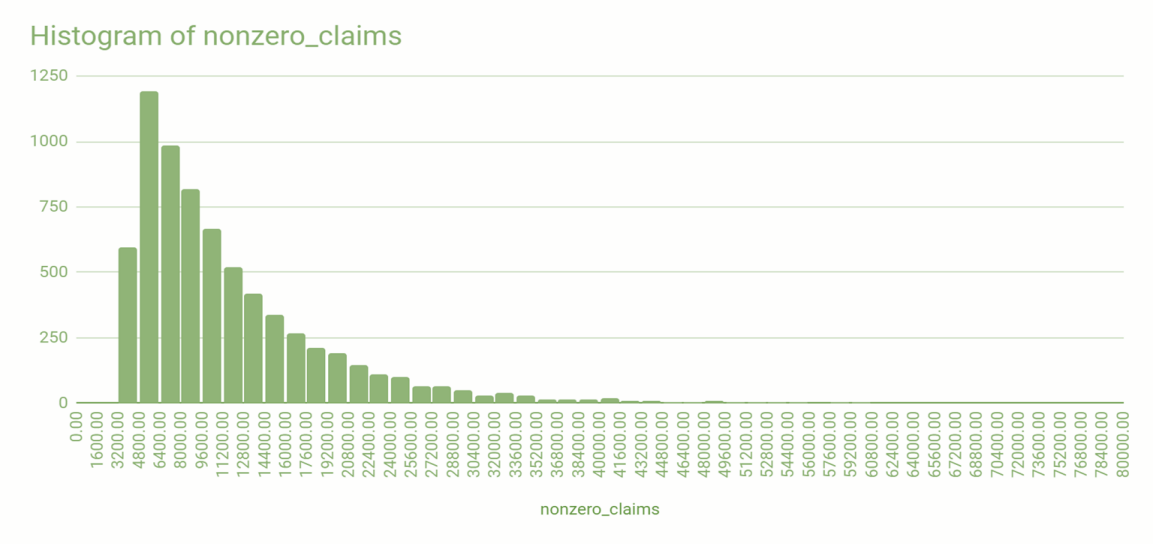
Nebraska: 160,355.8227 feet

Part 3:
Mathematical Modeling

12. What percentage of claims were paid out to policyholders in states bordering the Mississippi River?

3 out of 5 states in the Sphinx insurance claims border the Mississippi river, and there are an equal number of claims in each state, so 60%.

13. Create a histogram of the claim amounts for nonzero claims. Describe the distribution. Sphynx is especially interested in reducing the occurrence of large claims, what might be called outliers. You may have experience with the 1.5 IQR rule or with using Mean ± 2SD for identifying outliers in a distribution. Based on the shape the claim amount histogram, do you think either of these are reasonable methods?



The histogram of nonzero claim amounts is strongly skewed to the right. The bulk of the claims falls between $3200 and about $15,000. A few claims are greater than $50,000 and the maximum is $77,9201.60. Using either the 1.5 IQR rule or the mean ± 2sd to identify outliers does not seem reasonable with this distribution since is extremely non-symmetric.

14. Sphynx management decides that it would like to try to reduce the number of claims that are greater than $50,000. Does it appear that claims less than or equal to $50,000 are more likely to have taken preventative measures than those that are greater $50,000?

	Claims ≤ $50,000	Claims > $50,000
n	6,969	31
# with preventative measures	3,793	14
% with preventative measures	3793/6969 = 54.43%	14/31 = 45.16%

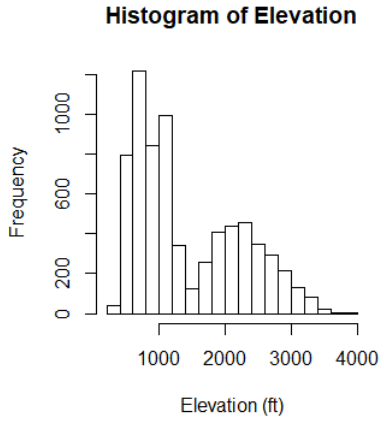
It does appear that it is more likely that claims under $50,000 have taken preventative measures, 54.43%, than those claims greater than $50,000 where only 45.16% did so.

Part 3:
Mathematical Modeling

15. With which variables can you identify a correlation coefficient with the size of the claims? Calculate the correlation coefficient between claims and these variables. What can you say about the strength of these correlations? Are there any issues in using correlation coefficients to identify relationships in your available data?

Correlation coefficients can only be calculated between quantitative variables, meaning that there is no value for the correlation between the State, or preventative countermeasures and nonzero claims, therefore the only coefficient that can be calculated is with elevation. This is: **0.0124.** However, this is a weak correlation, both Region and prev_measures seem to have stronger correlations when graphed against each other. Students can see this by creating boxplots of the data and should recognize that correlation coefficients are not the only way to measure relationships between variables.

16. Plot the distribution of elevation and explain some potential reasons of its shape.



The histogram of elevation has a bimodal distribution, which indicates that most states have a mean elevation less than 1000 feet, but others have means around 2000 feet. Overall it looks like a combination of two normal distributions. This makes sense because different states are naturally located at different elevations, and so their individual distributions are likely to vary from state to state. Not only this, but their proximity to the Mississippi River may help indicate overall elevation, so the 3 states that border the river are more likely to be at lower elevations than the 2 states that aren't bordering it.

17. Use elevation only to predict the expected loss of a policyholder living at an elevation of exactly 2,000 feet. Is this a good way to predict future claim sizes?

This involves simple linear regression to predict claims. The regression equation is as follows:

$\hat{y}$=11,680+.1254*[elevation]

Then, just plug in 2,000ft for elevation.

Expected loss = $11,933.79

Elevation and claim size are very poorly correlated as evidenced through the nonlinear scatterplot, as well as a correlation coefficient calculated out to be 0.0124.

Part 4:
Critical thinking & Risk Analysis

Questions from this part of the scenario build upon Part 4 of the Actuarial Process. It may be valuable to review this section of the Actuarial Process Guide before answering the questions below.

Additionally, the following questions reference information about insurance not previously discussed. For insurance companies there are several ways of making sure they will be able to cover all expected losses in a given year. The "premium" is the base amount a policyholder must pay (either annually or monthly). A loading charge is an additional percentage increase in a base premium designed to cover overhead expenses for the insurance company. A yearly fee is another method of recovering expenses which adds a standard fee for each policy. Companies may also consider having a co-pay for any claim – meaning the policyholder must pay a certain percentage of the claim. Companies also include deductibles which are dollar amounts that must be met before the insurance will pay the rest of the claim.

18. Assuming 35% of Nebraska policyholders had a loss last year, if all premiums are the same, what is the minimum premium that should be charged per month in the state to cover the expected losses if that is the only fee the insurance company gathers from the policy holder?

If 1400 = 35% of policy holders, then there are a total of 4000 policy holders in NE (1400/.35). This means that in order to cover they year's expenses ($15,363,917.6) the company must divide those across the 4000 policy holders to get a yearly premium of $3841. So monthly is divided by 12 which equals $320.08.

19. What is the minimum premium required if there is also a 10% annual charge on the expected losses in Nebraska and $200 yearly fee on every policy?

Min. Premium per year=1.1*[E(y)*.35]+200

Where E(y) is the mean claim in Nebraska. The subsequent monthly premium is the yearly premium divided by twelve, resulting in $368.76

20. If all other variables stayed the same and the likelihood of severe storms were to increase 10% throughout the region, in written, general terms, what would you expect to happen to:

1. **The average claim value:**

   The average claim value probably would not be affected very much by an increased likelihood of a storm, because it simply means there would be more storms so maybe more claims, but not more damaging claims.

2. **The number of claims:**

   This would be affected because more severe storms would likely lead to an increased amount of claims.

3. **The average premium Sphinx would need to charge to break-even.**

   This would also increase because more of their policies would be having a loss so they would have to charge more on average to cover for those loses.

Level 4 Questions: Critical thinking recommendations

21. Suppose Sphynx wants to start selling flood insurance in Idaho. Should they use their data on losses in the Midwest to determine the new rates? Why or why not? Use data to justify your answer.

They should not use their Midwest data to determine new rates because loss experience may be different in different regions of the country, leading to over/underpricing. This is evidenced by the large variance in average losses of each state (1745931) that indicates that loss experience will likely differ greatly between each state, especially for a state that is geographically different from the Midwest. It would be hard to qualitatively determine which state has losses most similar to Idaho without any prior loss experience, and the given data ranges between an average loss of $10,974 (Nebraska) and $13,321 (Kansas). This decision can have a large impact on premiums without data to provide evidence for the rate.

22. What bias may be present in a model based on this data alone? Explain your reasoning.

This data is from one year, may not accurately reflect all loss experience; elevation may be skewed towards areas of states that are more prone to floods which is why they purchase flood insurance. Additionally, other variables should be considered as well to more accurately depict loss experience because only 2.445% of the variation in claims are explained by the existing variables. [explain that it is r^2 calculation]

23. Sphynx is considering writing more policies. Does it make more sense to write them in states where they already write policies, or expanding their business into new states? Why? If they were to write more policies in Iowa, Illinois, Missouri, Kansas, or Nebraska, which state should they choose and why?

In general, expanding the geographical area covered will diversify the insurance risk and improve the risk profile overall, because floods that occur in one region will likely not affect states far away. If more policies are written in the same state, a flood experienced by a significant proportion of policyholders in a state will likely affect the new policyholders as well. If Sphynx were to write new policies in a new state it should be in Illinois because the variation in claims are the smallest here, meaning it's easier to predict future losses.

24. Based on this data, do you think it's more profitable to use a rate based on elevation instead of a flat rate for each state? What are the limitations of using a rate based on elevation alone?

Low correlation with elevation (r = .01) so an elevation-based rate would not be entirely accurate and would only catch extreme cases, not a generic variation in losses, evidenced by the R2 statistic of .01535% in a simple linear model. Furthermore, using elevation alone can be deceiving when aggregated in a model because high-elevation areas that are close to rivers or lakes can be just as prone to flooding as a low-elevation area that does not receive a lot of rain or is far from any sources of water.

25. An average sump pump costs a policyholder $500 to purchase and install. Sphynx wants to implement a nationwide discount program that reimburses policyholders for installing a pump. If 72.5% of policies did not have a loss last year and Sphynx charges a $220 flat annual fee and a 12% yearly loading charge for each policy. If the company is willing to reduce the loading charge by 8% for six months if a policyholder installs a pump and reduce the annual fee by $20 for the next six years, what percentage of the pump cost will policyholders ultimately have to pay?

Min. Premium per six months=(1.12*[E(y)*(1-.725)]+220)/2

Where E(y) is the mean/expected loss of policyholders with preventative measures, calculated out to be $11,654.31. Therefore, the normal minimum premium per six months is $1904.76. If the loading charge is reduced to 4% per year, then the new six-month minimum premium becomes $1776.57, and the subsequent difference is $128.20. But the annual fee reduction is $20 per year for six years, totaling $20*6 = $120, so the total savings for the policyholder is 128.20 + 120 = $248.20. Thus, the policyholder will have to pay (500-248.20)/500= **50.36% of the pump cost.**