1. For a second-order autoregressive process, you are given:

\[ \rho_1 = 0.53 \]
\[ \rho_2 = -0.22 \]

Determine \( \rho_3 \).

(A) Less than –0.70
(B) At least –0.70, but less than –0.30
(C) At least –0.30, but less than 0.10
(D) At least 0.10, but less than 0.50
(E) At least 0.50
2. You are given:

<table>
<thead>
<tr>
<th>Claim Size</th>
<th>Number of Claims</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-25</td>
<td>30</td>
</tr>
<tr>
<td>25-50</td>
<td>32</td>
</tr>
<tr>
<td>50-100</td>
<td>20</td>
</tr>
<tr>
<td>100-200</td>
<td>8</td>
</tr>
</tbody>
</table>

Assume a uniform distribution of claim sizes within each interval.

Estimate the second raw moment of the claim size distribution.

(A) Less than 3300
(B) At least 3300, but less than 3500
(C) At least 3500, but less than 3700
(D) At least 3700, but less than 3900
(E) At least 3900
3. You are given:

(i) The number of claims per auto insured follows a Poisson distribution with mean $\lambda$.

(ii) The prior distribution for $\lambda$ has the following probability density function:

$$f(\lambda) = \frac{(500\lambda)^{50} e^{-500\lambda}}{\lambda \Gamma(50)}$$

(iii) A company observes the following claims experience:

<table>
<thead>
<tr>
<th></th>
<th>Year 1</th>
<th>Year 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of claims</td>
<td>75</td>
<td>210</td>
</tr>
<tr>
<td>Number of autos insured</td>
<td>600</td>
<td>900</td>
</tr>
</tbody>
</table>

The company expects to insure 1100 autos in Year 3.

Determine the expected number of claims in Year 3.

(A) 178
(B) 184
(C) 193
(D) 209
(E) 224
4. Which of the following statements about the Product-Limit estimator is false?

(A) The Product-Limit estimator is based on the assumption that knowledge of a censoring time for an individual provides no further information about this person’s likelihood of survival at a future time had the individual continued in the study.

(B) If the largest study time corresponds to a death time, then the Product-Limit estimate of the survival function is undetermined beyond this death time.

(C) When there is no censoring or truncation, the Product-Limit estimator reduces to the empirical survival function.

(D) Under certain regularity conditions, the Product-Limit estimator is a nonparametric maximum likelihood estimator.

(E) The Product-Limit estimator is consistent.
5. You fit the following model to eight observations:

\[ Y = \alpha + \beta X + \epsilon \]

You are given:

\[
\hat{\beta} = -35.69 \\
\sum (X_i - \bar{X})^2 = 1.62 \\
\sum (Y_i - \hat{Y}_i)^2 = 2394
\]

Determine the symmetric 90-percent confidence interval for \( \beta \).

(A) \((-74.1, 2.7)\)

(B) \((-66.2, -5.2)\)

(C) \((-63.2, -8.2)\)

(D) \((-61.5, -9.9)\)

(E) \((-61.0, -10.4)\)
6. The graph below shows a $q-q$ plot of a fitted distribution compared to a sample.

Which of the following is true?

(A) The tails of the fitted distribution are too thick on the left and on the right, and the fitted distribution has less probability around the median than the sample.

(B) The tails of the fitted distribution are too thick on the left and on the right, and the fitted distribution has more probability around the median than the sample.

(C) The tails of the fitted distribution are too thin on the left and on the right, and the fitted distribution has less probability around the median than the sample.

(D) The tails of the fitted distribution are too thin on the left and on the right, and the fitted distribution has more probability around the median than the sample.

(E) The tail of the fitted distribution is too thick on the left, too thin on the right, and the fitted distribution has less probability around the median than the sample.
7. You are given the following information about six coins:

<table>
<thead>
<tr>
<th>Coin</th>
<th>Probability of Heads</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 4</td>
<td>0.50</td>
</tr>
<tr>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>6</td>
<td>0.75</td>
</tr>
</tbody>
</table>

A coin is selected at random and then flipped repeatedly. $X_i$ denotes the outcome of the $i$th flip, where “1” indicates heads and “0” indicates tails. The following sequence is obtained:

$$S = \{X_1, X_2, X_3, X_4\} = \{1, 1, 0, 1\}$$

Determine $E(X_3|S)$ using Bayesian analysis.

(A) 0.52
(B) 0.54
(C) 0.56
(D) 0.59
(E) 0.63
8. To study the effect of smoke alarms on the size of fire insurance claims, a proportional hazards model was used on a random sample of six claims. A single covariate $Z$ was used with $Z = 0$ indicating the absence and $Z = 1$ indicating the presence of a smoke alarm. The sizes of the claims (in standard units) were:

- Without a smoke alarm: 2, 5, 7
- With a smoke alarm: 1, 3, 6

The maximum likelihood estimate of the parameter $\beta$ was 0.6.

Using Breslow’s estimate of the baseline cumulative hazard rate, estimate the probability that a claim from a property without a smoke alarm will exceed 4 units.

(A) 0.44
(B) 0.50
(C) 0.57
(D) 0.64
(E) 0.67
9. Based on 100 observations of a time series, you determine the following sample autocorrelation coefficients:

<table>
<thead>
<tr>
<th>$k$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\rho}_k$</td>
<td>0.06</td>
<td>0.02</td>
<td>0.02</td>
<td>-0.16</td>
<td>-0.14</td>
<td>-0.01</td>
<td>0.03</td>
<td>-0.01</td>
<td>0.01</td>
<td>-0.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$k$</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\rho}_k$</td>
<td>0.03</td>
<td>0.10</td>
<td>-0.17</td>
<td>0.02</td>
<td>0.01</td>
</tr>
</tbody>
</table>

You also determine that the Box-Pierce $Q$ statistic, where $Q = T \sum_{k=1}^{15} \hat{\rho}_k^2$, is 9.38.

You must decide if the time series has been generated by a white noise process.

Which of the following is true?

(A) If the time series has been generated by a white noise process, then the sample autocorrelation coefficients are distributed approximately according to a normal distribution with mean zero and standard deviation 0.1.

(B) Because the absolute values of three of the fifteen sample autocorrelation coefficients exceed 0.1, the probability is 80% that the time series is not generated by a white noise process.

(C) Because none of the absolute values of the fifteen sample autocorrelation coefficients exceeds 0.2, the probability is 95% that all of the true autocorrelation coefficients are simultaneously zero.

(D) The $Q$ statistic is approximately chi-square distributed with 85 degrees of freedom.

(E) Because the $Q$ statistic does not exceed its critical value at the 0.05 level of significance, the probability is 95% that the true autocorrelation coefficients are all zero.
10. You observe the following five ground-up claims from a data set that is truncated from below at 100:

\[ 125 \quad 150 \quad 165 \quad 175 \quad 250 \]

You fit a ground-up exponential distribution using maximum likelihood estimation.

Determine the mean of the fitted distribution.

(A) 73
(B) 100
(C) 125
(D) 156
(E) 173
11. An insurer writes a large book of home warranty policies. You are given the following information regarding claims filed by insureds against these policies:

(i) A maximum of one claim may be filed per year.

(ii) The probability of a claim varies by insured, and the claims experience for each insured is independent of every other insured.

(iii) The probability of a claim for each insured remains constant over time.

(iv) The overall probability of a claim being filed by a randomly selected insured in a year is 0.10.

(v) The variance of the individual insured claim probabilities is 0.01.

An insured selected at random is found to have filed 0 claims over the past 10 years.

Determine the Bühlmann credibility estimate for the expected number of claims the selected insured will file over the next 5 years.

(A) 0.04
(B) 0.08
(C) 0.17
(D) 0.22
(E) 0.25
12. A study of the time to first claim includes only policies issued during 1996 through 1998 on which claims occurred by the end of 1999.

The table below summarizes the information about the 50 policies included in the study:

<table>
<thead>
<tr>
<th>Year of Issue</th>
<th>Number of Policies</th>
<th>Time to First Claim</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1 year</td>
</tr>
<tr>
<td>1996</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>1997</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>1998</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

Use the Product-Limit estimator to estimate the conditional probability that the first claim on a policy occurs less than 2 years after issue given that the claim occurs no later than 3 years after issue.

(A) Less than 0.20
(B) At least 0.20, but less than 0.25
(C) At least 0.25, but less than 0.30
(D) At least 0.30, but less than 0.35
(E) At least 0.35
13. You fit the following model to four observations:

\[ Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_i, \quad i = 1, 2, 3, 4 \]

You are given:

<table>
<thead>
<tr>
<th>( i )</th>
<th>( X_{2i} )</th>
<th>( X_{3i} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-3</td>
<td>-1</td>
</tr>
<tr>
<td>2</td>
<td>-1</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>-3</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

The least squares estimator of \( \beta_3 \) is expressed as

\[ \hat{\beta}_3 = \sum_{i=1}^{4} w_i Y_i. \]

Determine \((w_1, w_2, w_3, w_4)\).

(A) \((-0.15, -0.05, 0.05, 0.15)\)

(B) \((-0.05, 0.15, -0.15, 0.05)\)

(C) \((-0.05, 0.05, -0.15, 0.15)\)

(D) \((-0.3, -0.1, 0.1, 0.3)\)

(E) \((-0.1, 0.3, -0.3, 0.1)\)
14. For a group of insureds, you are given:

(i) The amount of a claim is uniformly distributed but will not exceed a certain unknown limit \( \theta \).

(ii) The prior distribution of \( \theta \) is \( \pi(\theta) = \frac{500}{\theta^2}, \theta > 500 \).

(iii) Two independent claims of 400 and 600 are observed.

Determine the probability that the next claim will exceed 550.

(A) 0.19

(B) 0.22

(C) 0.25

(D) 0.28

(E) 0.31
15. You are given the following information about a general liability book of business comprised of 2500 insureds:

(i) \( X_i = \sum_{j=1}^{N_i} Y_{ij} \) is a random variable representing the annual loss of the \( i^{th} \) insured.

(ii) \( N_1, N_2, \ldots, N_{2500} \) are independent and identically distributed random variables following a negative binomial distribution with parameters \( r = 2 \) and \( \beta = 0.2 \).

(iii) \( Y_{i1}, Y_{i2}, \ldots, Y_{iN_i} \) are independent and identically distributed random variables following a Pareto distribution with \( \alpha = 3.0 \) and \( \theta = 1000 \).

(iv) 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) 0.34
(B) 0.42
(C) 0.47
(D) 0.50
(E) 0.53
16. Which of the following statements about moving-average models is false?

(A) Both simple (unweighted) moving-average models and exponentially weighted moving-average (EWMA) models can be used to forecast future values of a time series.

(B) EWMA models always give greater weight to more recent observations of the time series.

(C) Forecasts using EWMA models represent true averages because the weights applied to the observations sum to one.

(D) Moving-average forecasts are adaptive because they automatically adjust themselves to the most recently available data.

(E) With an EWMA model, the differences between adjacent forecasted values of a time series from a fixed starting point increase as the number of steps ahead increases.
17. To estimate $E[X]$, you have simulated $X_1, X_2, X_3, X_4$ and $X_5$ with the following results:

\begin{align*}
1 & \quad 2 & \quad 3 & \quad 4 & \quad 5 \\
\end{align*}

You want the standard deviation of the estimator of $E[X]$ to be less than 0.05. Estimate the total number of simulations needed.

(A) Less than 150

(B) At least 150, but less than 400

(C) At least 400, but less than 650

(D) At least 650, but less than 900

(E) At least 900
18. You are given the following information about a book of business comprised of 100 insureds:

(i) \( X_i = \sum_{j=1}^{N_i} Y_{ij} \) is a random variable representing the annual loss of the \( i \)\(^{th} \) insured.

(ii) \( N_1, N_2, \ldots, N_{100} \) are independent random variables distributed according to a negative binomial distribution with parameters \( r \) (unknown) and \( \beta = 0.2 \).

(iii) Unknown parameter \( r \) has an exponential distribution with mean 2.

(iv) \( Y_{i1}, Y_{i2}, \ldots, Y_{iN_i} \) are independent random variables distributed according to a Pareto distribution with \( \alpha = 3.0 \) and \( \theta = 1000 \).

Determine the Bühlmann credibility factor, \( Z \), for the book of business.

(A) 0.000

(B) 0.045

(C) 0.500

(D) 0.826

(E) 0.905
19-20. Use the following information for questions 19 and 20.

For a mortality study of insurance applicants in two countries, you are given:

(i)

<table>
<thead>
<tr>
<th></th>
<th>Country A</th>
<th></th>
<th></th>
<th>Country B</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t_i$</td>
<td>$d_i$</td>
<td>$Y_i$</td>
<td>$\theta_i$</td>
<td>$d_i$</td>
<td>$Y_i$</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>200</td>
<td>0.05</td>
<td>15</td>
<td>100</td>
<td>0.10</td>
</tr>
<tr>
<td>2</td>
<td>54</td>
<td>180</td>
<td>0.10</td>
<td>20</td>
<td>85</td>
<td>0.10</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>126</td>
<td>0.15</td>
<td>20</td>
<td>65</td>
<td>0.10</td>
</tr>
<tr>
<td>4</td>
<td>22</td>
<td>112</td>
<td>0.20</td>
<td>10</td>
<td>45</td>
<td>0.10</td>
</tr>
</tbody>
</table>

(ii) $Y_i$ is the number at risk over the period $(t_{i-1}, t_i)$. Deaths during the period $(t_{i-1}, t_i)$ are assumed to occur at $t_i$.

(iii) $\theta_i$ is the reference hazard rate over the period $(t_{i-1}, t_i)$. Within a country, $\theta_i$ is the same for all study participants.

(iv) $S^T(t)$ is the Product-Limit estimate of $S(t)$ based on the data for all study participants.

(v) $S^B(t)$ is the Product-Limit estimate of $S(t)$ based on the data for study participants in Country B.

19. Determine $|S^T(4) - S^B(4)|$.

(A) 0.06
(B) 0.07
(C) 0.08
(D) 0.09
(E) 0.10
19-20. (Repeated for convenience) Use the following information for questions 19 and 20.

For a mortality study of insurance applicants in two countries, you are given:

(i) | Country A | Country B |
---|---|---|
| $t_i$ | $d_i$ | $Y_i$ | $\theta_i$ | $d_i$ | $Y_i$ | $\theta_i$ |
| 1 | 20 | 200 | 0.05 | 15 | 100 | 0.10 |
| 2 | 54 | 180 | 0.10 | 20 | 85 | 0.10 |
| 3 | 14 | 126 | 0.15 | 20 | 65 | 0.10 |
| 4 | 22 | 112 | 0.20 | 10 | 45 | 0.10 |

(ii) $Y_i$ is the number at risk over the period $(t_{i-1}, t_i)$. Deaths during the period $(t_{i-1}, t_i)$ are assumed to occur at $t_i$.

(iii) $\theta_i$ is the reference hazard rate over the period $(t_{i-1}, t_i)$. Within a country, $\theta_i$ is the same for all study participants.

(iv) $S^T(t)$ is the Product-Limit estimate of $S(t)$ based on the data for all study participants.

(v) $S^B(t)$ is the Product-Limit estimate of $S(t)$ based on the data for study participants in Country B.

20. Calculate $\hat{A}(4)$, the estimated cumulative excess mortality at time 4 under the additive mortality model, based on the data for all study participants.

(A) 0.21
(B) 0.31
(C) 0.36
(D) 0.41
(E) 0.52
21. Three models have been fit to 20 observations:

Model I: \[ Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon \]

Model II: \[ Y = \beta_1 + \beta_2 (X_2 + X_3) + \varepsilon \]

Model III: \[ Y - X_3 = \beta_1 + \beta_2 (X_2 - X_3) + \varepsilon \]

You are given:

<table>
<thead>
<tr>
<th>Model</th>
<th>( ESS )</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>484</td>
</tr>
<tr>
<td>II</td>
<td>925</td>
</tr>
<tr>
<td>III</td>
<td>982</td>
</tr>
</tbody>
</table>

Calculate the value of the \( F \) statistic used to test the hypothesis \( H_0: \beta_2 + \beta_3 = 1 \).

(A) Less than 15  
(B) At least 15, but less than 16  
(C) At least 16, but less than 17  
(D) At least 17, but less than 18  
(E) At least 18
22. You fit an exponential distribution to the following data:

<table>
<thead>
<tr>
<th></th>
<th>1000</th>
<th>1400</th>
<th>5300</th>
<th>7400</th>
<th>7600</th>
</tr>
</thead>
</table>

Determine the coefficient of variation of the maximum likelihood estimate of the mean, $\theta$.

(A) 0.33  
(B) 0.45  
(C) 0.70  
(D) 1.00  
(E) 1.21
23. You are given the following information on claim frequency of automobile accidents for individual drivers:

<table>
<thead>
<tr>
<th></th>
<th>Business Use</th>
<th>Pleasure Use</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Expected Claims</td>
<td>Claim Variance</td>
</tr>
<tr>
<td>Rural</td>
<td>1.0</td>
<td>0.5</td>
</tr>
<tr>
<td>Urban</td>
<td>2.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Total</td>
<td>1.8</td>
<td>1.06</td>
</tr>
</tbody>
</table>

You are also given:

(i) Each driver’s claims experience is independent of every other driver’s.

(ii) There are an equal number of business and pleasure use drivers.

Determine the Bühlmann credibility factor for a single driver.

(A) 0.05

(B) 0.09

(C) 0.17

(D) 0.19

(E) 0.27
24. When estimating a time-series model based on $T$ observations, which of the following is false?

(A) Assuming normally distributed errors and setting aside the problem of determining past unobservable process values, the maximum likelihood estimators are the same as the least-squares estimators.

(B) The Yule-Walker equations are sufficient to provide the initial guesses for the parameter values.

(C) If the model has been specified correctly, the residuals $\hat{e}_t$ constitute a white noise process.

(D) If the model has been specified correctly, the residual autocorrelations $\hat{r}_k$ for large displacements are themselves uncorrelated, normally distributed random variables with mean 0 and variance $1/T$.

(E) Several residuals $\hat{r}_k$ much larger than $2/\sqrt{T}$ indicate that the model should be respecified.
25. You are investigating insurance fraud that manifests itself through claimants who file claims with respect to auto accidents with which they were not involved. Your evidence consists of a distribution of the observed number of claimants per accident and a standard distribution for accidents on which fraud is known to be absent. The two distributions are summarized below:

<table>
<thead>
<tr>
<th>Number of Claimants per Accident</th>
<th>Standard Probability</th>
<th>Observed Number of Accidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.25</td>
<td>235</td>
</tr>
<tr>
<td>2</td>
<td>0.35</td>
<td>335</td>
</tr>
<tr>
<td>3</td>
<td>0.24</td>
<td>250</td>
</tr>
<tr>
<td>4</td>
<td>0.11</td>
<td>111</td>
</tr>
<tr>
<td>5</td>
<td>0.04</td>
<td>47</td>
</tr>
<tr>
<td>6+</td>
<td>0.01</td>
<td>22</td>
</tr>
<tr>
<td>Total</td>
<td>1.00</td>
<td>1000</td>
</tr>
</tbody>
</table>

Determine the result of a chi-square test of the null hypothesis that there is no fraud in the observed accidents.

(A) Reject at the 0.005 significance level.
(B) Reject at the 0.010 significance level, but not at the 0.005 level.
(C) Reject at the 0.025 significance level, but not at the 0.010 level.
(D) Reject at the 0.050 significance level, but not at the 0.025 level.
(E) Do not reject at the 0.050 significance level.
26. You are given the following data on large business policyholders:

(i) Losses for each employee of a given policyholder are independent and have a common mean and variance.

(ii) The overall average loss per employee for all policyholders is 20.

(iii) The variance of the hypothetical means is 40.

(iv) The expected value of the process variance is 8000.

(v) The following experience is observed for a randomly selected policyholder:

<table>
<thead>
<tr>
<th>Year</th>
<th>Average Loss per Employee</th>
<th>Number of Employees</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15</td>
<td>800</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>600</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>400</td>
</tr>
</tbody>
</table>

Determine the Bühlmann-Straub credibility premium per employee for this policyholder.

(A) Less than 10.5

(B) At least 10.5, but less than 11.5

(C) At least 11.5, but less than 12.5

(D) At least 12.5, but less than 13.5

(E) At least 13.5
27. You are given the following information about a group of 10 claims:

<table>
<thead>
<tr>
<th>Claim Size Interval</th>
<th>Number of Claims in Interval</th>
<th>Number of Claims Censored in Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0-15,000]</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(15,000-30,000]</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(30,000-45,000]</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

Assume that claim sizes and censorship points are uniformly distributed within each interval.

Estimate, using the life table methodology, the probability that a claim exceeds 30,000.

(A) 0.67
(B) 0.70
(C) 0.74
(D) 0.77
(E) 0.80
28. You fit the model \( Y = \alpha + \beta X + \varepsilon \).

The error variance is proportional to \( X^{-1/2} \).

Which of the following models corrects for this form of heteroscedasticity?

(A) \( Y X^{1/4} = \alpha.5 X^{1/4} + \beta.5 X^{3/4} + \varepsilon^* \)

(B) \( Y X^{1/4} = \alpha + \beta X^{3/4} + \varepsilon^* \)

(C) \( Y X^{1/2} = \alpha X^{1/2} + \beta X^{3/2} + \varepsilon^* \)

(D) \( Y X^{-1/4} = \alpha X^{-1/4} + \beta X^{3/4} + \varepsilon^* \)

(E) \( Y X^{-1/2} = \alpha X^{-1/2} + \beta X^{3/2} + \varepsilon^* \)
29. In order to simplify an actuarial analysis Actuary A uses an aggregate distribution
\[ S = X_1 + ... + X_N, \]
where \( N \) has a Poisson distribution with mean 10 and \( X_i = 1.5 \) for all \( i \).

Actuary A’s work is criticized because the actual severity distribution is given by

\[ \Pr(Y_i = 1) = \Pr(Y_i = 2) = 0.5, \text{ for all } i, \]

where the \( Y_i \)'s are independent.

Actuary A counters this criticism by claiming that the correlation coefficient between \( S \) and
\[ S^* = Y_1 + ... + Y_N \]
is high.

Calculate the correlation coefficient between \( S \) and \( S^* \).

(A) 0.75
(B) 0.80
(C) 0.85
(D) 0.90
(E) 0.95
30. You are making credibility estimates for regional rating factors. You observe that the Bühlmann-Straub nonparametric empirical Bayes method can be applied, with rating factor playing the role of pure premium.

\( X_{ij} \) denotes the rating factor for region \( i \) and year \( j \), where \( i = 1,2,3 \) and \( j = 1,2,3,4 \). Corresponding to each rating factor is the number of reported claims, \( m_{ij} \), measuring exposure.

You are given:

<table>
<thead>
<tr>
<th>( i )</th>
<th>( m_i = \sum_{j=1}^{4} m_{ij} )</th>
<th>( \bar{X}<em>i = \frac{1}{m_i} \sum</em>{j=1}^{4} m_{ij} X_{ij} )</th>
<th>( \hat{\nu}<em>i = \frac{1}{3} \sum</em>{j=1}^{4} m_{ij} (X_{ij} - \bar{X}_i)^2 )</th>
<th>( m_i (\bar{X}_i - \bar{X})^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>1.406</td>
<td>0.536</td>
<td>0.887</td>
</tr>
<tr>
<td>2</td>
<td>300</td>
<td>1.298</td>
<td>0.125</td>
<td>0.191</td>
</tr>
<tr>
<td>3</td>
<td>150</td>
<td>1.178</td>
<td>0.172</td>
<td>1.348</td>
</tr>
</tbody>
</table>

Determine the credibility estimate of the rating factor for region 1 using the method that preserves \( \sum_{i=1}^{3} m_i \bar{X}_i \).

(A) 1.31
(B) 1.33
(C) 1.35
(D) 1.37
(E) 1.39
31. A study of short-term disability claims produced the following information:

(i) The study period began January 1, 1999 and ended December 31, 2000.

(ii) A random sample was taken of 100 individuals who began short-term disability claims sometime during the study period.

(iii) The following results were observed:

<table>
<thead>
<tr>
<th>$t_i$</th>
<th>$d_i$</th>
<th>$c_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

where:

- $d_i =$ number of claimants who returned to work after spending $t_i$ months on disability
- $c_i =$ number of claimants who were still on disability after $t_i$ months as of December 31, 2000

You use the one-sample log-rank test to test whether these 100 individuals come from an exponential survival model with a hazard rate of 0.24.

Determine the value of the chi-square test statistic.

(A) 2.1
(B) 2.5
(C) 3.4
(D) 4.2
(E) 5.3
32. You are given:

\[ y_t = 0.5 y_{t-1} + 2.0 + \varepsilon_t \]

\[ y_T = 6.0 \]

Calculate \( \hat{y}_T(3) \), the three-period forecast.

(A) 4.00
(B) 4.25
(C) 4.50
(D) 4.75
(E) 5.00
33. You are given:

(i) Claim amounts follow a shifted exponential distribution with probability density function:

\[ f(x) = \frac{1}{\theta} e^{-(x-\delta)\theta}, \quad \delta < x < \infty \]

(ii) A random sample of claim amounts \( X_1, X_2, \ldots, X_{10} \):

\[ 5 \quad 5 \quad 5 \quad 6 \quad 8 \quad 9 \quad 11 \quad 12 \quad 16 \quad 23 \]

(iii) \( \sum X_i = 100 \) and \( \sum X_i^2 = 1306 \)

Estimate \( \delta \) using the method of moments.

(A) 3.0
(B) 3.5
(C) 4.0
(D) 4.5
(E) 5.0
You are given:

(i) The annual number of claims for each policyholder follows a Poisson distribution with mean $\theta$.

(ii) The distribution of $\theta$ across all policyholders has probability density function:

$$f(\theta) = \theta e^{-\theta}, \theta > 0$$

(iii) $\int_0^\infty \theta e^{-\theta} d\theta = \frac{1}{n^2}$

A randomly selected policyholder is known to have had at least one claim last year.

Determine the posterior probability that this same policyholder will have at least one claim this year.

(A) 0.70

(B) 0.75

(C) 0.78

(D) 0.81

(E) 0.86
35. You observe $N$ independent observations from a process whose true model is:

$$Y_i = \alpha + \beta X_i + \epsilon_i$$

You are given:

(i) $Z_i = X_i^2$, for $i = 1, 2, ..., N$

(ii) $b^* = \frac{\sum(Z_i - \bar{Z})(Y_i - \bar{Y})}{\sum(Z_i - \bar{Z})(X_i - \bar{X})}$

Which of the following is true?

(A) $b^*$ is a nonlinear estimator of $\beta$.

(B) $b^*$ is a heteroscedasticity-consistent estimator (HCE) of $\beta$.

(C) $b^*$ is a linear biased estimator of $\beta$.

(D) $b^*$ is a linear unbiased estimator of $\beta$, but not the best linear unbiased estimator (BLUE) of $\beta$.

(E) $b^*$ is the best linear unbiased estimator (BLUE) of $\beta$. 
36. For an insurance policy, you are given:

(i) The policy limit is 1,000,000 per loss, with no deductible.

(ii) Expected aggregate losses are 2,000,000 annually.

(iii) The number of losses exceeding 500,000 follows a Poisson distribution.

(iv) The claim severity distribution has

\[ \Pr(\text{Loss} > 500,000) = 0.0106 \]
\[ E[\min(\text{Loss}; 500,000)] = 20,133 \]
\[ E[\min(\text{Loss}; 1,000,000)] = 23,759 \]

Determine the probability that no losses will exceed 500,000 during 5 years.

(A) 0.01

(B) 0.02

(C) 0.03

(D) 0.04

(E) 0.05
37. A survival study gave (1.63, 2.55) as the 95% linear confidence interval for the cumulative hazard function $H(t_0)$.

Calculate the 95% log-transformed confidence interval for $H(t_0)$.

(A) (0.49, 0.94)  
(B) (0.84, 3.34)  
(C) (1.58, 2.60)  
(D) (1.68, 2.50)  
(E) (1.68, 2.60)
38. You are given:

(i) Claim size, \( X \), has mean \( \mu \) and variance 500.

(ii) The random variable \( \mu \) has a mean of 1000 and variance of 50.

(iii) The following three claims were observed: 750, 1075, 2000

Calculate the expected size of the next claim using Bühlmann credibility.

(A) 1025
(B) 1063
(C) 1115
(D) 1181
(E) 1266
39. For an ARMA(1,1) model, you are given:

\[ y_t = 0.8y_{t-1} + 3 + \epsilon_t - 0.3\epsilon_{t-1} \]

Calculate \( \rho_1 \).

(A) –0.6
(B) –0.3
(C) 0.0
(D) 0.3
(E) 0.6
40. Losses come from a mixture of an exponential distribution with mean 100 with probability $p$ and an exponential distribution with mean 10,000 with probability $1 - p$.

Losses of 100 and 2000 are observed.

Determine the likelihood function of $p$.

(A) $\left( \frac{pe^{-1}}{100} \cdot \frac{(1 - p)e^{-0.01}}{10,000} \right) \left( \frac{pe^{-20}}{100} \cdot \frac{(1 - p)e^{-0.2}}{10,000} \right)$

(B) $\left( \frac{pe^{-1}}{100} \cdot \frac{(1 - p)e^{-0.01}}{10,000} \right) + \left( \frac{pe^{-20}}{100} \cdot \frac{(1 - p)e^{-0.2}}{10,000} \right)$

(C) $\left( \frac{pe^{-1}}{100} + \frac{(1 - p)e^{-0.01}}{10,000} \right) \cdot \left( \frac{pe^{-20}}{100} + \frac{(1 - p)e^{-0.2}}{10,000} \right)$

(D) $\left( \frac{pe^{-1}}{100} + \frac{(1 - p)e^{-0.01}}{10,000} \right) + \left( \frac{pe^{-20}}{100} + \frac{(1 - p)e^{-0.2}}{10,000} \right)$

(E) $p \cdot \left( \frac{e^{-1}}{100} + \frac{e^{-0.01}}{10,000} \right) + (1 - p) \cdot \left( \frac{e^{-20}}{100} + \frac{e^{-0.2}}{10,000} \right)$

**END OF EXAMINATION**
Exam

<table>
<thead>
<tr>
<th>Test Item</th>
<th>Key</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>E</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
</tr>
<tr>
<td>5</td>
<td>B</td>
</tr>
<tr>
<td>6</td>
<td>E</td>
</tr>
<tr>
<td>7</td>
<td>C</td>
</tr>
<tr>
<td>8</td>
<td>D</td>
</tr>
<tr>
<td>9</td>
<td>A</td>
</tr>
<tr>
<td>10</td>
<td>A</td>
</tr>
<tr>
<td>11</td>
<td>D</td>
</tr>
<tr>
<td>12</td>
<td>A</td>
</tr>
<tr>
<td>13</td>
<td>B</td>
</tr>
<tr>
<td>14</td>
<td>E</td>
</tr>
<tr>
<td>15</td>
<td>C</td>
</tr>
<tr>
<td>16</td>
<td>E</td>
</tr>
<tr>
<td>17</td>
<td>E</td>
</tr>
<tr>
<td>18</td>
<td>E</td>
</tr>
<tr>
<td>19</td>
<td>B</td>
</tr>
<tr>
<td>20</td>
<td>B</td>
</tr>
<tr>
<td>21</td>
<td>D</td>
</tr>
<tr>
<td>22</td>
<td>B</td>
</tr>
<tr>
<td>23</td>
<td>D</td>
</tr>
<tr>
<td>24</td>
<td>C</td>
</tr>
<tr>
<td>25</td>
<td>A</td>
</tr>
<tr>
<td>26</td>
<td>C</td>
</tr>
<tr>
<td>27</td>
<td>C</td>
</tr>
<tr>
<td>28</td>
<td>A</td>
</tr>
<tr>
<td>29</td>
<td>E</td>
</tr>
<tr>
<td>30</td>
<td>C</td>
</tr>
<tr>
<td>31</td>
<td>D</td>
</tr>
<tr>
<td>32</td>
<td>B</td>
</tr>
<tr>
<td>33</td>
<td>D</td>
</tr>
<tr>
<td>34</td>
<td>D</td>
</tr>
<tr>
<td>35</td>
<td>D</td>
</tr>
<tr>
<td>36</td>
<td>A</td>
</tr>
<tr>
<td>37</td>
<td>E</td>
</tr>
<tr>
<td>38</td>
<td>B</td>
</tr>
<tr>
<td>39</td>
<td>E</td>
</tr>
<tr>
<td>40</td>
<td>C</td>
</tr>
</tbody>
</table>
Course 4 Solutions
November 2001 Exams
Question #1
Answer is B

From the Yule-Walker equations:
\[ \rho_1 = \phi_1 + \rho \phi_2 \]
\[ \rho_2 = \rho \phi_1 + \phi_2. \]

Substituting the given quantities yields:
\[ 0.53 = \phi_1 + 0.53\phi_2 \]
\[ -0.22 = 0.53\phi_1 + \phi_2. \]

The solution is \( \phi_1 = 0.90 \) and \( \phi_2 = -0.70. \)

The next Yule-Walker equation is:
\[ \rho_3 = \phi_1 \rho_2 + \phi_2 \rho_1 \]
\[ = 0.90(-0.22) + -0.70(0.53) \]
\[ = -0.57. \]

Question #2
Answer is E

For an interval running from \( c \) to \( d \), the uniform density function is \( f(x) = \frac{g}{n(d-c)} \) where \( g \) is the number of observations in the interval and \( n \) is the sample size. The contribution to the second raw moment for this interval is:
\[
\int_c^d x^2 \frac{g}{n(d-c)} dx = \frac{gx^3}{3n(d-c)} \bigg|_c^d = \frac{g(d^3-c^3)}{3n(d-c)}.
\]

For this problem, the second raw moment is:
\[
\frac{1}{90} \left[ \frac{30(25^3 - 0^3)}{3(25-0)} + \frac{32(50^3 - 25^3)}{3(50-25)} + \frac{20(100^3 - 50^3)}{3(100-50)} + \frac{8(200^3 - 100^3)}{3(200-100)} \right] = 3958.33.
\]
Question #3
Answer is B

Because the Bayes and Bühlmann results must be identical, this problem can be solved either way. For the Bühlmann approach, \( \mu(\lambda) = \nu(\lambda) = \lambda \). Then, noting that the prior distribution is a gamma distribution with parameters 50 and 1/500, we have:
\[
\begin{align*}
\mu &= E(\lambda) = 50/500 = 0.1 \\
\nu &= E(\lambda) = 0.1 \\
a &= \text{Var}(\lambda) = 50/500^2 = 0.0002 \\
k &= \nu/a = 500 \\
Z &= 1500/(1500 + 500) = 0.75 \\
\bar{X} &= \frac{75 + 210}{600 + 900} = 0.19.
\end{align*}
\]

The credibility estimate is \( 0.75(0.19) + 0.25(0.1) = 0.1675 \). For 1100 policies, the expected number of claims is \( 1100(0.1675) = 184.25 \).

For the Bayes approach, the posterior density is proportional to (because in a given year the number of claims has a Poisson distribution with parameter \( \lambda \) times the number of policies)
\[
\begin{align*}
\frac{e^{-600\lambda} (600\lambda)^{75} e^{-900\lambda} (900\lambda)^{210} (500\lambda)^{50} e^{-500\lambda}}{75! \cdot 210! \cdot \lambda \Gamma(50)} \propto \lambda^{335} e^{-2000\lambda} \end{align*}
\]
which is a gamma density with parameters 335 and 1/2000. The expected number of claims per policy is \( 335/2000 = 0.1675 \) and the expected number of claims in the next year is 184.25.

Question #4
Answer is B

All but B can be seen as true from various items on pages 91-96 of *Survival Analysis*. B is false because if the last observed time is a death time, then the number of deaths is equal to the number at risk (that is, \( d = Y \)). Thus the survival function will be multiplied by zero and will become zero.
Question #5
Answer is B

\[ s^2 = \frac{\sum \epsilon_i^2}{N - 2} = \frac{2394}{6} = 399 \]

\[ s^2_{\beta} = \frac{s^2}{\sum x_i^2} = \frac{399}{162} = 246.3 \]

\[ s_{\beta} = 15.69 \]

\[ t_{0.95} = 1.943 \]

The confidence interval is 

\[-35.69 \pm 1.943(15.69) = (-66.2, -5.2)\]

Question #6
Answer is E

The q-q plot takes the ordered values and plots the jth point at \( j/(n+1) \) on the horizontal axis and at \( F(x_j; \theta) \) on the vertical axis. For small values, the model assigns more probability to being below that value than occurred in the sample. This indicates that the model has a heavier left tail than the data. For large values, the model again assigns more probability to being below that value (and so less probability to being above that value). This indicates that the model has a lighter right tail than the data. Of the five answer choices, only E is consistent with these observations. In addition, note that as you go from 0.4 to 0.6 on the horizontal axis (thus looking at the middle 20% of the data), the q-q plot increases from about 0.3 to 0.4 indicating that the model puts only about 10% of the probability in this range, thus confirming answer E.

Question #7
Answer is C

The posterior probability of having one of the coins with a 50% probability of heads is proportional to \((.5)(.5)(.5)(.5)(4/6) = 0.04167\). This is obtained by multiplying the probabilities of making the successive observations 1, 1, 0, and 1 with the 50% coin times the prior probability of 4/6 of selecting this coin. The posterior probability for the 25% coin is proportional to \((.25)(.25)(.75)(.25)(1/6) = 0.00195\) and the posterior probability for the 75% coin is proportional to \((.75)(.75)(.25)(.75)(1/6) = 0.01758\). These three numbers total 0.06120. Dividing by this sum gives the actual posterior probabilities of 0.68088, 0.03186, and 0.28726. The expected value for the fifth toss is then \(.68088(.5) + .03186(.25) + .28726(.75) = 0.56385\).
Question #8
Answer is D

From Section 8.6 of *Survival Analysis*, the “times” to be considered are $t_1 = 1, t_2 = 2$ and $t_3 = 3$. $W(t_1) = 3 + 3e^b = 8.4664; \ W(t_2) = 3 + 2e^b = 6.6442; \ W(t_3) = 2 + 2e^b = 5.6442$.

By formula (8.6.2),
$$\hat{H}_0(4) = \frac{1}{8.4664} + \frac{1}{6.6442} + \frac{1}{5.6442} = 0.4458.$$ 

So $\hat{S}_0(4) = e^{-0.4458} = 0.6403$.

Question #9
Answer is A

This material is on page 496 of *Econometric Models*. Answer A is true because the standard deviation is $1/\sqrt{T} = 1/\sqrt{100} = 0.1$, while B is nonsense, C would be correct in reference to one autocorrelation coefficient, but not all fifteen, if the interpretation of a hypothesis test were corrected (95% is the probability of not rejecting a hypothesis, *given* that the null hypothesis is true), D would be correct if it stated 15 degrees of freedom, and E misinterprets the results of a hypothesis test.

Question #10
Answer is A

Because the exponential distribution is memoryless, the excess over the deductible is also exponential with the same parameter. So subtracting 100 from each observation yields data from an exponential distribution and noting that the maximum likelihood estimate is the sample mean gives the answer of 73.

Working from first principles,
$$L(\theta) = \frac{f(x_1)f(x_2)f(x_3)f(x_4)f(x_5)}{[1 - F(100)]^5} = \frac{\theta^{-1}e^{-125/\theta}\theta^{-1}e^{-150/\theta}\theta^{-1}e^{-165/\theta}\theta^{-1}e^{-175/\theta}\theta^{-1}e^{-250/\theta}}{(e^{-100/\theta})^5}$$
$$= \theta^{-5}e^{-365/\theta}.$$ 

Taking logarithms and then a derivative gives
$$l(\theta) = -5\ln(\theta) - 365/\theta, \ l'(\theta) = -5/\theta + 365/\theta^2 = 0.$$ 

The solution is $\hat{\theta} = 365/5 = 73$. 
Question #11
Answer is D

The number of claims for each insured has a binomial distribution with \( n = 1 \) and \( q \) unknown. We have

\[
\mu(q) = q, \quad \nu(q) = q(1-q)
\]

\[
\mu = E(q) = 0.1, \quad \text{given in item (iv)}
\]

\[
a = Var(q) = E(q^2) - E(q)^2 = E(q^2) - 0.01 = 0.01, \quad \text{given in item (v)}
\]

Therefore, \( E(q^2) = 0.02 \)

\[
\nu = E(q - q^2) = 0.1 - 0.02 = 0.08
\]

\[
k = \nu / a = 8, \quad Z = \frac{10}{10 + 8} = \frac{10}{18} = \frac{5}{9}.
\]

Then the expected number of claims in the next one year is \( (5/9)(0) + (4/9)(0.1) = 2/45 \) and the expected number of claims in the next five years is \( 5(2/45) = 2/9 = 0.22 \).

Question #12
Answer is A

Using the set-up as in the text, the solution proceeds as follows: Taking one year as the unit of time, we have \( \tau = 3 \), \( X \) is the time between the issue and the first claim on a policy, and we want to estimate \( P(X < 2 | X \leq 3) \).

| No. Of Policies | \( T_i \) | \( X_i \) | \( R_i \) | \( d_i \) | \( Y_i \) | \( P(X < x_i | X \leq 3) \) |
|-----------------|--------|--------|--------|--------|--------|-----------------------|
| 5               | 0      | 1      | 2      |        |        |                       |
| 6               | 1      | 1      | 2      |        |        |                       |
| 7               | 2      | 1      | 2      | 18     | 18     | 0                     |
| 9               | 0      | 2      | 1      |        |        |                       |
| 10              | 1      | 2      | 1      | 19     | 30     | \( \left( \frac{14}{27} \right) \times \left( \frac{11}{30} \right) = 0.1901 \) |
| 13              | 0      | 3      | 0      | 13     | 27     | \( \frac{14}{27} = 0.5185 \) |

The answer is the middle number in the last column, namely 0.1901.

Alternatively, perhaps all that one remembers is that for right-truncated data the Kaplan-Meier estimate can be used provided we work with, in this case, the variable \( R = 3 - X \). Then the observations become left truncated. The probability we seek is
\[
\Pr(X < 2 \mid X \leq 3) = \Pr(3 - R < 2 \mid 3 - R \leq 3) = \Pr(R > 1 \mid R \geq 0)
\]
and because \(R\) cannot be negative, this reduces to \(\Pr(R > 1)\).

The six entries in the original table can be identified as follows:

<table>
<thead>
<tr>
<th>Number of entries</th>
<th>Left truncation point</th>
<th>Value of (R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

The left truncation point is three minus the right truncation point. For example the entries in the second row of the original table could have \(X\) values of 1 or 2, but no higher. So they have a right truncation point 2 which for \(R\) is a left truncation point of 3 - 2 = 1. We then observe that the risk set at time 0 is 27 (the observations with a left truncation point at 0) and of them, there were 13 deaths. The Kaplan-Meier estimate of surviving past time 0 is then \((14/27)\). At time 1 the risk set has 30 members (the 43 who were left truncated at 0 or 1 less the 13 who died at time 0) of which 19 died (had an \(R\) value of 1). The Kaplan-Meier estimate of surviving past time 1 is \((14/27)(11/30) = 0.1901\).

**Question #13**

**Answer is B**

From the matrix formulas for multiple regression,

\[
X = \begin{bmatrix} 1 & -3 & -1 \\ 1 & -1 & 3 \\ 1 & 1 & -3 \\ 1 & 3 & 1 \end{bmatrix}, \quad X'X = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 20 & 0 \\ 0 & 0 & 20 \end{bmatrix}, \quad (X'X)^{-1} = \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/20 & 0 \\ 0 & 0 & 1/20 \end{bmatrix}
\]

and then

\[
\hat{\beta} = (X'X)^{-1}X'Y = \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/20 & 0 \\ 0 & 0 & 1/20 \end{bmatrix} \begin{bmatrix} 1/4 & 0 & 0 \\ 1/4 & 1 & 1 \\ 1/4 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ -3/20 & -1/20 & 1/20 & 3/20 \\ -1/20 & 3/20 & -3/20 & 1/20 \end{bmatrix} \end{bmatrix}
\]

The coefficients of \(\hat{\beta}_3\) are found in the third row of the matrix.

Alternatively, one may use formula (4.5) on page 86 of *Econometric Models*.
Question #14
Answer is E

The model distribution is \( f(x | \theta) = 1/\theta, 0 < x < \theta \). Then the posterior distribution is proportional to
\[
\pi(\theta | 400,600) \propto \frac{1}{\theta} \frac{1}{\theta} \frac{500}{\theta^2} \propto \theta^{-4}, \theta > 600.
\]
It is important to note the range. Being a product, the posterior density function is non-zero only when all three terms are non-zero. Because one of the observations was equal to 600, the value of the parameter must be greater than 600 in order for the density function at 600 to be positive. Or, by general reasoning, posterior probability can only be assigned to possible values. Having observed the value 600 we know that parameter values less than or equal to 600 are not possible.

The constant is obtained from \( \int_{600}^{\infty} \theta^{-4} d\theta = \frac{1}{3(600)^3} \) and thus the exact posterior density is
\[
\pi(\theta | 400,600) = 3(600)^3 \theta^{-4}, \theta > 600.
\]
The posterior probability of an observation exceeding 550 is
\[
Pr(X_3 > 550 | 400,600) = \int_{600}^{\infty} Pr(X_3 > 550 | \theta) \pi(\theta | 400,600) d\theta
\]
\[
= \int_{600}^{\infty} \frac{\theta - 550}{\theta^3(600)^3} \theta^{-4} d\theta = 0.3125
\]
where the first term in the integrand is the probability of exceeding 550 from the uniform distribution.

Question #15
Answer is C

\[
E(N) = r\beta = 0.40
\]
\[
Var(N) = r\beta(1 + \beta) = 0.48
\]
\[
E(Y) = \theta/(\alpha - 1) = 500
\]
\[
Var(Y) = \theta^2\alpha / \left[ (\alpha - 1)^2 (\alpha - 2) \right] = 750,000
\]

Therefore,
\[
E(X) = 0.40(500) = 200
\]
\[
Var(X) = 0.40(750,000) + 0.48(500)^2 = 420,000
\]

The full credibility standard is \( n = \left( \frac{1.645}{0.05} \right)^2 \frac{420,000}{200^2} = 11,365 \) and then
\[
Z = \sqrt{2500/11,365} = 0.47.
\]
Question #16
Answer is E

In an EWMA model, all forecasted values from a fixed starting point are the same. Therefore, the differences between all forecasted values are zero. The other four statements are true based on material from pages 476-477 of *Econometric Models*.

Question #17
Answer is E

The sample variance is \( s^2 = \frac{(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2}{4} = 2.5 \). The estimator of \( E[X] \) is the sample mean and the variance of the sample mean is the variance divided by the sample size, estimated here as \( 2.5/n \). Setting the standard deviation of the estimator equal to 0.05 gives the equation \( \sqrt{2.5/n} = 0.05 \) which yields \( n = 1000 \).

Question #18
Answer is E

\[
\begin{align*}
\mu(r) &= E(X \mid r) = E(N)E(Y) = r^\beta \theta / (\alpha - 1) = 100r \\
\nu(r) &= Var(X \mid r) = Var(N)E(Y)^2 + E(N)Var(Y) \\
&= r^\beta (1 + \beta) \theta^2 / (\alpha - 1)^2 + r^\beta \alpha \theta^2 / [(\alpha - 1)^2 (\alpha - 2)] = 210,000r.
\end{align*}
\]

\( v = E(210,000r) = 210,000(2) = 420,000 \)
\( a = Var(100r) = (100)^2 (4) = 40,000 \)
\( k = v/a = 10.5 \)
\( Z = 100/(100 + 10.5) = 0.905. \)

Question #19
Answer is B

Using all participants, \( S^T(4) = \left( \begin{array}{c} 1 - \frac{35}{300} \\ 1 - \frac{74}{265} \\ 1 - \frac{34}{191} \\ 1 - \frac{32}{157} \end{array} \right) = 0.41667 \).

Using only Country B, \( S^B(4) = \left( \begin{array}{c} 1 - \frac{15}{100} \\ 1 - \frac{20}{85} \\ 1 - \frac{20}{65} \\ 1 - \frac{10}{45} \end{array} \right) = 0.35 \).

The difference is \( S^T(4) - S^B(4) = 0.41667 - 0.35 = 0.0667 = 0.07 \).
Question #20
Answer is B

From page 168 of Survival Analysis,

\[ \Theta(4) = \frac{0.05(200)+0.10(100)}{300} + \frac{0.10(180)+0.10(85)}{265} + \frac{0.15(126)+0.10(65)}{191} + \frac{0.20(112)+0.10(45)}{157} = 0.47099. \]

Then,

\[ \hat{\Theta}(4) = \frac{35}{300} + \frac{74}{265} + \frac{34}{191} + \frac{32}{157} - 0.47099 = 0.30675. \]

Question #21
Answer is D

The unrestricted model is Model I with ESS = 484. To obtain the restricted model, substitute 1 – \( \beta_2 \) for \( \beta_3 \) to yield Model III with ESS = 982. Then,

\[ F = \frac{(982 - 484)/1}{484/17} = 17.49. \] The 17 in the denominator is the sample size of 20 less the 3 parameters in the unrestricted model. The 1 in the numerator is the 3 parameters in the unrestricted model less the 2 parameters in the restricted model.

Question #22
Answer is B

For an exponential distribution the maximum likelihood estimate of the mean is the sample mean. We have

\[ E(\bar{X}) = E(X) = \theta, \ Var(\bar{X}) = Var(X)/n = \theta^2/n. \]

\[ cv = SD(\bar{X})/E(\bar{X}) = [\theta/\sqrt{n}]/\theta = 1/\sqrt{n} = 1/\sqrt{5} = 0.447. \]

If the above facts are not known, the loglikelihood function can be used:

\[ L(\theta) = \theta^{-n} e^{-\bar{x}/\theta}, \ l(\theta) = -n \ln \theta - n \bar{X} / \theta, \ l'(\theta) = -n \theta^{-1} + n \bar{X} \theta^{-2} = 0 \Rightarrow \theta = \bar{X}. \]

\[ l''(\theta) = n \theta^{-2} - 2 n \bar{X} \theta^{-3}, \ l(\theta) = E[-n \theta^{-2} + 2 n \bar{X} \theta^{-3}] = n \theta^{-2}. \]

Then, \( Var(\hat{\theta}) = \theta^2/n. \)
Question #23
Answer is D

Because the total expected claims for business use is 1.8, it must be that 20\% of business users are rural and 80\% are urban. Thus the unconditional probabilities of being business-rural and business-urban are 0.1 and 0.4 respectively. Similarly the probabilities of being pleasure-rural and pleasure-urban are also 0.1 and 0.4 respectively. Then,

\[ \mu = 0.1(1.0) + 0.4(2.0) + 0.1(1.5) + 0.4(2.5) = 2.05 \]
\[ \nu = 0.1(0.5) + 0.4(1.0) + 0.1(0.8) + 0.4(1.0) = 0.93 \]
\[ a = 0.1(1.0^2) + 0.4(2.0^2) + 0.1(1.5^2) + 0.4(2.5^2) - 2.05^2 = 0.2225 \]
\[ k = \nu / a = 4.18 \]
\[ Z = 1/(1 + 4.18) = 0.193. \]

Question #24
Answer is C

(A) True – page 551 of *Econometric Models*
(B) True – page 553
(C) False – On page 555 the text says “Then, if the model has been specified correctly, the residuals \( \hat{e}_t \) should resemble a white noise process.” Answer C is almost the same, however the word “constitute” is used in place of “resemble.” This makes the answer false because the residuals are only approximately white noise. That is because they are calculated from the estimated parameter values, which in turn are calculated from the observed values. This causes the residuals to be slightly correlated and thus not white noise. The error terms, which are calculated from the true parameter values do constitute a white noise process, so another way to make statement C true is to replace “residuals” with “errors” and remove the “hat” from \( \hat{e}_t \).
(D) True – page 555
(E) True – page 556

Question #25
Answer is A

<table>
<thead>
<tr>
<th>No. claims</th>
<th>Hypothesized</th>
<th>Observed</th>
<th>Chi-square</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>250</td>
<td>235</td>
<td>15^2/250 = 0.90</td>
</tr>
<tr>
<td>2</td>
<td>350</td>
<td>335</td>
<td>15^2/350 = 0.64</td>
</tr>
<tr>
<td>3</td>
<td>240</td>
<td>250</td>
<td>10^2/240 = 0.42</td>
</tr>
<tr>
<td>4</td>
<td>110</td>
<td>111</td>
<td>1^2/110 = 0.01</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>47</td>
<td>7^2/40 = 1.23</td>
</tr>
<tr>
<td>6+</td>
<td>10</td>
<td>22</td>
<td>12^2/10 = 14.40</td>
</tr>
</tbody>
</table>

The last column sums to the test statistic of 17.60 with 5 degrees of freedom (there were no estimated parameters), so from the table reject at the 0.005 significance level.
**Question #26**  
**Answer is C**

In part (ii) you are given that \( \mu = 20 \). In part (iii) you are given that \( a = 40 \). In part (iv) you are given that \( \nu = 8000 \). Therefore, \( k = \nu / a = 200 \). Then,

\[
\overline{X} = \frac{800(15) + 600(10) + 400(5)}{1800} = \frac{100}{9}
\]

\[
Z = \frac{1800}{1800 + 200} = 0.9
\]

\[
P_c = 0.9(100/9) + 0.1(20) = 12.
\]

**Question #27**  
**Answer is C**

\[
\text{Pr}(X > 30,000) = S(30,000) = \left(1 - \frac{1}{10 - 2/2}\right)\left(1 - \frac{1}{7 - 2/2}\right) = 20/27 = 0.741.
\]

**Question #28**  
**Answer is A**

To correct for heteroscedasticity, divide the model by something proportional to the standard deviation of the error. In this case, divide by \( \sqrt{X^{-1/2}} = X^{-1/4} \) which is equivalent to multiplying by \( X^{1/4} \). Doing so produces the model in answer A.

**Question #29**  
**Answer is E**

For Actuary A,

\[
E(S) = E(N)E(X) = 10(1.5) = 15
\]

\[
Var(S) = E(N)Var(X) + Var(N)E(X)^2 = 10(0) + 10(1.5)^2 = 22.5.
\]

The true values are,

\[
E(S^*) = E(N)E(Y) = 10(1.5) = 15
\]

\[
Var(S^*) = E(N)Var(Y) + Var(N)E(Y)^2 = 10(0.25) + 10(1.5)^2 = 25.
\]

Also,

\[
E(SS^*) = E[E[1.5N(Y_1 + \cdots + Y_N) \mid N]] = E[1.5N(1.5N)] = E(2.25N^2)
\]

\[
= 2.25[Var(N) + E(N)^2] = 2.25(10 + 100) = 247.5.
\]

Then the correlation is \( \rho = \frac{247.5 - 15(15)}{\sqrt{22.5(25)}} = 0.95 \).
**Question #30**  
**Answer is C**

The formulas are from Section 5.5 of *Loss Models*.

\[
\hat{\nu} = \frac{3(0.536 + 0.125 + 0.172)}{3 + 3 + 3} = 0.27767.
\]

\[
\hat{a} = \frac{0.887 + 0.191 + 1.348 - 2(0.27767)}{500 - \frac{1}{500}(50^2 + 300^2 + 150^2)} = 0.00693.
\]

Then,

\[
k = \frac{0.27767}{0.00693} = 40.07, \quad Z_1 = \frac{50}{50 + 40.07} = 0.55512, \quad Z_2 = \frac{300}{300 + 40.07} = 0.88217,
\]

\[
Z_3 = \frac{150}{150 + 40.07} = 0.78918.
\]

The credibility weighted mean is,

\[
\hat{\mu} = \frac{0.55512(1.406) + 0.88217(1.298) + 0.78918(1.178)}{0.55512 + 0.88217 + 0.78918} = 1.28239.
\]

The credibility premium for state 1 is

\[
P_c = 0.55512(1.406) + 0.44488(1.28239) = 1.351.
\]

**Question #31**  
**Answer is D**

For the one-sample log-rank test, the test statistic is

\[
\left[ O(\tau) - E(\tau) \right]^2 / E(\tau).
\]

\[O(12) \text{ is the observed number of events at or prior to time } 12 = 67\]

The expected number of events at or prior to time 12 is

\[E(12) = \sum H_0(T_j) - H_0(L_j).\]

From the exponential model, \( H_0(t) = 0.24t \), and so \( E(12) \) is 0.24 times the total observation time for all subjects. From the table, there were 40 subjects observed for 2 months, 30 for 3, 20 for 5, 8 for 8, and 2 for 12 months for a total of 358 months. Thus, \( E(12) = 0.24(358) = 85.92 \). The test statistic is

\[
\left( \frac{(67 - 85.92)^2}{85.92} \right) = 4.17.
\]
Question #32
Answer is B

\[ \hat{y}_r(1) = 0.5(6) + 2.0 = 5 \]
\[ \hat{y}_r(2) = 0.5(5) + 2.0 = 4.5 \]
\[ \hat{y}_r(3) = 0.5(4.5) + 2.0 = 4.25. \]

Question #33
Answer is D

\[ E(X) = \int_{0}^{\infty} \frac{x}{\theta} e^{-(x-\delta)/\theta} \, dx = \int_{0}^{\infty} \frac{y + \delta}{\theta} e^{-y/\theta} \, dy = \theta + \delta \]
\[ E(X^2) = \int_{0}^{\infty} \frac{x^2}{\theta} e^{-(x-\delta)/\theta} \, dx = \int_{0}^{\infty} \frac{y^2 + 2y\delta + \delta^2}{\theta} e^{-y/\theta} \, dy = \theta^2 + 2\theta\delta + \delta^2. \]

Both derivations use the substitution \( y = x - \delta \) and then recognize that the various integrals are requesting moments from an ordinary exponential distribution. The method of moments solves the two equations
\[ \theta + \delta = 10 \]
\[ 2\theta^2 + 2\theta\delta + \delta^2 = 130.6 \]
producing \( \hat{\delta} = 4.468. \)

It is faster to do the problem if it is noted that \( X = Y + \delta \) where \( Y \) has an ordinary exponential distribution. Then \( E(X) = E(Y) + \delta = \theta + \delta \) and \( Var(X) = Var(Y) = \theta^2. \)

Question #34
Answer is D

The posterior density is proportional to the product of the probability of the observed value and the prior density. Thus, \( \pi(\theta \mid N > 0) \propto \Pr(N > 0 \mid \theta) \pi(\theta) = (1 - e^{-\theta})\theta e^{-\theta}. \)

The constant of proportionality is obtained from \( \int_{0}^{\infty} \theta e^{-\theta} - \theta e^{-2\theta} \, d\theta = \frac{1}{1^2} \cdot \frac{1}{2^2} = 0.75. \)

The posterior density is \( \pi(\theta \mid N > 0) = (4/3)(\theta e^{-\theta} - \theta e^{-2\theta}). \)

Then,
\[ \Pr(N_2 > 0 \mid N_1 > 0) = \int_{0}^{\infty} \Pr(N_2 > 0 \mid \theta) \pi(\theta \mid N_1 > 0) \, d\theta = \int_{0}^{\infty} (1 - e^{-\theta})(4/3)(\theta e^{-\theta} - \theta e^{-2\theta}) \, d\theta \]
\[ = \frac{4}{3} \int_{0}^{\infty} \theta e^{-\theta} - 2\theta e^{-2\theta} + \theta e^{-3\theta} \, d\theta = \frac{4}{3} \left( \frac{1}{1^2} - \frac{2}{2^2} + \frac{1}{3^2} \right) = 0.8148. \]
Question #35
Answer is D

Using deviations form \((z_i = Z_i - \bar{Z}, x_i = X_i - \bar{X}, y_i = Y_i - \bar{Y})\), we have
\[b^* = \sum w_i y_i, \quad w_i = \frac{z_i}{Z_j h_{ij}}\] and so \(b^*\) is a linear estimator, making (A) false. To check for bias,
\[E(b^*) = \sum w_i E(y_i) = \sum w_i \beta x_i = \beta \sum \frac{z_i}{Z_j h_{ij}} x_i = \beta,\] making (C) false. We know that the ordinary least squares estimator, \(b = \frac{\sum X_i Y_i}{\sum X_i^2}\) is BLUE, so \(b^*\) cannot be BLUE, making (E) false and (D) true. Finally, HCE estimators are concerned with estimating variances, so (B) must be false.

Question #36
Answer is A

Let \(S\) be the annual aggregate losses, \(N\) the number of losses, and \(X\) the distribution of an individual loss, limited to 1,000,000. Then \(E(S) = E(N)E(X)\) or \(2,000,000 = E(N)(23,759)\) and so \(E(N) = 84.1786\). Because the probability of a loss exceeding 500,000 is 0.0106, the number of losses in excess of 500,000 will have a Poisson distribution with mean \((0.0106)(84.1786) = 0.89229\) per year. That means the number of such losses in 5 years has a Poisson distribution with mean \(5(0.89229) = 4.46145\). The probability of no such losses is \(e^{-4.46145} = 0.01155\).

Question #37
Answer is E

The interval is centered at 2.09 and the plus/minus term is 0.46 which must equal 1.96\(\bar{c}\) and so \(\bar{c} = 0.2347\). For the log-transformed interval we need
\[\phi = e^{1.96(0.2347)/2.09} = 1.2462.\] The lower limit is 2.09/1.2462 = 1.68 and the upper limit is 2.09(1.2462) = 2.60.

Question #38
Answer is B

From item (ii), \(\mu = 1000\) and \(a = 50\). From item (i), \(v = 500\). Therefore, \(k = v/a = 10\) and \(Z = 3/(3+10) = 3/13\). Also, \(\bar{X} = (750 + 1075 + 2000)/3 = 1275\). Then
\[P_c = (3/13)(1275) + (10/13)(1000) = 1063.46.\]
Question #39
Answer is E

Using formula (17.58) on page 536 of *Econometric Models*:

\[
\rho_1 = \frac{(1-\phi_1)(\phi_1 - \theta_1)}{1+\theta_1^2 - 2\theta_1} \\
\rho_1 = \frac{[1-0.8(0.3)](0.8-0.3)}{1+0.3^2 - 2(0.8)(0.3)} = 0.623.
\]

Question #40
Answer is C

\[
f(x) = p \frac{1}{100} e^{-x/100} + (1-p) \frac{1}{10,000} e^{-x/10,000} \\
L(100,200) = f(100)f(200) \\
= \left( \frac{pe^{-1}}{100} + \frac{(1-p)e^{-0.01}}{10,000} \right) \left( \frac{pe^{-20}}{100} + \frac{(1-p)e^{-0.2}}{10,000} \right)
\]