An Integrated Component of Your Study Manual Program

Actuarial University
Your Path to Success

GOAL
Practice. Quiz. Test. PASS

Study Manual

Virtual eFlashcards

Videos & More!

# a/s/m

# Exam MAS-II Study Manual

6th Edition

Abraham Weishaus, Ph.D., FSA, CFA, MAAA

a/s/m
Actuarial Study Materials
Learning Made Easier

# a/s/m

# Exam MAS-II Study Manual

6th Edition

Abraham Weishaus, Ph.D., FSA, CFA, MAAA

**TO OUR READERS:**

Please check A.S.M.'s web site at **www.studymanuals.com** for errata and updates. If you have any comments or reports of errata, please e-mail us at **mail@studymanuals.com**.

# Welcome to Actuarial University

Actuarial University is a reimagined platform built around a more simplified way to study. It combines all the products you use to study into one interactive learning center.

**You can find integrated topics using this network icon.**

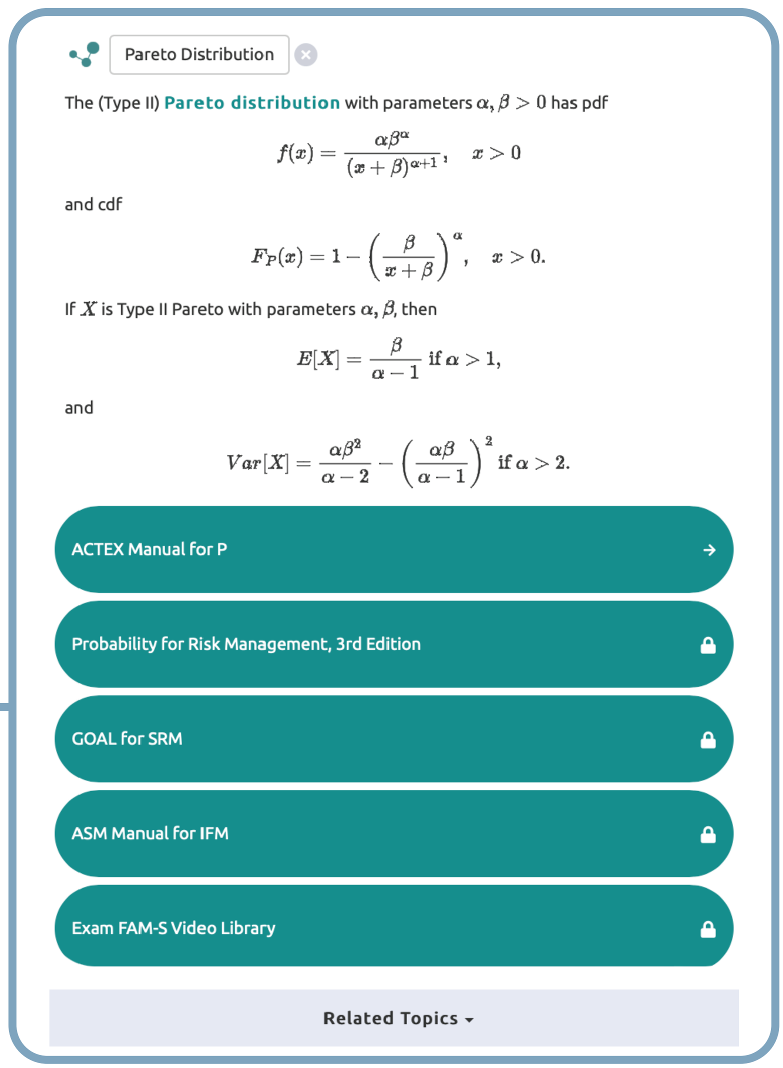
When this icon appears, it will be next to an important topic in the manual. Click the **link** in your digital manual, or search the <u>underlined topic</u> in your print manual.

1. Login to: www.actuarialuniversity.com

2. Locate the **Topic Search** on your exam dashboard and enter the word or phrase into the search field, selecting the best match.

3. A topic "**Hub**" will display a list of integrated products that offer more ways to study the material.

4. Here is an example of the topic **Pareto Distribution:**



Within the **Hub** there will be unlocked and locked products.

**Unlocked Products** are the products that you own.

ACTEX Manual for P →

**Locked Products** are products that you do not own, and are available for purchase.

Probability for Risk Management, 3rd Edition

Many of Actuarial University's features are already unlocked with your study program, including:

| | |
|---|---|
| **Instructional Videos*** | **Planner** |
| **Topic Search** | **Formula & Review Sheet** |

## Make your study session more efficient with our Planner!

**Planner**

**Template** ACTEX FM Study Manual - New 2022 syllabus

**Begin Study** 07/01/2023 **End Study** 11/14/2023

| | | |
|---|---|---|
| ✔ | 7/1/2023 - 7/16/2023 | Interest Rates and the Time Value of Money |
| ✔ | 7/16/2023 - 8/12/2023 | Annuities |
| ✔ | 8/12/2023 - 8/27/2023 | Loan Repayment |
| ✔ | 8/27/2023 - 9/15/2023 | Bonds |
| ✔ | 9/15/2023 - 9/22/2023 | Yield Rate of an Investment |
| ✔ | 9/22/2023 - 10/11/2023 | The Term Structure of Interest Rates |
| ✔ | 10/11/2023 - 10/30/2023 | Asset-Liability Management |

**Available standalone, or included with the Study Manual Program Video Bundle*

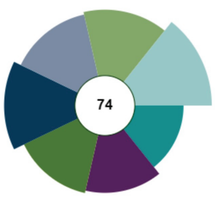
GOAL

# Track your exam readiness with GOAL Score!

**Available for P, FM, FAM, FAM-L, FAM-S, ALTAM, ASTAM, MAS-I, MAS-II, & CAS 5**

**GOAL Score** tracks your performance through GOAL Practice Sessions, Quizzes, and Exams, resulting in an aggregate weighted score that gauges your exam preparedness.

By measuring both your performance, and the consistency of your performance, **GOAL Score** produces a reliable number that will give you confidence in your preparation before you sit for your exam.




Your GOAL Score
74

Target GOAL Score
70

If your GOAL Score is a 70 or higher, you are well-prepared to sit for your exam!

| | | |
|---|---|---|
| 64<br>A. Multiple State Models<br>Weighting: 20% | 71<br>B. Multiple Decrement Models<br>Weighting: 10% | 73<br>C. Multiple Life Models<br>Weighting: 10% |
| 82<br>D. Profit Analysis<br>Weighting: 15% | 76<br>E. Universal Life Insurance<br>Weighting: 15% | 78<br>F. Embedded Options and Equity-Linked Insurance<br>Weighting: 15% |
| 87<br>G. Pension Plans and Retirement Benefits<br>Weighting: 15% | | |

See key areas where you can improve.

A. Multiple State Models — A1. Multiple State Probabilities

| Difficulty | Mean | Proportion |
|---|---|---|
| Core | 38 | 22 / 48 |
| Advanced | 38 | 18 / 48 |
| Mastery | 38 | 8 / 48 |

Detailed performance tracking.

GOAL Session Activity Summary — Filter: Quizzes — Complete

| Created | Last Accessed | Completed | Mode | Categories | Questions | Status | Action |
|---|---|---|---|---|---|---|---|
| 03/24/2023 | 03/24/2023 | 03/24/2023 | Quiz | G. Pension Plans and Retirement Benefits | 15 | Complete | Review |
| 03/24/2023 | 03/24/2023 | 03/24/2023 | Quiz | C1. Joint Life/Last Survivor Probabilities, C2. Contingent Probabilities, C3. Expectation and Variance, C6. ML Models with Transition Intensities, C4. ML Models: Premiums and Reserves (Discrete), C5. ML Models: Premiums and Reserves (Continuous) | 15 | Complete | Review |
| 03/24/2023 | 03/24/2023 | 03/24/2023 | Quiz | F. Embedded Options and Equity-Linked Insurance | 20 | Complete | Review |
| 03/24/2023 | 03/24/2023 | 03/24/2023 | Quiz | D. Profit Analysis | 10 | Complete | Review |
| 09/19/2022 | 09/19/2022 | 09/19/2022 | Quiz | G4. Pension Plans and Retirement Benefits (Written Answers) | 2 | Complete | Review |

Quickly return to previous sessions.

# Contents

# Introduction

Welcome to Exam MAS-II!

This exam covers four subjects: credibility, linear mixed models, statistical learning, and time series, The syllabus lists them in that order. You can study the subjects in any order. In this manual, time series comes right after credibility, because there are brief references to AR models in linear mixed models and statistical learning.

Credibility calculations have been done for a long time; the computations can be done on a calculator. ARIMA time series analysis has been around for more then 50 years. But linear mixed models and statistical learning methods have only become practical relatively recently with the availability of fast computers.

What are the weights of these subjects on the exam? According to the course outline for the course:

**Table 1:** Weights of the Four Subjects on MAS-II

| Subject | Syllabus Weight |
|---|---|
| Credibility | 15–25% |
| Linear mixed models | 10–20% |
| Statistical learning | 40–50% |
| Time series | 15–25% |

## About the exam

At this writing (December 2023), the course outline for the exam is at

https://www.casact.org/sites/default/files/2023-06/MASII_Content_Outline.pdf.

The tables will be the same as they were in the past:

https://www.casact.org/sites/default/files/2021-03/masii_tables.pdf.

except that normal and chi-square distributions are not needed since you will get a spreadsheet having those functions.

Monitor the CAS website for any additional information. Perhaps they'll post new sample questions or updated tables.

CAS exams frequently had defective questions when they were printed. But with the move to computer-based testing (CBT), questions will probably be proofread more closely, and defects may show up and be fixed while questions are still pilot questions. I expect defective questions to be less common in the future.

Talking about CBT, you get a one-sheet spreadsheet with Excel functions at the exam. See

https://wsr.pearsonvue.com/testtaker/common/StartTestLaunch.htm?clientCode=CAS&seriesCode=Spreadsheet_SAMPLE&languageCode=ENU

The Excel functions available are listed here:

https://home.pearsonvue.com/Clients/Casualty-Actuarial-Society-(CAS)/Spreadsheet-Function-List.aspx

There will probably not be questions with calculations so complex that they would require a spreadsheet.

As a result of CBT, "the order of questions will be randomly presented within the various sections of the syllabus". I interpret this to mean that the questions for each section of the syllabus will still be kept together, but they will be in a random order within each section.

## This manual

This manual has everything you need to score a 10 on your exam. But (with the exception of credibility) it does not have enough material to do the programming that would be necessary at your job. For that, you need the textbooks, plus a good knowledge of R or another statistical programming language.

Based on Table 1, credibility is officially about 20% of the exam. However, unlike the other subjects on this exam, credibility has been tested for many decades. As a result, there are tons of old exam questions available. I've taken the old material, discarded about 45% of it, and updated the notation to match the syllabus textbook, but it is still a lot of material relative to its weight on the exam. In the introduction to credibility, I give guidance on which lessons are most important. That way if you have a lot of time and want to make sure you get all 6–10 credibility questions right, you have everything you need to do that. But if you're aiming to get a good score on the exam and don't mind missing 1 or 2 credibility questions, you can just study the important parts, do maybe half the exercises, and move on.

In contrast, you'll want to spend a lot of time on Statistical Learning, the last but not least (in fact, most important) subject. You should work out every exercise there.

## Acknowledgements

I would like to thank the SOA and CAS for allowing me to use their old and sample exam questions.

The creators of TEX, LATEX, and its multitude of packages all deserve thanks for making possible the professional typesetting of this mathematical material.

I would like to thank Nao Mimoto for proofreading the linear mixed models part of this manual.

I would like to thank the readers who submitted errata: Pete Abbate, Chris Bellwood, Tiffany Chang, Carrie Cui, Kyle Cusick, Boaz Haberman, Calvin Haught, Chieng Hin, Michelle Lau, Tzong Her Lee, Akiva Levy, Jonathan Lim, Trevor McElhinney, Christopher Ng, Hannah Park, Isabella Peakes, Rachel Ruble, Casey Schnakenberg, Amy Schroeder, Yuliya Semibratova, Alex Shepard, Allen Sherpa, Roxanne Sylvain, Amanda Van Vleet, Yongkang Wang, Matthew Wessler, Joe Wooton, Joseph Zalot, Luke Zhang, Xiaolu Zhang.

## Errata

Please report all errors you find in this manual to the author. You may send them to the publisher at mail@studymanuals.com or directly to me at errata@aceyourexams.net. Please identify the manual, edition, and printing the error is in. This is the sixth edition of the Exam MAS-II manual.

An errata list will be posted at http://errata.aceyourexams.net. Check this errata list frequently.

# Lesson 2

# Classical Credibility: Non-Poisson Frequency

**Reading:** Tse 6.4

Undoubtedly you're scratching your head. What is this lesson doing here? Tse Section 6.4 is not on the syllabus! OK, feel free to skip this lesson, but be aware that the Fall 2019 exam had a question on it. (See exercise 2.10.) This question, in fact, is very similar to Tse Example 6.12, which is in Section 6.4.

As we saw in the previous lesson (formula (1.1)), the general formula for the **standard for full credibility** in ***exposure units*** is

$$e_F = \lambda_F \left(\frac{\sigma}{\mu}\right)^2$$

where $e_F$ is the standard *measured in exposure units, not claims*, $\mu$ is the mean and $\sigma$ is the standard deviation of the item you are measuring the credibility of. To obtain $n_F$, the standard measured in expected claims, we multiply $e_F$ by the mean frequency of claims, $\mu_N$. This general formula is used to determine the number of exposure units (e.g., policy years) needed for full credibility of the pure premium if you're only given the mean and variance of aggregate claims, but not the separate means and variances of frequency and severity.

If you are establishing a standard for full credibility of claim frequency in terms of the number of exposures, formula (1.1) translates into

$$e_F = \lambda_F \left(\frac{\sigma_N^2}{\mu_N^2}\right) \tag{2.1}$$

where $\mu_N$ is the mean of frequency and $\sigma_N^2$ is the variance of frequency. To express this standard in terms of number of expected claims, we multiply both sides by $\mu_N$ to obtain

$$n_F = \lambda_F \left(\frac{\sigma_N^2}{\mu_N}\right) \tag{2.2}$$

Notice how this generalizes the formula of the previous lesson for Poisson-distributed claim frequency, where $\sigma_N^2 = \mu_N$.

If you are establishing a standard for full credibility of aggregate loss or pure premium, assuming that claim counts and claim sizes are independent, the formula for the (expected) number of claims needed for full credibility can be derived as follows. The **mean of pure premium** is

$$\mathbf{E}[S] = \mathbf{E}[N]\,\mathbf{E}[X] = \mu_N \mu_X$$

By the **compound variance formula**, equation (1.5):

$$\mathrm{Var}(S) = \mathbf{E}[N]\,\mathrm{Var}(X) + \mathrm{Var}(N)\,\mathbf{E}[X]^2 = \mu_N \sigma_X^2 + \sigma_N^2 \mu_X^2$$

From formula (1.1), noting that $n_F = \mu_N e_F$:

$$\begin{aligned} n_F &= \mu_N \lambda_F \frac{\mu_N \sigma_X^2 + \sigma_N^2 \mu_X^2}{\mu_N^2 \mu_X^2} \\ &= \lambda_F \left(\frac{\sigma_N^2}{\mu_N} + C_X^2\right) \end{aligned} \tag{2.3}$$

**Table 2.1:** Classical credibility formulas

| | Credibility for | | |
|---|---|---|---|
| Experience expressed in | Number of claims | Claim size (severity) | Aggregate losses/ Pure premium |
| Policyholder-years | $\lambda_F\left(\dfrac{\sigma_N^2}{\mu_N^2}\right)$ | N/A | $\lambda_F\left(\dfrac{\sigma_N^2}{\mu_N^2}+\dfrac{\sigma_X^2}{\mu_X^2\mu_N}\right)$ |
| Number of claims | $\lambda_F\left(\dfrac{\sigma_N^2}{\mu_N}\right)$ | $\lambda_F\left(\dfrac{\sigma_X^2}{\mu_X^2}\right)$ | $\lambda_F\left(\dfrac{\sigma_N^2}{\mu_N}+\dfrac{\sigma_X^2}{\mu_X^2}\right)$ |

Notice the asymmetry: the denominator for frequency is $\mu_N$ not squared, whereas $C_X^2 = \sigma_X^2/\mu_X^2$. This standard can be expressed in terms of exposure units by dividing $n_F$ by $\mu_N$.

**Example 2A** You are given:

- Claim counts follow a negative binomial distribution with $r = 2$ and $\beta = 0.5$.
- Claim sizes have coefficient of variation equal to 2.5.
- Claim counts and claim sizes are independent.

The standard for full credibility of aggregate losses is set so that actual aggregate losses are within 5% of expected 95% of the time.

Determine the number of expected claims needed for full credibility. ■

**Solution:** The ratio of the variance of a negative binomial, $r\beta(1+\beta)$, to its mean, $r\beta$ is $1+\beta$, which is 1.5 here. Using formula (2.3),

$$n_F = \left(\frac{1.96}{0.05}\right)^2 (1.5 + 2.5^2) = \boxed{\mathbf{11{,}909}}$$
□

From this example, you see that $\sigma_N^2/\mu_N$ for a negative binomial is $1+\beta$. It may also be useful to know that for a binomial, $\sigma_N^2/\mu_N = 1-q$.

A summary of the formulas for all possible combinations of experience units used and what the credibility is for is shown in Table 2.1. To make the formulas parallel, I've avoided using the coefficient of variation. The two most common formulas are shaded.

## Exercises

**2.1.** The methods of classical credibility are used. The standard for full credibility is that the item measured should be within $100k\%$ of the true mean with probability $p$.

Order the following items from lowest to highest. In each case, assume the distribution is non-degenerate (in other words, that the random variable is not a constant).

I. Number of losses for full credibility of individual losses if losses follow an exponential distribution.

II. Number of expected claims for full credibility of claim counts if claim counts follow a binomial distribution with $m = 1$.

III. Number of expected claims for full credibility of aggregate losses if claim counts follow a Poisson distribution and loss sizes follow a two-parameter Pareto distribution.

A. I < II < III  B. I < III < II  C. II < I < III  D. II < III < I  E. III < II < I



Exercises continue on the next page . . .

**2.2.** Claim size follows a two parameter Pareto distribution with parameters $\alpha = 3.5$ and $\theta$. The full credibility standard for claim size is set so that actual average claim size is within 5% of expected claim size with probability 98%.

Determine the number of claims needed for full credibility.

**2.3.** Claim size follows an inverse Gaussian distribution with parameters $\mu = 1000$, $\theta = 500$. You set a standard for full credibility of claim size, using the methods of classical credibility, so that expected claim size is within $100k$% of actual claim size with probability 90%. Under this standard, 1500 claims are needed for full credibility.

Determine $k$.

**2.4.** You are given the following information for a risk.

| | |
|---|---|
| Claim frequency: | mean = 0.2, variance = 0.3 |
| Claim severity: | gamma distribution with $\alpha = 2$, $\theta = 10{,}000$. |

Using the methods of classical credibility, determine the number of expected claims needed so that aggregate claims experienced are within 5% of expected claims with probability 90%.

A. Less than 1800
B. At least 1800, but less than 1900
C. At least 1900, but less than 2000
D. At least 2000, but less than 2100
E. At least 2100

**2.5.** For a certain coverage, claim frequency has a negative binomial distribution with $\beta = 0.25$. The full credibility standard is set so that the actual number of claims is within 6% of the expected number with probability 95%.

Determine the number of expected claims needed for full credibility.

**2.6.** **[4B-F94:15]** (3 points) You are given the following:

- $Y$ represents the number of independent homogeneous exposures in an insurance portfolio.
- The claim frequency *rate* per exposure is a random variable with mean = 0.025 and variance = 0.0025.
- A full credibility standard is devised that requires the observed sample frequency *rate* per exposure to be within 5% of the expected population frequency rate per exposure 90% of the time.

Determine the value of $Y$ needed to produce full credibility for the portfolio's experience.

A. Less than 900
B. At least 900, but less than 1500
C. At least 1500, but less than 3000
D. At least 3000, but less than 4500
E. At least 4500



*Exercises continue on the next page . . .*

**2.7.** [**4-F04:21**, **STAM Sample Question #148**] You are given:

- The number of claims has probability function:

$$p(x) = \binom{m}{x} q^x (1-q)^{m-x} \qquad x = 0, 1, 2, \ldots, m$$

- The actual number of claims must be within 1% of the expected number of claims with probability 0.95.
- The expected number of claims for full credibility is 34,574.

Determine $q$.

A. 0.05 B. 0.10 C. 0.20 D. 0.40 E. 0.80

**2.8.** [**4-F00:14**] For an insurance portfolio, you are given:

- For each individual insured, the number of claims follows a Poisson distribution.
- The mean claim count varies by insured, and the distribution of mean claim counts follows a gamma distribution.
- For a random sample of 1000 insureds, the observed claim counts are as follows:

| Number Of Claims, $n$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Number Of Insureds, $f_n$ | 512 | 307 | 123 | 41 | 11 | 6 |

$$\sum n f_n = 750 \qquad \sum n^2 f_n = 1494$$

- Claim sizes follow a Pareto distribution with mean 1500 and variance 6,750,000.
- Claim sizes and claim counts are independent.
- The full credibility standard is to be within 5% of the expected aggregate loss 95% of the time.

Determine the minimum number of insureds needed for the aggregate loss to be fully credible.

A. Less than 8300
B. At least 8300, but less than 8400
C. At least 8400, but less than 8500
D. At least 8500, but less than 8600
E. At least 8600



*Exercises continue on the next page . . .*

**2.9.** [**C-S05:2**, **STAM Sample Question #173**] You are given:

- The number of claims follows a negative binomial distribution with parameters $r$ and $\beta = 3$.
- Claim severity has the following distribution:

| Claim Size | Probability |
|---|---|
| 1 | 0.4 |
| 10 | 0.4 |
| 100 | 0.2 |

- The number of claims is independent of the severity of claims.

Determine the expected number of claims needed for aggregate losses to be within 10% of expected aggregate losses with 95% probability.

A. Less than 1200
B. At least 1200, but less than 1600
C. At least 1600, but less than 2000
D. At least 2000, but less than 2400
E. At least 2400

**2.10.** [**MAS-II-F19:4**] You are given the following parameters.

- Assume the full-credibility standard using classical credibility is based on probability parameter 0.05 and range parameter 0.02.
- The expected claim frequency per exposure unit is 0.03.
- $W$ is the full-credibility standard for claim frequency in exposure units assuming Poisson claim frequency.
- $V$ is the full-credibility standard for claim frequency in exposure units assuming Bernoulli claim frequency.

Calculate $|W - V|$.

A. Less than 4,000
B. At least 4,000, but less than 6,000
C. At least 6,000, but less than 8,000
D. At least 8,000, but less than 10,000
E. At least 10,000

# Solutions

**2.1.** For exponential $X$ with mean $\theta$, the variance is $\theta^2$ so the coefficient of variation squared is $\theta^2/\theta^2 = 1$ and the standard for full credibility, using the general formula, is $n_0 C^2 = n_0$.

For a binomial with parameters $m = 1$ and $q$, the coefficient of variation squared is $q(1-q)/q^2 = (1-q)/q$. The number of exposures for full credibility, by the general formula, is $n_0(1-q)/q$, and the number of expected claims per exposure is $q$, so the number of expected claims needed for full credibility is $n_0(1-q)$. Since $q > 0$, this is less than I. (If $q = 0$, the distribution is degenerate; the random variable is the constant 0.)

For a two-parameter Pareto, the standard for full credibility in terms of expected claims is $2n_0(\alpha-1)/(\alpha-2)$, based on Table 1.2. This is greater than $2n_0$, hence greater than I. **(C)**

**2.2.** Since we want credibility for *severity* using number of claims (exposure), formula (1.7) is the appropriate one. We calculate $C_X^2$.

$$\mu_X = \frac{\theta}{2.5}$$

$$\sigma_X^2 = \frac{2\theta^2}{(2.5)(1.5)} - \frac{\theta^2}{2.5^2} = \frac{28}{75}\theta^2$$
$$C_X^2 = \left(\frac{\sigma}{\mu}\right)^2 = \frac{28}{75}\left(\frac{25}{4}\right) = \frac{7}{3}$$

Using the formula, we conclude

$$n_F = \left(\frac{2.326}{0.05}\right)^2\left(\frac{7}{3}\right) = \boxed{\mathbf{5049.6}}$$

**2.3.** Since we want credibility for *severity* using number of claims (exposure), formula (1.7) is the appropriate one. To back out the $k$, we must first calculate $C_X$.

$$\mu_X = 1000$$
$$\mu_X = \frac{10^9}{500}$$
$$C_X^2 = 2$$

Now we equate 1500 to $e_F$.

$$1500 = \left(\frac{1.645}{k}\right)^2(2)$$
$$1500k^2 = (1.645^2)(2)$$
$$750k^2 = 1.645^2$$
$$k = \frac{1.645}{\sqrt{750}} = \boxed{\mathbf{0.06}}$$

**2.4.** We want credibility for aggregate claims using number of expected claims as the basis. With separate information on frequency and severity, formula (2.3) applies. We must calculate $C_X$.

$$\mu_X = 2(10{,}000) = 20{,}000$$
$$\mu_X = 2(10{,}000^2) = 2\cdot 10^8$$
$$C_X^2 = \frac{2\cdot 10^8}{20{,}000^2} = \frac{1}{2}$$

We now apply formula (2.3).

$$n_F = 1082.41\left(\frac{0.3}{0.2} + \frac{1}{2}\right) = \boxed{\mathbf{2164.82}} \qquad \textbf{(E)}$$

**2.5.** We want credibility for frequency using number of expected claims as the basis. Formula (2.2) applies.

$$\frac{\sigma_N^2}{\mu_N} = 1.25$$
$$n_F = \left(\frac{1.96}{0.06}\right)^2(1.25) = \boxed{\mathbf{1333.89}}$$

**2.6.** We want credibility for frequency using exposures as the basis. Formula (2.2) applies.

$$n_F = 1082.41\left(\frac{0.0025}{0.025}\right) = 108.241$$
$$Y = e_F = \frac{108.241}{0.025} = \boxed{\mathbf{4329.64}} \qquad \textbf{(D)}$$

**2.7.** We have

$$\lambda_F = \left(\frac{1.96}{0.01}\right)^2\left(\frac{\sigma^2}{\mu}\right)$$

and for a binomial, $\sigma^2 = mq(1-q)$ and $\mu = mq$ so the quotient $\sigma^2/\mu = 1-q$. Then

$$\begin{aligned} 196^2(1-q) &= 34{,}574 \\ 38{,}416(1-q) &= 34{,}574 \\ q &= 1 - \frac{34{,}574}{38{,}416} = \boxed{0.1} \qquad \textbf{(B)} \end{aligned}$$

**2.8.** The coefficient of variation for severity squared is $6{,}750{,}000/1500^2 = 3$. For frequency, we use the summary statistics to estimate the variance over the mean. Estimated mean is $750/1000 = 0.75$. Estimated variance is $1494/1000 - 0.75^2 = 0.9315$. If you wish, you can multiply this by $1000/999$ (so that the sample variance is divided by $n-1$ instead of by $n$), but it hardly makes a difference. So we have

$$\begin{aligned} n_F &= \left(\frac{1.96}{0.05}\right)^2\left(\frac{0.9315}{0.75} + 3\right) = 6518.43 \\ e_F &= \frac{6518.43}{0.75} = \boxed{8691.24} \qquad \textbf{(E)} \end{aligned}$$

**2.9.** The variance of number of claims divided by the mean is $1+\beta = 4$.

The mean of claim size is $0.4(1)+0.4(10)+0.2(100) = 24.4$. The second moment is $0.4(1)+0.4(100)+0.2(10{,}000) = 2040.4$. The variance is then $2040.4 - 24.4^2 = 1445.04$, and the coefficient of variation squared is $1445.04/24.4^2 = 2.4272$. The answer is then

$$n_F = \left(\frac{1.96}{0.1}\right)^2(4 + 2.4272) = \boxed{2469} \qquad \textbf{(E)}$$

**2.10.** First calculate $n_0$.

$$n_0 = \left(\frac{z_{0.975}}{0.02}\right)^2 = \left(\frac{1.96}{0.02}\right)^2 = 9604$$

For Poisson frequency, this represents number of expected claims needed; divide it by 0.03 to obtain exposure units.

$$W = \frac{9604}{0.03}$$

For binomial (with $m = 1$, or Bernoulli) frequency, first multiply by variance divided by mean to obtain expected claims needed, then divide by 0.03 to obtain exposure units. So multiply by $(0.03)(1-0.03)/0.03^2$.

$$V = 9604\left(\frac{1-0.03}{0.03}\right) = \frac{9604}{0.03} - 9604$$

Then $W - V = \boxed{9604}$. **(D)**

# Lesson 3

# Classical Credibility: Partial Credibility

**Reading:** Tse 6.3

When there is inadequate experience for full credibility, then the credibility estimate, which we will sometimes call the *credibility premium*, or $P_C$, is calculated by

$$P_C = Z\bar{X} + (1 - Z)M \tag{3.1}$$

where $M$ is the manual premium, the premium initially assumed if there is no credibility, and $Z$ is the **credibility factor**. For calculator purposes, it is easier to use this formula in the form

$$P_C = M + Z(\bar{X} - M) \tag{3.2}$$

since you don't need any memory, and entering $M$ twice is likely to be easier than entering $Z$, which is usually a long decimal, twice. This alternative form is also intuitive; you are modifying $M$ by adding the difference between actual experience and $M$, multiplied by the credibility assigned to the experience.

We want to determine $Z$.

We saw in the story of Ventnor Manufacturing at the beginning of Lesson 1 that to multiply the variance of the results by $\alpha$, we must multiply the results by $\sqrt{\alpha}$. Therefore, the credibility factor for $n$ expected claims is

$$Z = \sqrt{\frac{n}{n_F}} \tag{3.3}$$

where $n_F$ is the number of expected claims needed for full credibility. The corresponding **square root rule** would apply to expressing credibility in exposures in terms of $e_F$, or credibility in terms of aggregate claims in terms of the amount needed for full credibility:

$$Z = \sqrt{\frac{e}{e_F}}$$

The partial credibility function is concave down; it grows rapidly for small numbers, then slows down. Figure 3.1 illustrates the curve if we assume 1082 claims are needed for full credibility.

Let's see how the Ventnor case fits into this formula. We established on page 7 that 1125 expected claims were needed for full credibility. We have 160 claims. Therefore $Z = \sqrt{160/1125} = 0.3771$, which matches the result we initially computed on page 3.

**Quiz 3-1**

If 250 expected claims result in 50% credibility, how many expected claims are needed for 20% credibility?

**Example 3A** [Version of **4B-S91:23**] (1 point) Claim counts for a group follow a Poisson distribution. The standard for full credibility is 19,544 expected claims. We observe 6000 claims and a total loss of 15,600,000 for a group of insureds.

If our prior estimate of the total loss is 16,500,000, determine the classical credibility estimate of the total loss for the group of insureds.

A. Less than 15,780,000

B. At least 15,780,000, but less than 15,870,000

C. At least 15,870,000, but less than 15,960,000

D. At least 15,960,000, but less than 16,050,000

E. At least 16,050,000

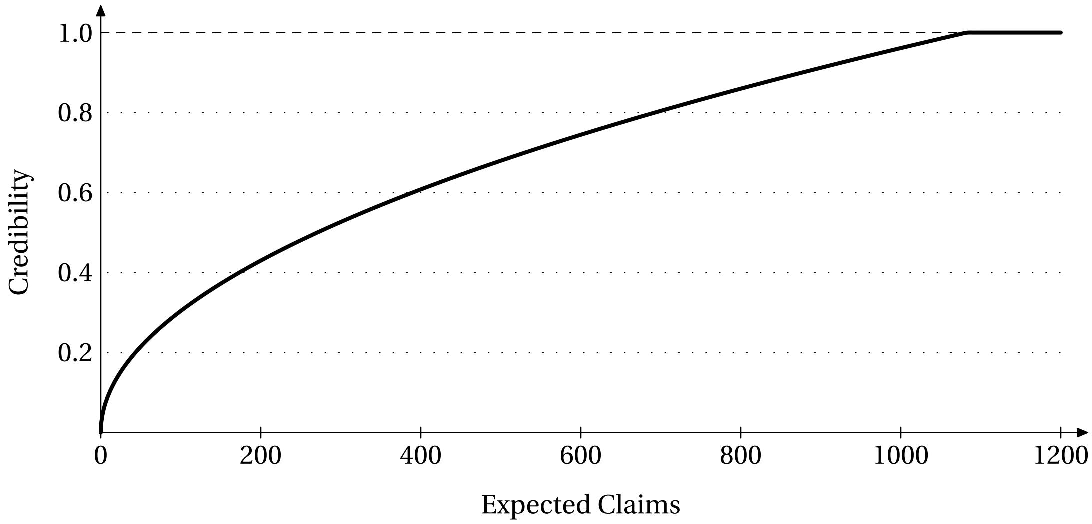



**Figure 3.1:** Partial credibility if $n_F = 1082$

**Solution:** The credibility factor is $\sqrt{n/n_F}$, with $n = 6000$ and $n_F = 19{,}544$, so $Z = \sqrt{\frac{6{,}000}{19{,}544}} = 0.55408$, and the estimate is:

$$P_C = 16{,}500{,}000 + \sqrt{\frac{6{,}000}{19{,}544}}(15{,}600{,}000 - 16{,}500{,}000) = \boxed{\mathbf{16{,}001{,}332}}. \qquad \textbf{(D)}$$

## Exercises

**3.1.** You are given the following:

- Number of claims follows a Poisson distribution.
- Classical credibility methods are used.
- The standard for credibility is set so that the actual aggregate losses are within 5% of expected losses 90% of the time.
- 605 expected claims are required for 50% credibility.

Determine the coefficient of variation for the claim size distribution.

A. Less than 1.50
B. At least 1.50, but less than 2.00
C. At least 2.00, but less than 2.50
D. At least 2.50, but less than 3.00
E. At least 3.00



*Exercises continue on the next page . . .*

**3.2.** **[4B-S92:6]** (1 point) You are given the following information for a group of insureds:

| | |
|---|---|
| Prior estimate of expected total losses | 20,000,000 |
| Observed total losses | 25,000,000 |
| Observed number of claims | 10,000 |
| Required number of claims for full credibility | 17,500 |

Using the methods of classical credibility, determine the estimate for the group's expected total losses based upon the latest observation.

A. Less than 21,000,000
B. At least 21,000,000, but less than 22,000,000
C. At least 22,000,000, but less than 23,000,000
D. At least 23,000,000, but less than 24,000,000
E. At least 24,000,000

**3.3.** **[4B-F93:20]** (2 points) You are given the following:

- $P$ = Prior estimate of pure premium for a particular class of business.
- $O$ = Observed pure premium during latest experience period for same class of business.
- $R$ = Revised estimate of pure premium for same class following observations.
- $F$ = Number of claims required for full credibility of pure premium.

Based on the methods of classical credibility, determine the number of claims used as the basis for determining $R$.

A. $F\left(\dfrac{R-P}{O-P}\right)$ B. $F\left(\dfrac{R-P}{O-P}\right)^2$ C. $\sqrt{F}\left(\dfrac{R-P}{O-P}\right)$ D. $\sqrt{F}\left(\dfrac{R-P}{O-P}\right)^2$ E. $F^2\left(\dfrac{R-P}{O-P}\right)$

**3.4.** **[4-F03:35, STAM Sample Question #27]** You are given:

- $X_{\text{partial}}$ = pure premium calculated from partially credible data.
- $\mu = \mathbf{E}[X_{\text{partial}}]$
- Fluctuations are limited to $\pm k\mu$ of the mean with probability $P$.
- $Z$ = credibility factor

Which of the following is equal to $P$?

A. $\Pr(\mu - k\mu \le X_{\text{partial}} \le \mu + k\mu)$
B. $\Pr(Z\mu - k\mu \le ZX_{\text{partial}} \le Z\mu + k)$
C. $\Pr(Z\mu - \mu \le ZX_{\text{partial}} \le Z\mu + \mu)$
D. $\Pr(1 - k \le ZX_{\text{partial}} + (1-Z)\mu \le 1 + k)$
E. $\Pr(\mu - k\mu \le ZX_{\text{partial}} + (1-Z)\mu \le \mu + k\mu)$



*Exercises continue on the next page . . .*

**3.5.** **[4B-F92:15]** (2 points) You are given the following:

- $X$ is the random variable for claim size.
- $N$ is the random variable for number of claims and has a Poisson distribution.
- $X$ and $N$ are independent.
- $\lambda_F$ is the standard for full credibility based only on number of claims.
- $n_F$ is the standard for full credibility based on total cost of claims.
- $n$ is the observed number of claims.
- $C$ is the random variable for total cost of claims.
- $Z$ is the amount of credibility to be assigned to total cost of claims.

According to the methods of classical credibility, which of the following are true?

1. $\operatorname{Var}(C) = \big(\mathbf{E}[N] \cdot \operatorname{Var}(X)\big) + \big(\mathbf{E}[X] \cdot \operatorname{Var}(N)\big)$
2. $n_F = \lambda_F \left(\dfrac{\mathbf{E}[X]^2 + \operatorname{Var}(X)}{\mathbf{E}[X]^2}\right)$
3. $Z = \sqrt{n/n_F}$

A. 1 only B. 2 only C. 1,2 only D. 2,3 only E. 1,2,3

**3.6.** The manual pure premium for an insurance coverage is 1,200. Average experience on a rate class is 1,000. Using the methods of classical credibility, a revised manual pure premium rate of 1,124 is used. The full credibility standard requires the actual number of claims to be within 5% of the expected number of claims 95% of the time. Claim counts on the coverage follow a Poisson distribution.

Determine the number of claims observed for the rate class.

**3.7.** **[4-S00:26]** You are given:

- Claim counts follow a Poisson distribution.
- Claim sizes follow a lognormal distribution with coefficient of variation 3.
- Claim sizes and claim counts are independent.
- The number of claims in the first year was 1000.
- The aggregate loss in the first year was 6.75 million.
- The manual premium for the first year was 5.00 million.
- The exposure in the second year is identical to the exposure in the first year.
- The full credibility standard is to be within 5% of the expected aggregate loss 95% of the time.

Determine the classical credibility net premium (in millions) for the second year.

A. Less than 5.5
B. At least 5.5, but less than 5.7
C. At least 5.7, but less than 5.9
D. At least 5.9, but less than 6.1
E. At least 6.1



*Exercises continue on the next page . . .*

**3.8.** **[MAS-II-S19:6]** An insurance company is currently using a limited-fluctuation credibility approach for a line of business with the following assumptions:

- The claim frequency follows a Poisson distribution.
- The mean of the claim frequency is large enough to justify the normal approximation to the Poisson.
- The square root rule is used to determine partial credibility.
- The standard for full credibility is the number of claims at which there is a 99% probability that the observed aggregate loss is within 5% of the mean.

You are given the following information about a block of 10,000 policies:

- The mean claim frequency is 0.12.
- The mean claim severity is 100.
- The variance of claim severity 14,400.

Calculate the credibility for this block of policies using the partial credibility method for aggregate loss.

A. Less than 0.45
B. At least 0.45, but less than 0.55
C. At least 0.55, but less than 0.65
D. At least 0.65, but less than 0.75
E. At least 0.75

**3.9.** **[MAS-II-F18:6]** You are given the following information:

- A block of insurance policies had 1,384 claims this period.
- The claims had a mean loss of 55 and a variance of loss of 6,010.
- The mean frequency of these claims is 0.085 per policy.
- Frequency follows a Poisson distribution.
- The block has 21,000 policies.
- Full credibility is based on a coverage probability of 98% for a range of within 5% deviation from the true mean.

You calculate the partial-credibility factor for severity, $Z_X$, and the partial-credibility factor for pure premium, $Z_P$, using the limited-fluctuation credibility method.

Calculate the absolute difference between $Z_X$ and $Z_P$.

A. Less than 0.05
B. At least 0.05, but less than 0.15
C. At least 0.15, but less than 0.25
D. At least 0.25, but less than 0.35
E. At least 0.35



Exercises continue on the next page . . .

**3.10.** **[4-F01:15, STAM Sample Question #65]** You are given the following information about a general liability book of business comprised of 2500 insureds:

- $X_i = \sum_{j=1}^{N_i} Y_{ij}$ is a random variable representing the annual loss of the $i^{\text{th}}$ insured.
- $N_1, N_2, \ldots, N_{2500}$ are independent and identically distributed random variables following a negative binomial distribution with parameters $r = 2$ and $\beta = 0.2$.
- $Y_{i1}, Y_{i2}, \ldots, Y_{iN_j}$ are independent and identically distributed random variables following a Pareto distribution with $\alpha = 3.0$ and $\theta = 1000$.
- The full credibility standard is to be within 5% of the expected aggregate losses 90% of the time.

Using classical credibility theory, determine the partial credibility of the annual loss experience for this book of business.

A. 0.35 B. 0.42 C. 0.47 D. 0.50 E. 0.53

## Solutions

**3.1.** $n_F = \lambda_F(1 + C_X^2) = 1082(1 + C_X^2)$, and $Z = 0.5 = \sqrt{\frac{605}{n_F}}$, so

$$\begin{aligned}
0.5 &= \sqrt{\frac{605}{1082(1 + C_X^2)}} \\
0.25\big(1082(1 + C_X^2)\big) &= 605 \\
C_X^2 &= 1.23575 \\
C_X &= \boxed{\mathbf{1.1116}} \qquad \textbf{(A)}
\end{aligned}$$

**3.2.** $Z = \sqrt{10{,}000/17{,}500}$, and $P_C = 20{,}000{,}000 + Z(25{,}000{,}000 - 20{,}000{,}000) = \boxed{\mathbf{23{,}779{,}645}}$. **(D)**

**3.3.** Let $n$ be the number of claims.

$$\begin{aligned}
R = P + Z(O - P) &= P + \sqrt{\frac{n}{F}}(O - P) \\
\frac{R - P}{O - P} &= \sqrt{\frac{n}{F}} \\
n &= \boxed{\boldsymbol{F\left(\frac{R - P}{O - P}\right)^2}} \qquad \textbf{(B)}
\end{aligned}$$

**3.4.** The idea of partial credibility is that we only take $Z$ times the observed mean ($X_{\text{partial}}$), and take $(1 - Z)$ times the prior mean ($\mu$), which has no variance. Thus $ZX_{\text{partial}} + (1 - Z)\mu$ is to be limited to being within $k\mu$ of the true mean $\mu$. That's exactly what **(E)** says.

**3.5.** Statement 1 should have $\mathbf{E}[X]^2$ instead of $\mathbf{E}[X]$. Statements 2 and 3, however, are correct. **(D)**

**3.6.** $Z = \dfrac{1200 - 1124}{1200 - 1000} = 0.38 = \sqrt{n/n_F}$, and $n_F = (1.96/0.05)^2 = 1536.64$, so $n = 1536.64(0.38^2) = \boxed{\mathbf{222}}$.

**3.7.** Expected claims needed for full credibility are:

$$n_F = \left(\frac{1.96}{0.05}\right)^2 (1 + 3^2) = 15{,}366.4$$

Credibility is therefore $Z = \sqrt{1000/15{,}366.4} = 0.2551$. The credibility premium is (in millions)

$$P_C = 5 + 0.2551(6.75 - 5) = \boxed{\mathbf{5.4464}} \qquad \textbf{(A)}$$

**3.8.** The second statement is strange; presumably "mean of the claim frequency" is the same as "mean claim frequency", but mean claim frequency is 0.12, which is not very large; nor is it clear why the size of the claim frequency should justify a normal approximation. What they meant is probably that the number of policies is large enough so that the normal approximation to the aggregate distribution may be used.

Note that you are given exposures (10,000), not expected claims, and the full credibility standard should be expressed in terms of exposures.

$$\begin{aligned}
\lambda_F &= \left(\frac{2.576}{0.05}\right)^2 = 2654 \\
C_X^2 &= \frac{14{,}000}{100^2} = 1.44 \\
n_F &= 2654(1 + 1.44) = 6475.76 \\
e_F &= \frac{6475.76}{0.12} = 53{,}965 \\
Z &= \sqrt{\frac{10{,}000}{53{,}965}} = \boxed{\mathbf{0.4305}} \qquad \textbf{(A)}
\end{aligned}$$

**3.9.**

$$\lambda_F = \left(\frac{\Phi^{-1}(0.99)}{0.05}\right)^2 = \left(\frac{2.326}{0.05}\right)^2 = 2164.11$$

For severity, the credibility standard is expressed in terms of number of exposures, which is number of claims. We had 1384 claims.

$$\begin{aligned}
e_X &= 2164.11\left(\frac{6{,}010}{55^2}\right) = 4{,}300 \\
Z_X &= \sqrt{\frac{1{,}384}{4{,}300}} = 0.567354
\end{aligned}$$

For pure premium, the credibility standard is expressed in terms of number of exposures, which is number of policies. We have 21,000 policies. We divide the usual formula for the credibility standard in terms of number of expected claims by 0.085 to express it in terms of number of policies

$$\begin{aligned}
e_P &= \frac{2164.11}{0.085}\left(1 + \frac{6{,}010}{55^2}\right) = 76{,}044 \\
Z_P &= \sqrt{\frac{21{,}000}{76{,}044}} = 0.525506
\end{aligned}$$

The absolute difference between credibility factors is $\boxed{\mathbf{0.0418}}$. **(A)**

**3.10.** The number of *expected claims* needed for full credibility is

$$\lambda_F = 1082\left(\frac{\sigma_N^2}{\mu_N} + \frac{\sigma_X^2}{\mu_X^2}\right)$$

and

$$
\begin{aligned}
\frac{\sigma_N^2}{\mu_N} &= \frac{r\beta(1+\beta)}{r\beta} = 1+\beta = 1.2\\
\mu_X &= 500\\
\sigma_X^2 &= 1000^2 - 500^2 = 750{,}000\\
n_F &= 1082(1.2+3) = 4546.122
\end{aligned}
$$

Expected claims from 2500 insureds is $2500r\beta = 1000$, so credibility is $Z = \sqrt{1000/4546.122} = \boxed{\mathbf{0.4690}}$. **(C)**

## Quiz Solutions

**3-1.** We want $\sqrt{n/n_F} = 0.2$ and are given that $\sqrt{250/n_F} = 0.5$, so $n_F = 250/0.5^2$ and $n = 250(0.2/0.5)^2 = \boxed{\mathbf{40}}$.

# Part V

# Practice Exams

Here are 6 practice exams.[2] All questions are original. But coming up with so many questions not requiring R programming was tough. So some are versions of released exam questions with numbers changed. And some are fixed versions of defective released exam questions! Others are similar to this manual's exercises. But you'll find enough new questions to challenge you.

I found it particularly difficult to come up with the large number of questions on statistical learning needed. Some of the questions on that topic are a bit repetitive.

The order of subjects on these exams is the same as on the released exams, although with computer based tests perhaps the questions will be randomly ordered. The weights of the topics are within the syllabus percentage ranges.

CAS exams will have five types of questions:

1. Traditional multiple choice. However, the number of choices won't always be 5.
2. Multiple choice where more than one choice may be correct.
3. Point and click questions on a graph.
4. Fill in the blank. For example, enter an integer as the solution (no choices offered).
5. Matching. For example, match 3 items on the left with 3 items on the right.

A short sample exam demonstrating these 5 types of questions can be found at

https://home.pearsonvue.com/cas/sample-exam/new-item-type

(Note: The questions on this sample exam are based mostly on Bayesian regression, a topic which constituted most of the Exam MAS-II syllabus before Fall 2023 but is no longer on the syllabus. So don't try to do them!) However, almost all of the questions in these practice exams are traditional multiple choice questions. I've written some new questions in the other styles, except for point and click. Point and click is really just a traditional multiple choice question; instead of having a map with the letters A, B, C, D, E where you choose the correct letter, you click on one of the areas of the map.

In the answer keys before the practice exam solutions, I've left the answer key for non-traditional questions blank.

[2]Practice Exam 7 is not a real practice exam. It is a collection of linear mixed models case studies, in case you'd like to try your hand at them.

# Practice Exam 1

1. For each exposure in a group, the hypothetical mean of aggregate losses is $\Theta$ and the process variance is $e^{0.3\Theta}$. $\Theta$ varies by group. Its probability distribution is exponential with mean 3.

For three years experience from a group, you have the following data:

| Year | Exposures | Aggregate losses |
|---|---|---|
| 1 | 20 | 70 |
| 2 | 25 | 90 |
| 3 | 30 | 110 |

There will be 35 exposures in the group next year.

Calculate the Bühlmann-Straub credibility estimate for the group.

A. Less than 119
B. At least 119, but less than 121
C. At least 121, but less than 123
D. At least 123, but less than 125
E. At least 125

2. For 4 policyholders, you have the following experience for number of claims over 4 years.

| | Year | | | |
|---|---|---|---|---|
| Policyholder | 1 | 2 | 3 | 4 |
| #1 | 0 | 0 | 1 | 0 |
| #2 | 0 | 1 | 0 | 0 |
| #3 | 1 | 1 | 1 | 0 |
| #4 | 0 | 0 | 0 | 0 |

Using empirical Bayes non-parametric credibility methods, calculate the estimated number of claims for policyholder #1 in year 5.

A. Less than 0.26
B. At least 0.26, but less than 0.27
C. At least 0.27, but less than 0.28
D. At least 0.28, but less than 0.29
E. At least 0.29

3. Claims sizes for a group follow a lognormal distribution with $\mu = 7$, $\sigma = 1.2$. Claim counts follow a Poisson distribution. The manual premium is 20,000 per insured. 3356 claims are observed, with average aggregate claims per insured of 18,000.

Classical credibility methods are used to determine the credibility premium. The standard for full credibility is set so that aggregate losses are within 5% of expected losses with probability $p$. Based on this standard, the credibility premium is 18,500.

Determine $p$.

A. 0.94    B. 0.95    C. 0.96    D. 0.97    E. 0.98

**4.** You are given:

- After 2 exposure periods with average aggregate losses of 240, the Bühlmann credibility estimate is 228.
- After 8 exposure periods with average aggregate losses of 210, the Bühlmann credibility estimate is 214.

Calculate the Bühlmann expectation after 12 exposure periods with average aggregate losses of 250. Answer to the nearest integer.

A. 241 B. 242 C. 243 D. 244 E. 245

**5.** On an insurance coverage, the probability that a loss exceeds the deductible is $p$. The prior distribution for $p$ has density function

$$\pi(p) = \frac{3}{2}\sqrt{p} \qquad 0 \le p \le 1$$

In a sample of 10 losses, 8 exceed the deductible.

Calculate the posterior probability that a loss exceeds the deductible.

A. Less than 0.625
B. At least 0.625, but less than 0.675
C. At least 0.675, but less than 0.725
D. At least 0.725, but less than 0.775
E. At least 0.775

**6.** For two classes of insureds, A and B, you are given the following probabilities for the number of claims per year.

| Number of claims | A | B |
|---|---|---|
| 0 | 0.8 | 0.7 |
| 1 | 0.1 | 0.2 |
| 2 | 0.1 | 0.1 |

Each class has an equal number of insureds.

For a randomly selected insured, there are 2 claims in 3 years.

Calculate the expected number of claims in the next year for this insured.

A. Less than 0.32
B. At least 0.32, but less than 0.34
C. At least 0.34, but less than 0.36
D. At least 0.36, but less than 0.38
E. At least 0.38

**7.** Claim sizes follow an inverse exponential distribution with parameter $\theta$. The distribution of $\theta$ over the entire population of policyholders is a gamma distribution with $\alpha = 5$, scale parameter 1000.

A randomly selected policyholder submits claims for 1000, 2000, and 5000.

Calculate the mean of the posterior distribution for $\theta$.

A. Less than 1900
B. At least 1900, but less than 2200
C. At least 2200, but less than 2500
D. At least 2500, but less than 2800
E. At least 2800



Exam questions continue on the next page . . .

**8.** You are given:

- Losses on an insurance coverage follow a lognormal distribution with parameters $\mu = \Theta$ and $\sigma = 2$.
- $\Theta$ varies by policyholder in accordance with a normal distribution with mean 5 and variance 3.
- A policyholder submits 20 claims for losses. The average size of the claims is 10,000.

Calculate the Bühlmann prediction of the size of the next loss for this policyholder.

A. Less than 6000
B. At least 6000, but less than 7000
C. At least 7000, but less than 8000
D. At least 8000, but less than 9000
E. At least 9000

**9.** Production of agents is modeled with a Linear Mixed Model. The model is a 3-level hierarchical model, with agents within agencies within regions. There are random intercepts for agency and region.

The fitted values of variance components are 82 for variance between agencies, 154 for variance between regions, and 58 for residual variance.

Based on these fitted values, estimate the intraclass correlation coefficient for agents working in the same agency.

A. Less than 0.2
B. At least 0.2, but less than 0.4
C. At least 0.4, but less than 0.6
D. At least 0.6, but less than 0.8
E. At least 0.8

**10.** While fitting a Linear Mixed Model, problems arose with the estimation of the covariance parameters. The following alternative approaches for the estimation of the covariance parameters were considered:

I. Rescale the covariates.
II. Remove random effects that may not be necessary.
III. Fit the implied marginal model.

Determine which of the alternative approaches above are appropriate for the estimation of the covariance parameters.

A. None of I, II, or III is true
B. I and II only
C. I and III only
D. II and III only
E. The answer is not given by (A), (B), (C), or (D)



*Exam questions continue on the next page . . .*

**11.** Consider the following statements regarding the use of the AR(1) structure for a homogeneous $\mathbf{R}_i$ matrix in a Linear Mixed Model.

I. If $\mathbf{R}_i$ has an AR(1) structure, then the variances of all observations in a group are equal.

II. The correlation parameter $\rho$ in an AR(1) structure is nonnegative.

III. The AR(1) structure is appropriate for repeated trials under the same conditions of an experiment .

Determine which of the above statements are true.

A. I only
B. II only
C. III only
D. I, II, and III
E. The answer is not given by (A), (B), (C), or (D)

**12.** The research team at Roadrunner Oil is studying gasoline sales at their 156 franchised stations. The model includes an intercept and fixed factors for

- CARS: Number of cars owned or leased within 10 miles of the station.
- PRICE: An index comparing the price of gasoline at the station to the average price of gasoline in the town in which the station is located.

The model includes a random intercept for the station and a random intercept for day of week for each of the 7 days of a week.

The duration of the study is 4 weeks. Thus there are 28 sales numbers for each station, or a total of $28 \times 156 = 4368$ observations.

Determine the size of the $\mathbf{D}$ matrix for this Linear Mixed Model.

A. $1 \times 1$ B. $2 \times 2$ C. $156 \times 156$ D. $163 \times 163$ E. $4368 \times 4368$

**13.** Consider the following statements regarding the residuals of a marginal linear model:

I. They are multivariate normally distributed.

II. They are uncorrelated.

III. They are homoskedastic. (They have the same variance.)

Determine which of these statements are true.

A. I only
B. II only
C. III only
D. I, II, and III
E. The answer is not given by (A), (B), (C), or (D)



Exam questions continue on the next page . . .

**14.** $\hat{\varepsilon}_i$ are the residuals of a Linear Mixed Model based on 50 observations. You are given

- $\hat{\varepsilon}_{10} = 3$
- $\sum_{i=1}^{50} \hat{\varepsilon}_i^2 = 100$

Consider the following statements:

I. The standardized residual for the tenth observation cannot be computed from the information given.

II. The internally studentized residual for the tenth observation is 2.100.

III. The externally studentized residual for the tenth observation is 2.179.

Determine which of these statements are true.

A. I only
B. II only
C. III only
D. I, II, and III
E. The answer is not given by (A), (B), (C), or (D)

**15.** An insurance company evaluates its model for loss cost using the Gini index.

The company plots a Lorenz curve of percentage of losses on percentage of exposures. The equation for the curve, based on the model, is $y = x^{1.5}$.

Determine the Gini index.

A. Less than 15%
B. At least 15%, but less than 20%
C. At least 20%, but less than 25%
D. At least 25%, but less than 30%
E. At least 30%

**16.** A model for loss cost is analyzed using a simple quantile plot. The plot is based on these ten cases:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Actual loss cost | 8 | 10 | 5 | 6 | 15 | 20 | 12 | 11 | 9 | 7 |
| Predicted loss cost | 9 | 7 | 5 | 4 | 18 | 22 | 10 | 11 | 8 | 6 |

The plot uses quintiles. The numbers are not divided by average predicted loss cost.

Calculate the lift. Answer to the nearest 0.5.

**17.** Determine which of the following statements are true regarding $K$-nearest neighbors (KNN) classification.

I. KNN error rate is no more than the Bayes error rate.

II. As a function of $K$, there is a strong relationship between the training error rate and the test error rate.

III. KNN becomes more flexible as $1/K$ increases.

A. None B. I and II only C. I and III only D. II and III only
E. The correct answer is not given by A. , B. , C. , or D.

**18.** The XYZ Insurance Company's claims department has come up with an index to predict whether a claim is fraudulent. The higher the index, the more likely the claim is fraudulent. XYZ sets the threshold above which a claim is investigated by using KNN.

The training data is

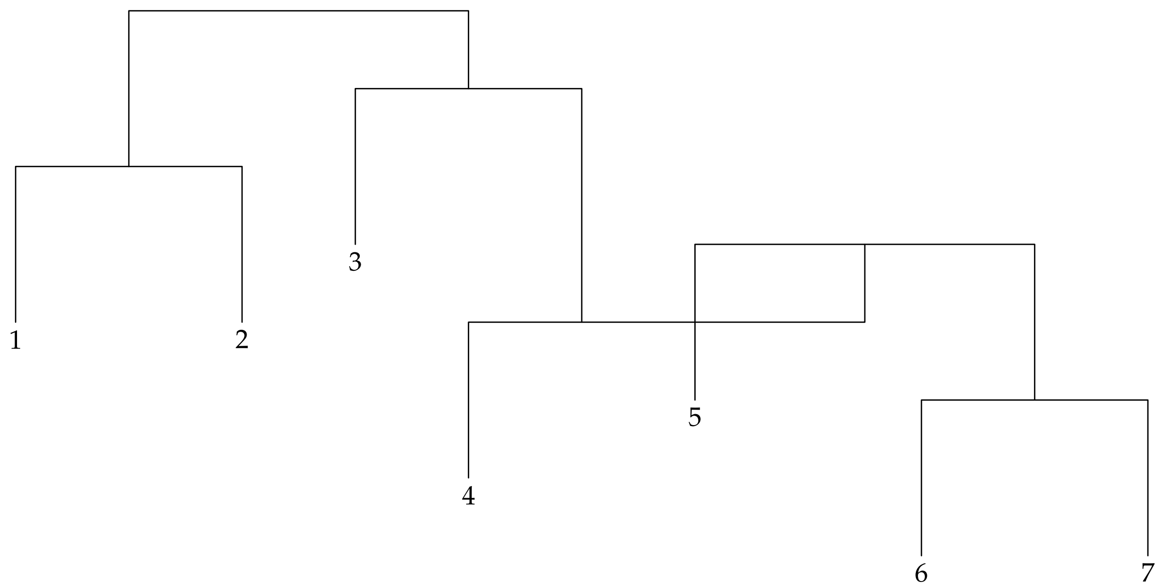
| Index | 82 | 105 | 116 | 127 | 131 | 144 | 151 | 168 | 180 | 211 | 219 | 222 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fraud? | No | No | No | No | Yes | No | Yes | No | Yes | Yes | Yes | Yes |

The company uses $K = 3$.

Calculate the error rate on the training data.

A. Less than 0.2
B. At least 0.2, but less than 0.3
C. At least 0.3, but less than 0.4
D. At least 0.4, but less than 0.5
E. At least 0.5

**19.** Hierarchical clustering is performed on 7 observations, resulting in the following dendrogram:




You are given the following statements:

I. Centroid linkage was used.
II. Observation 3 is closer to observation 4 than to observation 7.
III. Observations 3 and 4 are closer to each other than observations 1 and 2.

Determine which of these statements is/are true.



Exam questions continue on the next page . . .

**20.** You are given the following statements regarding boosting.

I. Selecting $B$ too high can result in overfitting.

II. Selecting a low shrinkage parameter tends to lead to selecting a lower $B$.

III. If $d = 1$, the model is an additive model.

Determine which of these statements are true.

A. None of I, II, or III is true
B. I and II only
C. I and III only
D. II and III only
E. The answer is not given by (A), (B), (C), or (D)

**21.** You are given the following statements regarding cost complexity pruning.

I. A higher $\alpha$ corresponds to higher MSE for the training data.

II. A higher $\alpha$ corresponds to higher bias for the test data.

III. A higher $\alpha$ corresponds to a higher $|T|$.

Determine which of the statements are true.

A. None of I, II, or III is true
B. I and II only
C. I and III only
D. II and III only
E. The answer is not given by (A), (B), (C), or (D)

**22.** You are given the following statements regarding classification trees.

I. Classification error is not sensitive enough for growing trees.

II. Classification error is not sensitive enough for pruning trees.

III. The predicted values of two terminal nodes coming out of a split are different.

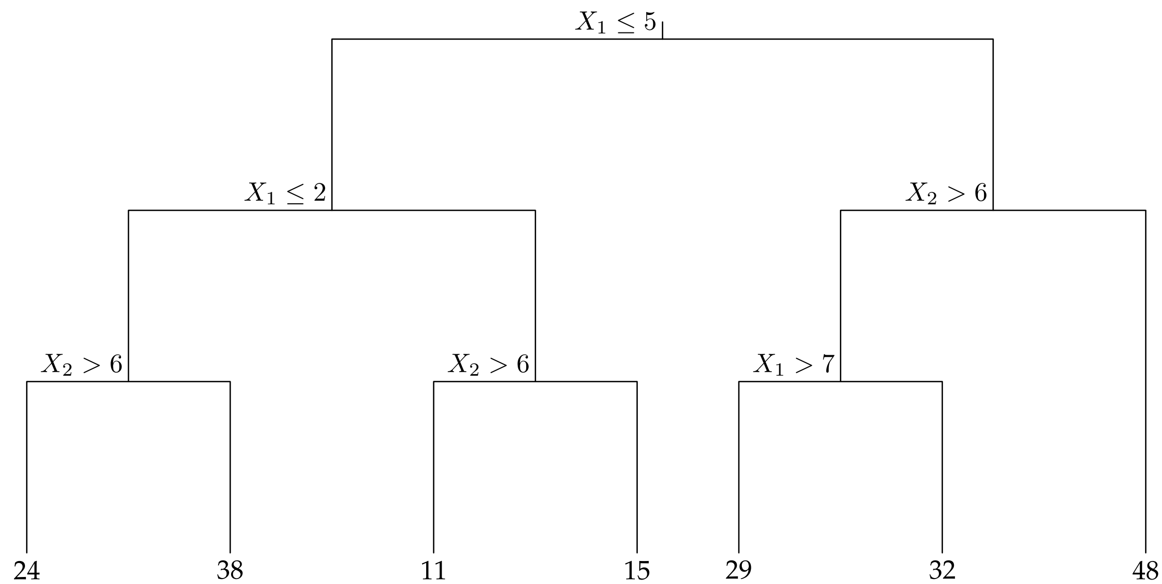
Determine which of these statements are true.

A. I only
B. II only
C. III only
D. I, II, and III
E. The answer is not given by (A), (B), (C), or (D)



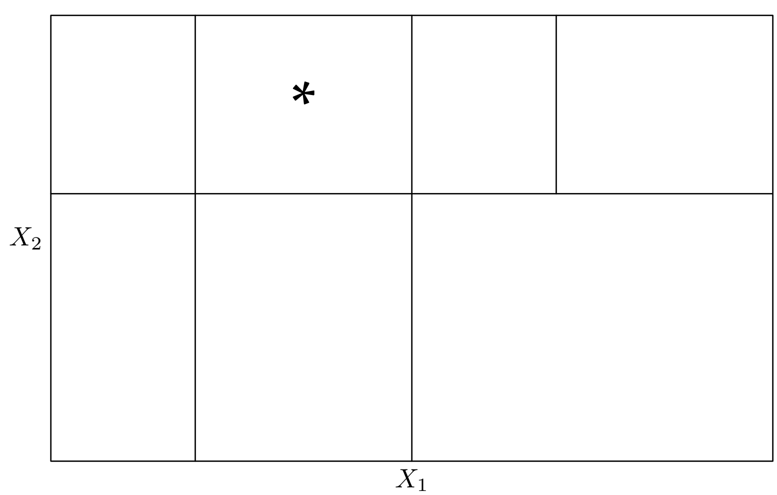
*Exam questions continue on the next page . . .*

**23.** You are given the following decision tree:




At each node, the left branch is taken if the condition specified at the node is satisfied.

Consider the following rectangle:




Determine the fitted value of the response variable in the area with an asterisk.

A. 11 B. 15 C. 24 D. 29 E. 32



Exam questions continue on the next page . . .

**24.** You are given the following data regarding students who sat for Exam MAS-II:

| Hours of study | 252 | 274 | 300 | 324 | 360 | 400 |
|---|---|---|---|---|---|---|
| Passed MAS-II? | No | No | No | Yes | Yes | No |

You build a classification decision tree. At the initial node, you will split it based on hours of study, using the Gini index as a criterion.

Determine the decrease in classification error as a result of the split.

A. Less than 0.05
B. At least 0.05, but less than 0.15
C. At least 0.15, but less than 0.25
D. At least 0.25, but less than 0.35
E. At least 0.35

**25.** In a convolutional neural network, you are given the following encoding of an image:

$$\begin{pmatrix} 2 & 4 & 0 \\ 3 & 0 & 5 \\ 1 & 7 & 11 \\ 12 & 4 & 2 \\ 0 & 1 & 2 \\ 13 & 3 & 0 \end{pmatrix}$$

Use the convolution filter

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{pmatrix}$$

Determine the sum of the first column of the convolved image. (Answer to nearest whole number.)

**26.** You are given the following statements regarding principal component analysis:

I. Principal components analysis is a method to visualize data.
II. Principal components are in the direction in which the data is most variable.
III. Principal components are orthogonal.

Determine which of these statements are true.

A. I only
B. II only
C. III only
D. I, II, and III
E. The answer is not given by (A), (B), (C), or (D)



Exam questions continue on the next page . . .

**27.** In a principal components analysis, there are 2 variables. The loading of the first principal component on the first variable is $-0.6$ and the loading of the first principal component on the second variable is positive. The variables have been centered at 0.

For the observation $(0.4, x_2)$, the first principal component score is 0.12.

Determine $x_2$.

A. Less than 0.30
B. At least 0.30, but less than 0.35
C. At least 0.35, but less than 0.40
D. At least 0.40, but less than 0.45
E. At least 0.45

**28.** Principal component methods may be used to fill in missing data.

Select the statements that are false. (There may be more than one false statement.)

I. These methods may only be used for data missing at random.
II. These methods should only be used if no more than 10% of the data is missing.
III. These methods may not produce a global minimum for the objective function.
IV. These methods work best when there are a small number of features.

**29.** A department store is conducting a cluster analysis to help focus its marketing. The store sells many different products, including food, clothing, furniture, and computers. Management would like the clusters to group customers with similar shopping patterns together.

A colleague of yours makes the following statements:

I. The clusters will depend on whether the input data is units sold or dollar amounts sold.
II. Hierarchical clustering would be preferable to $K$-means clustering.
III. If a correlation-based dissimilarity measure is used, frequent and infrequent shoppers will be grouped together.

Determine which of these statements are correct.

A. I only
B. II only
C. III only
D. I, II, and III
E. The answer is not given by (A), (B), (C), or (D)

**30.** You are given the following two clusters:

$$\{(8,2),(9,7),(12,5)\} \text{ and } \{(10,3),(11,1)\}$$

Calculate the dissimilarity measure between the clusters using Euclidean distance and average linkage.

A. 3.6 B. 3.7 C. 3.8 D. 3.9 E. 4.0



*Exam questions continue on the next page . . .*

**31.** The $K$-means clustering algorithm, using $K = 2$ and squared Euclidean distance, is used to split the following 6 observations into 2 clusters:

$$(1,3) \qquad (2,2) \qquad (3,4) \qquad (1,1) \qquad (3,1) \qquad (5,1)$$

A local minimum for the objective function occurs when the first three observations are put into the first cluster and the second three observations are put into the second cluster.

Determine the value of the objective function for this split. (Answer to the nearest whole number.)

**32.** You are given the following statements regarding hierarchical clustering.

I. Complete linkage is based on minimal intercluster dissimilarity.

II. Single linkage is based on maximal intercluster dissimilarity.

III. Single linkage may result in extended trailing clusters.

Determine which of these statements are true.

A. I only
B. II only
C. III only
D. I, II, and III
E. The answer is not given by (A), (B), (C), or (D)

**33.** For a time series $x_t$, you are given $x_1 = 18$, $x_2 = 35$, $x_3 = 27$ $x_4 = 21$, $x_5 = 19$.

Calculate the sample autocorrelation at lag 2.

A. Less than −0.4
B. At least −0.4, but less than −0.3
C. At least −0.3, but less than −0.2
D. At least −0.2, but less than −0.1
E. At least −0.1



*Exam questions continue on the next page . . .*

**34.** A monthly time series has seasonal patterns. Seasonality of the series is modeled with an additive model. When the 12-month centered moving average is subtracted from the series, the average result by month is

| | |
|---|---|
| January | −5.3 |
| February | −8.5 |
| March | −3.2 |
| April | 1.0 |
| May | 1.0 |
| June | 4.4 |
| July | 2.1 |
| August | 0.8 |
| September | 0.6 |
| October | −3.5 |
| November | −1.1 |
| December | 6.9 |

For January 2014, the unadjusted value of the time series is 102.8.

Calculate the seasonally adjusted value.

A. Less than 99
B. At least 99, but less than 102
C. At least 102, but less than 105
D. At least 105, but less than 108
E. At least 108

**35.** For a random walk with variance parameter $\sigma^2$, which of the following are true?

I. The random walk is stationary in the mean.

II. At time 50, the variance is $50\sigma^2$.

III. At time 50, the lag 1 autocorrelation is 0.99.

A. I only B. II only C. III only D. I, II, and III
E. The correct answer is not given by A. , B. , C. , or D.

**36.** For an AR(1) time series, you are given:

- The variance of the terms is 25.
- The autocovariance of the terms at lag 2 is 16.

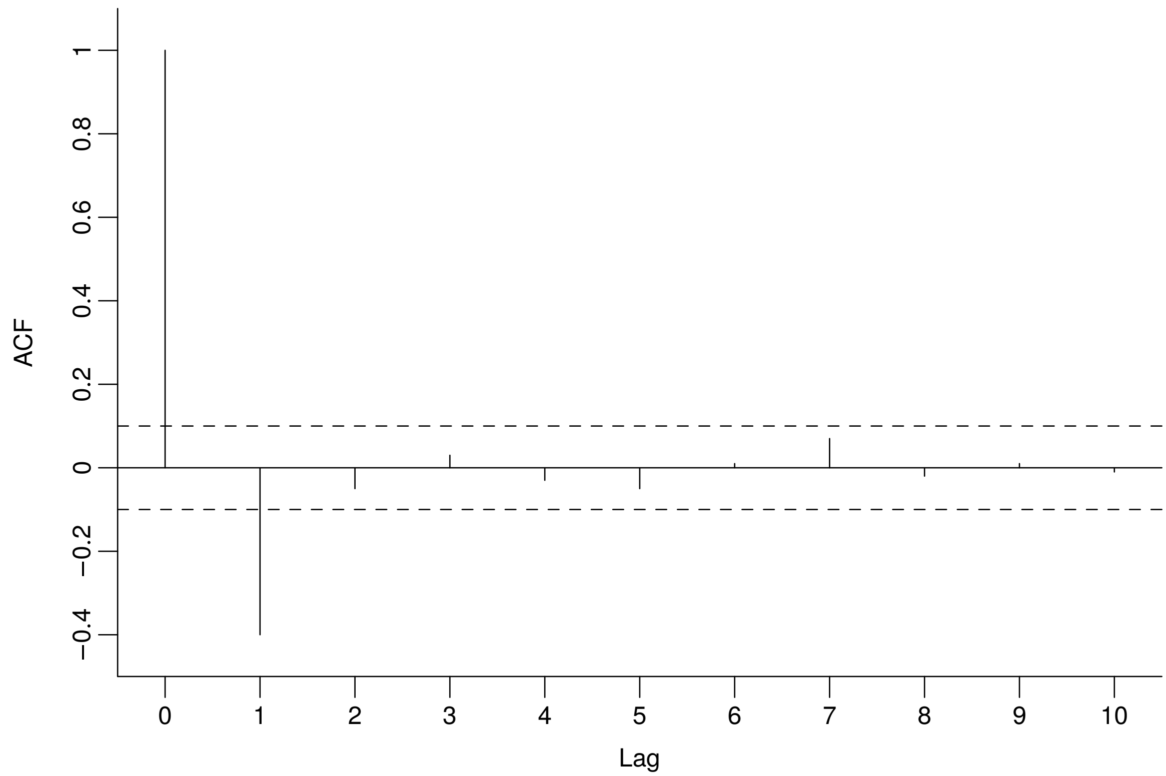
Determine the variance of the error term $w_t$.

A. Less than 6
B. At least 6, but less than 8
C. At least 8, but less than 10
D. At least 10, but less than 12
E. At least 12



Exam questions continue on the next page . . .

**37.** The correlogram for a time series is




An invertible MA model is fitted to this time series.

Determine the model.

A. $x_t = 0.4w_{t-1} + w_t$

B. $x_t = -0.4w_{t-1} + w_t$

C. $x_t = 0.5w_{t-1} + w_t$

D. $x_t = -0.5w_{t-1} + w_t$

E. $x_t = 2w_{t-1} + w_t$

**38.** Determine which of the following processes have redundant parameters. (There may be more than one.)

Model I: $x_t = 1.4x_{t-1} - 0.48x_{t-2} - 0.6w_{t-1} + w_t$

Model II: $x_t = 1.4x_{t-1} - 0.48x_{t-2} + 0.6w_{t-1} + w_t$

Model III: $x_t = 1.4x_{t-1} + 0.48x_{t-2} - 0.6w_{t-1} + w_t$



*Exam questions continue on the next page . . .*

**39.** A time series $\{x_t\}$ can be expressed as

$$x_t = \alpha_0 + \alpha_1 t + w_t$$

where $w_t$ is Gaussian white noise.

Determine the type of process followed by $\nabla x_t$, the differences of $x_t$.

A. White noise
B. Random walk
C. AR(1)
D. MA(1)
E. ARMA(1,1)

**40.** `R` provides the following estimate for the coefficients of an MA(3) time series:

```
  ma1    ma2    ma3  intercept
0.842  0.621  0.200       -3.5
```

You are given that the residuals for periods 18, 19, and 20 are 6, −4, and 10 respectively.

Forecast the value of the time series in period 21.

A. Less than 3.7
B. At least 3.7, but less than 3.8
C. At least 3.8, but less than 3.9
D. At least 3.9, but less than 4.0
E. At least 4.0

*Solutions to the above questions begin on page 689.*

# Appendices

# Appendix A. Solutions to the Practice Exams

## Answer Key for Practice Exam 1

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | E | 11 | A | 21 | B | 31 | |
| 2 | C | 12 | D | 22 | A | 32 | C |
| 3 | A | 13 | A | 23 | A | 33 | B |
| 4 | | 14 | D | 24 | C | 34 | D |
| 5 | D | 15 | C | 25 | | 35 | D |
| 6 | C | 16 | | 26 | D | 36 | C |
| 7 | E | 17 | E | 27 | E | 37 | D |
| 8 | B | 18 | B | 28 | | 38 | |
| 9 | E | 19 | | 29 | D | 39 | D |
| 10 | E | 20 | C | 30 | C | 40 | A |

## Practice Exam 1

**1.** [**Lesson 12**] $\mu_{\mathrm{HM}} = \mathbf{E}[\Theta] = 3$. $\sigma^2_{\mathrm{HM}} = \mathrm{Var}(\Theta) = 9$. The expected value of the process variance is

$$\mu_{\mathrm{PV}} = \int e^{0.3\theta} f_\Theta(\theta)\mathrm{d}\theta = \int_0^\infty e^{0.3\theta}\left(\frac{e^{-\theta/3}}{3}\right)\mathrm{d}\theta = \frac{\int_0^\infty e^{-\theta/30}\mathrm{d}\theta}{3} = 10$$

So $K = \frac{10}{9}$, $Z = \frac{75}{75+\frac{10}{9}} = \frac{675}{685}$. There are 75 exposures and 270 aggregate losses, so $\bar{x} = \frac{270}{75} = \frac{54}{15}$ and the credibility premium per exposure is

$$\frac{675(54/15) + 10(3)}{685} = 3.5912$$

The credibility premium for the group is $35(3.5912) = \boxed{\mathbf{125.69}}$. **(E)**

**2.** [**Section 14.1**] Use formulas (14.1), (14.2), and (14.3).

$$\begin{aligned}
\hat{\mu} &= \frac{5}{16} = 0.3125\\
\bar{x}_1 = \bar{x}_2 &= \frac{1}{4} = 0.25\\
\bar{x}_3 &= \frac{3}{4} = 0.75\\
\bar{x}_4 &= 0\\
\bar{x} &= \frac{1}{4}\left(\frac{1}{4} + \frac{1}{4} + \frac{3}{4} + 0\right) = \frac{5}{16}\\
v_1 = v_2 = v_3 &= \frac{1}{3}\left(3(0-0.25)^2 + (1-0.25)^2\right) = \frac{0.75}{3} = 0.25
\end{aligned}$$

$v_3$ is the same as $v_1$ and $v_2$, since the variance of three 1's and one 0 is the same as the variance three 0's and one 1; the latter random variable is 1 minus the former.

$$\begin{aligned}
v_4 &= 0\\
\widehat{\mu}_{\mathrm{PV}} &= \frac{1}{4}[3(0.25)] = 0.1875
\end{aligned}$$