

**\*\*BEGINNING OF EXAMINATION\*\***

1. You fit the following model to 20 observations:

$$Y = \mathbf{a} + \mathbf{b}X + \mathbf{e}$$

You determine that  $R^2 = 0.64$ .

Calculate the value of the  $F$  statistic used to test for a linear relationship.

- (A) Less than 30
- (B) At least 30, but less than 33
- (C) At least 33, but less than 36
- (D) At least 36, but less than 39
- (E) At least 39

2. You are given the following random sample of ten claims:

46	121	493	738	775
1078	1452	2054	2199	3207

Determine the smoothed empirical estimate of the 90<sup>th</sup> percentile, as defined in Klugman, Panjer and Willmot.

- (A) Less than 2150
- (B) At least 2150, but less than 2500
- (C) At least 2500, but less than 2850
- (D) At least 2850, but less than 3200
- (E) At least 3200

3. You are given the following information about two classes of business, where  $X$  is the loss for an individual insured:

	Class 1	Class 2
Number of insureds	25	50
$E(X)$	380	23
$E(X^2)$	365,000	----

You are also given that an analysis has resulted in a Bühlmann  $k$  value of 2.65.

Calculate the process variance for Class 2.

- (A) 2,280
- (B) 2,810
- (C) 7,280
- (D) 28,320
- (E) 75,050

4. For a mortality study with right-censored data, you are given:

Time	Number of Deaths	Number at Risk
$t_i$	$d_i$	$Y_i$
5	2	15
7	1	12
10	1	10
12	2	6

Calculate  $\tilde{S}(12)$  based on the Nelson-Aalen estimate  $\tilde{H}(12)$ .

- (A) 0.48
- (B) 0.52
- (C) 0.60
- (D) 0.65
- (E) 0.67

5. You are given the following information about an MA(4) model:

$$m = 0$$

$$q_1 = 1.800$$

$$q_2 = -1.110$$

$$q_3 = 0.278$$

$$q_4 = -0.024$$

$$s_e^2 = 8.000$$

Determine the standard deviation of the forecast error three steps ahead.

- (A) 2.8
- (B) 3.6
- (C) 4.9
- (D) 5.8
- (E) 6.6

6. A jewelry store has obtained two separate insurance policies that together provide full coverage.

You are given:

- (i) The average ground-up loss is 11,100.
- (ii) Policy A has an ordinary deductible of 5,000 with no policy limit.
- (iii) Under policy A, the expected amount paid per loss is 6,500.
- (iv) Under policy A, the expected amount paid per payment is 10,000.
- (v) Policy B has no deductible and a policy limit of 5,000.

Given that a loss less than or equal to 5,000 has occurred, what is the expected payment under policy B?

- (A) Less than 2,500
- (B) At least 2,500, but less than 3,000
- (C) At least 3,000, but less than 3,500
- (D) At least 3,500, but less than 4,000
- (E) At least 4,000

7. You are given the following information about two classes of risks:
- (i) Risks in Class A have a Poisson claim count distribution with a mean of 1.0 per year.
  - (ii) Risks in Class B have a Poisson claim count distribution with a mean of 3.0 per year.
  - (iii) Risks in Class A have an exponential severity distribution with a mean of 1.0.
  - (iv) Risks in Class B have an exponential severity distribution with a mean of 3.0.
  - (v) Each class has the same number of risks.
  - (vi) Within each class, severities and claim counts are independent.

A risk is randomly selected and observed to have two claims during one year. The observed claim amounts were 1.0 and 3.0.

Calculate the posterior expected value of the aggregate loss for this risk during the next year.

- (A) Less than 2.0
- (B) At least 2.0, but less than 4.0
- (C) At least 4.0, but less than 6.0
- (D) At least 6.0, but less than 8.0
- (E) At least 8.0

8. You are given the following data on time to death:

(i)	Time $t_i$	Number of Deaths $d_i$	Number of Risks $Y_i$	$\tilde{H}(t_i)$
	10	1	20	0.0500
	34	1	19	0.1026
	47	1	18	0.1582
	75	1	17	0.2170
	156	1	16	0.2795
	171	1	15	0.3462

(ii)  $\tilde{H}(t_i)$  is the Nelson-Aalen estimate of the cumulative hazard function.

(iii)  $\hat{h}(t)$ , the kernel-smoothed estimate of the hazard rate, is determined using bandwidth 60 and the uniform kernel

$$K(x) = \begin{cases} \frac{1}{2}, & -1 \leq x \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Determine  $\hat{h}(100)$ .

- (A) 0.0010
- (B) 0.0015
- (C) 0.0029
- (D) 0.0590
- (E) 0.0885



9. The following models are fitted to 30 observations:

$$\text{Model I: } Y = \mathbf{b}_1 + \mathbf{b}_2 X_2 + \mathbf{e}$$

$$\text{Model II: } Y = \mathbf{b}_1 + \mathbf{b}_2 X_2 + \mathbf{b}_3 X_3 + \mathbf{b}_4 X_4 + \mathbf{e}$$

You are given:

$$(i) \quad \sum (Y - \bar{Y})^2 = 160$$

$$(ii) \quad \sum (X_2 - \bar{X}_2)^2 = 10$$

$$(iii) \quad \text{For Model I, } \hat{\mathbf{b}}_2 = -2$$

$$(iv) \quad \text{For Model II, } R^2 = 0.70$$

Determine the value of the  $F$  statistic used to test that  $\mathbf{b}_3$  and  $\mathbf{b}_4$  are jointly equal to zero.

- (A) Less than 15
- (B) At least 15, but less than 18
- (C) At least 18, but less than 21
- (D) At least 21, but less than 24
- (E) At least 24

**10-11.** Use the following information for questions 10 and 11.

The size of a claim for an individual insured follows an inverse exponential distribution with the following probability density function:

$$f(x|\mathbf{q}) = \frac{\mathbf{q}e^{-\mathbf{q}/x}}{x^2}, \quad x > 0$$

The parameter  $\mathbf{q}$  has a prior distribution with the following probability density function:

$$g(\mathbf{q}) = \frac{e^{-\mathbf{q}/4}}{4}, \quad \mathbf{q} > 0$$

**10.** For question 10 only, you are also given:

One claim of size 2 has been observed for a particular insured.

Which of the following is proportional to the posterior distribution of  $\mathbf{q}$ ?

(A)  $\mathbf{q}e^{-\mathbf{q}/2}$

(B)  $\mathbf{q}e^{-3\mathbf{q}/4}$

(C)  $\mathbf{q}e^{-\mathbf{q}}$

(D)  $\mathbf{q}^2e^{-\mathbf{q}/2}$

(E)  $\mathbf{q}^2e^{-9\mathbf{q}/4}$

**10-11.** (Repeated for convenience) Use the following information for questions 10 and 11.

The size of a claim for an individual insured follows an inverse exponential distribution with the following probability density function:

$$f(x|\mathbf{q}) = \frac{\mathbf{q}e^{-\mathbf{q}/x}}{x^2}, \quad x > 0$$

The parameter  $\mathbf{q}$  has a prior distribution with the following probability density function:

$$g(\mathbf{q}) = \frac{e^{-\mathbf{q}/4}}{4}, \quad \mathbf{q} > 0$$

**11.** For question 11 only, you are also given:

For a particular insured, the following five claims are observed:

1   2   3   5   13

Determine the value of the Kolmogorov-Smirnov statistic to test the goodness of fit of  $f(x|\mathbf{q}=2)$ .

- (A) Less than 0.05
- (B) At least 0.05, but less than 0.10
- (C) At least 0.10, but less than 0.15
- (D) At least 0.15, but less than 0.20
- (E) At least 0.20

12. For a mortality study, you are given:
- (i) 100 newborn mice are observed for four months.
  - (ii)  $H_0$  is a hypothesized cumulative hazard function.
  - (iii) Deaths during the month are assumed to occur at the end of the month.
  - (iv) The 95<sup>th</sup> percentile of the chi-square distribution with one degree of freedom is 3.84.
  - (v) The observed mortality experience and the values of  $H_0$  are given below.

Month	Number at beginning of month	Number of deaths	$H_0$ , beginning of month	$H_0$ , end of month
1	100	10	0.00	0.08
2	90	12	0.08	0.15
3	78	10	0.15	0.25
4	68	$d_4$	0.25	0.40

Calculate the smallest value of  $d_4$  needed for the one-sample log-rank test to yield the conclusion, at the 0.05 significance level, that the true cumulative hazard function differs from  $H_0$ .

- (A) 0
- (B) 3
- (C) 6
- (D) 9
- (E) 12

13. You smooth a time series  $y_t$  using Holt's two-parameter exponential smoothing method.

You calculate:

$t$	$y_t$	$\tilde{y}_t$	$r_t$
1995	120.5	120.0	10.0
1996	135.0	131.5	11.2
1997	147.7	144.2	12.4
1998	146.6	$\tilde{y}_{1998}$	$r_{1998}$

Determine the two-period forecast  $\hat{y}_{2000}$  by first completing Holt's two-parameter exponential smoothed series.

- (A) Less than 166
- (B) At least 166, but less than 168
- (C) At least 168, but less than 170
- (D) At least 170, but less than 172
- (E) At least 172

14. Which of the following statements about evaluating an estimator is false?

(A) Modeling error is not possible with empirical estimation.

(B)  $MSE(\hat{\mathbf{q}}) = Var(\hat{\mathbf{q}}) + [Bias(\hat{\mathbf{q}})]^2$

(C)  $S_n^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2$  is an asymptotically unbiased estimator of variance.

(D) If  $\hat{\mathbf{q}}_n$  is asymptotically unbiased and  $\lim_{n \rightarrow \infty} Var(\hat{\mathbf{q}}_n) = 0$ , then  $\hat{\mathbf{q}}_n$  is weakly consistent.

(E) A robust estimator is one that performs well even with sampling error.

15. An insurer has data on losses for four policyholders for seven years.  $X_{ij}$  is the loss from the  $i^{\text{th}}$  policyholder for year  $j$ .

You are given:

$$\sum_{i=1}^4 \sum_{j=1}^7 (X_{ij} - \bar{X}_i)^2 = 33.60$$

$$\sum_{i=1}^4 (\bar{X}_i - \bar{X})^2 = 3.30$$

Calculate the Bühlmann credibility factor for an individual policyholder using nonparametric empirical Bayes estimation.

- (A) Less than 0.74
- (B) At least 0.74, but less than 0.77
- (C) At least 0.77, but less than 0.80
- (D) At least 0.80, but less than 0.83
- (E) At least 0.83

16. You are given:

(i)  $x_1 = -2$   
 $x_2 = -1$   
 $x_3 = 0$   
 $x_4 = 1$   
 $x_5 = 2$

(ii) The true model for the data is  $y = 10x + 3x^2 + \mathbf{e}$ .

(iii) The model fitted to the data is  $y = \mathbf{b}^*x + \mathbf{e}^*$ .

Determine the expected value of the least-squares estimator of  $\mathbf{b}^*$ .

- (A) 6  
(B) 7  
(C) 8  
(D) 9  
(E) 10



17. You are given a random sample of two values from a distribution function  $F$ :

1 3

You estimate  $q(F) = \text{Var}(X)$  using the estimator  $g(X_1, X_2) = \frac{1}{2} \sum_{i=1}^2 (X_i - \bar{X})^2$ , where  $\bar{X} = \frac{X_1 + X_2}{2}$ .

Determine the bootstrap approximation to the mean square error.

- (A) 0.0
- (B) 0.5
- (C) 1.0
- (D) 2.0
- (E) 2.5

18. You are given two independent estimates of an unknown quantity  $m$ :

(i) Estimate A:  $E(\mathbf{m}_A) = 1000$  and  $S(\mathbf{m}_A) = 400$

(ii) Estimate B:  $E(\mathbf{m}_B) = 1200$  and  $S(\mathbf{m}_B) = 200$

Estimate C is a weighted average of the two estimates A and B, such that:

$$\mathbf{m}_C = w \cdot \mathbf{m}_A + (1 - w) \cdot \mathbf{m}_B$$

Determine the value of  $w$  that minimizes  $S(\mathbf{m}_C)$ .

- (A) 0
- (B) 1/5
- (C) 1/4
- (D) 1/3
- (E) 1/2

19. For a mortality study with right-censored data, the cumulative hazard rate is estimated using the Nelson-Aalen estimator.

You are given:

- (i) No deaths occur between times  $t_i$  and  $t_{i+1}$ .
- (ii) A 95% linear confidence interval for  $H(t_i)$  is (0.07125, 0.22875).
- (iii) A 95% linear confidence interval for  $H(t_{i+1})$  is (0.15607, 0.38635).

Calculate the number of deaths observed at time  $t_{i+1}$ .

- (A) 4
- (B) 5
- (C) 6
- (D) 7
- (E) 8

20. For the time series  $y_t$ , you are given:

$t$	$y_t$	$y_t - \bar{y}$
1	984	-16
2	1023	23
3	965	-35
4	1040	40
5	988	-12

Estimate the partial autocorrelation function at time displacement  $k = 2$ .

- (A) -0.46
- (B) -0.16
- (C) 0.00
- (D) 0.51
- (E) 0.84

21. You are given the following five observations:

521    658    702    819    1217

You use the single-parameter Pareto with cumulative distribution function

$$F(x) = 1 - \left(\frac{500}{x}\right)^a, \quad x > 500, \quad \mathbf{a} > 0.$$

Calculate the maximum likelihood estimate of the parameter  $\mathbf{a}$ .

- (A) 2.2
- (B) 2.5
- (C) 2.8
- (D) 3.1
- (E) 3.4

22. You are given:

- (i) A portfolio of independent risks is divided into two classes, Class A and Class B.
- (ii) There are twice as many risks in Class A as in Class B.
- (iii) The number of claims for each insured during a single year follows a Bernoulli distribution.
- (iv) Classes A and B have claim size distributions as follows:

Claim Size	Class A	Class B
50,000	0.60	0.36
100,000	0.40	0.64

- (v) The expected number of claims per year is 0.22 for Class A and 0.11 for Class B.

One insured is chosen at random. The insured's loss for two years combined is 100,000.

Calculate the probability that the selected insured belongs to Class A.

- (A) 0.55
- (B) 0.57
- (C) 0.67
- (D) 0.71
- (E) 0.73

23. You test the effect of gender and age on survival of patients receiving kidney transplants.

(i) You use a Cox proportional hazards model with the indicator variable  $Z_1$  equal to 1 when the subject is a male, and the indicator variable  $Z_2$  equal to 1 when the subject is an adult.

(ii) The resulting partial maximum likelihood parameter estimates are:

$$\begin{aligned}b_1 &= 0.25 \\b_2 &= -0.45\end{aligned}$$

(iii) The variance-covariance matrix of  $b_1$  and  $b_2$  is given by:

$$\begin{pmatrix} 0.36 & 0.10 \\ 0.10 & 0.20 \end{pmatrix}$$

Which of the following is a 95% confidence interval for the relative risk of a male child subject compared to a female adult subject?

(A) (-0.5, 1.9)

(B) (0.0, 1.4)

(C) (0.6, 6.5)

(D) (1.0, 4.1)

(E) (1.2, 3.2)

24. You are given the following linear regression results:

$t$	<u>Actual</u>	<u>Fitted</u>
1	77.0	77.6
2	69.9	70.6
3	73.2	70.9
4	72.7	72.7
5	66.1	67.1

Estimate the lag 1 serial correlation coefficient for the residuals, using the Durbin-Watson statistic.

- (A) Less than  $-0.2$
- (B) At least  $-0.2$ , but less than  $-0.1$
- (C) At least  $-0.1$ , but less than  $0.0$
- (D) At least  $0.0$ , but less than  $0.1$
- (E) At least  $0.1$



25. You model a loss process using a lognormal distribution with parameters  $\boldsymbol{m}$  and  $\boldsymbol{s}$ .

You are given:

(i) The maximum likelihood estimates of  $\boldsymbol{m}$  and  $\boldsymbol{s}$  are:

$$\hat{\boldsymbol{m}} = 4.215$$

$$\hat{\boldsymbol{s}} = 1.093$$

(ii) The estimated covariance matrix of  $\hat{\boldsymbol{m}}$  and  $\hat{\boldsymbol{s}}$  is:

$$\begin{pmatrix} 0.1195 & 0 \\ 0 & 0.0597 \end{pmatrix}$$

(iii) The mean of the lognormal distribution is  $\exp(\boldsymbol{m} + \frac{\boldsymbol{s}^2}{2})$ .

Estimate the variance of the maximum likelihood estimate of the mean of the lognormal distribution, using the delta method.

- (A) Less than 1500
- (B) At least 1500, but less than 2000
- (C) At least 2000, but less than 2500
- (D) At least 2500, but less than 3000
- (E) At least 3000

26. You are given:

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

Determine the limited fluctuation 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

27. You are analyzing the time between the occurrence and payment of claims. You are given the following data on four claims that were paid during the time period  $t = 0$  to  $t = 6$  (in months):

Time of occurrence (t)	Time between occurrence and payment
3	1
5	1
3	2
2	3

You have no information on claims that occurred during the period  $t = 0$  to  $t = 6$  but were not paid until after  $t = 6$ .

Given that a claim is paid no later than six months after its occurrence, using the appropriate form of the Product-Limit estimator, estimate the probability that the claim is paid less than two months after its occurrence.

- (A) 0
- (B)  $1/4$
- (C)  $1/3$
- (D)  $1/2$
- (E)  $2/3$

28. You are given two time series,  $x_t$  and  $y_t$ . Each time series is assumed to be a random walk.

Which of the following statements about these series is correct?

- (A) No linear combination of these two time series can be stationary.
- (B) The time series  $z_t = x_t - \mathbf{I}y_t$  is always stationary for some value  $\mathbf{I}$ .
- (C) The time series  $z_t = x_t - \mathbf{I}y_t$  may be stationary for some value  $\mathbf{I}$  that can be determined precisely using regression techniques.
- (D) The time series  $z_t = x_t - \mathbf{I}y_t$  may be stationary for some value  $\mathbf{I}$  that can be estimated by running an ordinary least-squares regression of  $x_t$  on  $y_t$ .
- (E) None of (A), (B), (C) or (D) is correct.

29. You are given the following observed claim frequency data collected over a period of 365 days:

<u>Number of Claims per Day</u>	<u>Observed Number of Days</u>
0	50
1	122
2	101
3	92
4+	0

Fit a Poisson distribution to the above data, using the method of maximum likelihood.

Group the data by number of claims per day into four groups:

0      1      2      3 or more

Apply the chi-square goodness-of-fit test to evaluate the null hypothesis that the claims follow a Poisson distribution.

Determine the result of the chi-square test.

- (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.

**30.** You are given:

- (i) An individual automobile insured has an annual claim frequency distribution that follows a Poisson distribution with mean  $\lambda$ .
- (ii)  $\lambda$  follows a gamma distribution with parameters  $\mathbf{a}$  and  $\mathbf{q}$ .
- (iii) The first actuary assumes that  $\mathbf{a} = 1$  and  $\mathbf{q} = 1/6$ .
- (iv) The second actuary assumes the same mean for the gamma distribution, but only half the variance.
- (v) A total of one claim is observed for the insured over a three year period.
- (vi) Both actuaries determine the Bayesian premium for the expected number of claims in the next year using their model assumptions.

Determine the ratio of the Bayesian premium that the first actuary calculates to the Bayesian premium that the second actuary calculates.

- (A) 3/4
- (B) 9/11
- (C) 10/9
- (D) 11/9
- (E) 4/3

31. You fit the following model to 48 observations:

$$Y = \mathbf{b}_1 + \mathbf{b}_2 X_2 + \mathbf{b}_3 X_3 + \mathbf{b}_4 X_4 + \mathbf{e}$$

You are given:

Source of Variation	Degrees of Freedom	Sum of Squares
Regression	3	103,658
Error	44	69,204

Calculate  $\bar{R}^2$ , the corrected  $R^2$ .

- (A) 0.57
- (B) 0.58
- (C) 0.59
- (D) 0.60
- (E) 0.61

**32.** You are given the following information about a sample of data:

- (i) Mean = 35,000
- (ii) Standard deviation = 75,000
- (iii) Median = 10,000
- (iv) 90<sup>th</sup> percentile = 100,000
- (v) The sample is assumed to be from a Weibull distribution.

Determine the percentile matching estimate of the parameter  $t$ .

- (A) Less than 0.25
- (B) At least 0.25, but less than 0.35
- (C) At least 0.35, but less than 0.45
- (D) At least 0.45, but less than 0.55
- (E) At least 0.55



33. The number of claims a driver has during the year is assumed to be Poisson distributed with an unknown mean that varies by driver.

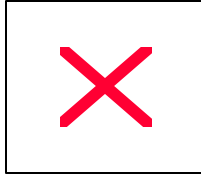
The experience for 100 drivers is as follows:

<u>Number of Claims during the Year</u>	<u>Number of Drivers</u>
0	54
1	33
2	10
3	2
4	1

Determine the credibility of one year's experience for a single driver using semiparametric empirical Bayes estimation.

- (A) 0.046
- (B) 0.055
- (C) 0.061
- (D) 0.068
- (E) 0.073

34. In a mortality study, the Weibull distribution with parameters  $\mathbf{I}$  and  $\mathbf{a}$  was used as the survival model, and log time  $Y$  was modeled as  $Y = \mathbf{m} + \mathbf{s}W$ , with  $W$  having the standard extreme value distribution. The parameters satisfy the relations:



The maximum likelihood estimates of the parameters are  $\hat{\mathbf{m}} = 4.13$  and  $\hat{\mathbf{s}} = 1.39$ .

The estimated variance-covariance matrix of  $\hat{\mathbf{m}}$  and  $\hat{\mathbf{s}}$  is:

$$\begin{pmatrix} 0.075 & 0.016 \\ 0.016 & 0.048 \end{pmatrix}$$

Use the delta method to estimate the covariance of  $\hat{\mathbf{a}}$  and  $\ln(\hat{\mathbf{I}})$ .

- (A) Less than  $-0.054$
- (B) At least  $-0.054$ , but less than  $-0.018$
- (C) At least  $-0.018$ , but less than  $0.018$
- (D) At least  $0.018$ , but less than  $0.054$
- (E) At least  $0.054$

35. You fit the following model to 30 observations:

$$Y = \mathbf{b}_1 + \mathbf{b}_2 X_2 + \mathbf{b}_3 X_3 + \mathbf{e}$$

- (i)  $s^2 = 10$
- (ii)  $r_{x_2 x_3} = 0.5$
- (iii)  $\sum (X_2 - \bar{X}_2)^2 = 4$
- (iv)  $\sum (X_3 - \bar{X}_3)^2 = 8$

Determine the estimate of the standard deviation of the least-squares estimate of the difference between  $\mathbf{b}_2$  and  $\mathbf{b}_3$ .

- (A) 1.7
- (B) 2.2
- (C) 2.7
- (D) 3.2
- (E) 3.7

**36.** You are given the following sample of five claims:

4      5      21      99      421

You fit a Pareto distribution using the method of moments.

Determine the 95<sup>th</sup> percentile of the fitted distribution.

- (A) Less than 380
- (B) At least 380, but less than 395
- (C) At least 395, but less than 410
- (D) At least 410, but less than 425
- (E) At least 425

37. You are given:

- (i)  $X_i$  is the claim count observed for driver  $i$  for one year.
- (ii)  $X_i$  has a negative binomial distribution with parameters  $b = 0.5$  and  $r_i$ .
- (iii)  $m_i$  is the expected claim count for driver  $i$  for one year.
- (iv) The  $m_i$ 's have an exponential distribution with mean 0.2.

Determine the Bühlmann credibility factor for an individual driver for one year.

- (A) Less than 0.05
- (B) At least 0.05, but less than 0.10
- (C) At least 0.10, but less than 0.15
- (D) At least 0.15, but less than 0.20
- (E) At least 0.20

38. A mortality study is conducted on 50 lives observed from time zero.

You are given:

(i)	Time t	Number of Deaths d <sub>t</sub>	Number Censored c <sub>t</sub>
	15	2	0
	17	0	3
	25	4	0
	30	0	c <sub>30</sub>
	32	8	0
	40	2	0

(ii)  $\hat{S}(35)$  is the Product-Limit estimate of  $S(35)$ .

(iii)  $\hat{V}[\hat{S}(35)]$  is the estimate of the variance of  $\hat{S}(35)$  using Greenwood's formula.

(iv) 
$$\frac{\hat{V}[\hat{S}(35)]}{[\hat{S}(35)]^2} = 0.011467$$

Determine  $c_{30}$ , the number censored at time 30.

- (A) 3
- (B) 6
- (C) 7
- (D) 8
- (E) 11

39. You fit an AR(1) model to the following data:

$$y_1 = 2.0$$

$$y_2 = -1.7$$

$$y_3 = 1.5$$

$$y_4 = -2.0$$

$$y_5 = 1.5$$

You choose as initial values  $\mathbf{e}_1 = 0$ ,  $\mathbf{m} = 0$  and  $\mathbf{r}_1 = 0.5$ .

Determine the value of the sum of squares function  $S = \sum [\mathbf{e}_t | \mathbf{e}_1 = 0, \mathbf{m} = 0, \mathbf{r}_1 = 0.5]^2$ .

- (A) 2
- (B) 12
- (C) 15
- (D) 21
- (E) 27

40. You are given the following accident data from 1000 insurance policies:

<u>Number of accidents</u>	<u>Number of policies</u>
0	100
1	267
2	311
3	208
4	87
5	23
6	4
7+	0
Total	<u>1000</u>

Which of the following distributions would be the most appropriate model for this data?

- (A) Binomial
- (B) Poisson
- (C) Negative Binomial
- (D) Normal
- (E) Gamma

**\*\*END OF EXAMINATION\*\***



Course 4  
Answer Key

May 2000

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

Solutions to May 2000 Course 4 Exam

When referencing textbooks, KM = *Survival Analysis*, KPW = *Loss Models*, PR = *Econometric Models and Economic Forecasts*, R = *Simulation*

1. From (4.12) on p. 91 of PR,  $F = \frac{R^2}{1-R^2} \frac{N-k}{k-1} = \frac{.64}{1-.64} \frac{20-2}{2-1} = 32$ . (B)

Alternatively, from the three definitions,  $F = \frac{RSS/(k-1)}{ESS/(N-k)}$ ,  $R^2 = RSS/TSS$ , and  $TSS = RSS + ESS$ , it is possible to use the given information to solve for  $F$ .

2. From Definition 2.14 on p. 35 of KPW,  $g = [11(.9)] = 9$ ,  $h = 11(.9) - 9 = .9$  and so  $\hat{p}_{.9} = .1x_{(.9)} + .9x_{(.10)} = .1(2199) + .9(3207) = 3106.2$ . (D)

3. From (5.60) on p. 437 of KPW, begin with  $E[E(X_j | \Theta)] = (1/3)(380) + (2/3)(23) = 142$ . Then,  $Var[E(X_j | \Theta)] = (1/3)380^2 + (2/3)23^2 - 142^2 = 28,322$ . From the definition of  $k$ ,  $E[Var(X_j | \Theta)] = 2.65(28,322) = 75,053.3 = (1/3)(365,000 - 380^2) + (2/3)Var(X_2)$  and so  $Var(X_2) = 2,280$ . (A)

4. From (4.2.3) on p. 86 of KM,  $\tilde{H}(12) = \frac{2}{15} + \frac{1}{12} + \frac{1}{10} + \frac{2}{6} = .65$  and then also from p. 86,  $\tilde{S}(12) = e^{-.65} = .52$ . (B)  
Note that in (4.2.3) the sum is over all death times less than *or equal to* the time at which the survival function is to be evaluated.

5. For an MA model the  $\mathbf{y}$ 's are equal to the negatives of the  $\mathbf{q}$ 's. From (18.42) on p. 563 of PR, the variance for the three-step-ahead forecast is  $(1 + \mathbf{q}_1^2 + \mathbf{q}_2^2)\mathbf{s}_e^2 = (1 + 1.8^2 + 1.11^2)(8) = 43.7768$ . The standard deviation is the square root, 6.6164. (E)

6. Using Theorem 2.5 and Corollary 2.6 on p. 74 of KPW,  $6,500 = E(X) - E(X \wedge 5,000) = 11,100 - E(X \wedge 5,000)$  and so  $E(X \wedge 5,000) = 4,600$ .

Also,  $10,000 = \frac{E(X) - E(X \wedge 5,000)}{1 - F(5,000)}$  and so  $F(5,000) = .35$ .

Using the definition of conditional expectation,

$$E(X | X < 5,000) = \frac{\int_0^{5,000} xf(x)dx}{F(5,000)} = \frac{E(X \wedge 5,000) - 5,000[1 - F(5,000)]}{F(5,000)} = 3,857. \quad (D)$$

7. See Example 5.30 on p. 428 of KPW. The calculations use the fact that the number of claims and the claim amounts are independent. The goal is to use Bayes' Theorem to obtain  $\Pr(A | N = 2, X_1 = 1, X_2 = 3)$ . To do this, we need

$$\Pr(N = 2, X_1 = 1, X_2 = 3 | A) = \frac{1^2 e^{-1}}{2!} e^{-1} e^{-3} = .5e^{-5} \text{ and}$$

$$\Pr(N = 2, X_1 = 1, X_2 = 3 | B) = \frac{3^2 e^{-3}}{2!} \frac{e^{-1/3}}{3} \frac{e^{-3/3}}{3} = .5e^{-13/3}. \text{ The answer is then}$$

$$\Pr(A | N = 2, X_1 = 1, X_2 = 3) = \frac{.5(.5e^{-5})}{.5(.5e^{-5}) + .5(.5e^{-13/3})} = .33924.$$

For class A, the expected cost is  $1(1) = 1$  and for class B it is  $3(3) = 9$ . The answer is  $.33924(1) + .66076(9) = 6.286$ . (D)

8. From (6.2.4) on p. 153 of KM,  $\hat{h}(t) = \frac{1}{b} \sum K \left( \frac{t - t_i}{b} \right) \Delta \tilde{H}(t)$ . Also,

$K(x) = .5, -1 \leq x \leq 1$ . The calculation is in the following table:

$t_i$	$(100 - t_i)/60$	$K$	$\Delta \tilde{H}(t_i)$
10	1.5	0	.05
34	1.1	0	.0526
47	.8833	.5	.0556
75	.4167	.5	.0588
156	-.9333	.5	.0625
171	-1.183	0	.0667

Then,  $\hat{h}(100) = \frac{1}{60} (.5)(.0556 + .0588 + .0625) = .00147$ . (B)

9. Method I: From (i),  $TSS = 160$  (using (3.26) on p. 71 of PR), from (iv),  $RSS_{II} = .7(160) = 112$  (using (4.9) on p. 89), and from (ii) and (iii),  $RSS_I = (-2)^2(10) = 40$  (using p. 73). Then we have  $ESS_{II} = 160 - 112 = 48$  and  $ESS_I = 160 - 40 = 120$ .

From (5.20) on p. 129,  $F = \frac{(120 - 48)/2}{48/(30 - 4)} = 19.5$ . (C)

Method II: Use (5.21) on p. 130 along with p. 73 to get  $R_I^2 = (-2)^2(10)/160 = .25$  and then  $F = \frac{(.7 - .25)/2}{(1 - .7)/26} = 19.5$ .

10. From Theorem 2.16 on p. 108 of KPW (also as (5.20) on p. 404),

$$p(\mathbf{q} | 2) \propto f(2 | \mathbf{q}) p(\mathbf{q}) = \frac{\mathbf{q} e^{-\mathbf{q}/2}}{4} \frac{e^{-\mathbf{q}/4}}{4} \propto \mathbf{q} e^{-.75\mathbf{q}}. \text{ (B)}$$

11. From pp. 123-5 of KPW and following Example 2.64,

Observation	Empirical cdf-	Empirical cdf+	Model cdf	Max difference
1	0	.2	.135	.135
2	.2	.4	.368	.168
3	.4	.6	.513	.113
5	.6	.8	.670	.130
13	.8	1	.857	.143

The model cdf is  $F(x) = \int_0^x q e^{-q/t} t^{-2} dt = e^{-q/x}$ . The overall maximum is the test statistic, .168. (D)

12. From p. 189 of KM begin with  $O(\mathbf{t}) =$  number of observed events  $= 32 + d_4$ . The other key quantity is  $E(\mathbf{t})$  which is the sum over all lives of the difference in  $H$  from when they were first seen until when they were last seen. All the lives were first seen at age 0. 10 lives were seen from 0 to 1 and contribute  $10(.08 - 0) = 8$ . The next 12 lives each contribute .15. All 68 lives in the final row were last seen at time 4 (some last seen alive, some dead). Thus,  $E(\mathbf{t}) = 10(.08) + 12(.15) + 10(.25) + 68(.40) = 32.3$  and so the test statistic is  $\frac{(32 + d_4 - 32.3)^2}{32.3} > 3.84$ . Solving this inequality leads to  $d_4 > 11.437$ . (E)

13. Formulas (15.34) and (15.35) on p. 480 of PR are  $\tilde{y}_t = \mathbf{a}y_t + (1 - \mathbf{a})(\tilde{y}_{t-1} + r_{t-1})$  and  $r_t = \mathbf{g}(\tilde{y}_t - \tilde{y}_{t-1}) + (1 - \mathbf{g})r_{t-1}$ . Using the first two rows (the second two rows could also be used),  $131.5 = \mathbf{a}(135.0) + (1 - \mathbf{a})(120.0 + 10.0)$ , for  $\mathbf{a} = .3$  and  $11.2 = \mathbf{g}(131.5 - 120) + (1 - \mathbf{g})(10.0)$ , for  $\mathbf{g} = .8$ . Then,  $\tilde{y}_{1998} = .3(146.6) + .7(144.2 + 12.4) = 153.6$  and  $r_{1998} = .8(153.6 - 144.2) + .2(12.4) = 10$ . From (15.36),  $\hat{y}_{2000} = \tilde{y}_{1998} + 2r_{1998} = 153.6 + 2(10) = 173.6$ . (E)

14. (A) is true because there is no model (p. 40 of KPW); (B) is true by (2.9) of KPW; (C) is true by p. 43 of KPW; (D) is true by p. 43 of KPW; (E) is false because robustness refers to model error (p. 48 of KPW). (E)

15. From (5.75) on p. 464 of KPW,  $\hat{v} = \frac{1}{4(6)} 33.60 = 1.4$  and from (5.76) on p. 465,  $\hat{a} = \frac{1}{3} 3.30 - \frac{1}{4(7)(6)} 33.6 = .9$ . Then,  $Z = \frac{7}{7 + 1.4/.9} = .818$ . (D)

16. Method I: Use (7.10) from p. 185 of PR.  $E(\hat{\mathbf{b}}^*) = \mathbf{b}_2 + \mathbf{b}_3 \sum x_2 x_3 / \sum x_2^2 = 10 + 3(0) / 10 = 10$ . (E)

Method II: From basic principles,  $\hat{\mathbf{b}}^* = \frac{\sum (x_i - \bar{x})y_i}{\sum (x_i - \bar{x})^2} = \frac{1}{10} \sum x_i y_i$ . Under the true model,

$$E(y_i) = 10x_i + 3x_i^2 \text{ and so } E(\hat{\mathbf{b}}^*) = \frac{1}{10} \sum x_i (10x_i + 3x_i^2) = 10.$$

17. The bootstrap is discussed in both KPW and R. In Section 7.3 of R it is used to estimate the mean square error. Following the discussion in R, there are four equally likely bootstrap samples, (1,1), (1,3), (3,1), and (3,3). Inserting each sample into the estimator yields the four bootstrap estimates, 0, 1, 1, and 0. The estimate using the actual sample of (1,3) is 1 and so the estimate of the mean square error is

$$\frac{1}{4}[(0-1)^2 + (1-1)^2 + (1-1)^2 + (0-1)^2] = .5. \quad (\text{B})$$

18. Because the estimators are independent,

$$\begin{aligned} \text{Var}(\mathbf{m}_c) &= w^2 \text{Var}(\mathbf{m}_a) + (1-w)^2 \text{Var}(\mathbf{m}_b) \\ &= 160,000w^2 + 40,000(1-w)^2 \\ &= 200,000w^2 - 80,000w + 40,000. \end{aligned}$$

The derivative with respect to  $w$  is  $400,000w - 80,000$  and setting it equal to zero gives  $w = .2$ . A similar example can be found in Exercise 2.21 on p. 160 of KPW. (B)

19. The intervals are  $.15 \pm .07875$  and  $.27121 \pm .11514$ . The standard deviations are the plus/minus terms divided by 1.96, or .04018 and .05874. From (4.2.4) on p. 86 of KM,

$$.04018^2 = \sum_{j=1}^i d_j / Y_j^2 \text{ and } .05874^2 = \sum_{j=1}^{i+1} d_j / Y_j^2 \text{ and therefore } .001836 = d_{i+1} / Y_{i+1}^2. \text{ Also,}$$

from (4.2.3) on p. 86,  $.27121 - .15 = .12121 = d_{i+1} / Y_{i+1}$ . Solving the two equations yields  $d_{i+1} = .12121^2 / .001836 = 8$ . (E)

20. From Section 17.2.2 on pp. 532-3 of PR, the partial autocorrelation function at  $k = 2$  is the estimate of  $\mathbf{f}_2$  from an AR(2) model using the Yule-Walker equations. The equations require the sample autocorrelations, which can be calculated using (16.23) on p. 495. They are:

$$\begin{aligned} \hat{\mathbf{r}}_1 &= \frac{-16(23) + 23(-35) - 35(40) + 40(-12)}{16^2 + 23^2 + 35^2 + 40^2 + 12^2} = \frac{-3053}{3754} = -.813266 \\ \hat{\mathbf{r}}_2 &= \frac{-16(-35) + 23(40) - 35(-12)}{3754} = \frac{1900}{3754} = .506127. \end{aligned}$$

The Yule-Walker equations are:

$$\mathbf{r}_1 = \mathbf{f}_1 + \mathbf{f}_2 \mathbf{r}_1, \quad -.813266 = \mathbf{f}_1 - .813266 \mathbf{f}_2$$

$$\mathbf{r}_2 = \mathbf{f}_1 \mathbf{r}_1 + \mathbf{f}_2, \quad .506127 = -.813266 \mathbf{f}_1 + \mathbf{f}_2$$

and solving the equations yields  $\mathbf{f}_2 = -.46$ . (A)

21. Following Example 2.22 on p. 57 of KPW,  $f(x) = \mathbf{a}500^{\mathbf{a}}x^{-\mathbf{a}-1}$  and so  
 $L = \mathbf{a}^5 500^{5\mathbf{a}} \left( \prod x_i \right)^{-\mathbf{a}-1}$ ,  $l = 5 \ln \mathbf{a} + 5\mathbf{a} \ln 500 - (\mathbf{a} + 1) \sum \ln x_i$ ,  
 $l = 5 \ln \mathbf{a} + 31.0730 \mathbf{a} - 33.1111(\mathbf{a} + 1)$ . Taking the derivative and setting it equal to zero  
yields  $5\mathbf{a}^{-1} - 2.0381 = 0$ ,  $\hat{\mathbf{a}} = 2.453$ . (B)

22. The solution is taken from formulas in Sections 2.8 and 5.2.5 of KPW as illustrated on pp. 425-6.

$$\Pr(100,000 | A) = .22(.4)(.78) + .78(.22)(.4) + .22(.6)(.22)(.6) = .154704$$

$$\Pr(100,000 | B) = .11(.64)(.89) + .89(.11)(.64) + .11(.36)(.11)(.36) = .126880.$$

From Bayes' Theorem,

$$\Pr(A | 100,000) = \frac{(2/3)(.154704)}{(2/3)(.154704) + (1/3)(.126880)} = .709. \quad (D)$$

23. Following p. 248 of KM, a confidence interval for  $\mathbf{b}_1 - \mathbf{b}_2$  is given by

$$.25 - (-.45) \pm 1.96 \sqrt{(1 \quad -1) \begin{pmatrix} .36 & .10 \\ .10 & .20 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix}}$$

$$= .7 \pm 1.96 \sqrt{.36} = .7 \pm 1.176.$$

The quantity of interest is  $e^{\mathbf{b}_1 - \mathbf{b}_2}$  and so a confidence interval is obtained by exponentiating the endpoints of the interval just obtained. That is,

$$\exp(.7 - 1.176) = .62 \text{ to } \exp(.7 + 1.176) = 6.53. \quad (C)$$

24. From (6.22) on p. 165 of PR,

$$DW = \frac{\sum (\hat{\mathbf{e}}_t - \hat{\mathbf{e}}_{t-1})^2}{\sum \hat{\mathbf{e}}_t^2} = \frac{(-.7+.6)^2 + (2.3+.7)^2 + (0-2.3)^2 + (-1-0)^2}{(-.6)^2 + (-.7)^2 + 2.3^2 + 0^2 + (-1)^2} = \frac{15.3}{7.14} = 2.143.$$

From text material on p. 165,  $DW$  is approximately  $2(1 - \hat{\mathbf{r}})$  and so

$$\hat{\mathbf{r}} = 1 - 2.143/2 = -.0715. \quad (C)$$

25. Use Theorem 2.3 on p. 67 of KPW as illustrated in Example 2.26 on that page. The reference to the "delta method" is in the second paragraph.

$$g(\mathbf{m}, \mathbf{s}) = e^{\mathbf{m} + \mathbf{s}^2/2}$$

$$\partial g / \partial \mathbf{m} = e^{\mathbf{m} + \mathbf{s}^2/2} = 123.017$$

$$\partial g / \partial \mathbf{s} = \mathbf{s} e^{\mathbf{m} + \mathbf{s}^2/2} = 134.458.$$

Then the variance is

$$(123.017 \quad 134.458) \begin{pmatrix} .1195 & 0 \\ 0 & .0597 \end{pmatrix} \begin{pmatrix} 123.017 \\ 134.458 \end{pmatrix} = 2887.73. \quad (D)$$

26. From pp. 412-3 of KPW, the credibility standard for expected claims is

$$\left(\frac{1.96}{.05}\right)^2 (1+3^2) = 15,366.4. \text{ Using the square root rule from pp. 416-7,}$$

$$Z = \sqrt{1000/15366.4} = .255. \text{ The credibility premium is } .255(6.75) + .745(5) = 5.45. \text{ (A)}$$

27. This data is right-truncated, so following the format of Table 5.5 on p. 136 of KM:

$T_i$	$X_i$	$R_i$	$d_i$	$Y_i$	$\hat{\Pr}[X < x_i   X \leq 6]$
3	1	5			
5	1	5	2	2	0
3	2	4	1	2	1/3
2	3	3	1	3	2/3

The key to the problem is to observe that the second observation cannot be included in the last two entries of the  $Y_i$  column because the time of the occurrence is too late in the observation period. From the last column of the table, the answer is 1/3. (C)

28. See pp. 513-4 of PR.

- (A) is incorrect because some time series are co-integrated.
- (B) is incorrect because a co-integrating parameter will not exist in all cases.
- (C) is incorrect because, even if two series are co-integrated, the co-integrating parameter cannot be determined precisely.
- (D) is correct.
- (E) is incorrect because (D) is correct. (D)

29. From Section 3.2.3 on p. 205 of KPW, the mle of  $\lambda$  is the sample mean which is  $600/365 = 1.6438$ . Following Example 3.4 on pp. 206-7,

No. of claims	No. of obs.	No. expected	Chi-square
0	50	$365e^{-1.6438} = 70.53$	$20.53^2/70.53 = 5.98$
1	122	$1.6438(70.53) = 115.94$	$6.06^2/115.94 = .32$
2	101	$1.6438(115.94)/2 = 95.29$	$5.71^2/95.29 = .34$
3+	92	$365 - 70.53 - 115.94 - 95.29 = 83.24$	$8.76^2/83.24 = .92$

The test statistic is 7.56. The critical values with two degrees of freedom (four rows less one estimated parameter less one) are 7.38 at 2.5% and 9.21 at 1%. (C)

30. Method I: Use Bayes' Theorem as used in Section 5.4.1 of KPW.

$$p(\mathbf{I} | x_1, x_2, x_3) \propto e^{-3\mathbf{I}} \mathbf{I}^{x_1+x_2+x_3} \mathbf{I}^{a-1} e^{-1/\mathbf{q}} = \mathbf{I}^a e^{-(3+1/\mathbf{q})\mathbf{I}} \text{ (noting that the } x\text{'s add to 1) which is a gamma distribution with parameters } \mathbf{a}+1 \text{ and } (3+1/\mathbf{q})^{-1} \text{ for a mean of } \frac{\mathbf{a}+1}{3+1/\mathbf{q}}.$$

Actuary 1's estimate is  $(1 + 1)/(3 + 6) = 2/9$ .

For Actuary 2, the parameters are 2 and 1/12 (to keep the mean at 1/6 but reduce the variance to 1/72 from 1/36) and so the estimate is  $(2 + 1)/(3 + 12) = 1/5$ . The ratio is 10/9. (C)

Method II: Use Section 5.4.5 by noting that this is a case where Bayesian and Bühlmann credibility are identical. The Bühlmann solution follows from Section 5.4.3. We have,  $\mathbf{m}(\mathbf{I}) = v(\mathbf{I}) = \mathbf{I}$ ,  $\mathbf{m} = v = E(\mathbf{I}) = \mathbf{aq}$ ,  $a = Var(\mathbf{I}) = \mathbf{aq}^2$ .

For Actuary 1,  $\mathbf{m} = v = 1/6$ ,  $a = 1/36$ ,  $Z = \frac{3}{3 + \frac{1/6}{1/36}} = 1/3$  for a premium of

$$(1/3)(1/3) + (2/3)(1/6) = 2/9.$$

Similarly, for Actuary 2,  $\mathbf{m} = v = 1/6$ ,  $a = 1/72$ ,  $Z = \frac{3}{3 + \frac{1/6}{1/72}} = 1/5$  for a premium of

$$(1/5)(1/3) + (4/5)(1/6) = 1/5.$$

31. From (4.11) on p. 90 of PR,  $\bar{R}^2 = 1 - (1 - R^2) \frac{N-1}{N-k}$  and from (4.9) on p. 89,

$$R^2 = RSS/TSS. \text{ Then } \bar{R}^2 = 1 - \left(1 - \frac{103,658}{172,862}\right) \frac{48-1}{48-4} = .5724. \quad (\text{A})$$

32. For the Weibull distribution,  $F(x) = 1 - e^{-(x/q)^t}$ . The method of percentile matching is defined on p. 47 of KPW. The equations to solve are

$$.5 = F(10,000) = 1 - e^{-(10,000/q)^t}$$

$$.9 = F(100,000) = 1 - e^{-(100,000/q)^t}.$$

Rearranging and taking logarithms produces the two equations

$$\left(\frac{10,000}{q}\right)^t = -\ln(.5) = .69315$$

$$\left(\frac{100,000}{q}\right)^t = -\ln(.1) = 2.30259.$$

Taking the ratio of the second equation to the first yields

$$10^t = 3.32192, \quad t \ln(10) = \ln(3.32192), \quad t = .5214. \quad (\text{D})$$

33. Following Example 5.49 on p. 472 of KPW,

$$\hat{\mathbf{m}} = \hat{v} = \bar{x} = 63/100 = .63, \quad \hat{a} = s^2 - \hat{v} = .68 - .63 = .05. \text{ Then } Z = \frac{1}{1 + .63/.05} = .073. \quad (\text{E})$$

$$\text{Note that } s^2 = \frac{54(0 - .63)^2 + 33(1 - .63)^2 + 10(2 - .63)^2 + 2(3 - .63)^2 + 1(4 - .63)^2}{99} = .68.$$



34. The formula is on p. 381 of KM. Begin by noting

$\ln \mathbf{l} = g_1(\mathbf{m}, \mathbf{s}) = -\mathbf{m}/\mathbf{s}$  and  $\mathbf{a} = g_2(\mathbf{m}, \mathbf{s}) = 1/\mathbf{s}$ . The required derivatives are  $g_1^1 = -1/\mathbf{s} = -.719$ ,  $g_1^2 = \mathbf{m}/\mathbf{s}^2 = 2.138$ ,  $g_2^1 = 0$ ,  $g_2^2 = -1/\mathbf{s}^2 = -.518$  and then

$$\begin{aligned} \text{Cov}(g_1, g_2) &= g_1^1 g_2^1 \text{Var}(\hat{\mathbf{m}}) + (g_1^1 g_2^2 + g_1^2 g_2^1) \text{Cov}(\hat{\mathbf{m}}, \hat{\mathbf{s}}) + g_1^2 g_2^2 \text{Var}(\hat{\mathbf{s}}) \\ &= -.719(0)(.075) + [-.719(-.518) + 2.138(0)](.016) + 2.138(-.518)(.048) \\ &= -.047. \end{aligned} \tag{B}$$

A quick way to do the last calculation is by writing it in matrix form:

$$(-.719 \quad 2.138) \begin{pmatrix} .075 & .016 \\ .016 & .048 \end{pmatrix} \begin{pmatrix} 0 \\ -.518 \end{pmatrix} = -.047.$$

35. Method I: From basic principles, begin with the definition of correlation,

$$r_{x_2, x_3}^2 = \frac{[\sum (x_2 - \bar{x}_2)(x_3 - \bar{x}_3)]^2}{\sum (x_2 - \bar{x}_2)^2 \sum (x_3 - \bar{x}_3)^2} \text{ and so } \sum (x_2 - \bar{x}_2)(x_3 - \bar{x}_3) = \sqrt{4(8)(.25)} = \sqrt{8}. \text{ Then,}$$

using Appendix 4.3, we have

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 30 & 0 & 0 \\ 0 & 4 & \sqrt{8} \\ 0 & \sqrt{8} & 8 \end{pmatrix}, \quad (\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 1/30 & 0 & 0 \\ 0 & 8/24 & -\sqrt{8}/24 \\ 0 & -\sqrt{8}/24 & 4/24 \end{pmatrix}.$$

We then have

$$\begin{aligned} \text{Var}(\hat{\mathbf{b}}_2 - \hat{\mathbf{b}}_3) &= \text{Var}(\hat{\mathbf{b}}_2) - 2\text{Cov}(\hat{\mathbf{b}}_2, \hat{\mathbf{b}}_3) + \text{Var}(\hat{\mathbf{b}}_3) \\ &= 10 \left[ \frac{8}{24} + 2 \frac{\sqrt{8}}{24} + \frac{4}{24} \right] = 7.357 \end{aligned}$$

for a standard deviation of 2.71. (C)

Method II: Use (4.6)-(4.8) on p. 88 to obtain the variances and covariance.

36. The method of moments is defined on p. 46 of KPW. For the first moment,

$$(4 + 5 + 21 + 99 + 421) / 5 = 110 = \mathbf{q} / (\mathbf{a} - 1) \text{ and for the second moment,}$$

$$(4^2 + 5^2 + 21^2 + 99^2 + 421^2) / 5 = 37,504.8 = 2\mathbf{q}^2 / (\mathbf{a} - 1)(\mathbf{a} - 2).$$

Dividing the second equation by the square of the first equation yields

$$\frac{37,504.8}{110^2} = \frac{2(\mathbf{a} - 1)}{\mathbf{a} - 2} \text{ and the solution is } \hat{\mathbf{a}} = 3.8189.$$

$$\text{Then } \hat{\mathbf{q}} = 110(3.8189 - 1) = 310.079.$$

$$\text{For the 95}^{\text{th}} \text{ percentile, } .95 = F(\mathbf{p}_{.95}) = 1 - \left( \frac{310.079}{310.079 + \mathbf{p}_{.95}} \right)^{3.8189} \text{ and the solution is}$$

$$\hat{\mathbf{p}}_{.95} = 369.37. \tag{A}$$

37. From Section 5.4.3 of KPW,  $\mathbf{m}(r) = r\mathbf{b} = 5r$  has an exponential distribution with mean 0.2. Because the exponential distribution is a scale family (KPW p. 83),  $r$  has an exponential distribution with mean 0.4.

Also,  $v(r) = r\mathbf{b}(\mathbf{b} + 1) = .75r$ .

For Bühlmann credibility,

$$v = E[v(r)] = E(.75r) = .75(.4) = .3, a = \text{Var}[\mathbf{m}(r)] = \text{Var}(.5r) = .25(.4)^2 = .04.$$

$$\text{Then, } Z = \frac{1}{1 + .3/.04} = .1176. \quad (\text{C})$$

38. The formula is on p. 84 of KM. For this problem,

$$\hat{V}[\hat{S}(35)] = \hat{S}(35)^2 \left[ \frac{2}{50(48)} + \frac{4}{45(41)} + \frac{8}{(41-c)(33-c)} \right]. \text{ The required ratio}$$

is  $.011467 = .000833 + .002168 + \frac{8}{(41-c)(33-c)}$  leading to the quadratic

$$c^2 - 74c + 408 = 0 \text{ and the solution is } c = 6. \quad (\text{B})$$

39. From (17.28) on p. 529 of PR,  $\mathbf{r}_1 = \mathbf{f}_1$  and so  $\mathbf{f}_1 = .5$ . From (17.20) on p. 527,  $\mathbf{m} = \mathbf{d}/(1 - \mathbf{f}_1)$  and so  $\mathbf{d} = 0$ . The model is  $y_t = .5y_{t-1} + \mathbf{e}_t$  (from (17.18) on p. 527). We have  $\mathbf{e}_t = y_t - .5y_{t-1}$ . Then,  $S = (-1.7 - 1)^2 + (1.5 + .85)^2 + (-2 - .75)^2 + (1.5 + 1)^2 = 26.625$ . (E)

40. Method I: The sample mean is 2 and the sample variance is 1.495. Of the three discrete distribution choices (the last two choices are continuous and so cannot be the model for this data set), only the binomial has a variance that is less than the mean. (A)

Method II: Following p. 222 of KPW, compute successive values of  $kn_k / n_{k-1}$ . They are 2.67, 2.33, 2.01, 1.67, 1.32, and 1.04. This sequence is linear, indicating that a member of the (a,b,0) family is appropriate. Because it is decreasing, a binomial model is preferred.