Basic Training 3: Mathematical Modeling

Teacher Workbook

A Program of The Actuarial Foundation
Modeling the Future Challenge

The Modeling The Future Challenge

As part of the Scenario Phase of the MTFC, teams will be demonstrating and applying their mathematical analysis skills to a scenario response paper as well as identifying a potential project and writing a proposal. The Actuarial Process Guide to be an invaluable resource.

- The Actuarial Process Guide

How to Use this MTFC Mathematical Modeling Scaffolding Guide

When all of the potential topics in the world are at your fingertips, identifying a topic, identifying possible risks, finding sources, mathematically modeling, and studying risks can seem overwhelming to begin. This guide will help scaffold the process and guide participants through the process of writing an MTFC Project Proposal. Each task refers to a specific section of The Actuarial Process Guide for more in-depth information.

Content: The process is scaffolded into 1 total task.

Suggested pacing: 1 task per week for 1 week. Completing the *Basic Training 2. Data Identification & Analysis* resource first will provide context to the scenario and scaffolding procedure.

Common Core Standards for Mathematical Practice

The Common Core Standards for Mathematical Practice. The MTFC Project Proposal specifically addresses the following standards:

- ❏ CCSS.MATH.PRACTICE.MP1 **Make sense of problems and persevere in solving them.**
- ❏ CCSS.MATH.PRACTICE.MP2 **Reason abstractly and quantitatively.**
- ❏ CCSS.MATH.PRACTICE.MP3 **Construct viable arguments and critique the reasoning of others.**
- ❏ CCSS.MATH.PRACTICE.MP4 **Model with mathematics.**

3.1 What Do Actuaries Want Their Models To Do?

The goal of a mathematical model is to (as accurately as possible) determine the frequency and severity of and other characterizations of potential risks. A mathematical model does not have to be a large complex system of equations and code. The model can be the answer (ie, values or numbers) to a small series of questions that lead to the ability to draw conclusions and make recommendations.

Some things to consider when creating a model:

- expected value
- confidence intervals
- variability
- effects and sensitivity to changes

Computational (code/algorithms) or purely mathematical. Goal is to examine the expected values of risks for a scenario. Can also extend analysis to examine what impact a risk mitigation strategy will have and how each strategy could lower risk; compare and make recommendations.

Practice making some assumptions? Consider what effect the assumptions have on the model

- identify frequencies (likelihoods) that each potential outcome of scenario will happen
- identify the severities (size) of possible losses
- identify expected values of potential loss for scenario
- understand the distribution of potential outcomes
- identify trends in the data and understand how the potential outcomes and associated risks may be changing over time
- identify possible risk mitigation strategies and quantify their effects

3.2 Multiple models?

Frequency (measure of the number of occurrences of an event over a period of time) and severity (measure of the impact of an event) are both examples of random variables (variables that take on random values based on underlying probability distributions) that describe the probability and potential value of each random variable.

In the course of creating your model, you may find that it may be beneficial to develop two models: one for severity and one for frequency. Used in tandem, you may then use calculate expected value. While each insurance company decide on a different formula for their determination of a premium, an example of how it could be approached is given by:

Expected claims amount for each customer =

expected number of incidents * **expected amount paid with each incident**

(ie, expected frequency) (ie, expected severity)

The premium would be set as the expected claims amount for a customer.

The example below uses known distributions of frequency and severity.

Example: A pricing actuary must calculate the premium for car insurance for a customer.

Premium=

Expected number of claims * **expected amount of claim**

(ie, expected frequency) (ie, expected severity)

The pricing actuary has distributions for the number and cost of claims for a driver in a 6-month period covered by the policy. *Note: creating distributions like this would be an example of quantifying frequency and quantifying severity for your project.*

Claims Frequency Distribution		Claims Severity Distribution	
Number of Claims	Probability	Cost of Claim	Probability
0	65%	$3,000	20%
1	20%	$6,000	55%
2	10%	$9,000	15%
3	5%	$12,000	10%

According to these distributions, there is a 0% that a customer has more than 3 accidents in a 6-month period and the probability that the severity is more than $12,000 is also 0%. Computing the expected value:

Expectation of number of claims

(ie, expected frequency)

=(0*0.65)+(1*0.20)+(2*0.10)+(3*0.05)
= 0.55

Expectation of amount of claims

(ie, expected severity)

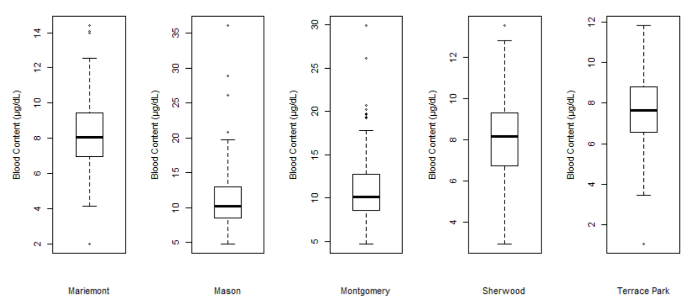
= ($3,000*0.20) + ($6,000*0.55)
+($9,000*0.15)+($12,000*0.10)
=$6,450

The formula used for premium by the pricing actuary is: premium = E(# of claims) * E(amount of claim)

Thus, the 6-month premium for a customer would be (0.55)*($6,450)=$3547.50, or $591.25 per month.

Task 3.1.1: Projecting Trends

Using the same Water Quality Scenario (See 2. Data Identification & Analysis Resource), consider the way the questions posed by the actuary exploring the dataset seeks to explore the effect changes in the lead content lead to

Questions	Response
The EPA wants to further understand the relationship between water quality and the lead content of a child's blood, based on a model using only these variables. What is the expected lead content of a child's blood if there were **no lead** in the water? What does a **1ppm increase** in the lead content of water mean for a household?	Regression equation: Blood content = 3.00683 + 0.44817 *[Water content] $\beta 0$ interpretation: When 0 ppm of lead is present in the water of a given household, the expected amount of lead in a child is expected to be 3.00683 µg/dL $\beta 1$ interpretation: On average, each 1.0 ppm increase in the water content of a household is associated with a .44817 µg/dL lead content increase in the blood of a child living there.
Is 'suburb' a significant variable in predicting lead content in blood? How do you know?	 If a partial F-test is conducted it will significance (p-value < 2.2E-16): However, a partial F-test is not required to understand significance. Students can mention difference in means and variances across different and variance as evidence of differences in blood lead concentration. For example, this box plot demonstrates the difference in distributions across each suburb, and the length of each box indicates varying variance as well as variable mean (demonstrated by the middle line on each plot).

Create models relating water quality to lead content in blood, this time creating a separate model for each suburb instead of using suburb as a variable. What's the new relationship between water quality and lead content in blood for each suburb?	This involves creating 5 equations using simple linear regression. Mariemont: Blood content = 0.4008*[Water Content] + 3.3218 Mason: Blood content = 0.2574*[Water Content] + 6.7351 Montgomery: Blood content = 0.1932*[Water Content] + 7.6944 Sherwood: Blood content = 0.2397*[Water Content] + 5.2155 Terrace Park: Blood content = 0.2449*[Water Content] + 4.9014
Using a model of the Montgomery data points, what's the expected lead content in a child's blood if there is 0 ppm of lead in the household's water?	Using the equation for Montgomery blood content, the expected lead content in a child's blood with 0 ppm of lead in the water is simply the intercept of the equation, β_0. β_0 = 7.6944 μg/dL

Suppose that the EPA evaluated the individual blood content data of these children across all suburbs and determined that 59 children (out of 750) needed to go to the hospital because of lead poisoning. Based on this fact, in what range did the EPA set the maximum allowed lead concentration for a child before they must be hospitalized?	59 children needed to be hospitalized, which means that 691 children didn't need to be hospitalized. If the children are ordered from lowest to highest based on the lead content in their blood, the 691st child has a lead content of 13.58509 µg/dL in their blood and does not have to go to the hospital, while the 692nd child has a lead content of 13.6509 µg/dL and must be hospitalized. Thus, we know that the cutoff must be greater than the lead content in the blood of the 691st child and less than or equal to the lead content in the blood of the 692nd child. **13.58509 µg/dL < EPA cutoff ≤ 13.6509 µg/dL** Or: **Cutoff range = (13.58509, 13.6509] µg/dL**