Exam S

CASUALTY ACTUARIAL SOCIETY
AND THE
CANADIAN INSTITUTE OF ACTUARIES

Exam S

Statistics and Probabilistic Models

October 26, 2017

4 HOURS

INSTRUCTIONS TO CANDIDATES

1. This 90 point examination consists of 45 multiple choice questions worth 2 points.

2. To answer the multiple choice questions, use the short-answer card provided and a number 2 or HB pencil only.

   - Fill in that it is Fall 2017 and that the exam name is S.
   - Darken the spaces corresponding to your Candidate ID number. Five rows are available. If your Candidate ID number is fewer than 5 digits, include leading zeros. For example, if your Candidate ID number is 987, consider that your Candidate ID number is 00987, enter a zero on the first row, a zero on the second row, 9 on the third row, 8 on the fourth row, and 7 on the fifth [last] row. Write in your Candidate ID number next to the place where you darken the spaces for your Candidate ID number. Your name, or any other identifying mark, must not appear on the short-answer card.
   - Mark your short-answer card during the examination period. No additional time will be allowed for this after the exam has ended. Make your marks dark and fill in the spaces completely.
   - For each of the multiple choice questions, select the one best answer and fill in the corresponding letter. One quarter of the point value of the question will be subtracted for each incorrect answer. No points will be added or subtracted for responses left blank.

3. Do all problems until you reach the last page of the examination where "END OF EXAMINATION" is marked.

CONTINUE TO NEXT PAGE OF INSTRUCTIONS

4. Prior to the start of the exam you will have a **fifteen-minute reading period** in which you can silently read the questions and check the exam booklet for missing or defective pages. Writing will NOT be permitted during this time and you will not be permitted to hold pens or pencils. You will also not be allowed to use calculators. The supervisor has additional exams for those candidates who have defective exam booklets.

   - Verify that you have a copy of "Tables for CAS Exam S" included in your exam packet.
   - Candidates should not interpolate in the tables unless explicitly instructed to do so in the problem, rather, a candidate should round the result that would be used to enter a given table to the same level of precision as shown in the table and use the result in the table that is nearest to that indicated by rounded result. For example, if a candidate is using the Tables of the Normal Distribution to find a significance level and has a Z value of 1.903, the candidate should round to 1.90 to find cumulative probability in the Normal table.

   - Candidates should employ a non-parametric test unless otherwise specified in the problem, or when there is a standard distribution that is logically or commonly associated with the random variable in question. Examples of problems with a logical or commonly associated distribution would include exponential wait times for Poisson processes and applications of the Central Limit Theorem.

5. Your Examination Envelope is pre-labeled with your Candidate ID number, name, exam number, and test center. Do not remove this label. Keep a record of your Candidate ID number for future inquiries regarding this exam.

6. Candidates must remain in the examination center until two hours after the start of the examination. The examination starts after the reading period is complete. You may leave the examination room to use the restroom with permission from the supervisor. To avoid excessive noise during the end of the examination, candidates may not leave the exam room during the last fifteen minutes of the examination.

7. At the end of the examination, place the short-answer card in the Examination Envelope. Nothing written in the examination booklet will be graded. Only the short-answer card will be graded. Also place any included reference materials in the Examination Envelope. BEFORE YOU TURN THE EXAMINATION ENVELOPE IN TO THE SUPERVISOR, BE SURE TO SIGN IT IN THE SPACE PROVIDED ABOVE THE CUT-OUT WINDOW.

8. If you have brought a self-addressed, stamped envelope, you may put the examination booklet and scrap paper inside and submit it separately to the supervisor. It will be mailed to you. Do not put the self-addressed stamped envelope inside the Examination Envelope. Interoffice mail is not acceptable.

   If you do not have a self-addressed, stamped envelope, please place the examination booklet in the Examination Envelope and seal the envelope. You may not take it with you. Do not put scrap paper in the Examination Envelope. The supervisor will collect your scrap paper.

   Candidates may obtain a copy of the examination from the CAS Web Site.

   All extra answer sheets, scrap paper, etc. must be returned to the supervisor for disposal.

9. Candidates must not give or receive assistance of any kind during the examination. Any cheating, any attempt to cheat, assisting others to cheat, or participating therein, or other improper conduct will result in the Casualty Actuarial Society and the Canadian Institute of Actuaries disqualifying the candidate's paper, and such other disciplinary action as may be deemed appropriate within the guidelines of the CAS Policy on Examination Discipline.

10. The exam survey is available on the CAS Web Site in the "Admissions/Exams" section. Please submit your survey by November 9, 2017.

END OF INSTRUCTIONS

1.

You are given the following information:

- Coins are tossed into a fountain according to a Poisson process at a rate of one every two minutes.
- The coin denominations are independently distributed as follows:

Coin Denomination	**Probability**
Penny	0.50
Nickel	0.20
Dime	0.20
Quarter	0.10

Calculate the probability that two Quarters will be tossed into the fountain before four non-Quarter coins.

A. Less than 0.06
B. At least 0.06, but less than 0.07
C. At least 0.07 but less than 0.08
D. At least 0.08, but less than 0.09
E. At least 0.09

EXAM CONTINUED ON NEXT PAGE

2.

You are given the following information:

- A reinsurer classifies three types of events as catastrophes: hurricanes, earthquakes and wildfires.
- Each type of catastrophe follows a homogeneous Poisson process with the following rates of occurrence:

Hurricanes:	5 per year
Earthquakes:	1 per 5 years
Wildfires:	1 per year

Calculate the probability that 2 or more catastrophes occur in a six-month period.

A. Less than 0.80
B. At least 0.80, but less than 0.81
C. At least 0.81, but less than 0.82
D. At least 0.82, but less than 0.83
E. At least 0.83

3.

You are given:

- A Poisson process N has a rate function:. $\lambda(t) = 3t^2$
- You've already observed 50 events by time t = 2.1.

Calculate the conditional probability, Pr[N(3)=68 | N(2.1)=50].

A. Less than 5%
B. At least 5%, but less than 10%
C. At least 10%, but less than 15%
D. At least 15%, but less than 20%
E. At least 20%

EXAM CONTINUED ON NEXT PAGE

4.

You are given the following information:

- Taxicabs leave a hotel with a group of passengers at a Poisson rate of 10 per hour.
- The number of people in each group taking a cab is independent and has the following probability:

Number of people	Probability
1	0.60
2	0.25
3	0.10
4	0.05

Calculate the probability that at least 800 people leave the hotel in a cab during a 48-hour period using the normal approximation.

A. Less than 0.200
B. At least 0.200, but less than 0.210
C. At least 0.210, but less than 0.220
D. At least 0.220, but less than 0.230
E. At least 0.230

EXAM CONTINUED ON NEXT PAGE

5.

You are given:

- The failure (hazard) rate function, $\lambda(t)$
$$\lambda(t) = \frac{\alpha}{(t+\theta)}; \text{for } t > 0$$

- The density function, $f(t)$
$$f(t) = \frac{\alpha\theta^{\alpha}}{(t+\theta)^{\alpha+1}}; \text{for } t > 0$$

- $\alpha = 3$
- $\theta = 5$

Calculate $S(15)$, the probability a life age 0 lives to age 15.

A. Less than 0.020
B. At least 0.020, but less than 0.040
C. At least 0.040, but less than 0.060
D. At least 0.060, but less than 0.080
E. At least 0.080

EXAM CONTINUED ON NEXT PAGE

6.

A customer calls a customer service center with two servers. Both servers are busy at the time of the call, therefore the customer is placed in a queue that is currently empty. The next available server will handle this customer's call.

You are given:

- Service times are exponentially distributed and independent.
- Server 1 is more experienced, so she handles calls in 5 minutes on average.
- Server 2 is less experienced, so she handles calls in 7 minutes on average.

Calculate the expected total waiting plus service time, in minutes, until this customer is finished.

A. Less than 7
B. At least 7, but less than 8
C. At least 8, but less than 9
D. At least 9, but less than 10
E. At least 10

EXAM CONTINUED ON NEXT PAGE

7.

A produce vendor must decide how many oranges to order per week in order to maximize her profit. You are given:

- The cost of oranges to the vendor is 0.50 per orange.
- The vendor is able to sell oranges at 1.00 per orange.
- Quantity demanded follows an exponential distribution with a θ of 0.001 per week.
- Any inventory left over at the end of the week will rot and is worthless.
- There is no penalty to the vendor if she cannot meet all of the demand.

Calculate the optimal amount of oranges the vendor should purchase in a week to maximize profit.

A. Less than 650
B. At least 650 but less than 675
C. At least 675 but less than 700
D. At least 700 but less than 725
E. At least 725

EXAM CONTINUED ON NEXT PAGE

8.

A 3-out-of-50 system is placed in parallel with a 48-out-of-50-system to form a combined system.

Calculate the number of minimal path sets for the combined system.

A. Fewer than 20,000
B. At least 20,000, but fewer than 30,000
C. At least 30,000, but fewer than 40,000
D. At least 40,000, but fewer than 50,000
E. At least 50,000

EXAM CONTINUED ON NEXT PAGE

9.

You are given:

- A 999-out-of-1000 system of independent identically distributed exponential components.
- The mean lifetime of each component is $\theta = 5$.

Calculate the expected system lifetime.

A. Less than 0.05
B. At least 0.05, but less than 0.10
C. At least 0.10, but less than 1.00
D. At least 1.00, but less than 10.00
E. At least 10.00

EXAM CONTINUED ON NEXT PAGE

10.

You are given the following information:

- System X is a series system with two components.
- System Y is a parallel system with two components.
- All components are independent and function for an amount of time, uniformly distributed over (0, 1).

Calculate the absolute value of the difference in expected system lifetimes between system X and system Y.

A. Less than 0.10
B. At least 0.10, but less than 0.20
C. At least 0.20, but less than 0.30
D. At least 0.30, but less than 0.40
E. At least 0.40

EXAM CONTINUED ON NEXT PAGE

11.

You are given the following transition matrix for a Markov chain with two states 0, 1:

$$A = \begin{bmatrix} 0.25 & 0.75 \\ 0.01 & 0.99 \end{bmatrix}$$

At time $t = 0$, the Markov chain is in state 0.

Calculate the expected number of steps needed to return to state 0.

A. Fewer than 10
B. At least 10, but fewer than 30
C. At least 30, but fewer than 50
D. At least 50, but fewer than 70
E. At least 70

12.

You are given:

- A branching process has an initial population consisting of 16 individuals.
- There are three states:
  - State 0 for zero offspring
  - State 1 for one offspring
  - State 2 for two offspring
- $P_0 = \frac{2}{5}, P_1 = \frac{1}{10}, P_2 = \frac{1}{2}$

Calculate the probability that the population will die out.

A. Less than 0.02
B. At least 0.02, but less than 0.03
C. At least 0.03, but less than 0.04
D. At least 0.04, but less than 0.05
E. At least 0.05

EXAM CONTINUED ON NEXT PAGE

13.

You are given the following information about a homogeneous Markov chain:

- Planes move among three states servicing flights:
  - State 0: On-time
  - State 1: Delayed
  - State 2: Canceled
- $P = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.7 & 0.0 & 0.3 \\ 0.8 & 0.2 & 0.0 \end{bmatrix}$
- A plane is currently in the state Canceled.

Calculate the probability that the plane will be in the state Canceled after two transitions.

A. Less than 0.07
B. At least 0.07, but less than 0.09
C. At least 0.09, but less than 0.11
D. At least 0.11, but less than 0.13
E. At least 0.13

14.

You are given the following information about a Markov chain with a transition probability matrix:

- $P = \begin{bmatrix} 0.60 & 0.30 & 0.10 \\ 0.70 & 0.30 & 0.00 \\ 0.30 & 0.00 & 0.70 \end{bmatrix}$
- The three states are 0, 1, and 2.

Calculate the long-run proportion of time in State 2.

A. Less than 0.200
B. At least 0.200, but less than 0.220
C. At least 0.220, but less than 0.240
D. At least 0.240, but less than 0.260
E. At least 0.260

15.

A life annuity has the following values:

- $\ddot{a}_{20} = 13.9294$
- $\ddot{a}_{21} = 13.8529$
- The interest rate used is 0.05.
- The force of mortality is constant between the ages of 20 and 30.

From a sample of 500 independent twenty-year-olds, you wish to calculate a 95% confidence interval for the number of people who will die within five years implied by the annuity values above, using a normal approximation.

Calculate the upper bound of this confidence interval.

A. Less than 60.0
B. At least 60.0, but less than 65.0
C. At least 65.0, but less than 70.0
D. At least 70.0, but less than 75.0
E. At least 75.0

16.

You are given the following information about a machine warranty:

- A warranty will provide payment at the end of the year in which the machine fails.
- The annual interest rate, i, is 0.05.
- The number of functioning machines at the start of the year and the number failing during the year are denoted by letters l and d respectively.

Age x	l_x	d_x	**Payment**
0	500		100
1	300		200
2		50	300
3	50	50	400

Calculate the actuarial present value of the warranty payment per unit at time 0.

A. Less than 100
B. At least 100, but less than 120
C. At least 120, but less than 140
D. At least 140, but less than 160
E. At least 160

17.

You are given the following information:

- $0 \leq \theta \leq 1$
- X has the probability distribution:

X	Pr(X)
0	$3\theta/4$
1	$\theta/4$
2	$2(1-\theta)/3$
3	$(1-\theta)/3$

- You observe the following 36 values from this distribution:

X	Obs
0	7
1	12
2	13
3	4

Calculate the maximum likelihood estimate of the parameter θ.

A. Less than 0.20
B. At least 0.20, but less than 0.40
C. At least 0.40, but less than 0.60
D. At least 0.60, but less than 0.80
E. At least 0.80

18.

You are given the following information:

- X has the probability mass function:
$$p(X = x|\alpha) = \frac{e^{-\alpha/2}\alpha^x}{2^x x!}, \text{where } \alpha > 0 \text{ and } X \in 0,1,2,3,\ldots$$
- You observe the following 7 values from this distribution:
0, 0, 1, 1, 3, 4, 5

Calculate the maximum likelihood estimate of the parameter α.

A. Less than 0.5
B. At least 0.5, but less than 1.5
C. At least 1.5, but less than 2.5
D. At least 2.5, but less than 3.5
E. At least 3.5

19.

$\hat{\alpha}_n$ is an estimator of a parameter α for a sample size of n. You are given:

- $\mathrm{E}[\hat{\alpha}_n] = \frac{7n}{n+4}$ for $n = 2,3,4,\ldots$
- $\mathrm{Var}[\hat{\alpha}_n] = \frac{140(n-1)}{2n-1}$ for $n = 2,3,4,\ldots$
- The true value of α is 7.

A list of potentially true statements about $\hat{\alpha}_n$ is given below.

I. $\hat{\alpha}_n$ is an asymptotically unbiased estimator of α.
II. The mean squared error of $\hat{\alpha}_8$ is greater than the variance of the estimator of $\hat{\alpha}_n$ for $n \geq 2$.
III. $\hat{\alpha}_n$ is a consistent estimator of α.

Determine which of the above statements are true.

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

20.

You are given the following information about a random sample of size 20 from a normal distribution:

- A two-sided 95% confidence interval for the distribution variance is estimated, with an equal probability in each tail.
- The width of the confidence interval is 150.

Calculate the unbiased sample variance for the 20 random draws.

A. Less than 25
B. At least 25, but less than 50
C. At least 50, but less than 75
D. At least 75, but less than 100
E. At least 100

EXAM CONTINUED ON NEXT PAGE

21.

Let X be a single observation from the distribution:

$$F(x) = 1 - \exp\left(-\left(\frac{x}{\theta}\right)^2\right).$$

You are testing the null hypothesis $H_0: \theta = y$ against the alternative hypothesis $H_1: \theta = 10y$. The null hypothesis is rejected if $X > k$. The probability of a Type I error is 5.0%.

Calculate the probability of a Type II error.

A. Less than 4.0%
B. At least 4.0%, but less than 8.0%
C. At least 8.0%, but less than 12.0%
D. At least 12.0%, but less than 16.0%
E. At least 16.0%

22.

You are given the claim activity for a group of 5,000 policyholders in years 2015 and 2016, as shown in the table below:

Group	2015 Claims	2016 Claims	Policyholders
I	0	0	3,500
II	0	>0	900
III	>0	0	500
IV	>0	>0	100

Your null hypothesis is that claim activity in year 2015 is independent of claim activity in 2016. The alternative hypothesis is that claim activity in 2015 is not independent of claim activity in 2016.

Calculate the p-value of this test using the Neyman-Pearson Lemma.

A. Less than 0.5%
B. At least 0.5%, but less than 1.0%
C. At least 1.0%, but less than 2.5%
D. At least 2.5% but less than 5.0%
E. At least 5.0%

23.

Let X_1 and X_2 be a random sample from a distribution with density function:

$$f(x) = \begin{cases} \theta x^{\theta-1}, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}, \text{ where } \theta > 0.$$

The null hypothesis $H_0: \theta = 4$ is tested against the alternative hypothesis $H_1: \theta = 3$ using the statistic $Y = \max(X_1, X_2)$.

Calculate the critical value given the probability of a Type I error is 0.001822.

A. Less than 0.30
B. At least 0.30, but less than 0.40
C. At least 0.40, but less than 0.50
D. At least 0.50, but less than 0.60
E. At least 0.60

EXAM CONTINUED ON NEXT PAGE

24.

You are given the following information:

- Data follows a Poisson distribution with mean λ.
- Under the null hypothesis $H_0: \lambda = 5$, you observe one value from the distribution and reject the null hypothesis if that observation is either 0 or 1. The probability of committing a Type I error in this situation is α.
- Assuming that the alternative hypothesis is $H_1: \lambda = 1.5$, under the same rejection rule as above the probability of a Type II error is β.

Calculate $\beta - \alpha$.

A. Less than 0.15
B. At least 0.15, but less than 0.25
C. At least 0.25, but less than 0.35
D. At least 0.35, but less than 0.45
E. At least 0.45

EXAM CONTINUED ON NEXT PAGE

25.

You are given the following information:

- X has the following probability density function:

$$f(x) = \frac{1}{7}e^{-\frac{x}{7}}$$

- $X > 0$
- You draw a sample of size 30 from this distribution.

Calculate the probability that the two smallest values in your sample are both smaller than 0.5.

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

EXAM CONTINUED ON NEXT PAGE

26.

Losses were separated into two types of claims and analyzed.

- The values of 12 randomly selected claims are provided in the following table:

Claim Type W		Claim Type X	
Loss Amount	Rank	Loss Amount	Rank
100,000	1	100,500	2
101,000	3	101,500	4
102,000	5	102,500	6
103,000	7	103,500	8
104,000	9	104,500	10
		105,500	11
		106,500	12

- H_0: $Median_{Type\ W} = Median_{Type\ X}$
- H_1: $Median_{Type\ W} \neq Median_{Type\ X}$

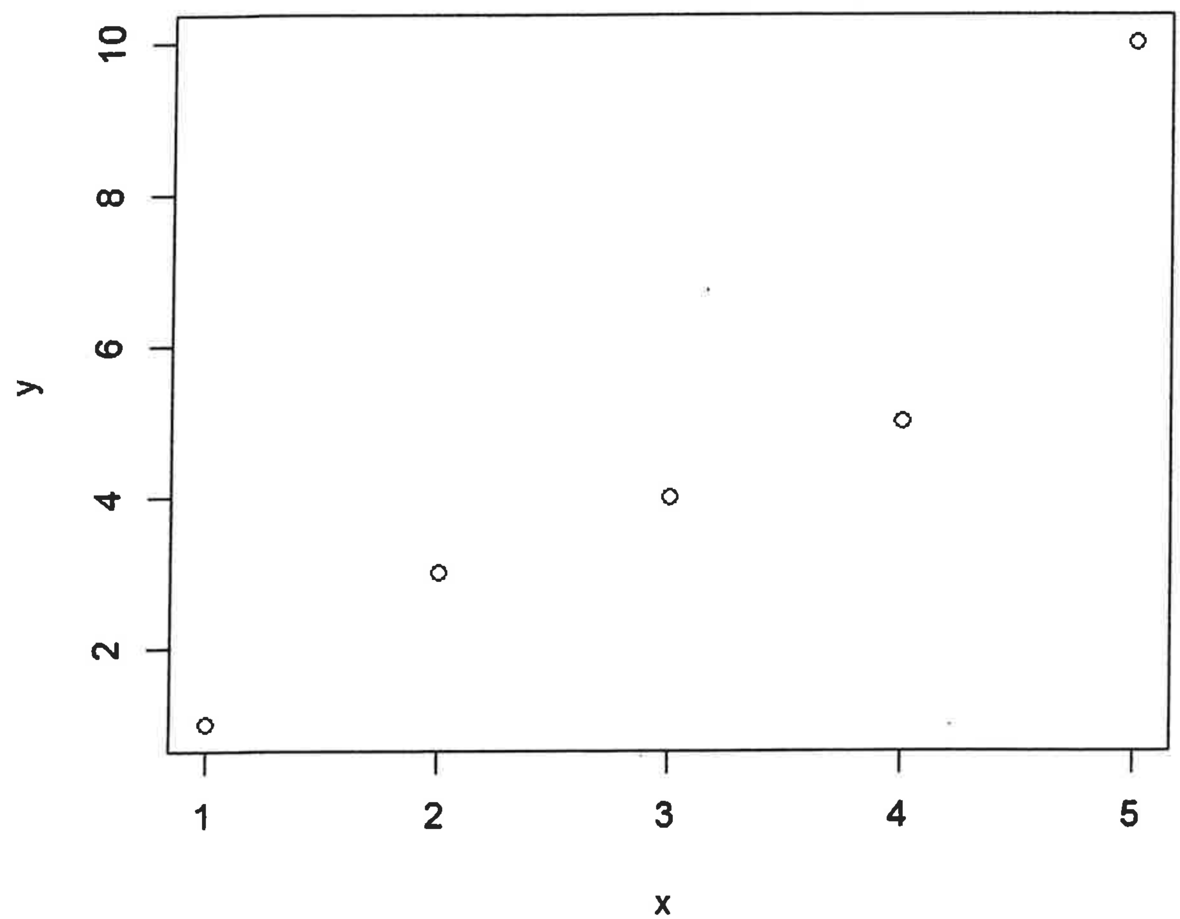
Calculate the smallest value for α for which H_0 can be rejected using the Mann-Whitney U Test.

A. Less than 0.010
B. At least 0.010, but less than 0.200
C. At least 0.200, but less than 0.400
D. At least 0.400, but less than 0.600
E. At least 0.600

EXAM CONTINUED ON NEXT PAGE

27.

You have calculated the Pearson correlation coefficient, Spearman's Rho, and Kendall's Tau using the following five observations:



Determine the correct statement.

A. Pearson's correlation coefficient < Kendall's Tau < Spearman's Rho
B. Pearson's correlation coefficient < Spearman's Rho < Kendall's Tau
C. Pearson's correlation coefficient < Kendall's Tau = Spearman's Rho
D. Spearman's Rho = Kendall's Tau < Pearson's correlation coefficient
E. Kendall's Tau < Pearson's correlation coefficient < Spearman's Rho

28.

You are given the following information:

- The probability that a randomly selected person has a specific rare disease is $\frac{1}{10,000}$.
- If they have the disease, they will have a positive test 90% of the time.
- If they do not have the disease, they will have a positive test 1% of the time.

You gather a sample of 10 people who all have positive test results.

Calculate the probability that no one in your sample actually has the disease.

A. Less than 0.6
B. At least 0.6, but less than 0.7
C. At least 0.7, but less than 0.8
D. At least 0.8, but less than 0.9
E. At least 0.9

29.

You are given the following information:

- Annual claim frequency follows a Poisson distribution with mean λ.
- λ follows a gamma distribution with mean 3 and variance 5.
- You observe 4 policies with 0, 0, 2, and 3 claims in the previous year.

Calculate the mean of the posterior distribution of λ.

A. Less than 1.3
B. At least 1.3, but less than 1.8
C. At least 1.8, but less than 2.3
D. At least 2.3, but less than 2.8
E. At least 2.8

30.

For a large portfolio of insurance contracts, suppose θ represents the proportion of policies for which there were claims made within one year. The value of θ is a random variable with the prior distribution expressed as:

$$p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}; \quad \text{for } 0 < \theta < 1$$

You choose a and b such that:

- The prior estimate of the mean of θ is 10%
- The posterior mean of θ gives twice as much weight to the observed average from a sample of 500 policies as the prior mean.

Calculate the value of a implied for the prior distribution.

A. Less than 22
B. At least 22, but less than 30
C. At least 30, but less than 38
D. At least 38, but less than 46
E. At least 46

31.

You are given the following output from a GLM to estimate an insured's Pure Premium:

Response variable	Pure Premium
Response distribution	Gamma
Link	Log

Parameter	df	$\hat{\beta}$
Intercept	1	5.26
Risk Group	2	
Group 1	0	0.00
Group 2	1	0.18
Group 3	1	0.37
Territory Code	2	
Region 1	0	0.00
Region 2	1	0.12
Region 3	1	0.25

Calculate the predicted Pure Premium for an insured in Risk Group 3 from Region 2.

A. Less than 250
B. At least 250, but less than 275
C. At least 275, but less than 300
D. At least 300, but less than 325
E. At least 325

EXAM CONTINUED ON NEXT PAGE

32.

Given a family of distributions where the variance is related to the mean through a power function:

$$\mathrm{Var}[Y] = a\mathrm{E}[Y]^p$$

One can characterize members of the exponential family of distributions using this formula.

You are given the following statements on the value of p for a given distribution:

I. Normal (Gaussian) distribution, $p = 0$
II. Compound Poisson–gamma distribution, $1 < p < 2$
III. Inverse Gaussian distribution, $p = -1$

Determine which of the above statements are correct.

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

EXAM CONTINUED ON NEXT PAGE

33.

A distribution belongs to the exponential family if it can be written in the canonical form:

$$f(y;\theta) = \exp[a(y)b(\theta) + c(\theta) + d(y)]$$

Determine the values of $a(y), b(\theta), c(\theta),$ and $d(y)$ for the Poisson distribution.

A. $a(y) = y, b(\theta) = \ln(\theta), c(\theta) = \theta,$ and $d(y) = \ln(y!)$
B. $a(y) = -y, b(\theta) = -\ln(\theta), c(\theta) = \theta,$ and $d(y) = \ln(y!)$
C. $a(y) = y, b(\theta) = \ln(\theta), c(\theta) = \theta,$ and $d(y) = 1/(y!)$
D. $a(y) = y, b(\theta) = \ln(\theta), c(\theta) = -\theta,$ and $d(y) = -\ln(y!)$
E. The answer is not given by (A), (B), (C), or (D)

EXAM CONTINUED ON NEXT PAGE

34.

For an ordinary linear regression with 5 parameters and 50 observations, you are given:

- The total sum of squares, $S_{yy} = 996$.
- The unbiased estimate for the constant variance, σ^2, is $s^2 = 2.47$.

Calculate the coefficient of determination.

A. Less than 0.65
B. At least 0.65, but less than 0.75
C. At least 0.75, but less than 0.85
D. At least 0.85, but less than 0.95
E. At least 0.95

EXAM CONTINUED ON NEXT PAGE

35.

Consider the following 2 models, which were fit to the same 30 observations using ordinary least squares:

Model 1

Response variable	Y	
Response distribution	Normal	
Link	Identity	
SS Total	19851	
SS Error	2781	
Parameter	$\hat{\beta}$	**df**
Intercept	130.2	1
X_1	5.1	1
X_2	-4.2	1

Model 2

Response variable	Y	
Response distribution	Normal	
Link	Identity	
SS Total	19851	
SS Error	2104	
Parameter	$\hat{\beta}$	**df**
Intercept	121.9	1
X_1	4.5	1
X_2	-5.9	1
X_3	3.5	1
X_4	-2.3	1

You test H_0: $\beta_3 = \beta_4 = 0$ against the alternative hypothesis that at least one of $\beta_3, \beta_4 \neq 0$.

Calculate the smallest significance level at which you reject H_0.

A. Less than 0.01
B. At least 0.01, but less than 0.02
C. At least 0.02, but less than 0.05
D. At least 0.05, but less than 0.10
E. At least 0.10

36.

You are analyzing a dataset and have fit a multiple regression with 12 continuous explanatory variables and one intercept. You are given the following:

- Sum of square errors $= 618$
- The 10th diagonal entry of the hat matrix, $h_{10,10} = 0.35$
- The 10th residual value = $\widehat{\epsilon_{10}} = 9.5$
- Cook's Distance = $D_{10} = 5.0$

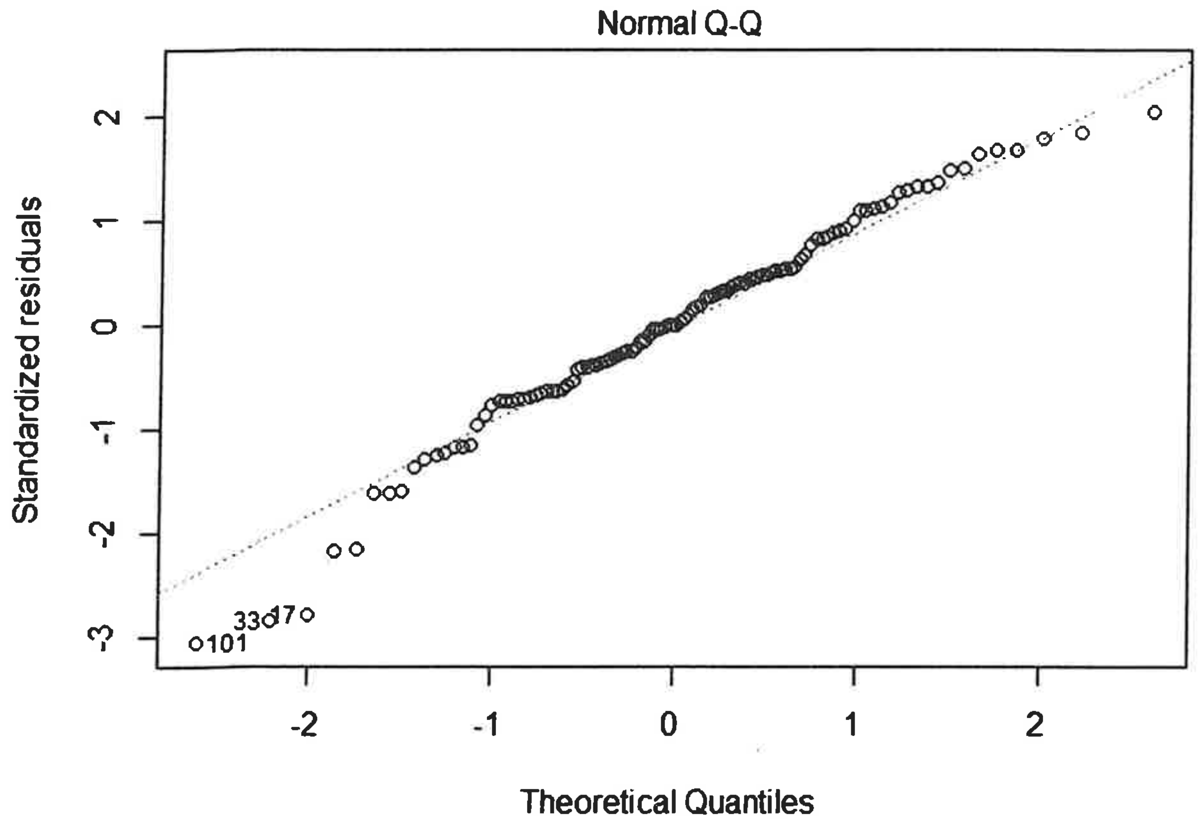
Calculate the number of data points this model was fit to.

A. Fewer than 200
B. At least 200, but fewer than 400
C. At least 400, but fewer than 600
D. At least 600, but fewer than 800
E. At least 800

EXAM CONTINUED ON NEXT PAGE

37.

You are given the following residual Q-Q plot for a fitted GLM:



I. The distribution of residuals is skewed to the left.
II. The residuals are serially correlated.
III. Observations 33 and 101 are influential points.

Determine which of the above statements can be concluded to be true from the above chart.

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

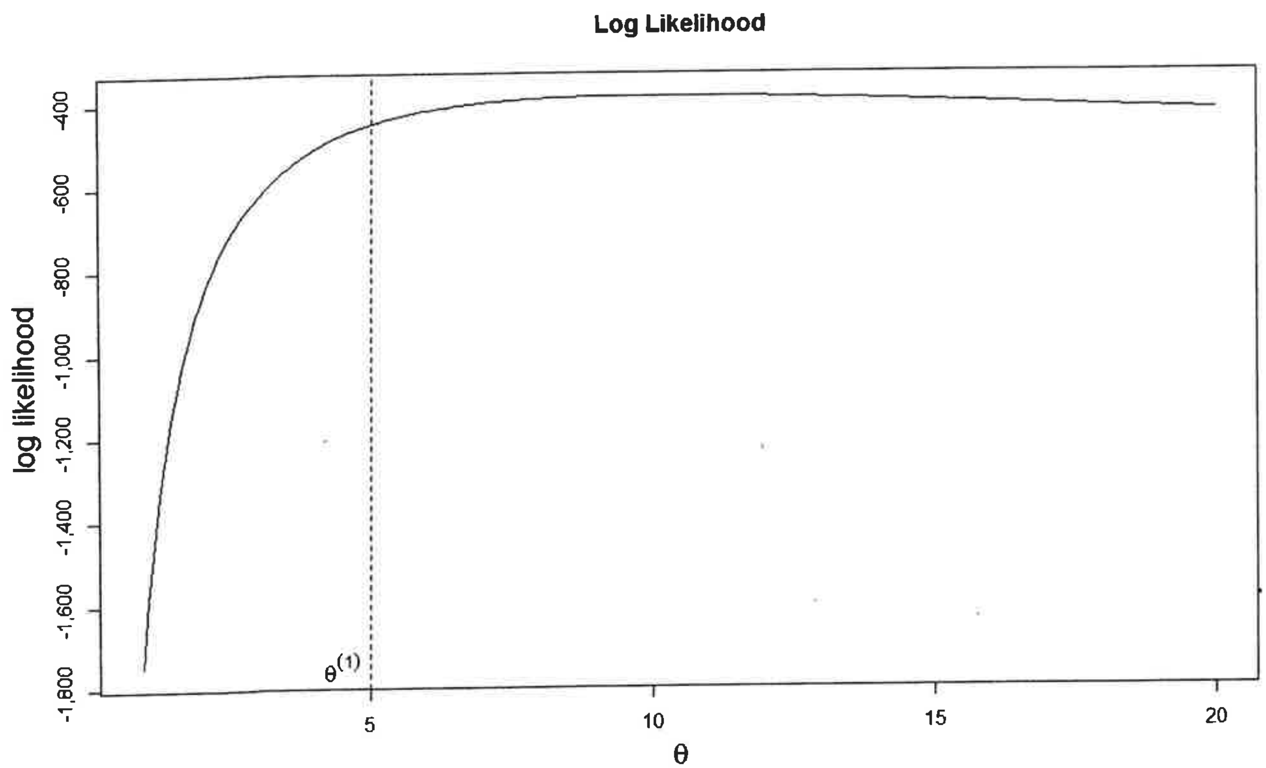
38.

A sample of 100 is taken from the following distribution:

$$f(y_i) = \frac{y_i e^{-\left(\frac{y_i}{\theta}\right)}}{\theta^2} \quad i = 1,2,\ldots,100$$

You are told that $\sum_{i=1}^{100} y_i = 2{,}025.24$

You are provided with a plot of the log likelihood as a function of θ for this data.



The method of scoring is used to estimate θ from the sample. Assume $\theta^{(1)} = 5$.

Calculate $\theta^{(2)}$.

A. Less than or equal to 2.0
B. Greater than 2.0 but less than or equal to 6.5
C. Greater than 6.5, but less than or equal to 11.0
D. Greater than 11.0, but less than or equal to 15.5
E. Greater than 15.5

39.

A set of n observations, $y_1, y_2, \dots, y_n$, are assumed to be exponentially distributed,

$$f(y_i) = \frac{e^{-\left(\frac{y_i}{\theta_i}\right)}}{\theta_i} \quad i = 1,2,\dots,n$$

A GLM is fit to the data with the following model specification:

$$ln(\theta_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{2i}^2 + \beta_4 x_{3i}$$

The vector of observed responses, Y and the design matrix, X, are given for the first four observations as well as the vector of all estimated parameters, $\hat{\beta}$:

$$Y = \begin{pmatrix} 14.8 \\ 137.6 \\ 0.4 \\ 38.3 \end{pmatrix} \qquad X = \begin{pmatrix} 1 & 1 & 2.7 & 7.29 & 1 \\ 1 & 0 & 0.6 & 0.36 & 1 \\ 1 & 1 & 2.9 & 8.41 & 1 \\ 1 & 1 & 3.0 & 9.00 & 0 \end{pmatrix} \qquad \hat{\beta} = \begin{pmatrix} 2.99 \\ -0.27 \\ -0.67 \\ 0.16 \\ 0.91 \end{pmatrix}$$

The first four rows and columns of the hat matrix, H, are:

$$H = \begin{pmatrix} 0.039 & -0.008 & 0.038 & 0.014 \\ -0.008 & 0.091 & -0.008 & -0.023 \\ 0.038 & -0.008 & 0.039 & 0.017 \\ 0.014 & -0.023 & 0.017 & 0.039 \end{pmatrix}$$

Calculate the standardized residual for the second observation, r_2.

A. Less than 1
B. At least 1, but less than 2
C. At least 2, but less than 3
D. At least 3 but less than 4
E. At least 4

EXAM CONTINUED ON NEXT PAGE

40.

You are fitting a Poisson regression model of the form:

$$E(Y_i) = \beta_1 + \beta_2 x_i$$

Maximum likelihood estimates of the beta coefficients are obtained using iterative weighted least squares procedure. You are given three matrixes:

- W is the weight matrix.
- X is the design matrix.
- Z has the beta values from the prior iteration applied to the explanatory variables as well as the correction term.

From the estimates of the first iteration, the following matrices are calculated using the matrixes as defined above:

$$(X^T W X)^{(1)} = \begin{bmatrix} 2.452 & -0.540 \\ -0.540 & 0.740 \end{bmatrix}$$
$$\left[(X^T W X)^{(1)}\right]^{-1} = \begin{bmatrix} 0.486 & 0.355 \\ 0.355 & 1.610 \end{bmatrix}$$

$$(X^T W z)^{(1)} = \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix}$$

Calculate $b_2^{(2)}$, the estimate for β_2 on the second iteration.

A. Less than 0
B. At least 0, but less than 0.40
C. At least 0.40, but less than 0.80
D. At least 0.80, but less than 1.20
E. At least 1.20

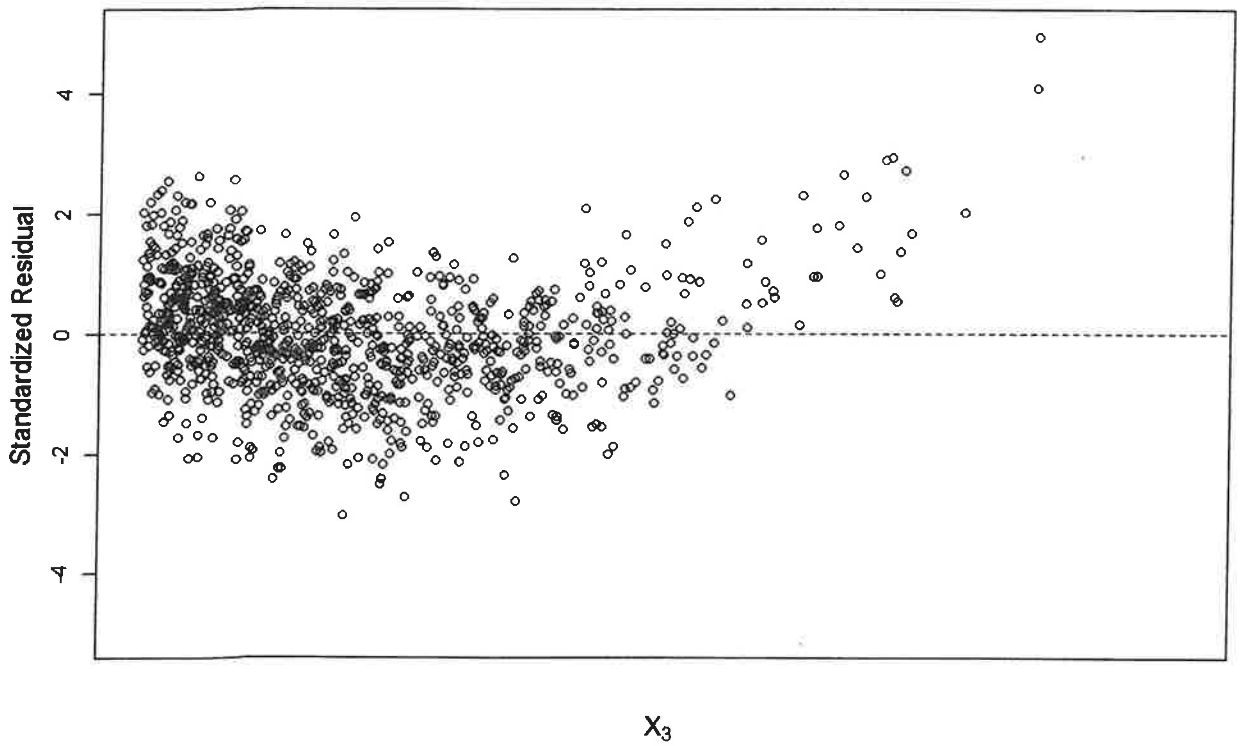
EXAM CONTINUED ON NEXT PAGE

41.

A GLM has been fit to a data set with 1,000 observations with the following model form:

$$E(y_i) = g^{-1}(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i})$$

The standardized residuals are plotted against the values of the variable represented by X_3 in the model. The plot is shown below.



Determine the best alternate model parameterization based on the residual plot above.

A. $E(y_i) = g^{-1}(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{3i}^2)$
B. $E(y_i) = g(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i})$
C. $E(y_i) = g^{-1}(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_4 X_{2i}^2 + \beta_3 X_{3i})$
D. $E(y_i^2) = g^{-1}(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i})$
E. $E[\ln(y_i)] = g^{-1}(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i})$

42.

A practitioner built a GLM to predict claim frequency. She used a Poisson error structure with a log link. You are given the following information regarding the model summary statistics:

- The model has 5 parameters.
- The likelihood ratio chi-squared statistic C is 11.601.

Determine the best statement of what we can we conclude from the value of C.

A. At least one of the non-intercept coefficients is likely to be non-zero using a p-value of .05, though we cannot conclude that more than one of the non-intercept coefficients are non-zero.
B. The model does not fit well, when the parsimony of the model is taken into account.
C. The model fit is significantly better with 5 parameters compared to 4 using a p-value of .05.
D. Each of the coefficients is likely to be non-zero.
E. The model fits well, even when the parsimony of the model is taken into account.

EXAM CONTINUED ON NEXT PAGE

43.

You are given the following information from a time series:

t	x_t
1	4.0
2	3.5
3	2.5
4	5.5
5	4.5
6	4.0
$\bar{x}$	4.0

Calculate the sample lag 4 autocorrelation.

A. Less than -0.30
B. At least -0.30, but less than -0.10
C. At least -0.10, but less than 0.10
D. At least 0.10, but less than 0.30
E. At least 0.30

EXAM CONTINUED ON NEXT PAGE

44.

You are given the following statements about stationarity:

I. Linear models for time series are stationary when they include functions of time.
II. All moving average processes are stationary.
III. All random walk processes are non-stationary.

Determine which of the above statements are true.

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

45.

You are given the following information:

- Data follows this model:
$$Y_t = 16 - y_{t-1} + 0.5(y_{t-1} - y_{t-2}) + w_t$$
where $w_t \sim N(0, 1)$.
- We have observed the first four observations of the time series:

t	y_t
1	3
2	7
3	11
4	7

Calculate the variance of Y_7 given y_1, y_2, y_3, y_4.

A. Less than 0.75
B. At least 0.75, but less than 1.25
C. At least 1.25, but less than 1.75
D. At least 1.75, but less than 2.25
E. At least 2.25

END OF EXAMINATION

Fall 2017 Exam S Final Answer Key

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

45 C