Empirical Bayes Analysis
For Safety

Larry Hagen, P.E., PTOE

“Driving Down Fatalities Through Knowledge Sharing”
Empirical Bayes for Dummies

De-mystify the Empirical Bayes method as often applied to traffic safety data.
My goal for this webinar:

Theory  Practice
If you really want to learn something...

TEACH IT
Answer the BIG questions:

- Who developed it?
- What does it do?
- When should I use it?
- Where should it not be used?
- Why should I use it?
- How do I use it?
Fact or Fiction?

Dr. Ezra Hauer developed the Empirical Bayes Method

FICTION
Who developed EB method?

Rev. Thomas Bayes
(c. 1702 – April 17, 1761)
Rev. Thomas Bayes publications:

He is known to have published two works in his lifetime, one theological and one mathematical:

1. Divine Benevolence, or an Attempt to Prove That the Principal End of the Divine Providence and Government is the Happiness of His Creatures (1731)

2. An Introduction to the Doctrine of Fluxions, and a Defence of the Mathematicians Against the Objections of the Author of The Analyst (published anonymously in 1736)
So where did EB come from?

Rev Bayes took a deep interest in probability late in his life.


It was published in *the Philosophical Transactions of the Royal Society of London*
"The purpose I mean is, to shew what reason we have for believing that there are in the constitution of things fixt laws according to which things happen, and that, therefore, the frame of the world must be the effect of the wisdom and power of an intelligent cause; and thus to confirm the argument taken from final causes for the existence of the Deity. It will be easy to see that the converse problem solved in this essay is more directly applicable to this purpose; for it shews us, with distinctness and precision, in every case of any particular order or recurrency of events, what reason there is to think that such recurrency or order is derived from stable causes or regulations in nature, and not from any irregularities of chance."
In recognition of Thomas Bayes's important work in probability,
the vault was restored in 1963 with contributions received
from statisticians throughout the world.

Mrs. Ann West, widow of the said Mr. Thomas West
and daughter of the said Joshua and Ann Bayes
21 Dec. 1758
Theodosia Bayes, wife of Samuel Bayes Esq., (68)
22 Sept. 1769
Samuel Bayes Esq., son of the said

Why use the EB method?

- If you don’t, you are considered naïve
- Presumed to improve precision of estimates
- Presumed to correct for regression-to-the-mean bias
What is the EB method?

Magic Pixie Dust!
Why use the EB method?

“The Empirical Bayes (EB) method for the estimation of safety increases the precision of estimation and corrects for the regression-to-mean bias. It is based on the recognition that accident counts are not the only clue to the safety of an entity. Another clue is in what is known about the safety of similar entities.”

Ezra Hauer, et al.
Note: Dr. Hauer uses “accident” rather than “crash” throughout his paper. I know better, but am just quoting his work verbatim where you see “accident.”
Why use the EB method?

Crashes are typically rare and random discrete events. They are typically infrequent for most intersections or roadway segments.
Why use the EB method?

In other words:

EB helps account for where we have a very small data size.
Example:

Consider a road segment where the expected crash frequency is 100 per year and you have 3 years of data.

The average yearly crash frequency can be estimated with a standard deviation computed as follows:

$$\sqrt{\frac{100}{3}} = \pm 5.8 \text{ crashes per year}$$

Thus, the standard deviation is 5.8% of the mean. Good precision.
Consider a road segment where the expected crash frequency is 1 every 10 years, you have 3 years of data.

The average yearly crash frequency can be estimated with a standard deviation computed as follows:

\[ \sqrt{\frac{0.1}{3}} = \pm 0.18 \text{ crashes per year} \]

Thus, the standard deviation is 180% of the mean. Poor precision.
What is “Regression to the Mean”

“The other shortcoming of safety estimates that are based only on accident counts is that they are subject to a common bias. For practical reasons one is often interested in the safety of entities that either require attention because they seem to have too many accidents, or merit attention because they have fewer accidents than expected. In both cases, were one to estimate safety using accident counts only, the estimate would be biased. The existence of this ‘regression-to-mean’ bias has been long recognized; it is known to produce inflated estimates of countermeasure effectiveness.”

Ezra Hauer, et al.
Low Cost Treatment Study

- Three intersections in Detroit were treated
- Crash reductions were 44%, 48% and 57%
- However, they were chosen due to high crashes
- Crash frequency, rate, or severity higher than 95% of the intersections
- No correction for RTM bias was applied
Why use the EB method?

“The Empirical Bayes (EB) method for the estimation of safety increases the precision of estimation and corrects for the regression-to-mean bias. It is based on the recognition that accident counts are not the only clue to the safety of an entity. Another clue is in what is known about the safety of similar entities.”

Ezra Hauer, et al.
Sensible Estimate of Safety

“A sensible estimate must be a mixture of the two clues. Similarly, to estimate the safety of a specific segment of, say a rural two-lane road, one should use not only the accident counts for this segment, but also the knowledge of the typical accident frequency of such roads in the same jurisdiction.”

Ezra Hauer, et al.
So what does that really mean?
Basic statistics tells us...

A sample size of at least 30 is needed to be statistically significant.
What do we typically have?

- 3 – 5 years of crash data?
- 3 – 5 years of ADT data?
- 3 – 5 years of geometric data?

3 – 5 data points

Can we draw statistically valid conclusions?
How do we overcome that?

Magic Pixie Dust!
Add More Data
Where do we get more data?

- Wait a while and collect more data
- Gather data from similar sites
EB Method assumes...

The crash experience for thirty years at one site is approximately equal to the crash experience for one year at thirty similar sites.
EB Method assumes...

\[
\frac{\text{Crash Experience } 100 \text{ years}}{2} = \text{Crash Experience } 100 \text{ sites}
\]
EB Method assumes...

The crash experience is really **not** random, but is instead predictable and deterministic.
Short Term vs Long Term

Expected Average Crash Frequency

Short Term Average Crash Frequency

Observed Crash Frequency

Year
Regression to the Mean

- Site Selected for Treatment Based on Short Term Trend
- Expected Average Crash Frequency (Without Treatment)
- Observed Crash Frequency
- Year

Perceived Effectiveness of Treatment

RTM Reduction

Actual Reduction Due to Treatment

“Driving Down Fatalities Through Knowledge Sharing”
How do we use EB?

“The task is to make joint use of two clues to the safety of an entity: the accident record of that entity and the accident frequency expected at similar entities.”

Ezra Hauer, et al.

The expected crash frequency at similar entities is determined by the Safety Performance Function (SPF)
The EB Procedure

\[ N_{\text{expected}} = \omega \times N_{\text{predicted}} + (1-\omega) \times N_{\text{observed}} \]

Where:

- \( N_{\text{expected}} \) = expected average crashes
- \( \omega \) = weighting adjustment to SPF prediction
- \( N_{\text{predicted}} \) = predicted average crash frequency (SPF)
- \( N_{\text{observed}} \) = observed crash frequency
Weighting adjustment factor

\[ W = \frac{1}{1 + k \times \sum N_{\text{predicted}}} \]

Where: \( k = \text{overdispersion parameter from the SPF} \)
Example:

- Two-lane rural road segment
- Segment length is 1.7 miles
- AADT for the last 3 years is 6,500
- Crashes are 6, 9, and 4 for this period
- Assume base conditions for SPF
- What is $N_{\text{expected}}$?
The EB Procedure

\[ N_{\text{expected}} = w \times N_{\text{predicted}} + (1-w) \times N_{\text{observed}} \]

Where:
- \( N_{\text{expected}} \) = expected average crashes
- \( w \) = weighting adjustment to SPF prediction
- \( N_{\text{predicted}} \) = predicted average crash frequency (SPF)
- \( N_{\text{observed}} \) = observed crash frequency
Weighting adjustment factor

\[ w = \frac{1}{1 + k \times \sum_{\text{predicted}} N_{\text{all study years}}} \]

Where: \( k = \) overdispersion parameter from the SPF
Find the Predicted N (SPF)

\[ N_{\text{spf rs}} = AADT \times L \times 365 \times 10^{-6} \times e^{(-0.312)} \]

Where:
- \( N_{\text{spf rs}} \) = expected average crashes
- \( AADT \) = Average Annual Daily Traffic
- \( L \) = length of roadway segment (miles)
Find the Predicted N (SPF)

\[ N_{spf\ rs} = AADT \times L \times 365 \times 10^{-6} \times 0.732 \]

Where:
- \( N_{spf\ rs} \) = expected average crashes
- \( AADT \) = Average Annual Daily Traffic
- \( L \) = length of roadway segment (miles)

\[ N_{spf\ rs} = 6,500 \times 1.7 \times 365 \times 10^{-6} \times 0.732 \]

\[ N_{spf\ rs} = 2.95 \text{ crashes per year} \]
We’ll skip this step...

\[ N_{predicted} = N_{spf rs} \times (CMF_1 \times CMF_2 \times \ldots) \times C_x \]

\[ N_{predicted} = N_{spf rs} = 2.95 \]
Overdispersion parameter:

\[ k = \frac{0.236}{L} \]

Source: Highway Safety Manual

\[ k = \frac{0.236}{1.7} \]

\[ k = 0.139 \]
Compute weighting factor

\[ w = \frac{1}{1 + k \times \sum_{\text{all study years}} N_{\text{predicted}}} \]

\[ w = \frac{1}{1 + 0.139 \times [3 \times 2.95]} = 0.449 \]
Compute expected crashes

\[ N_{\text{expected}} = w \times N_{\text{predicted}} + (1-w) \times N_{\text{observed}} \]

\[ N_{\text{expected}} = 0.449 \times 2.95 + (1-0.449) \times (6+9+4)/3 \]

\[ N_{\text{expected}} = 4.81 \text{ crashes per year} \]
Where should EB not be used?

“The EB Method is only applicable when both predicted and observed crash frequencies are available for the specific roadway network conditions for which the estimate is being made.”

“Where only a predicted or only observed crash data are available, the EB Method is not applicable.”

Source: Highway Safety Manual
What SPF’s are available?

- Rural Two-Lane Roads
  - Segments
  - Intersections
- Rural Multilane Highways
  - Segments
  - Intersections
- Urban and Suburban Arterials
  - Segments
  - Intersections
- Freeways

Source: Highway Safety Manual
Rural Multilane Highways

“The term “multilane” refers to facilities with four through lanes.”

“Facilities with six or more lanes are not covered in Chapter 11.”

Source: Highway Safety Manual
Urban and Suburban Arterials

- Two-lane undivided arterials
- Three-lane arterials with center TWLTL
- Four-lane undivided arterials
- Four-lane divided arterials
- Five-lane arterials with center TWLTL

Source: Highway Safety Manual
Answer the BIG questions:

- Who developed it?  Rev Thomas Bayes
- What does it do?  Correct for RTM
- When should I use it?  When data and SPF available
- Where should it not be used?  Where no SPF exists
- Why should I use it?  Better precision
- How do I use it?  Very carefully!
References:

- Estimating Safety by the Empirical Bayes Method: A Tutorial
  http://www.ctre.iastate.edu/educweb/ce552/docs/Bayes_tutor_hauer.pdf

- Rev Thomas Bayes
  http://en.wikipedia.org/wiki/Thomas_Bayes

- Weighting Using the Empirical Bayes Method
  Highway Safety Manual, Volume 1, Section 3.5.5
"That's all Folks!"
Don’t forget your PDH form...

- Email completed form to:
  Larry@HagenConsultingServices.com

- Fax completed form to
  866-426-5153 (toll free)
PDH Request Form

Please fill out this form to receive one (1) Professional Development Hour for attending one session of the District 7 Local Agency Traffic Safety Academy. Your PDH will be reported to the Florida Board of Professional Engineers.

By filling out and submitting this form, on my honor as a licensed Florida Professional Engineer, I hereby certify that I have attended this workshop session for which I am requesting PDH credit.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Florida P.E. number:</td>
<td>43968</td>
</tr>
<tr>
<td>Name as it appears on your license:</td>
<td>Lawrence T. Hagen</td>
</tr>
<tr>
<td>E-mail address</td>
<td><a href="mailto:Larry@HagenConsultingServices.com">Larry@HagenConsultingServices.com</a></td>
</tr>
</tbody>
</table>

“Driving Down Fatalities Through Knowledge Sharing”
Questions / Comments?

Larry@HagenConsultingServices.com

“Driving Down Fatalities Through Knowledge Sharing”