Exam MAS-II

CASUALTY ACTUARIAL SOCIETY
AND THE
CANADIAN INSTITUTE OF ACTUARIES

Expertise. Insight. Solutions.

Canadian Institute of Actuaries / Institut canadien des actuaires

MAS-II

Modern Actuarial Statistics II

April 24, 2019

4 HOURS

INSTRUCTIONS TO CANDIDATES

1. This 84 point examination consists of 42 multiple choice questions each 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 Spring 2019 and that the exam name is MAS-II.
   - 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. <u>No additional time will be allowed for this after the exam has ended.</u> 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 the case study, "Systolic Blood Pressure Case Study", included in your exam packet.

   - Verify that you have a copy of "Tables for CAS MAS-II" 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.

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.

CONTINUE TO NEXT PAGE OF INSTRUCTIONS

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 May 8, 2019.

END OF INSTRUCTIONS

1.

In the supplemental material, you have been given a case study, "Systolic Blood Pressure Case Study", showing the results of different treatment options and the description of how that study was set up. You have also been asked to evaluate three statements regarding the results of the study.

I. None of the treatments significantly reduce systolic blood pressure.
II. The age of the patient has a significant effect on the change in blood pressure expected over a one-year time period.
III. The age of the patient interacts significantly with the treatment.

Determine which of the above statements are true.

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

CONTINUED ON NEXT PAGE

2.

In the supplemental material, you have been given a case study, "Systolic Blood Pressure Case Study", showing the results of different treatment options and the description of how that study was set up. You have also been asked to evaluate three statements regarding the results of the study.

I. Using a linear mixed effects model is not appropriate, since the dependent variable is not normally distributed.
II. The age effect should have been modeled with a quadratic term.
III. One should assume that variance in the change in systolic pressure is a function of the age of the patient rather than assuming constant variance.

Determine which of the above statements are true.

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

CONTINUED ON NEXT PAGE

3.

An insurance company is determining limited-fluctuation credibility standards for its automobile losses. You are given the following information:

- The company selects all of its credibility standards to be the number of claims at which there is a 99% probability that the observed amount is within 10% of the mean.
- The standard for full credibility for aggregate loss is 4,800 claims.
- Claim frequency follows a Poisson distribution.
- Claim frequency and claim severity are independent.

Calculate the limited-fluctuation credibility standard for claim severity.

A. Less than 4,100
B. At least 4,100, but less than 4,300
C. At least 4,300, but less than 4,500
D. At least 4,500, but less than 4,700
E. At least 4,700

CONTINUED ON NEXT PAGE

4.

Below is a table with the physical damage claim experience of commercial auto fleets for three insured companies for the most recent period.

Company	Number of vehicles	Mean claim amount per vehicle	Standard deviation of claim amount
X	150	3130.20	579.62
Y	272	2278.53	139.10
Z	511	2950.50	472.08

Company X will have 175 vehicles in its fleet in the next period.

Calculate the Bühlmann credibility prediction of Company X's aggregate claim amount for the next period.

A. Less than 500,000
B. At least 500,000, but less than 515,000
C. At least 515,000, but less than 530,000
D. At least 530,000, but less than 545,000
E. At least 545,000

CONTINUED ON NEXT PAGE

5.

You are provided with the following claims frequency data about the portfolio of a large commercial auto insurer:

- The claims frequency of each vehicle is independently and identically distributed.
- The overall average claims frequency is 0.2 claims per earned car year.
- The variance of the hypothetical means is 0.3.
- The expected value of the process variance is 1200.
- The limited-fluctuation credibility standard is 1083 claims.

You are asked to change the credibility methodology from limited-fluctuation to Bühlmann-Straub for a policyholder with 200 claims in 1800 earned car years in its prior loss experience.

Calculate the percentage change in the estimate of claims frequency for this policyholder due to the change in methodology.

A. Less than -5%
B. At least -5%, but less than 0%
C. At least 0%, but less than 5%
D. At least 5%, but less than 10%
E. At least 10%

CONTINUED ON NEXT PAGE

6.

An insurance company is currently using a limited-fluctuation credibility approach for a line of business with the following assumptions:

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

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

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

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

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

CONTINUED ON NEXT PAGE

7.

You are given:

- Claims follow a Poisson distribution with mean λ.
- The prior distribution of λ follows gamma distribution with $\alpha = 4$ and $\theta = 0.06$.
- Claims are independent and identically distributed between risks.

In Year 1, the following claim experience was observed for a portfolio:

Cohort	Risks	Claim Frequency per Risk
X	500	0.12
Y	300	0.20
Z	200	0.15
Total	1,000	0.15

In Year 2, the portfolio has the same number of risks.

Calculate the expected number of claims for the portfolio in Year 2 using the Bayesian approach.

A. Less than 153
B. At least 153, but less than 158
C. At least 158, but less than 163
D. At least 163, but less than 168
E. At least 168

CONTINUED ON NEXT PAGE

8.

You are using the Bühlmann-Straub approach to determine the credibility of the historical observations in order to estimate the number of claims in Year 3.

You are given:

- The following table of insureds and claim counts per year:

Year	Number of Insureds	Number of Claims
1	P_1	N_1
2	P_2	N_2

- The number of claims incurred per year for each insured is distributed as a Poisson random variable with mean λ.
- λ follows a gamma distribution with shape parameter α and rate parameter θ.
- The following three statements:

I. The credibility factor, Z, of the historical observations will change if α changes and all other values remain constant.

II. The credibility factor, Z, of the historical observations will change if θ changes and all other values remain constant.

III. The credibility factor, Z, of the historical observations will change if N_1 changes and all other values remain constant.

Determine which of the preceding statements are true.

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

CONTINUED ON NEXT PAGE

9.

You are given the following three statements:

I. Linear Mixed Models use random effects explicitly to model between-subject variance.

II. Marginal Models are population-averaged models.

III. Implied Marginal Models do not provide the distribution of the population mean.

Determine which of the above statements are true.

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

CONTINUED ON NEXT PAGE

10.

You are building a Linear Mixed Model with constant variance at each time point. The covariance matrix for the residuals for a given subject, $\boldsymbol{R}_i$, has a constant diagonal structure. Consider the following statements about $\boldsymbol{R}_i$.

I. This structure is often used in an experiment with repeated trials under the same condition.
II. The residuals associated with observations of the same subject are assumed to be uncorrelated.
III. This matrix requires only one parameter to be specified.

Determine which of the preceding statements are true.

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

CONTINUED ON NEXT PAGE

11.

Joe wants to use a Linear Mixed Model to determine if the use of any fertilizer results in a significantly different apple weight. After recording his observations for ten trees within the orchard, he has a dataset with the following form:

Tree ID	Treatment	Apple Count	Trunk Diameter (cm)	Apple ID	Apple Weight (g)	Red Apple	Green Apple
1	Control	20	14	1	83	Y	N
1	Control	20	14	2	65	Y	N
1	Control	20	14	3	76	Y	N
1	Control	20	14	4	69	N	Y
1	Control	20	14	5	87	N	Y
...	...	...	...	...	...	...	...
10	Brand K	18	18	178	94	Y	N
10	Brand K	18	18	179	102	Y	N

He has provided some notes to accompany his dataset:

- *TREE ID* and *APPLE ID* are both identifiers of the observation and therefore not considered to be covariates.
- The trees were either left untreated or treated with one of two brands (J or K).
- *APPLE COUNT* is the number of apples on a given tree.
- The apple can be only one of two colors: red or green.

You are given:

I. The apple dataset is an example of a two-level clustered dataset.

II. Including the Control Treatment in the intercept estimate and modeling Brand K and Brand J effect as the difference between the Control and the effect for Brand K or the difference between the intercept and Brand J will result in intrinsic aliasing.

III. If β_1 and β_2 represent the coefficients of the fixed effects associated with treatment, *TREAT_J* and *TREAT_K*, respectively, then the null and alternative hypotheses are as follows:

Null Hypothesis: $H_0\colon \beta_1 = 0$ and $\beta_2 = 0$

Alternative Hypothesis: $H_1\colon \beta_1 \neq 0$ or $\beta_2 \neq 0$

Determine which of the preceding statements are true.

A. I only

B. II only

C. III only

D. I, II and III

E. The answer is not given by (A), (B), (C), or (D)

CONTINUED ON NEXT PAGE

12.

Fitting a Linear Mixed Model requires estimating the fixed-effect parameters, $\boldsymbol{\beta}$, and the covariance parameters for the $\boldsymbol{D}$ and $\boldsymbol{R}_i$ matrices, $\boldsymbol{\theta}$. Consider the following statements regarding maximum likelihood (ML) estimation and restricted maximum likelihood (REML) estimation.

I. If $\boldsymbol{\theta}$ is known, ML produces the best linear unbiased estimator of $\boldsymbol{\beta}$.
II. REML accounts for the degrees of freedom used to estimate $\boldsymbol{\beta}$ in order to produce an unbiased estimate of $\boldsymbol{\theta}$.
III. ML and REML estimation both use the same formulas to estimate $\boldsymbol{\beta}$.

Determine which of the preceding statements are true.

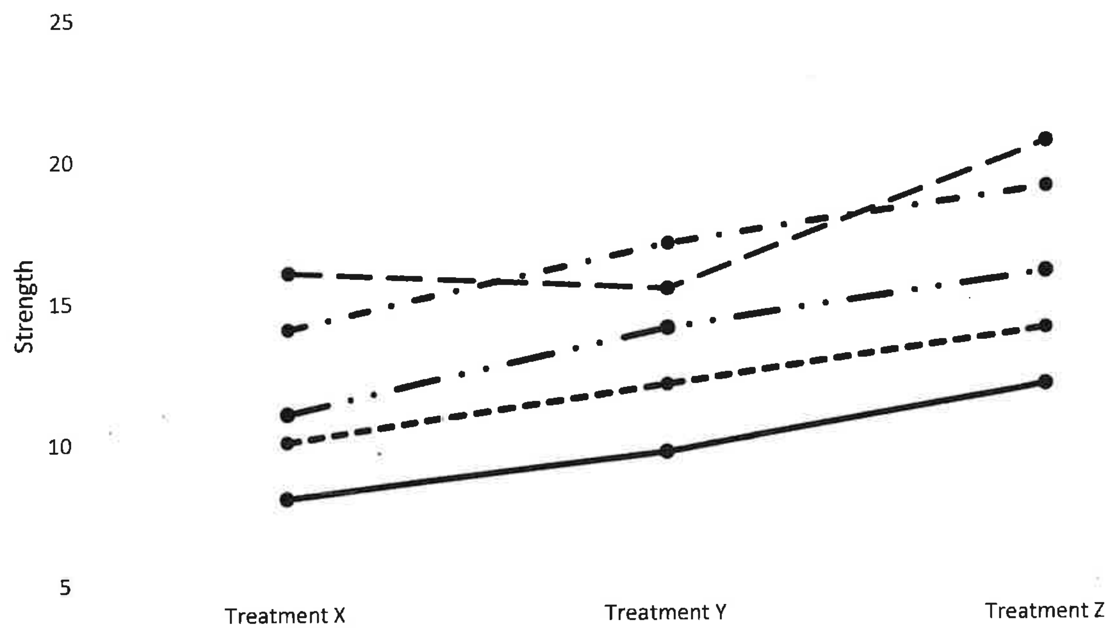
A. I only
B. II only
C. III only
D. I, II, and III
E. The correct answer is not given by (A), (B), (C), or (D)

CONTINUED ON NEXT PAGE

13.

The following study was conducted to examine the effect of three treatments on wood strength.

There are five wood planks. Each is cut into three pieces, and one of the three treatments is randomly assigned to each piece. The change in wood strength is measured and graphed as follows, where each line represents the measurements of the strength of a single plank:



You are given the following observations from the graph:

I. Wood strength is not affected by treatment.

II. Variation between planks is noticeable, suggesting a random effect associated with treatment for each plank.

III. Variation between planks is noticeable, suggesting a random effect associated with the intercept for each plank.

Determine which of the preceding statements are true.

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

CONTINUED ON NEXT PAGE

14.

You are fitting a Linear Mixed Model to determine the effect of different preservatives on the shelf life of apples. You are given the following information:

- Each apple is cut into three pieces and treated with a different preservative.
- The measured between subject variance is 1.43 and the within subject variance is 0.87.

Determine which of the following matrices is the variance-covariance matrix for a randomly selected subject for the Implied Marginal Model.

A. $\begin{bmatrix} 1.43 & 1.43 & 1.43 \\ 1.43 & 1.43 & 1.43 \\ 1.43 & 1.43 & 1.43 \end{bmatrix}$

B. $\begin{bmatrix} 0.87 & 0 & 0 \\ 0 & 0.87 & 0 \\ 0 & 0 & 0.87 \end{bmatrix}$

C. $\begin{bmatrix} 2.30 & 1.43 & 1.43 \\ 1.43 & 2.30 & 1.43 \\ 1.43 & 1.43 & 2.30 \end{bmatrix}$

D. $\begin{bmatrix} 0.87 & 1.43 & 1.43 \\ 1.43 & 0.87 & 1.43 \\ 1.43 & 1.43 & 0.87 \end{bmatrix}$

E. $\begin{bmatrix} 2.30 & 2.30 & 2.30 \\ 2.30 & 2.30 & 2.30 \\ 2.30 & 2.30 & 2.30 \end{bmatrix}$

CONTINUED ON NEXT PAGE

15.

You are fitting a Linear Mixed Model on a clustered dataset and considering the following two models:

Model	Model Specification
I	$Y_{ij} = \beta_0 + \beta_1 X_{ij}^{(1)} + \beta_2 X_{ij}^{(2)} + \beta_3 X_{ij}^{(3)} + u_j + \varepsilon_{ij}$ u_j follows a normal distribution with mean 0 and variance σ_c^2 ε_{ij} follows a normal distribution with mean 0 and variance σ^2 All ε_{ij} are independent
II	$Y_{ij} = \beta_0 + \beta_1 X_{ij}^{(1)} + \beta_2 X_{ij}^{(2)} + \beta_3 X_{ij}^{(3)} + \varepsilon_{ij}$ ε_{ij} follows a normal distribution with mean 0 and variance σ^2 All ε_{ij} are independent

The fit statistics are as follows:

	Model I	Model II
-2 log-likelihood	401.1	489.9

You are using the test statistic to choose between Model I and Model II. Determine which of the following statements are true:

A. Based on referring the test statistic to a chi-squared distribution with 1 degree of freedom, we select Model I at a confidence level of 0.005

B. Based on referring the test statistic to an average of chi-squared distributions with 0 and 1 degrees of freedom, we select Model I at a confidence level of 0.005

C. Based on referring the test statistic to a chi-squared distribution with 1 degree of freedom, we select Model II at a confidence level of 0.005

D. Based on referring the test statistic to an average of chi-squared distributions with 0 and 1 degrees of freedom, we select Model II at a confidence level of 0.005

E. The correct answer is not given by (A), (B), (C), or (D)

CONTINUED ON NEXT PAGE

16.

Your company implements a new marketing strategy for agents. It is implemented under the direction of two marketing directors who each lead the agents in two different states. You build a Linear Mixed Model to model the policy growth over the next year after this new strategy is implemented. Below is the model output from this Linear Mixed Model.

Solution for Fixed Effects			
Effect	**Location**	**Estimate**	**Standard Error**
Intercept		86.75	61.39
Location	Rural	-272.71	63.82
Location	Town	-179.56	47.62
Location	City	0.00	0.00
Tenure		-14.66	2.88
Beginning Policies		0.33	0.07

Solution for Random Effects				
Effect	**Director**	**State**	**Estimate**	**Standard Error Prediction**
Intercept	D1		-21.66	32.39
Intercept	D1	S1	5.85	15.94
Intercept	D1	S2	1.89	15.93
Intercept	D2		21.66	32.39
Intercept	D2	S3	-3.13	15.93
Intercept	D2	S4	7.10	15.92

Consider an agent with the following characteristics:

Location	**Tenure**	**Beginning Policies**	**Director**	**State**
Town	2	921	D1	S2

Calculate the expected policy growth for this agent over this time period according to this Linear Mixed Model.

A. Less than 140
B. At least 140 but less than 150
C. At least 150 but less than 160
D. At least 160 but less than 170
E. At least 170

CONTINUED ON NEXT PAGE

17.

Frank, the frequentist, fits a model to claim frequency using frequentist methods. Bob, a Bayesian, fits a model to claim frequency using MCMC.

- Frank reports a 95% confidence interval for claim frequency of (1.45, 2.75).
- Bob reports a 95% percentile interval for claim frequency of (1.50, 2.40).

You are provided with the following three statements:

I. Frank's analysis indicates that claim frequency is between 1.45 and 2.75 with probability 95%.

II. 95% of Bob's posterior sample is between 1.50 and 2.40.

III. The value of the 96th percentile of Bob's posterior sample is not within Bob's interval.

Determine which of the statements above are true.

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

CONTINUED ON NEXT PAGE

18.

You are building the following model:

$$y_i \sim Exponential(\theta)$$

$$\theta \sim Uniform(0, \omega)$$

$$\omega \sim Uniform(7, 10)$$

Your buddy, Sarah Jane, tells you that the posterior distribution of $\theta|\omega$ is:

$$\begin{cases} \dfrac{\frac{1}{\theta}e^{-\Sigma y_i/\theta}}{\int_0^{\omega}\frac{1}{\theta}e^{-\Sigma y_i/\theta}\,d\theta} & if\ 0 < \theta < \omega \\ 0 & otherwise \end{cases}$$

Determine which of the following statements is correct.

A. Sarah Jane is correct.
B. Sarah Jane is incorrect because the bounds of the integral are incorrect.
C. Sarah Jane is incorrect because she didn't include the prior density.
D. Sarah Jane is incorrect because the variable of integration ($d\theta$) is incorrect.
E. Sarah Jane is incorrect for another reason not listed in (B), (C), or (D).

CONTINUED ON NEXT PAGE

19.

You are modeling how many claims, Y_i, each individual policy i will have next year. You are using the following model:

$$\Pr(Y_i = y_i) = \binom{10}{y_i} p^{y_i}(1-p)^{10-y_i}$$

$$f(p) = \frac{\Gamma(10)}{\Gamma(4)\Gamma(6)} p^3(1-p)^5$$

Given the following five data points:

0, 0, 0, 1, 3

Calculate the mean of the posterior distribution of p.

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

CONTINUED ON NEXT PAGE

20.

You are analyzing the claim counts for n individual policyholders. Let Y_i denote the claim count for the i^{th} policyholder. You build your model as follows:

$$\Pr(Y_i = y_i | P = p) = (1-p)^{y_i} p$$

$$f(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

$$f(\alpha) = \frac{1}{5} e^{-\frac{\alpha}{5}}$$

$$f(\beta) = \frac{1}{8} e^{-\frac{\beta}{8}}$$

Assuming that p is constrained to be between 0 and 1, the conditional posterior distribution of $p|(\alpha, \beta, y_1, \ldots, y_n)$ is proportional to K.

Identify which of the following could be K.

A. $p^{\alpha-1}(1-p)^{\beta-1}$

B. $p^{\alpha}(1-p)^{y_i+\beta-1}$

C. $\frac{\beta}{8}\frac{\alpha}{5}\, p^{\alpha+n-1}(1-p)^{\Sigma y_i+\beta-1}$

D. $p^{4+n}(1-p)^{\Sigma y_i+7}$

E. None of the above could be K.

CONTINUED ON NEXT PAGE

21.

You are given the following:

- h is normally distributed with mean μ and variance 1.
- μ is normally distributed with mean 25 and variance 9.
- You observe a value of h of 24.
- The posterior distribution of h has a mean of 24.1 and a variance of 0.81.

Calculate the width of the 50% highest posterior density interval of h.

A. Less than 1.10
B. Greater than or equal to 1.10 and less than 1.15
C. Greater than or equal to 1.15 and less than 1.20
D. Greater than or equal to 1.20 and less than 1.25
E. Greater than or equal to 1.25

CONTINUED ON NEXT PAGE

22.

Data are collected on 10 players on a baseball team. For each individual player, i, let the number of hits be denoted as $hits_i$ and the number of at-bats be denoted n_i. The batting average is the ratio of hits to at-bats ($hits_i/n_i$).

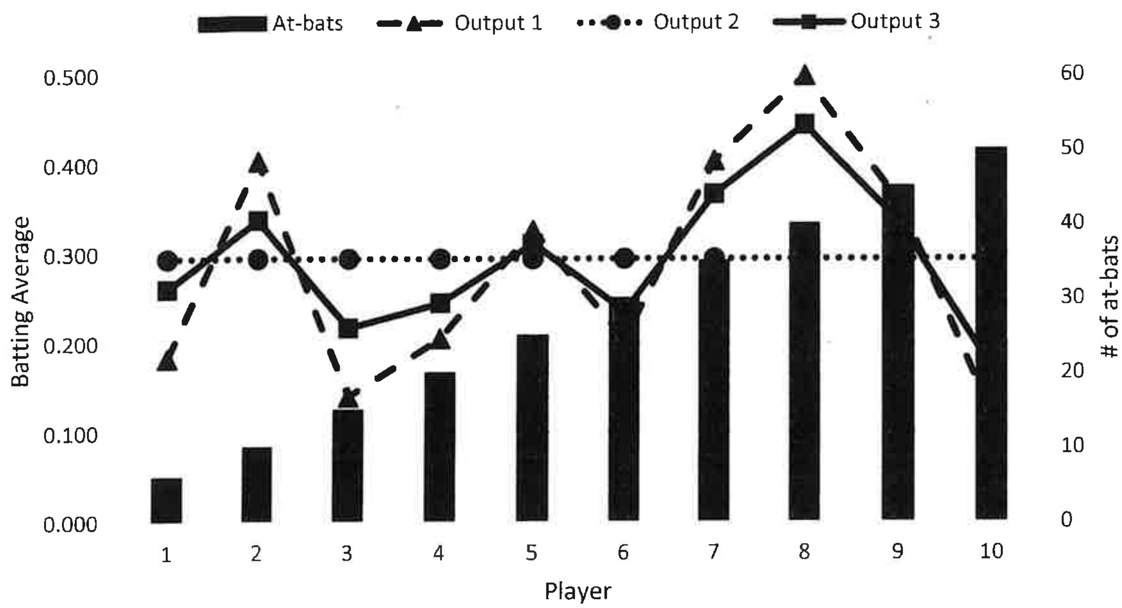
The following three models are proposed:

Model X
$hits_i \sim Binomial(n_i, p_i)$
$logit(p_i) = \alpha_i$
$\alpha_i \sim Normal(0, 25)$

Model Y
$hits_i \sim Binomial(n_i, p_i)$
$logit(p_i) = \alpha_i$
$\alpha_i \sim Normal(\alpha, \sigma^2)$
$\alpha \sim Normal(0, 1)$
$\sigma \sim HalfCauchy(0, 1)$

Model Z
$hits_i \sim Binomial(n_i, p_i)$
$logit(p_i) = \alpha$
$\alpha \sim Normal(0, 25)$

The models are fit and the posterior means of batting average for each player are plotted below:



Determine which of the following are true.

A. Model X produced Output 1, Model Y produced Output 2, and Model Z produced Output 3.
B. Model X produced Output 2, Model Y produced Output 1, and Model Z produced Output 3.
C. Model X produced Output 1, Model Y produced Output 3, and Model Z produced Output 2.
D. Model X produced Output 3, Model Y produced Output 1, and Model Z produced Output 2.
E. None of the above statements are true.

CONTINUED ON NEXT PAGE

23.

A process is assumed to follow a Binomial distribution with probability of success equal to π. The process is observed for 10 outcomes and all 10 of the observed outcomes were successes. A model is fit to this observed data and the prior distribution for π in this model is given below:

$$\pi^{-1}(1-\pi)^{-1}; \quad \pi \in [0,1]$$

Calculate the posterior mean of π.

A. Less than or equal to 0.60
B. Greater than 0.60 but less than or equal to 0.75
C. Greater than 0.75 but less than or equal to 0.90
D. Greater than 0.90
E. No posterior mean exists. The posterior distribution is improper.

CONTINUED ON NEXT PAGE

24.

You are building a model to understand student test scores. Your model is as follows:

$$y_{ij} = \beta_0 + x_{ij}\beta_1 + \tau_i + \epsilon_{ij}$$

$$\beta_0 \sim Normal(\mu, \sigma_0^2)$$
$$\beta_1 \sim Normal(0, \sigma_1^2)$$
$$\tau_i \sim Normal(0, \sigma_2^2)$$
$$\mu \sim Normal(0, \sigma_3^2)$$
$$\epsilon_{ij} \sim Normal(0, \sigma_4^2)$$

where student j in classroom i has a final exam score of y_{ij} and a midterm score of x_{ij}.

You feel that the classroom data is not dependable and, therefore, you want to shrink the classroom-specific effects toward the overall mean.

Determine how you would adjust the hyperparameters for this model.

A. Reduce the value of σ_0^2
B. Reduce the value of σ_1^2
C. Reduce the value of σ_2^2
D. Reduce the value of σ_3^2
E. Reduce the value of σ_4^2

CONTINUED ON NEXT PAGE

25.

You fit the following model:

$$y_i \sim Normal(\mu, 4)$$
$$\mu \sim Normal(0, 100)$$

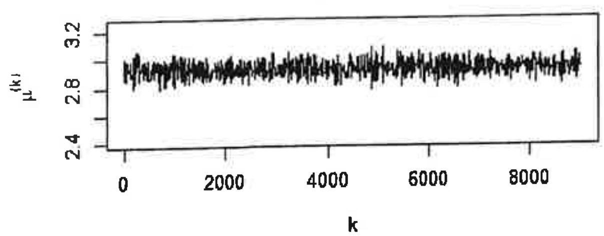
You propose new values for $\mu^{(k)}$ by generating a random value from the following proposal distribution:

$$\mu^* \sim Normal(\mu^{(k-1)}, \sigma^2)$$

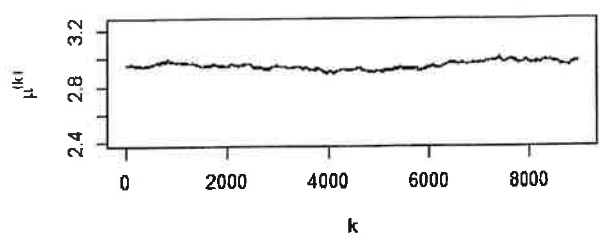
where $\mu^{(j)}$ is the j^{th} value of μ in your chain. The trace plot of $\mu^{(k)}$ from the above proposal distribution is labeled "First Pass" (after sufficient warm-up). You then re-run the sampling process after changing the proposal distribution to:

$$\mu^{(k)} \sim Normal(\mu^{(k-1)}, 100\sigma^2)$$

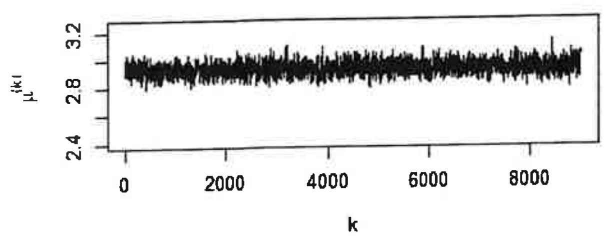
First Pass



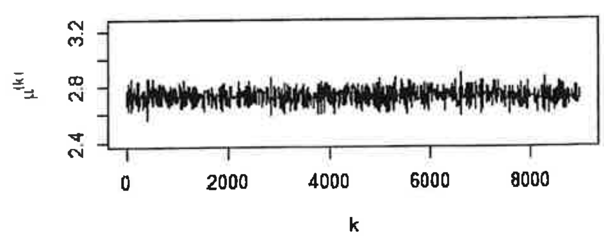
Plot I



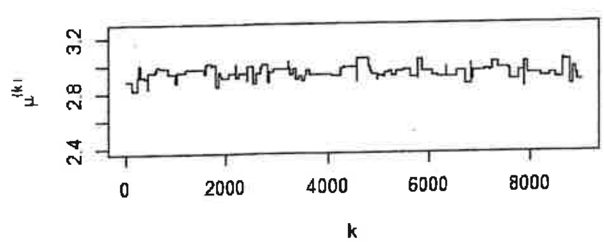
Plot II



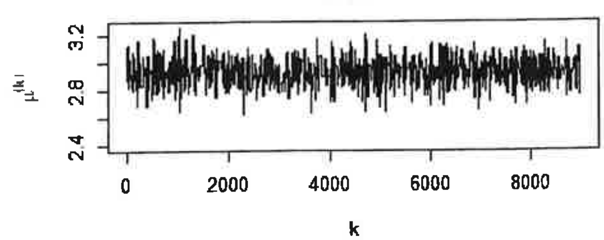
Plot III



Plot IV



Plot V



Determine which of the plots above is most likely the resulting trace plot with the new proposal distribution.

A. Plot I
B. Plot II
C. Plot III
D. Plot IV
E. Plot V

CONTINUED ON NEXT PAGE

26.

You plan to fit the following model:

$$y \sim Poisson(\lambda)$$

$$\lambda \sim Exponential(5)$$

You have the following three data points:

0 1 4

You use Metropolis-Hastings, and a proposal distribution for $\lambda^{(k)}$ of:

$$\lambda^* \sim Exponential(\lambda^{(k-1)})$$

Your current value of λ is 4.2 and the proposed value is 4.3.

Calculate the probability of accepting this proposal.

A. Less than 0.25
B. Greater than 0.25, but less than 0.45
C. Greater than 0.45, but less than 0.65
D. Greater than 0.65, but less than 0.85
E. Greater than 0.85

CONTINUED ON NEXT PAGE

27.

A multilevel model with varying intercepts for $class$ is fit to a dataset that includes two other independent variables, X_1 and X_2. The model is specified below.

$$\begin{array}{lcl} y_i & \sim & Poisson(\lambda_i) \\ ln(\lambda_i) & = & \alpha + \alpha_{class[i]} + \beta_1 X_{1i} + \beta_2 ln(X_{2i}) \\ \alpha & \sim & Normal(0,\ 2) \\ \alpha_{class} & \sim & Normal(0,\ \sigma_{class}^2) \\ \beta_1 & \sim & Normal(0,\ 1) \\ \beta_2 & \sim & Normal(0,\ 10) \\ \sigma_{class} & \sim & HalfCauchy(0,\ 1) \end{array}$$

An actuary wants to obtain a prediction for an unobserved, average $class$ when $x_1 = 0.28$ and $x_2 = 122.4$.

Determine the minimum number of posterior distributions that the actuary needs to consider to obtain a prediction.

A. 2 or fewer
B. 3
C. 4
D. 5
E. 6 or more

CONTINUED ON NEXT PAGE

28.

You have two random variables ϕ and θ where

$$\begin{aligned} \phi|\theta &\sim Binomial(n, \theta) \\ \theta &\sim Beta(a, b) \end{aligned}$$

and n is a fixed positive integer. Thus, the prior and the conditional distributions are given by

$$\begin{aligned} \pi(\theta) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1} \\ f(\phi|\theta) &= \binom{n}{\phi}\theta^{\phi}(1-\theta)^{n-\phi} \end{aligned}$$

where $\Gamma(.)$ denotes the Gamma function and $\int \theta^{a-1}(1-\theta)^{b-1}d\theta = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

An actuary wants to use a Gibbs sampler to sample from the joint distribution of θ and ϕ. She sets an initial value for ϕ and θ denoted by (ϕ_0, θ_0) and subsequently in the i^{th} step samples $\phi_i \sim f(\phi|\theta_{i-1})$ and $\theta_i \sim f(\theta|\phi_i)$

Determine which conditional distribution for $\theta|\phi_i$ correctly implements the Gibbs sampler.

A. $Beta(\phi_i, n-\phi_i)$
B. $Beta(b+\phi_i, a+n-\phi_i)$
C. $Beta(a+b+\phi_i, n-\phi_i)$
D. $Beta(a+\phi_i, b+n-\phi_i)$
E. None of the above

CONTINUED ON NEXT PAGE

29.

A random variable X follows a binomial distribution:

$$X \sim Binomial(n, \theta)$$

The prior probability distribution of θ is:

$$p(\theta) = 2\cos^2(4\pi\theta)$$

An actuary uses the Metropolis algorithm to sample from the posterior distribution of θ. At each iteration of the algorithm let θ_{curr} and θ_{prop} denote the current and proposed value of θ.

Determine the acceptance probability of θ_{prop}.

A. $\min\left\{\frac{\theta_{curr}^x(1-\theta_{curr})^{n-x}}{\theta_{prop}^x(1-\theta_{prop})^{n-x}}, 1\right\}$

B. $\min\left\{\frac{\theta_{prop}^x(1-\theta_{prop})^{n-x}cos^2(4\pi\theta_{prop})}{\theta_{curr}^x(1-\theta_{curr})^{n-x}cos^2(4\pi\theta_{curr})}, 1\right\}$

C. $\min\left\{\frac{\theta_{curr}^x(1-\theta_{curr})^{n-x}cos^2(4\pi\theta_{curr})}{\theta_{prop}^x(1-\theta_{prop})^{n-x}cos^2(4\pi\theta_{prop})}, 1\right\}$

D. $\min\left\{\frac{\theta_{prop}^x(1-\theta_{prop})^{n-x}cos^2(4\pi\theta_{curr})}{\theta_{curr}^x(1-\theta_{curr})^{n-x}cos^2(4\pi\theta_{prop})}, 1\right\}$

E. $\min\left\{\frac{\theta_{curr}^x(1-\theta_{curr})^{n-x}cos^2(4\pi\theta_{prop})}{\theta_{prop}^x(1-\theta_{prop})^{n-x}cos^2(4\pi\theta_{curr})}, 1\right\}$

CONTINUED ON NEXT PAGE

30.

Insurance company XYZ writes personal auto insurance. You are given:

- XYZ classifies drivers into two categories: risky and non-risky.
- Risky drivers have a 50% probability of filing an auto claim in a policy year.
- Non-risky drivers have a 5% probability of filing an auto claim in a policy year.
- Risky drivers comprise 20% of the population. The remaining 80% are non-risky.

A new driver submits a personal auto insurance application to ABC and mentions that he filed a claim with his previous insurance company in the most recent policy year.

Calculate the probability that the new driver will file an auto claim in the new policy year.

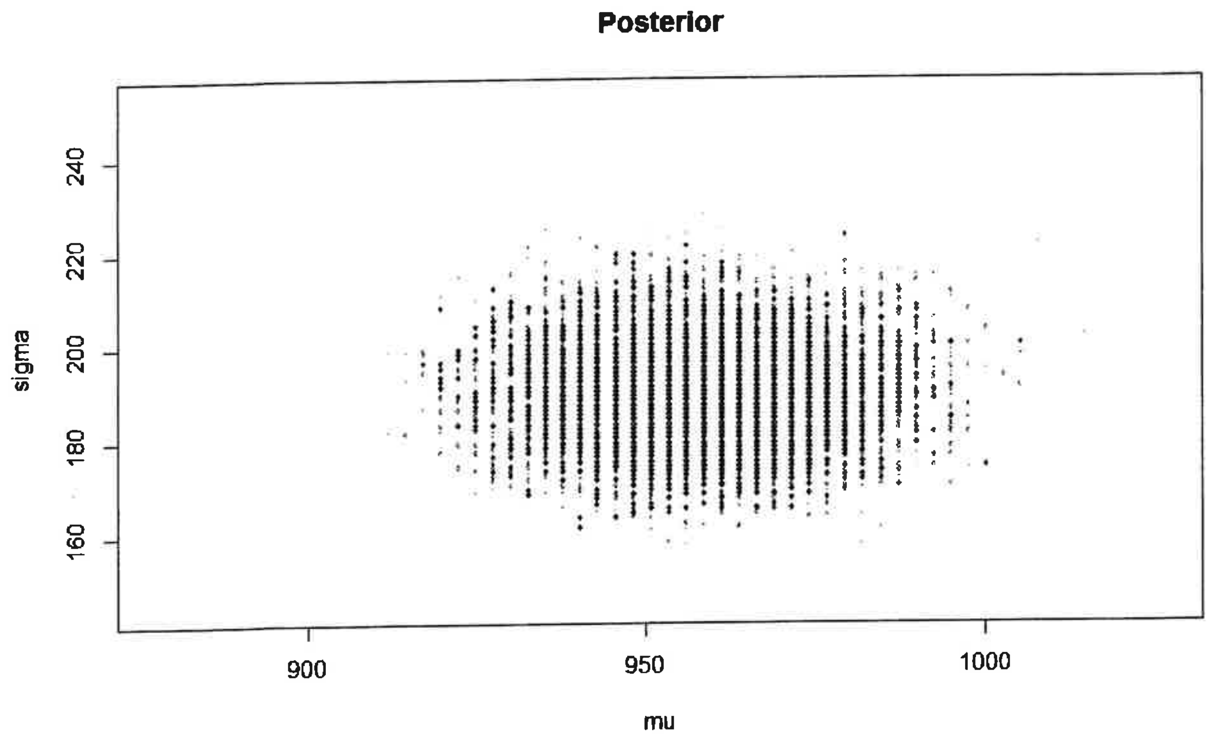
A. Less than 10%
B. At least 10%, but less than 20%
C. At least 20%, but less than 30%
D. At least 30%, but less than 40%
E. At least 40%

CONTINUED ON NEXT PAGE

31.

An actuary is reviewing a portfolio of auto claims, and is interested in analyzing the claim severity distribution using a Bayesian model. The portfolio's observed auto claim severity sample follows a normal distribution with mean 1,000 and standard deviation 200.

The graph below represents the samples from the posterior distribution for the mean (mu) and standard deviation (sigma).



Determine which of the following most likely represents the prior distributions used by the actuary.

A. $\mu \sim Normal(600, 100^2)$; $\sigma \sim Uniform(0, 200)$
B. $\mu \sim Normal(600, 100^2)$; $\sigma \sim Uniform(0, 400)$
C. $\mu \sim Normal(1000, 100^2)$; $\sigma \sim Uniform(0, 400)$
D. $\mu \sim Normal(1400, 100^2)$; $\sigma \sim Uniform(0, 400)$
E. $\mu \sim Normal(2000, 100^2)$; $\sigma \sim Uniform(0, 400)$

CONTINUED ON NEXT PAGE

32.
A Bayesian logistic regression model that predicts if a prospective customer will purchase an insurance policy is fit to a dataset with two cluster variables. The first cluster variable is the *channel* through which the insurance policy was sold. The *channel* variable can only take values from one of the following categories:

- Agency
- Broker
- Internet

The second cluster variable is the *type* of insurance policy. The *type* variable can only take values from one of the following categories:

- Personal Automobile
- Homeowners
- Workers Compensation

The model also includes two binary categorical variables that are represented as slopes. The first indicates if the offered insurance policy included a discount to the price. The second indicates if the prospective customer had a prior insurance policy.

In the model, let y_i indicate if the i^{th} prospective customer purchased the insurance policy. Let the *channel* and *type* of policy from the i^{th} prospective customer be represented by the $channel[i]$ and $type[i]$ subscripts respectively. Let D_i indicate if the offered insurance policy to the i^{th} customer included a discount to the price, and let P_i indicate if the prospective customer had a prior insurance policy.

The systematic portion of the model is specified below.

$$
\begin{aligned}
y_i &\sim Bernoulli(p_i) \\
logit(p_i) &= \gamma_i + \delta_i D_i + \Lambda_i P_i \\
\gamma_i &= \alpha + \alpha_{channel[i]} + \alpha_{type[i]} \\
\delta_i &= \beta_D + \beta_{D,channel[i]} + \beta_{D,type[i]} \\
\Lambda_i &= \beta_P + \beta_{P,channel[i]} + \beta_{P,type[i]}
\end{aligned}
$$

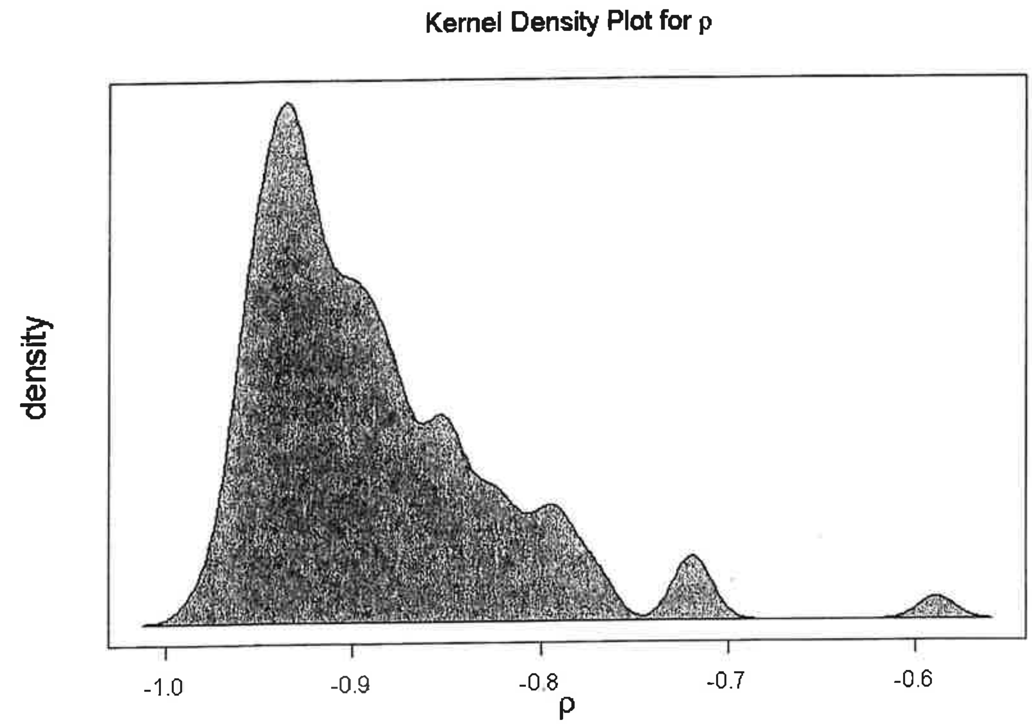
Multivariate Gaussian prior distributions are employed to model the covariance between the intercept and slope parameters for each cluster. In particular, an LKJ prior distribution for the correlation parameters is selected for this model.

Determine the number of correlation parameters that need to be estimated in this model.

A. Less than or equal to 2
B. Greater than 2 but less than or equal to 5
C. Greater than 5 but less than or equal to 8
D. Greater than 8 but less than or equal to 11
E. Greater than 11

CONTINUED ON NEXT PAGE

33.
An MCMC sampling algorithm with 10,000 samples post warm-up is employed to estimate the posterior distribution of two parameters. The kernel density estimate of the posterior distribution of the correlation, ρ, between the two parameters is plotted below:



The effective number of samples from the sampling algorithm is 202 for the first parameter and 331 for the second parameter.

Determine the most likely MCMC sampling algorithm employed for this analysis.

A. Gibbs
B. Hamiltonian Monte Carlo
C. Metropolis
D. Metropolis-Hastings
E. Quadratic Approximation

CONTINUED ON NEXT PAGE

34.

You want to check your software's implementation of the Hamiltonian Monte Carlo algorithm (HMC) to ensure is working properly for a model with a single parameter, θ. The software's HMC implementation has the following properties:

- The derivative of the potential function with respect to θ is -2.2
- Step size, ϵ, is 0.02
- Number of steps, L, is 50
- The standard deviation, s, of the normal momentum distribution is 0.05

At the end of the L steps for a particular HMC iteration, you compute the table of densities below. The *current* column contains densities at the start of the iteration and the *proposed* column contains densities at the end of the iteration.

	current	*proposed*
likelihood	0.025	0.021
prior	0.150	0.100
momentum	0.070	0.068

Calculate the probability of rejecting the proposed θ at the end of this iteration.

A. Less than 0.25
B. Greater than or equal to 0.25 but less than 0.50
C. Greater than or equal to 0.50 but less than 0.75
D. Greater than or equal to 0.75 but less than 1.00
E. 1.00

CONTINUED ON NEXT PAGE

35.

Suppose we want to model the standard deviation σ_i of a Gaussian distribution for each observation such that it is a function of a predictor variable x_i:

$$y_i \sim Normal(\mu, \sigma_i^2)$$
$$\mu \sim Normal(0, 25)$$
$$g(\sigma_i) = \alpha + \beta x_i$$
$$\alpha \sim Uniform(-5, 5)$$
$$\beta \sim Normal(0, 2)$$

where g is a link function.

Three candidates for g are considered:

I. $g(z) = \log(z)$

II. $g(z) = (z - 1)^{\frac{1}{\tau}}$, where $\tau = 2$

III. $g(z) = e^z$

Determine which of the candidate link functions would be appropriate for this model.

A. None
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)

CONTINUED ON NEXT PAGE

36.

A farmer with a giant basket of apples and oranges is trying to determine the probability of picking an orange. Consider the following:

- The basket only contains apples and oranges.
- After a piece of fruit is picked from the basket, it is then returned to the basket before another piece is picked.
- The farmer thinks the prior probability of picking an orange follows a beta distribution with probability density function of $p(x) = 12x^2(1-x)$.
- The farmer picks five fruits and one of them is an orange.
- The farmer assumes a Binomial likelihood function to construct a 50% highest posterior density interval (HPDI) using grid approximation with grid points of 0.2, 0.4, 0.5, 0.6, and 0.8.

Determine the lower bound of the farmer's 50% HPDI.

A. Less than or equal to 0.25
B. Greater than 0.25 but less than or equal to 0.35
C. Greater than 0.35 but less than or equal to 0.45
D. Greater than 0.45 but less than or equal to 0.55
E. Greater than 0.55

CONTINUED ON NEXT PAGE

37.

An actuary fits two models, X and Y, to the same dataset.

- Model X contains 10 parameters, resulting in a DIC of 1,020.18.
- The actuary drops two parameters from Model X – the resulting model, Model Y, contains 8 parameters, resulting in a DIC of 1,016.18.
- In both models, the actuary selects flat and non-informative prior distributions for the parameters.

Consider the following statements:

I. Model Y fits the data more closely than model X.

II. Model Y is preferred over Model X to make out-of-sample predictions.

III. Model Y is not overfit.

Determine which of the above statements must be true.

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

CONTINUED ON NEXT PAGE

38.

You are provided with training and test data samples consisting of a single variable X, and an observation Y consisting of two possible classes, T and F.

Training Data

i	x_i	y_i
1	-1.60	T
2	-1.50	F
3	-0.60	T
4	-0.30	T
5	0.40	F
6	0.40	F
7	0.70	T
8	1.20	T
9	1.30	T
10	2.10	F

Test Data

i	x_i	y_i
1	-1.3	F
2	0.9	F
3	1.2	T

The Bayes classifier $\Pr[Y = \text{F}|X = x] = e^{-x^2}$.

Calculate the amount by which the K-nearest neighbors test error rate with K=3 exceeds the Bayes error rate on the test data.

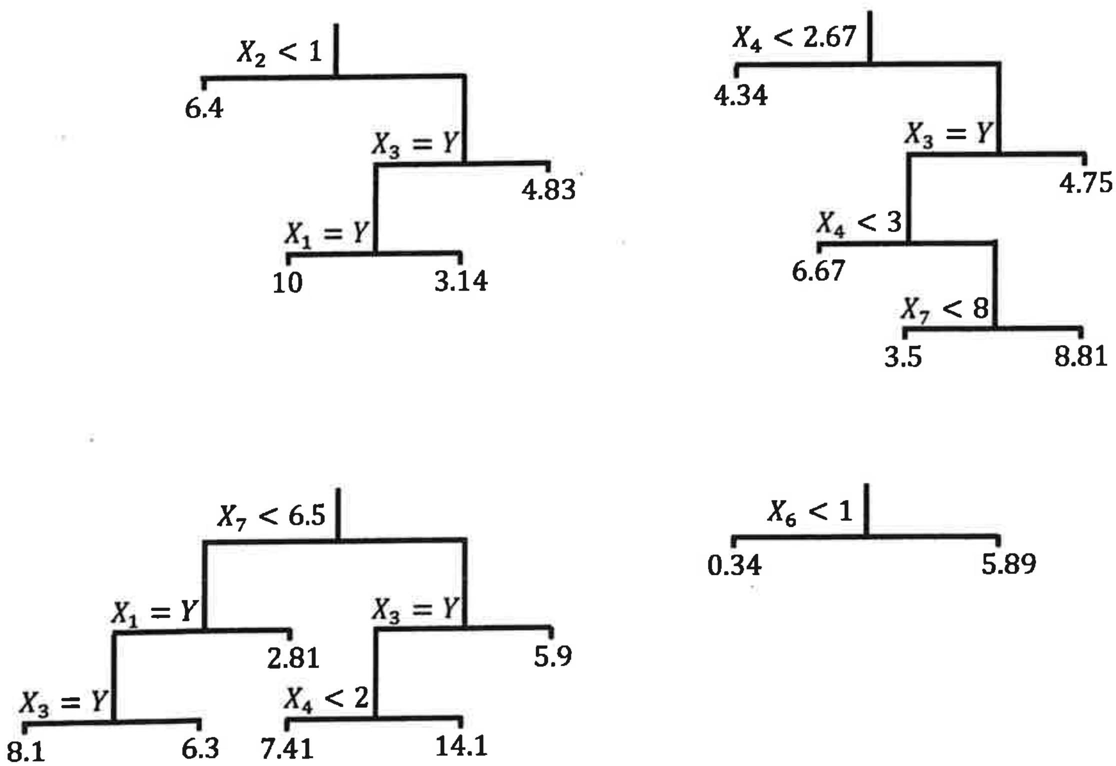
A. Less than 0.35
B. At least 0.35 but less than 0.40
C. At least 0.40 but less than 0.45
D. At least 0.45 but less than 0.50
E. Greater than or equal to 0.50

CONTINUED ON NEXT PAGE

39.

A boosted tree model is defined by:

- $\lambda = 0.2$
- The following four trees:



You are given the following record:

X_1	X_2	X_3	X_4	X_5	X_6	X_7
N	6	Y	4	0.5	0.25	6

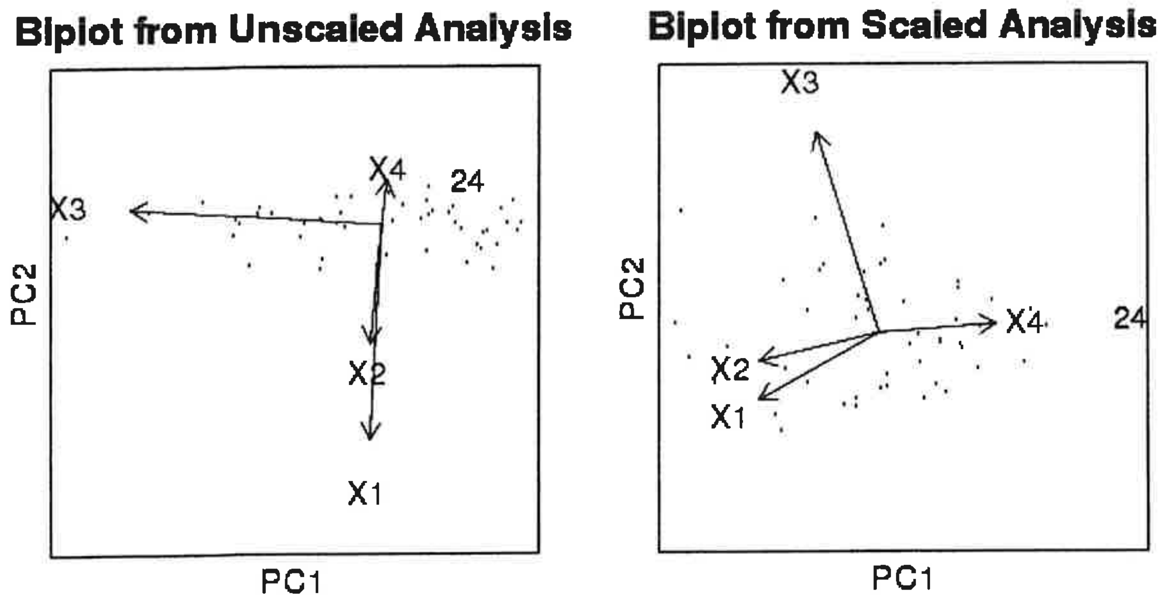
Calculate the prediction of the boosted tree model for this record.

A. Less than 2
B. At least 2, but less than 5
C. At least 5, but less than 8
D. At least 8, but less than 11
E. At least 11

CONTINUED ON NEXT PAGE

40.

You perform two separate principal component analyses on the same four variables in a particular data set: X1, X2, X3, and X4. The first analysis centers but does not scale the variables, and the second analysis centers and scales the variables. The biplots of the first two principal components produced from these analyses are shown below. The location of Observation 24 is labeled on the plots as well.



Given the following statements:

I. X1 is more highly correlated with X2 than with X3.
II. X3 has the highest variance of these four variables.
III. Observation 24 has a relatively large, positive value for X4.

Determine which of the preceding statements are demonstrated in the biplots shown above.

A. None of I, II, or II are demonstrated in the biplots
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)

CONTINUED ON NEXT PAGE

41.

You are reviewing a dataset with 100 observations in four variables: X_1, X_2, X_3, and X_4. You analyze this data using two principal components:

$$Z_1 = \varphi_{11}X_1 + \varphi_{21}X_2 + \varphi_{31}X_3 + \varphi_{41}X_4$$

$$Z_2 = \varphi_{12}X_1 + \varphi_{22}X_2 + \varphi_{32}X_3 + \varphi_{42}X_4$$

You are given the following statements:

I. $\sum_{i=1}^{100}\left(\sum_{j=1}^{4}\varphi_{j1}x_{ij}\right)^2 = \sum_{i=1}^{100}\left(\sum_{j=1}^{4}\varphi_{j2}x_{ij}\right)^2$

II. $\sum_{j=1}^{4}\varphi_{j1}\varphi_{j2} = 0$

III. $\sum_{j=1}^{4}\varphi_{j1}^2 + \sum_{j=1}^{4}\varphi_{j2}^2 = 1$

Determine which of the preceding statements are always true.

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

CONTINUED ON NEXT PAGE

42.

You have decided to perform k-means clustering with K=2 on the following dataset and have already randomly assigned clusters as follows:

Observation	x_1	x_2	Initial Cluster
1	5	5	2
2	4	6	2
3	3	0	1
4	5	3	1
5	5	1	2
6	3	6	1
7	2	5	2

- The centroid of the initial cluster 1 is $x_1 = 3.667$, $x_2 = 3.000$
- The centroid of the initial cluster 2 is $x_1 = 4.00$, $x_2 = 4.250$

Calculate the Euclidean distance of Observation 5 from the final centroid of Cluster 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

END OF EXAMINATION

Exam MAS-II Spring 2019

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

Systolic Blood Pressure Case Study

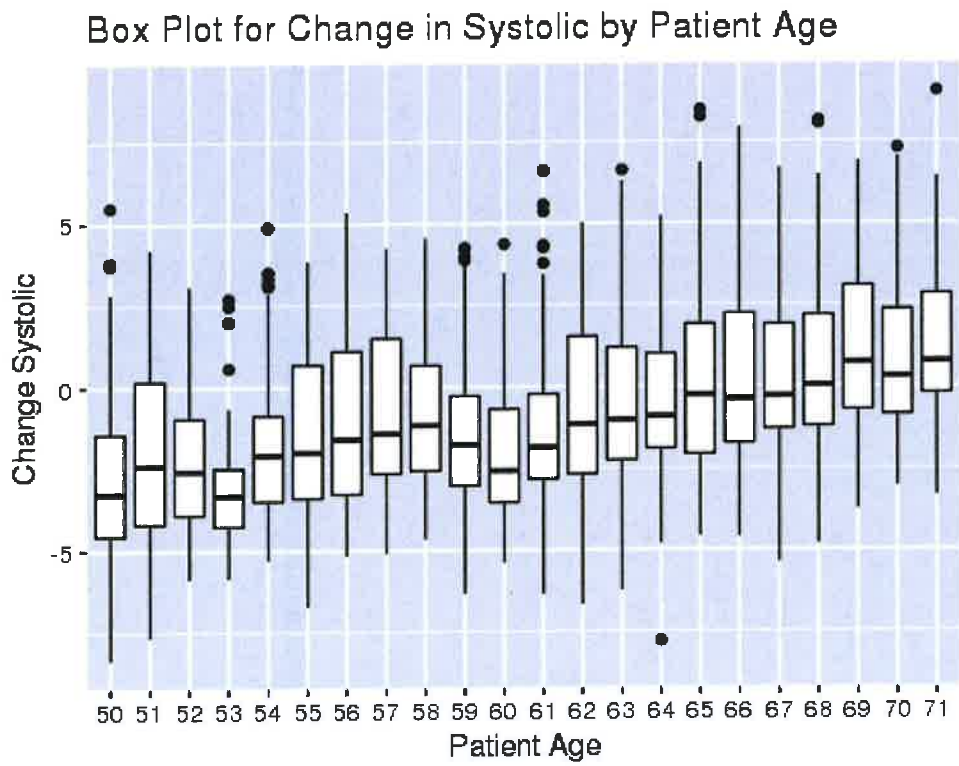
MAS II Spring 2019

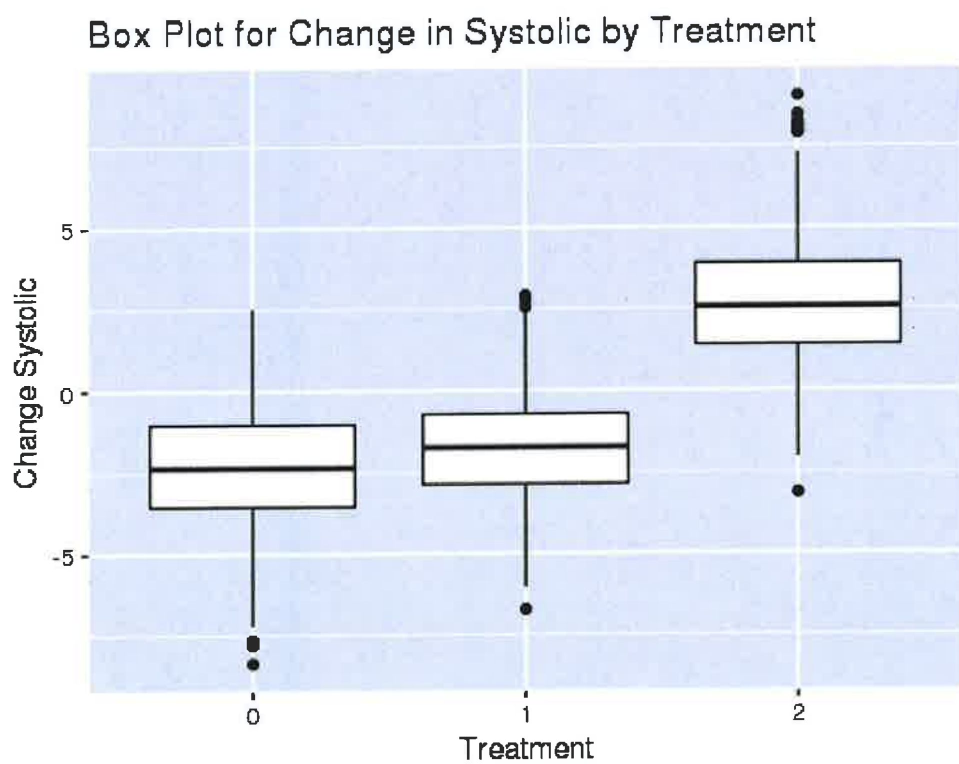
Table of Contents

B. Create Exploratory Data Output

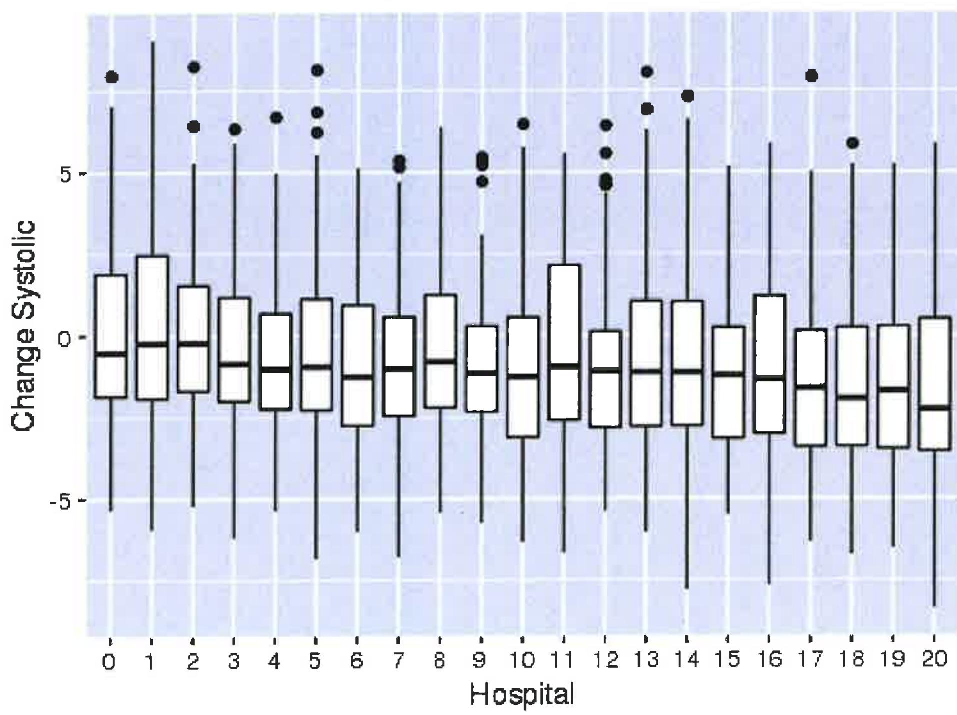
The case study examines the change in systolic blood pressure from three different treatments with the study conducted in 20 different hospitals. The patients included in the study had varying ages, but in all cases the study was conducted over a one year time period and all patients that began the study were included in the final study results. The question the study was intended to respond to is what is the effect of the three different treatments on systolic blood pressure and does that effect vary by age of the patient.

```
##     Treament          hospital         Age          end_systolic
##  Min.   :0.0000   1       : 100   Min.   :50.00   Min.   : 58.18
##  1st Qu.:0.0000   2       : 100   1st Qu.:55.00   1st Qu.:111.30
##  Median :1.0000   3       : 100   Median :62.00   Median :131.29
##  Mean   :0.9709   4       : 100   Mean   :60.96   Mean   :131.32
##  3rd Qu.:2.0000   5       : 100   3rd Qu.:65.00   3rd Qu.:149.86
##  Max.   :2.0000   6       : 100   Max.   :71.00   Max.   :224.16
##                   (Other):1499
##   beg_systolic    change_systolic   Treatment norm_change_systolic
##  Min.   : 62.69   Min.   :-8.3653   0:639     Min.   :-2.7802
##  1st Qu.:111.73   1st Qu.:-2.6882   1:882     1st Qu.:-0.7140
##  Median :131.56   Median :-1.1343   2:578     Median :-0.1484
##  Mean   :132.05   Mean   :-0.7265             Mean   : 0.0000
##  3rd Qu.:149.42   3rd Qu.: 0.9967             3rd Qu.: 0.6272
##  Max.   :216.95   Max.   : 9.0491             Max.   : 3.5579
##

## # A tibble: 2,099 x 8
##    Treament hospital   Age end_systolic beg_systolic change_systolic
##       <int> <fct>    <int>        <dbl>        <dbl>           <dbl>
##  1        0 0           71         188.         189.          -1.29
##  2        1 0           61         124.         126.          -1.81
##  3        2 0           64         143.         140.           3.10
##  4        0 0           54         103.         105.          -2.56
##  5        2 0           67         162.         158.           4.13
##  6        1 0           65         146.         146.           0.333
##  7        1 0           64         143.         143.           0.839
##  8        2 0           61         131.         128.           2.64
##  9        0 0           70         174.         175.          -1.28
## 10        1 0           70         181.         179.           1.33
##    Treatment norm_change_systolic
##    <fct>                    <dbl>
##  1 0                       -0.204
##  2 1                       -0.395
##  3 2                        1.39
##  4 0                       -0.668
##  5 2                        1.77
##  6 1                        0.386
##  7 1                        0.570
##  8 2                        1.22
##  9 0                       -0.200
## 10 1                        0.748
## # ... with 2,089 more rows
```

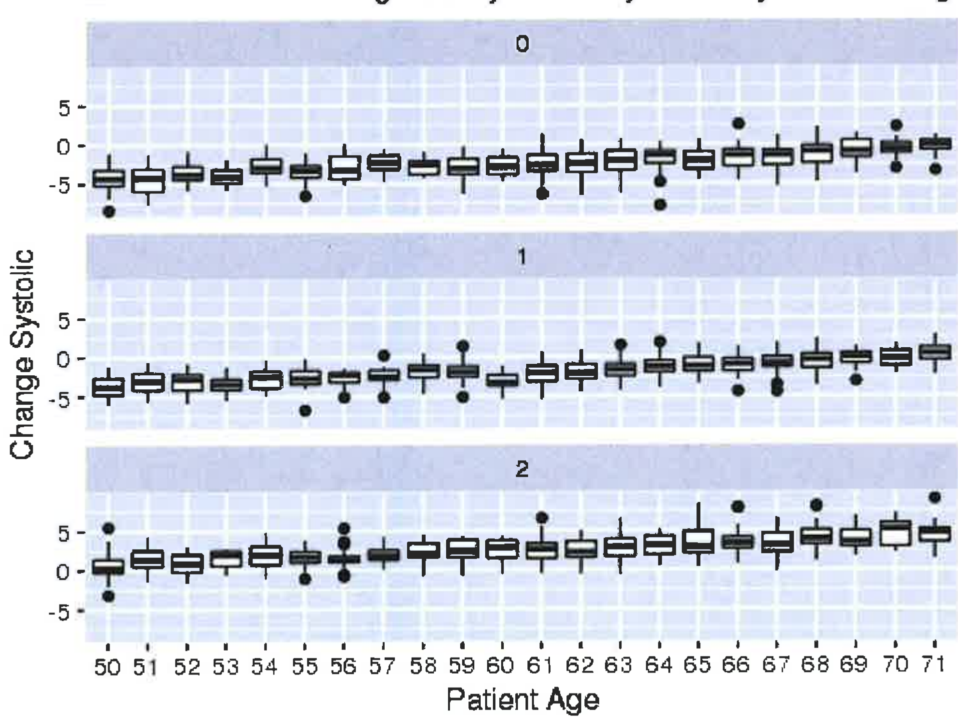

Box Plot for Change in Systolic by Patient Age
Change Systolic
5
0
-5
50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71
Patient Age


Box Plot for Change in Systolic by Treatment
Change Systolic
5
0
-5
0
1
2
Treatment

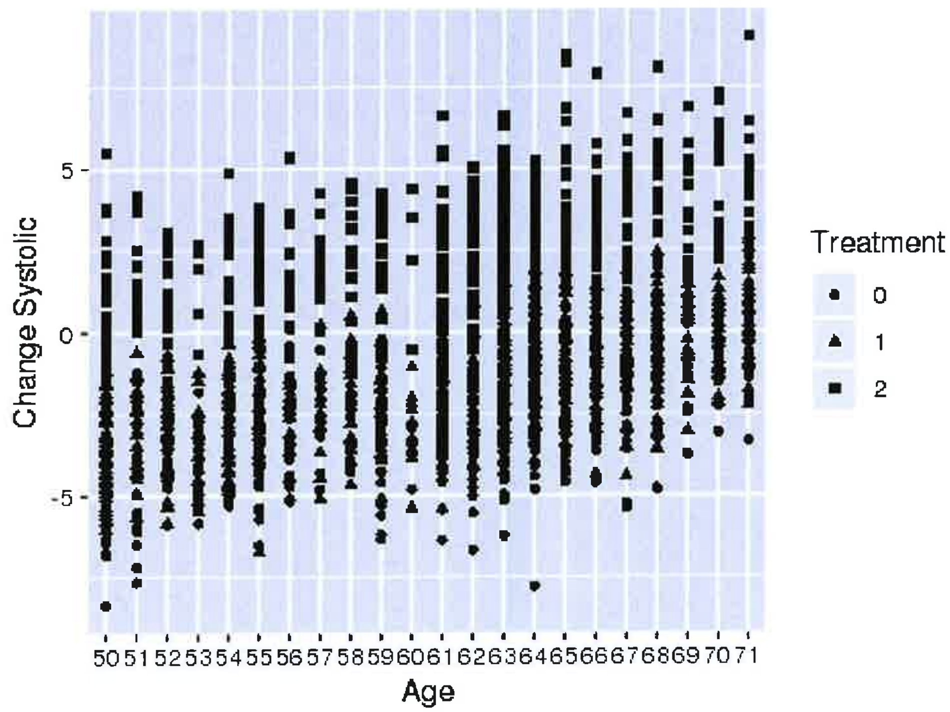
Box Plot for Change in Systolic by Hospital



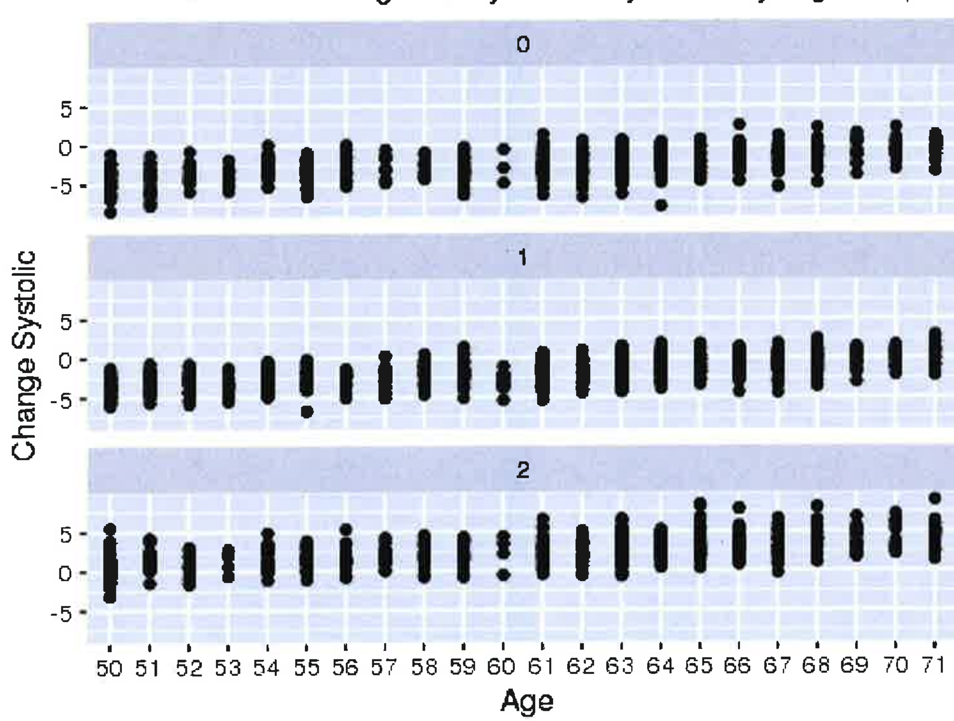
Box Plot for Change in Systolic Systolic by Patient Age

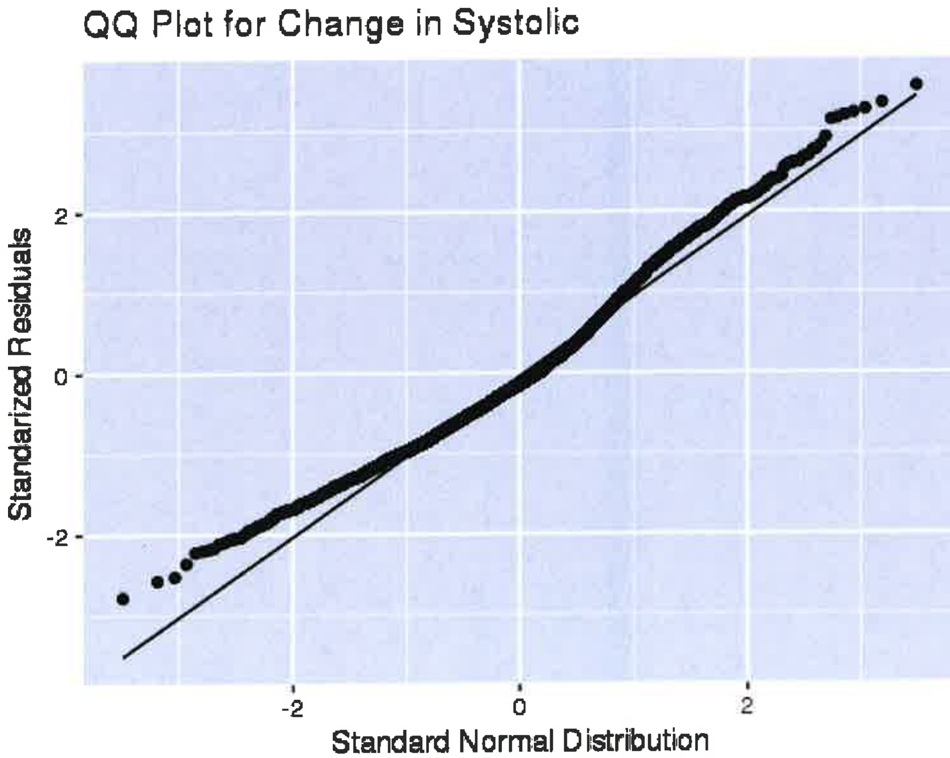


Scatter Plot for Change in Systolic Systolic by Age & Tı



Box Plot for Change in Systolic Systolic by Age Separa



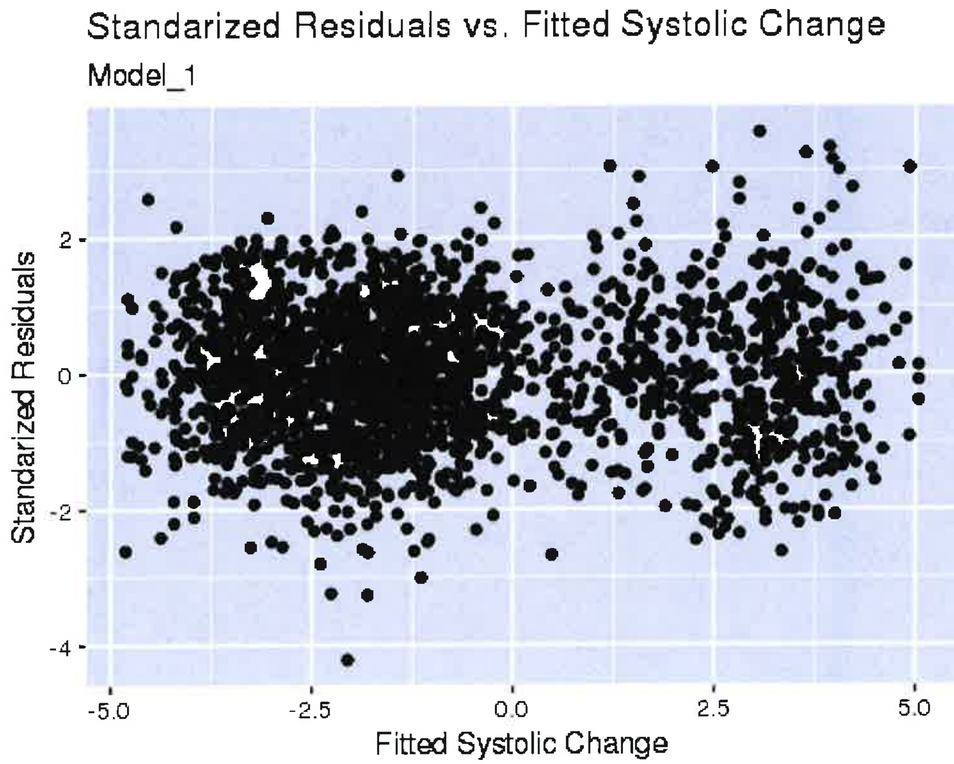

QQ Plot for Change in Systolic
Standarized Residuals
2
0
-2
-2
0
2
Standard Normal Distribution

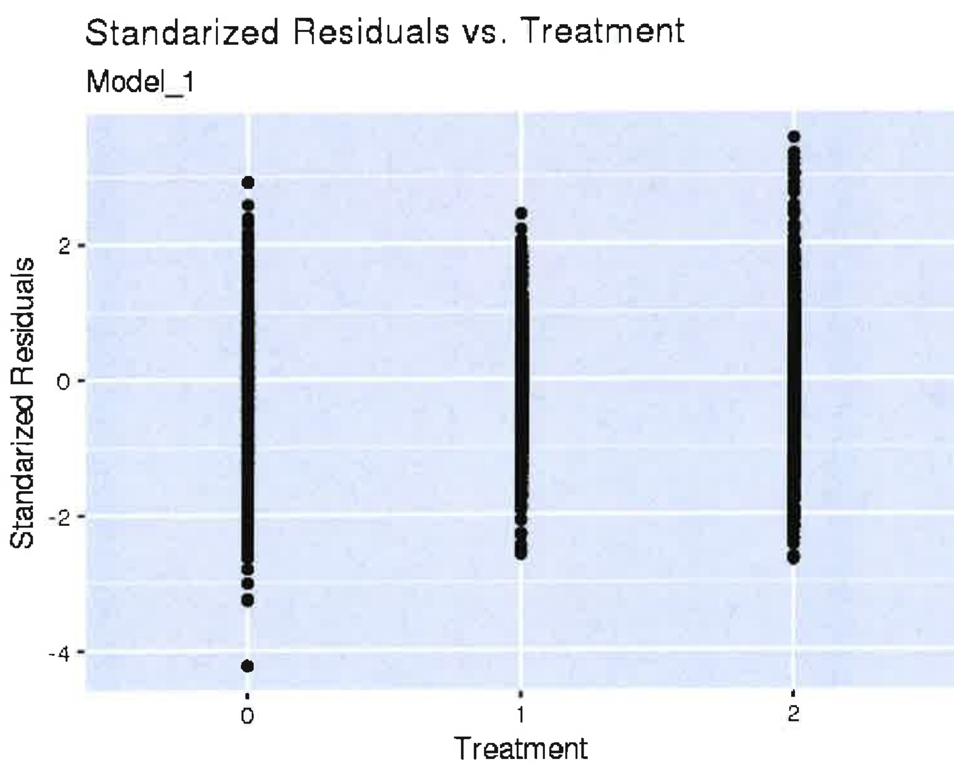
C. Results of Fitting Models

Create Model 1

Fit model via REML and assume that change in systolic is a function of two main effects: patient age and treatment. Include hospital as a random effect. Assume constant variance.

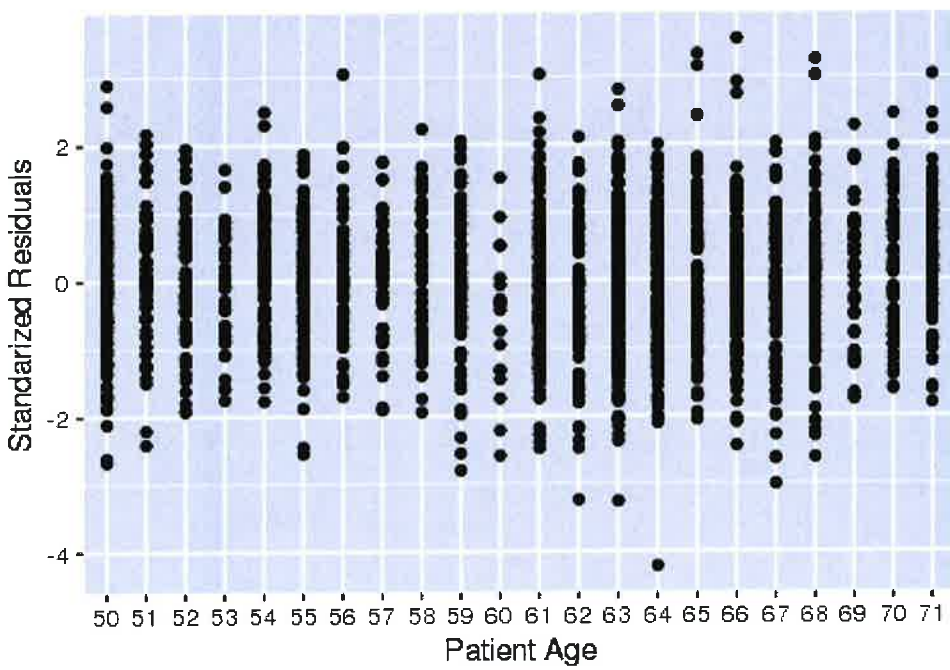
```
## Linear mixed-effects model fit by REML
##  Data: systolic_g
##        AIC      BIC    logLik
##   7327.429 7361.313 -3657.715
##
## Random effects:
##  Formula: ~1 | hospital
##         (Intercept) Residual
## StdDev:   0.4262212 1.361067
##
## Fixed effects: change_systolic ~ Treatment + Age
##                 Value Std.Error   DF   t-value p-value
## (Intercept) -12.487130 0.3181821 2075 -39.24523       0
## Treatment1    0.654929 0.0709604 2075   9.22950       0
## Treatment2    5.024052 0.0784480 2075  64.04309       0
## Age           0.165733 0.0049154 2075  33.71726       0
##  Correlation:
##             (Intr) Trtmn1 Trtmn2
## Treatment1 -0.137
## Treatment2 -0.100  0.525
## Age        -0.941  0.008 -0.018
##
## Standardized Within-Group Residuals:
##          Min           Q1          Med           Q3          Max
## -4.216974684 -0.669620174  0.004961872  0.693805437  3.558259774
##
## Number of Observations: 2099
## Number of Groups: 21

##             numDF denDF   F-value p-value
## (Intercept)     1  2075   55.3061  <.0001
## Treatment       2  2075 2521.6816  <.0001
## Age             1  2075 1136.8534  <.0001
```


Standarized Residuals vs. Fitted Systolic Change
Model_1
Standarized Residuals
2
0
-2
-4
-5.0
-2.5
0.0
2.5
5.0
Fitted Systolic Change


Standarized Residuals vs. Treatment
Model_1
Standarized Residuals
2
0
-2
-4
0
1
2
Treatment

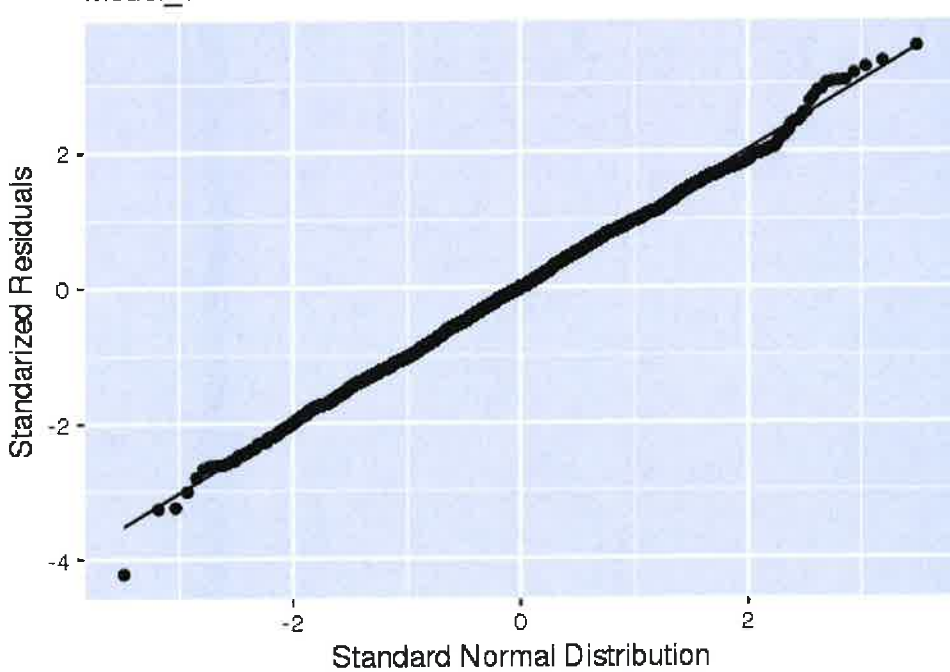
Standarized Residuals vs. Patient Age

Model_1



QQ Plot for Standarized Residuals

Model_1

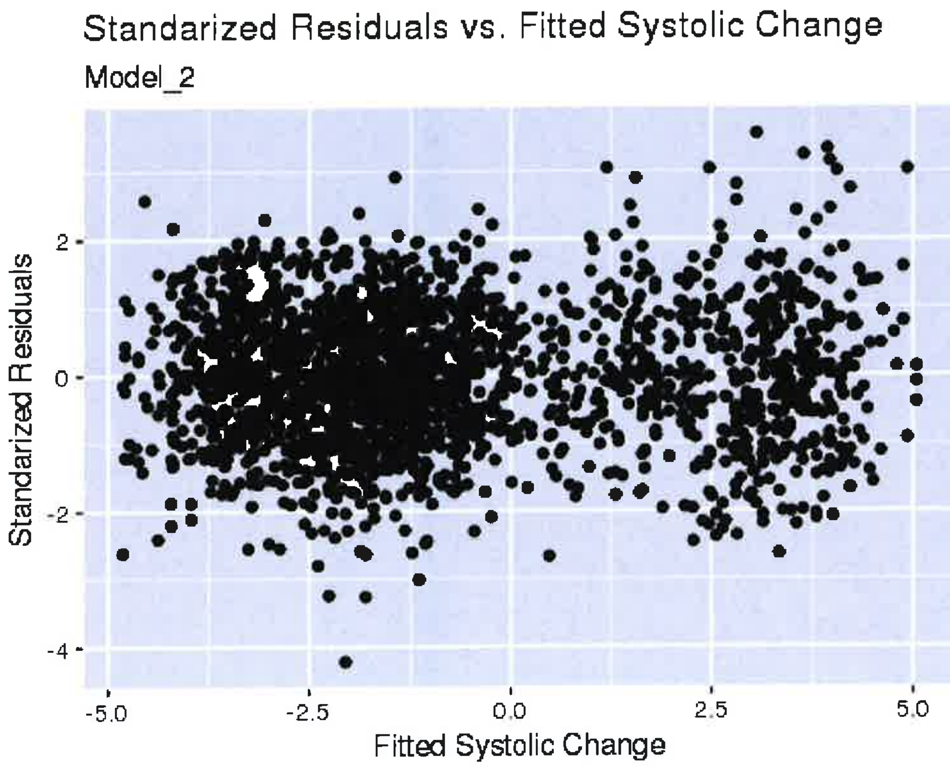


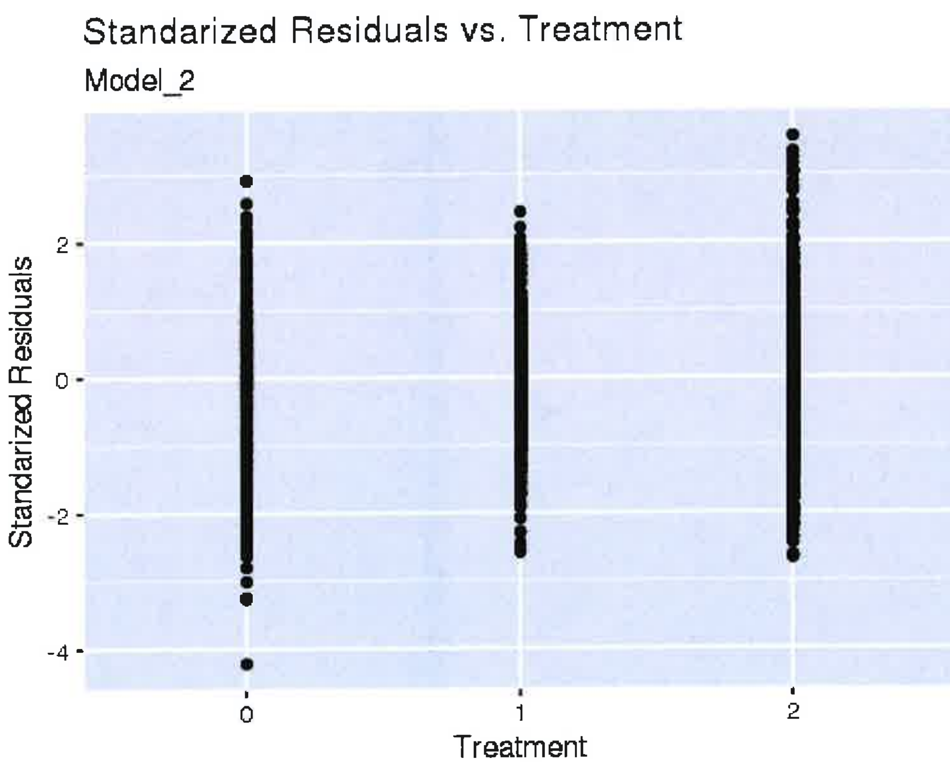
Create Model 2

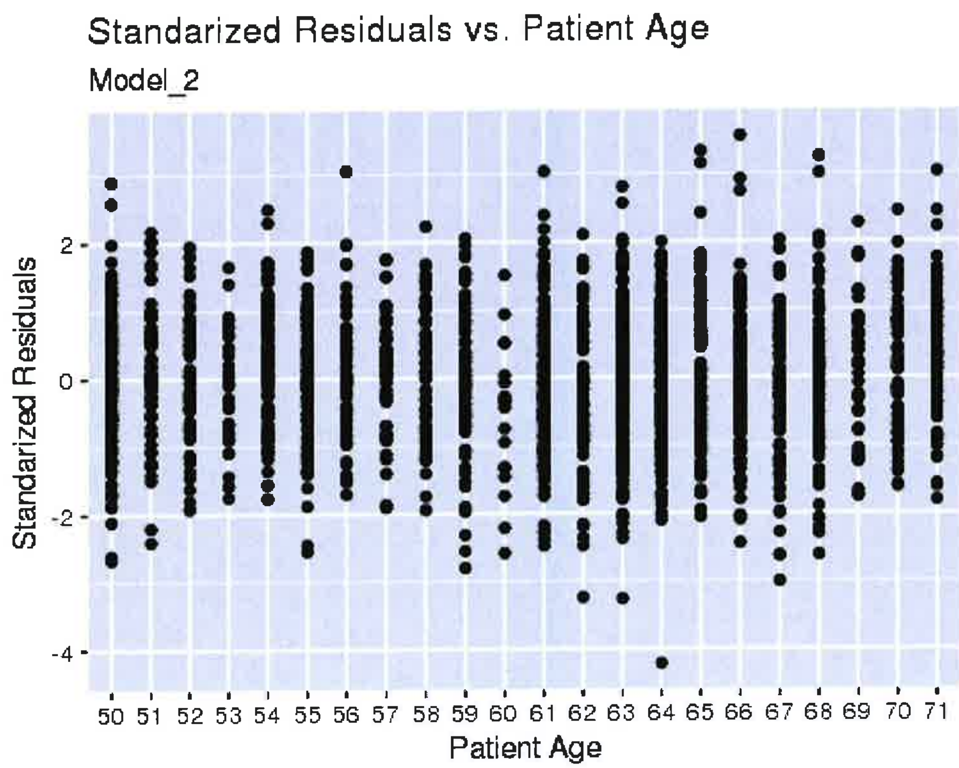
Fit model via restricted maximum likelihood and assume that change in systolic is a function of two main effects: patient age and treatment. Include hospital as a random effect. Assume constant variance.

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: systolic_g
##        AIC     BIC    logLik
##   7308.765 7342.66 -3648.382
##
## Random effects:
##  Formula: ~1 | hospital
##         (Intercept) Residual
## StdDev:   0.4148813 1.360086
##
## Fixed effects: change_systolic ~ Treatment + Age
##                  Value Std.Error   DF   t-value p-value
## (Intercept) -12.488019 0.3175521 2075 -39.32589       0
## Treatment1    0.654856 0.0709757 2075   9.22648       0
## Treatment2    5.023998 0.0784647 2075  64.02879       0
## Age           0.165748 0.0049164 2075  33.71350       0
##  Correlation:
##            (Intr) Trtmn1 Trtmn2
## Treatment1 -0.137
## Treatment2 -0.100  0.525
## Age        -0.943  0.008 -0.018
##
## Standardized Within-Group Residuals:
##          Min           Q1          Med           Q3          Max
## -4.220719085 -0.669274492  0.004643611  0.695385460  3.559238300
##
## Number of Observations: 2099
## Number of Groups: 21
```

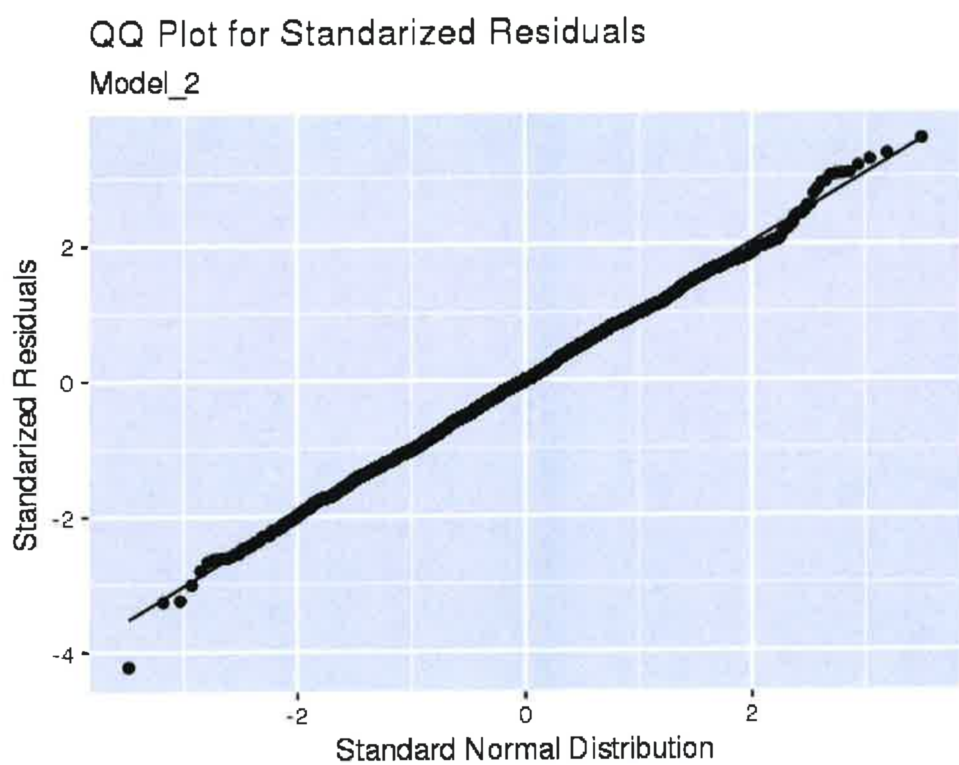
```
##             numDF denDF   F-value p-value
## (Intercept)     1  2075   57.9706  <.0001
## Treatment       2  2075 2520.5690  <.0001
## Age             1  2075 1136.6000  <.0001
```


Standarized Residuals vs. Fitted Systolic Change
Model_2
Standarized Residuals
2
0
-2
-4
-5.0
-2.5
0.0
2.5
5.0
Fitted Systolic Change


Standarized Residuals vs. Treatment
Model_2
Standarized Residuals
2
0
-2
-4
0
1
2
Treatment



Standarized Residuals vs. Patient Age
Model_2
Standarized Residuals
2
0
-2
-4
50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71
Patient Age



QQ Plot for Standarized Residuals
Model_2
Standarized Residuals
2
0
-2
-4
-2
0
2
Standard Normal Distribution

Create Model 3

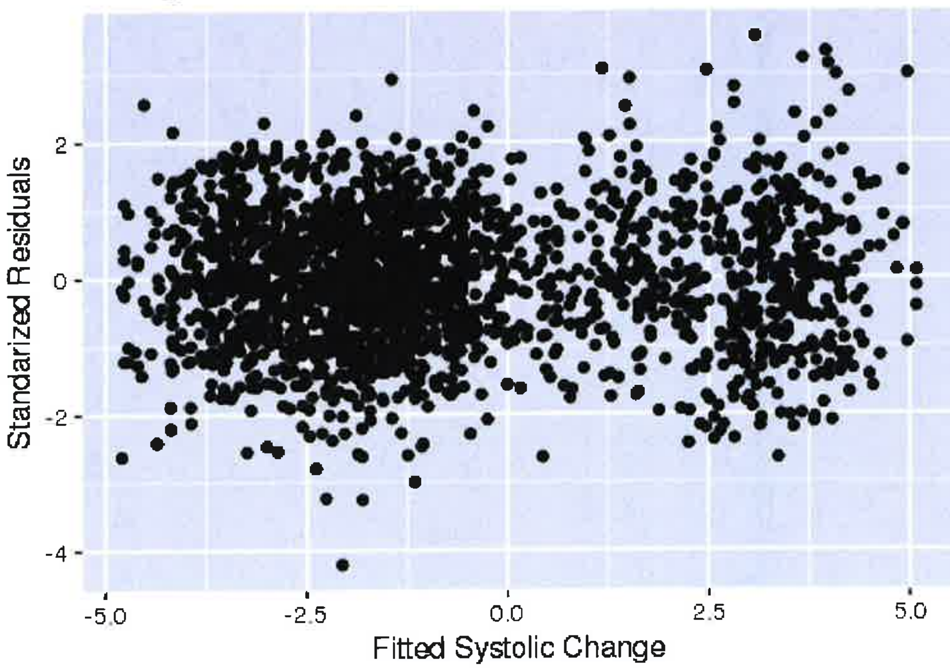
Fit model via maximum likelihood and assume that change in systolic is a function of two main effects and include an interaction effect: patient age and treatment. Include hospital as a random effect. Assume constant variance.

```
## Linear mixed-effects model fit by REML
##  Data: systolic_g
##        AIC      BIC   logLik
##   7345.421 7390.592 -3664.71
##
## Random effects:
##  Formula: ~1 | hospital
##         (Intercept) Residual
## StdDev:   0.4261245  1.36164
##
## Fixed effects: change_systolic ~ Treatment + Age + Treatment * Age
##                     Value Std.Error   DF    t-value p-value
## (Intercept)    -12.383800 0.5482496 2073 -22.587888  0.0000
## Treatment1       0.638400 0.7128276 2073   0.895588  0.3706
## Treatment2       4.672223 0.7881578 2073   5.928030  0.0000
## Age              0.164038 0.0088235 2073  18.591054  0.0000
## Treatment1:Age   0.000267 0.0116514 2073   0.022920  0.9817
## Treatment2:Age   0.005757 0.0128419 2073   0.448268  0.6540
##  Correlation:
##                (Intr) Trtmn1 Trtmn2 Age    Trt1:A
## Treatment1     -0.746
## Treatment2     -0.674  0.518
## Age            -0.981  0.753  0.680
## Treatment1:Age  0.742 -0.995 -0.514 -0.756
## Treatment2:Age  0.672 -0.516 -0.995 -0.685  0.518
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -4.21106656 -0.66702337  0.00195741  0.69141317  3.54297491
##
## Number of Observations: 2099
## Number of Groups: 21
```

```
##               numDF denDF   F-value p-value
## (Intercept)       1  2073   55.3245  <.0001
## Treatment         2  2073 2519.5614  <.0001
## Age               1  2073 1135.9036  <.0001
## Treatment:Age     2  2073    0.1305  0.8777
```

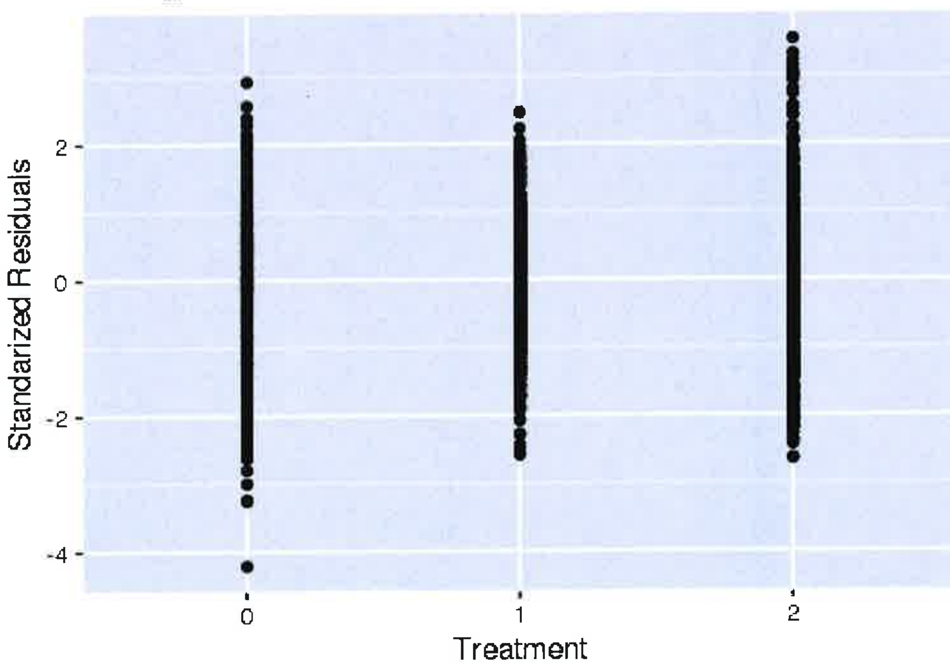
Standarized Residuals vs. Fitted Systolic Change

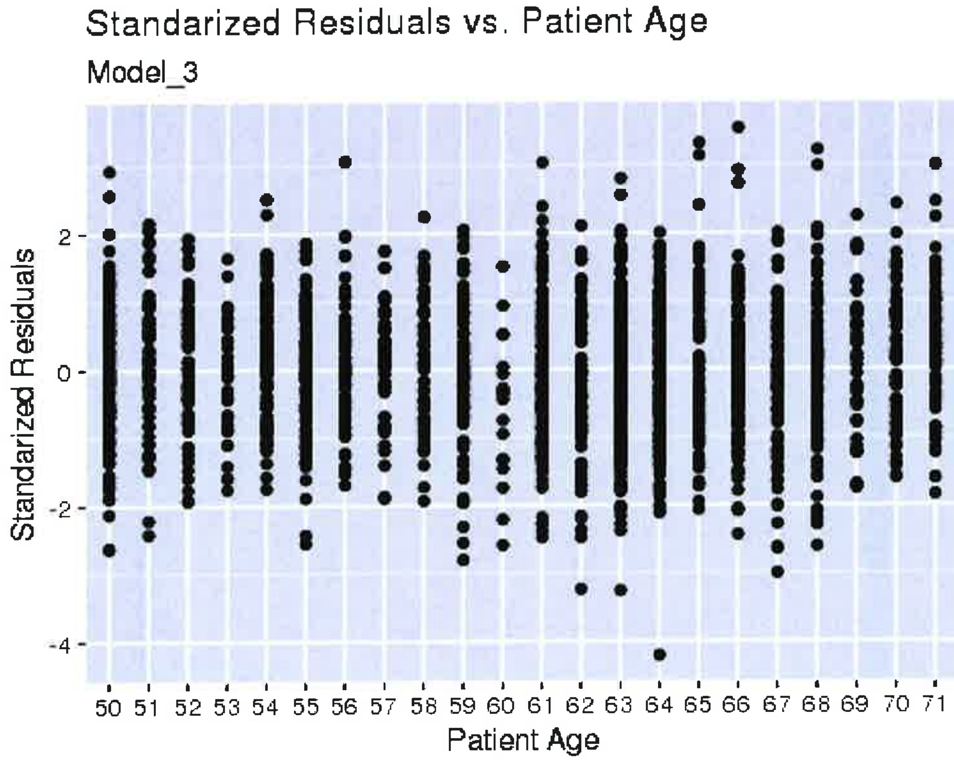
Model_3

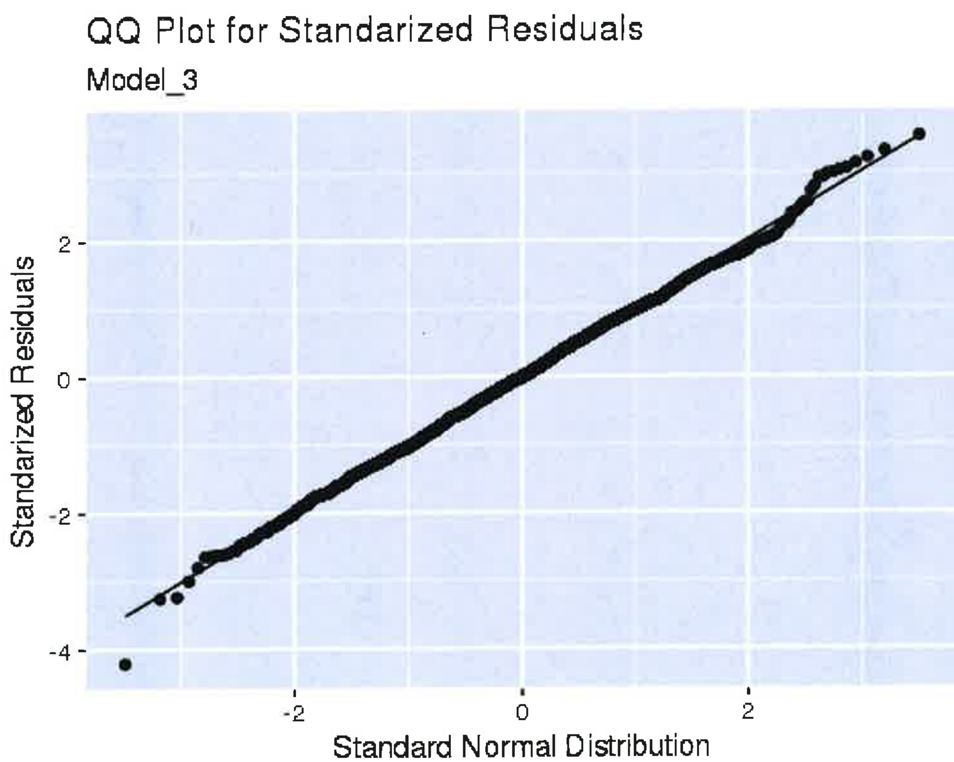


Standarized Residuals vs. Treatment

Model_3



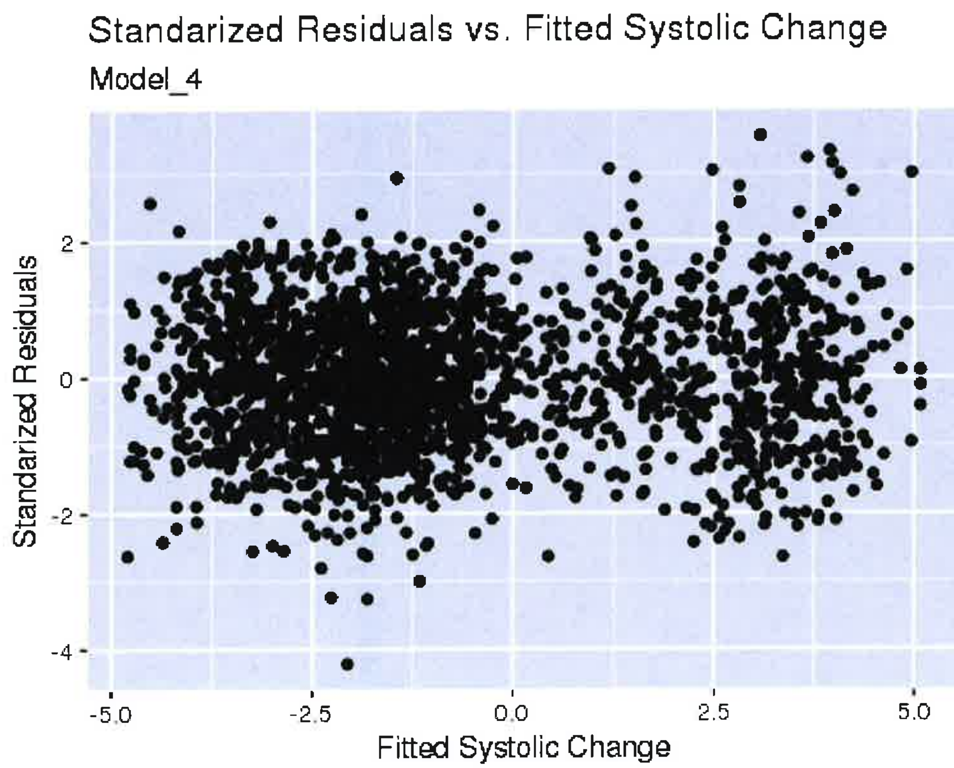

Standarized Residuals vs. Patient Age
Model_3
Standarized Residuals
2
0
-2
-4
50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71
Patient Age

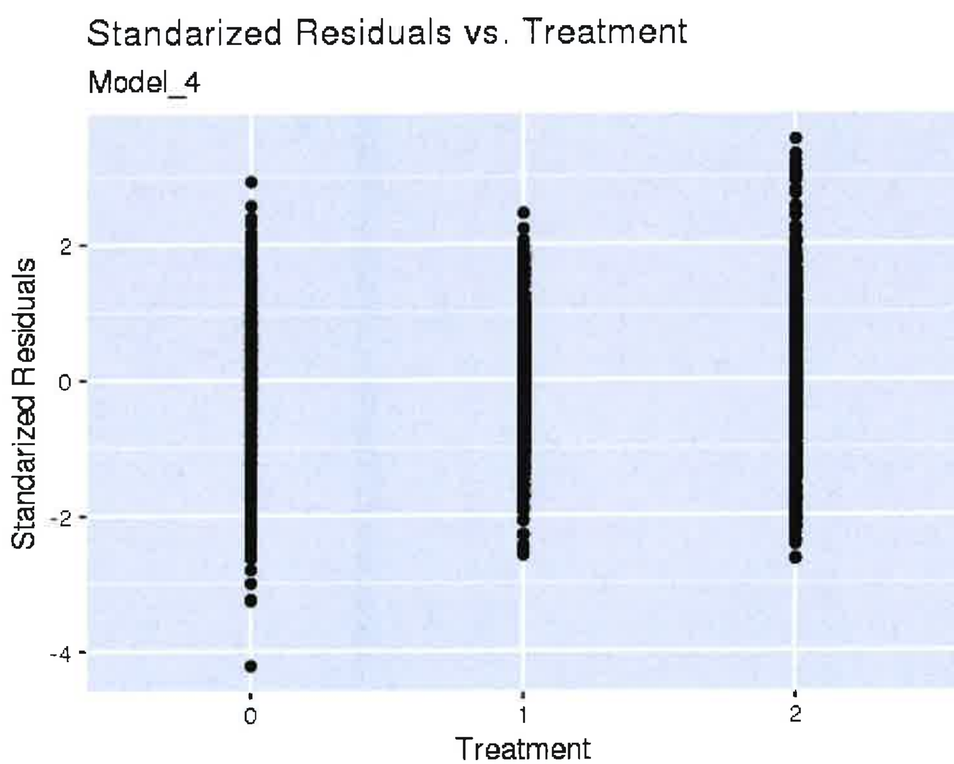

QQ Plot for Standarized Residuals
Model_3
Standarized Residuals
2
0
-2
-4
-2
0
2
Standard Normal Distribution

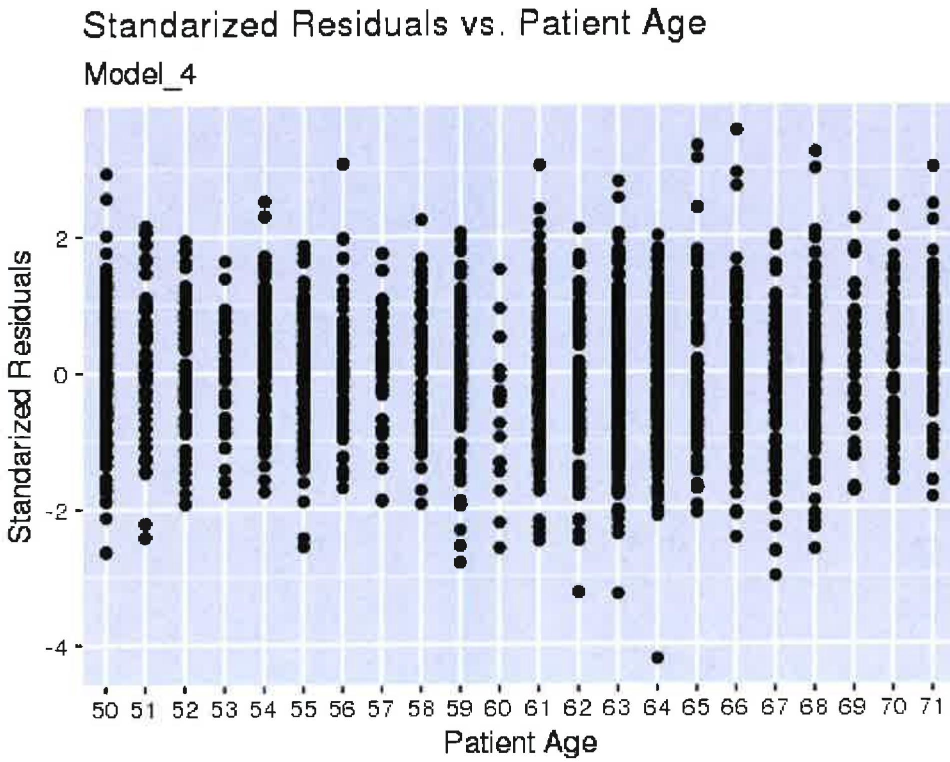
Create Model 4

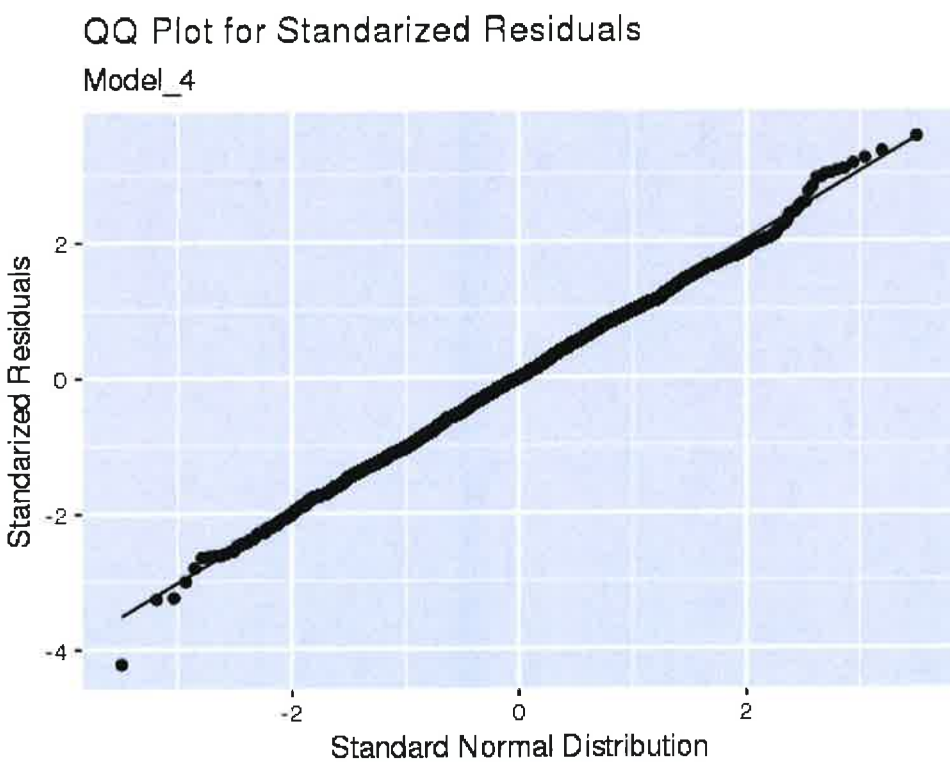
Fit model via maximum likelihood and assume that change in systolic is a function of two main effects: patient age and treatment. Include hospital as a random effect. Assume constant variance.

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: systolic_g
##        AIC      BIC    logLik
##   7312.503 7357.697 -3648.252
##
## Random effects:
##  Formula: ~1 | hospital
##         (Intercept) Residual
## StdDev:   0.4147901 1.360003
##
## Fixed effects: change_systolic ~ Treatment + Age + Treatment * Age
##                     Value Std.Error   DF    t-value p-value
## (Intercept)    -12.385084 0.5479691 2073 -22.601792  0.0000
## Treatment1       0.639184 0.7129802 2073   0.896496  0.3701
## Treatment2       4.672374 0.7883255 2073   5.926961  0.0000
## Age              0.164059 0.0088253 2073  18.589618  0.0000
## Treatment1:Age   0.000253 0.0116539 2073   0.021708  0.9827
## Treatment2:Age   0.005753 0.0128446 2073   0.447909  0.6543
##  Correlation:
##                (Intr) Trtmn1 Trtmn2 Age    Trt1:A
## Treatment1     -0.747
## Treatment2     -0.674  0.518
## Age            -0.981  0.753  0.680
## Treatment1:Age  0.742 -0.995 -0.514 -0.756
## Treatment2:Age  0.672 -0.516 -0.995 -0.685  0.518
##
## Standardized Within-Group Residuals:
##          Min           Q1          Med           Q3          Max
## -4.216841053 -0.668376991  0.001482204  0.692266958  3.545667320
##
## Number of Observations: 2099
## Number of Groups: 21

##                numDF denDF   F-value p-value
## (Intercept)        1  2073   57.9389  <.0001
## Treatment          2  2073 2518.4700  <.0001
## Age                1  2073 1135.6550  <.0001
## Treatment:Age      2  2073    0.1306  0.8776
```


Standarized Residuals vs. Fitted Systolic Change
Model_4
Standarized Residuals
2
0
-2
-4
-5.0
-2.5
0.0
2.5
5.0
Fitted Systolic Change


Standarized Residuals vs. Treatment
Model_4
Standarized Residuals
2
0
-2
-4
0
1
2
Treatment


Standarized Residuals vs. Patient Age
Model_4
Standarized Residuals
2
0
-2
-4
50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71
Patient Age


QQ Plot for Standarized Residuals
Model_4
Standarized Residuals
2
0
-2
-4
-2
0
2
Standard Normal Distribution

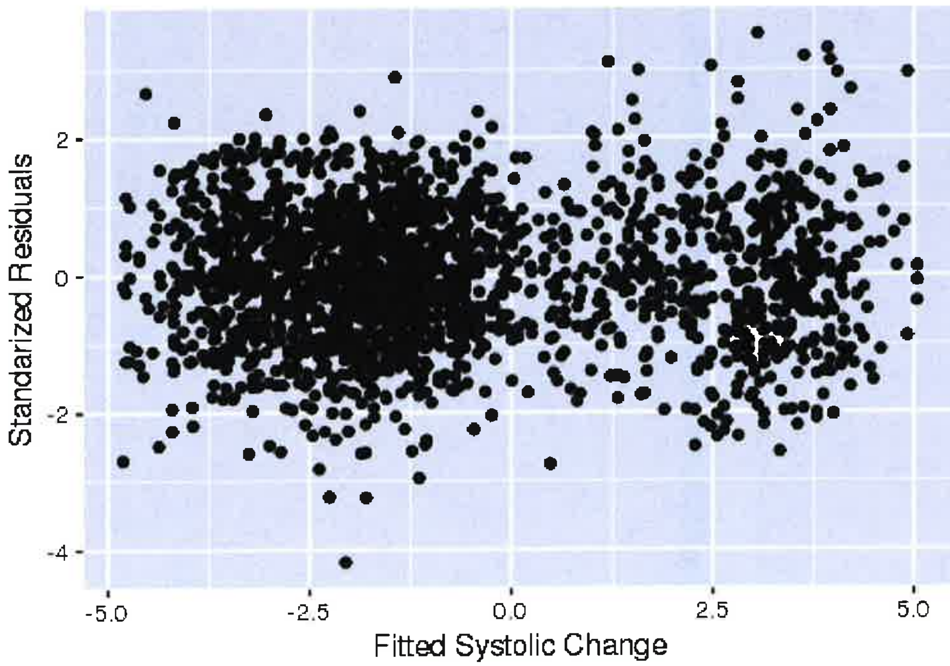
Create Model 5

Fit model via restricted maximum likelihood and assume that change in systolic is a function of two main effects: patient age and treatment. Include hospital as a random effect. Assume variance is a function of patient age.

```
## Linear mixed-effects model fit by REML
##  Data: systolic_g
##        AIC      BIC    logLik
##   7328.156 7367.687 -3657.078
##
## Random effects:
##  Formula: ~1 | hospital
##         (Intercept) Residual
## StdDev:    0.427332 1.139657
##
## Combination of variance functions:
##  Structure: Exponential of variance covariate
##  Formula: ~Age/2
##  Parameter estimates:
##      expon
## 0.00581435
## Fixed effects: change_systolic ~ Treatment + Age
##                  Value  Std.Error   DF   t-value p-value
## (Intercept) -12.453354 0.31578612 2075 -39.43604       0
## Treatment1    0.654950 0.07089656 2075   9.23811       0
## Treatment2    5.023061 0.07842213 2075  64.05157       0
## Age           0.165182 0.00488987 2075  33.78050       0
##  Correlation:
##            (Intr) Trtmn1 Trtmn2
## Treatment1 -0.137
## Treatment2 -0.101  0.525
## Age        -0.940  0.007 -0.018
##
## Standardized Within-Group Residuals:
##          Min           Q1          Med           Q3          Max
## -4.179745856 -0.679512321  0.002767709  0.698735726  3.512537373
##
## Number of Observations: 2099
## Number of Groups: 21

##             numDF denDF   F-value p-value
## (Intercept)     1  2075   60.8281  <.0001
## Treatment       2  2075 2521.8727  <.0001
## Age             1  2075 1141.1222  <.0001
```

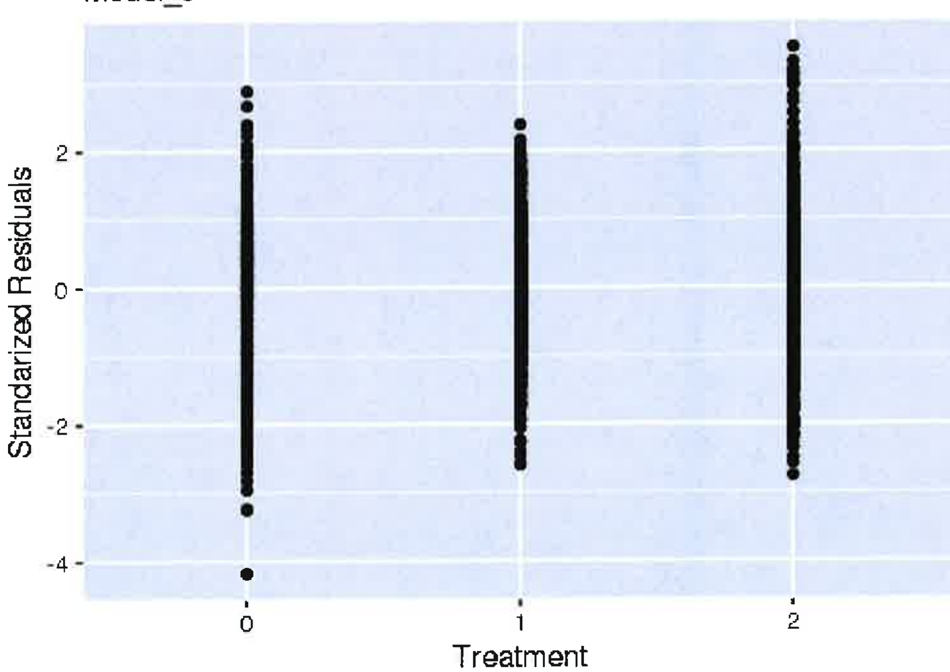
Standarized Residuals vs. Fitted Systolic Change

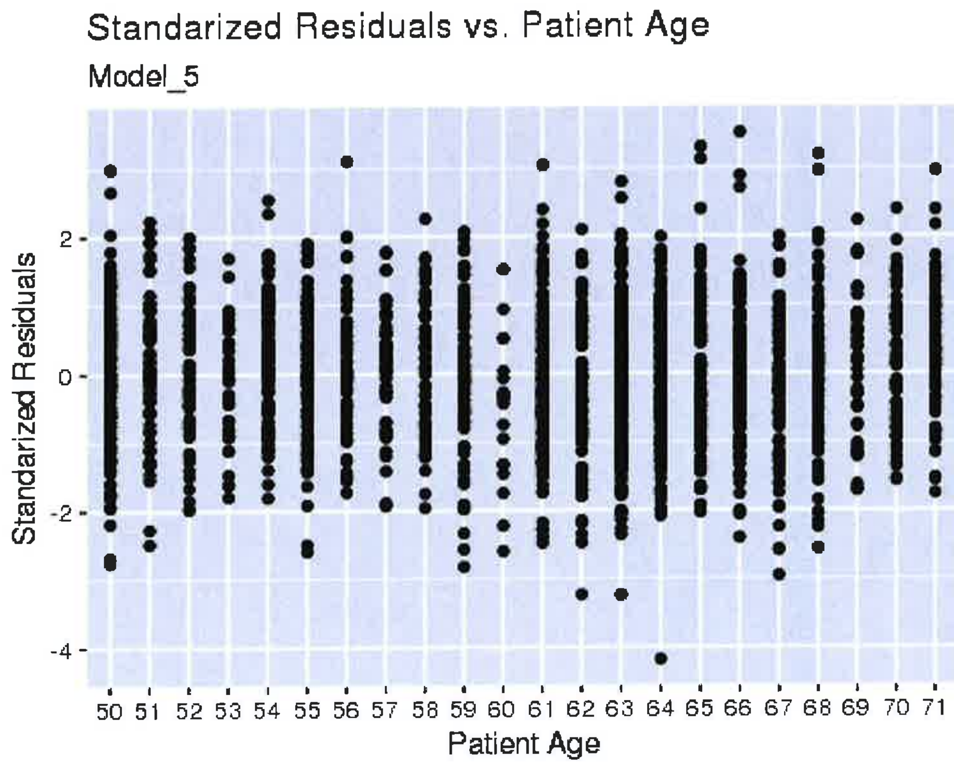
Model_5

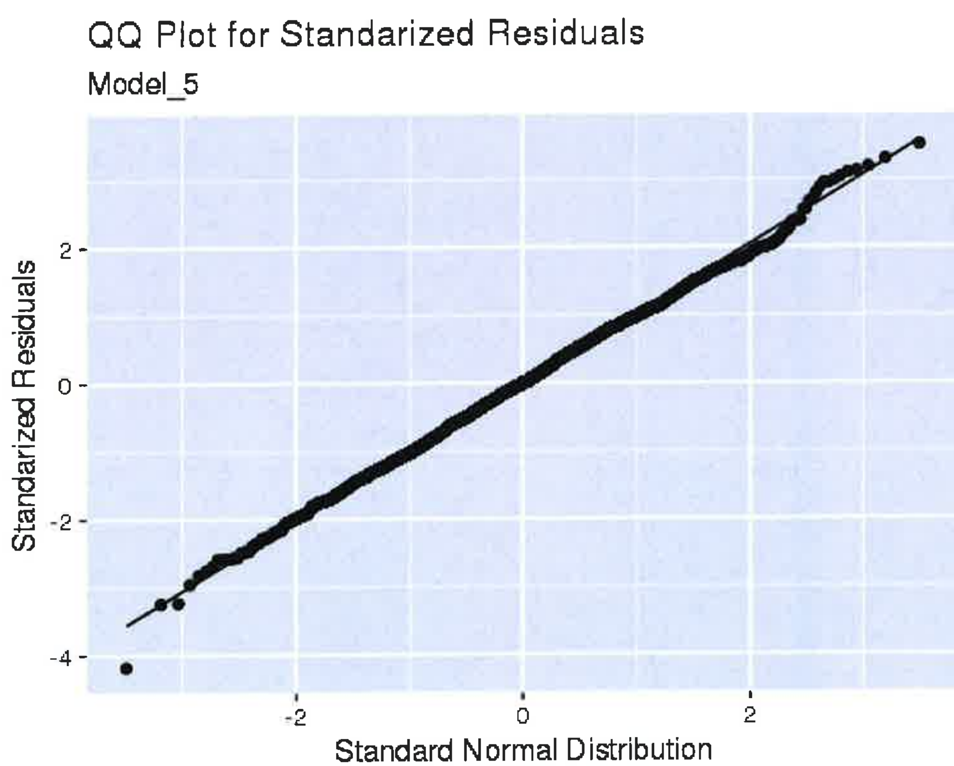


Standarized Residuals vs. Treatment

Model_5




Standarized Residuals vs. Patient Age
Model_5
Standarized Residuals
2
0
-2
-4
50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71
Patient Age


QQ Plot for Standarized Residuals
Model_5
Standarized Residuals
2
0
-2
-4
-2
0
2
Standard Normal Distribution

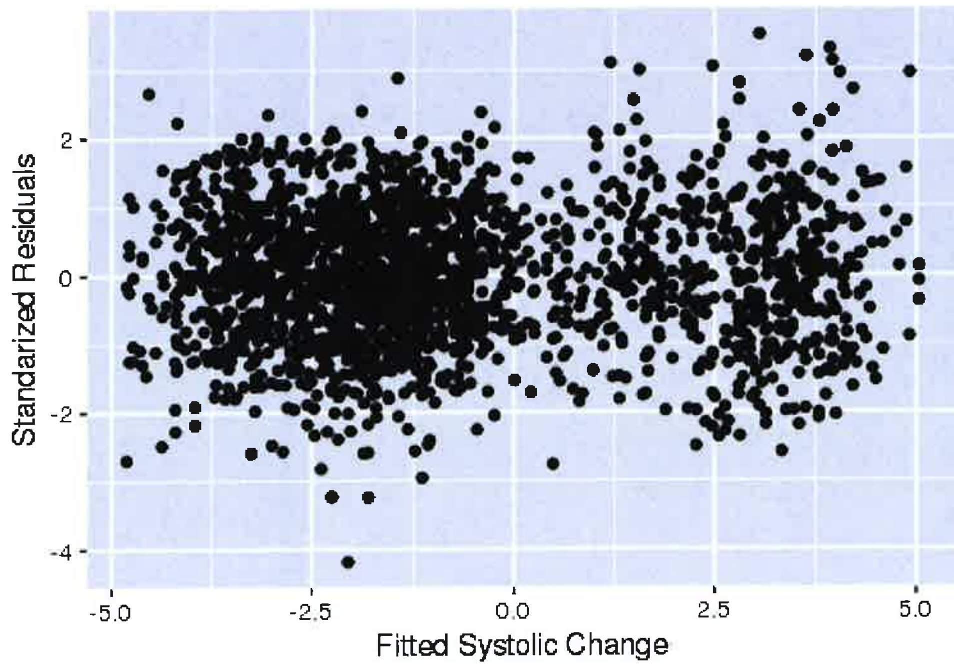
Create Model 6

Fit model via maximum likelihood and assume that change in systolic is a function of two main effects: patient age and treatment. Include hospital as a random effect. Assume variance is a function of patient age

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: systolic_g
##        AIC      BIC    logLik
##   7309.484 7349.028 -3647.742
##
## Random effects:
##  Formula: ~1 | hospital
##         (Intercept) Residual
## StdDev:   0.4159726 1.138249
##
## Combination of variance functions:
##  Structure: Exponential of variance covariate
##  Formula: ~Age/2
##  Parameter estimates:
##       expon
## 0.005831218
## Fixed effects: change_systolic ~ Treatment + Age
##                  Value  Std.Error   DF   t-value p-value
## (Intercept) -12.454165 0.31514020 2075 -39.51944       0
## Treatment1    0.654879 0.07091161 2075   9.23514       0
## Treatment2    5.023004 0.07843874 2075  64.03728       0
## Age           0.165196 0.00489079 2075  33.77698       0
##  Correlation:
##            (Intr) Trtmn1 Trtmn2
## Treatment1 -0.137
## Treatment2 -0.101  0.525
## Age        -0.942  0.007 -0.018
##
## Standardized Within-Group Residuals:
##          Min           Q1          Med           Q3          Max
## -4.183346230 -0.680588398  0.003433337  0.698196584  3.513372530
##
## Number of Observations: 2099
## Number of Groups: 21

##             numDF denDF   F-value p-value
## (Intercept)     1  2075   63.7766  <.0001
## Treatment       2  2075 2520.7572  <.0001
## Age             1  2075 1140.8843  <.0001
```

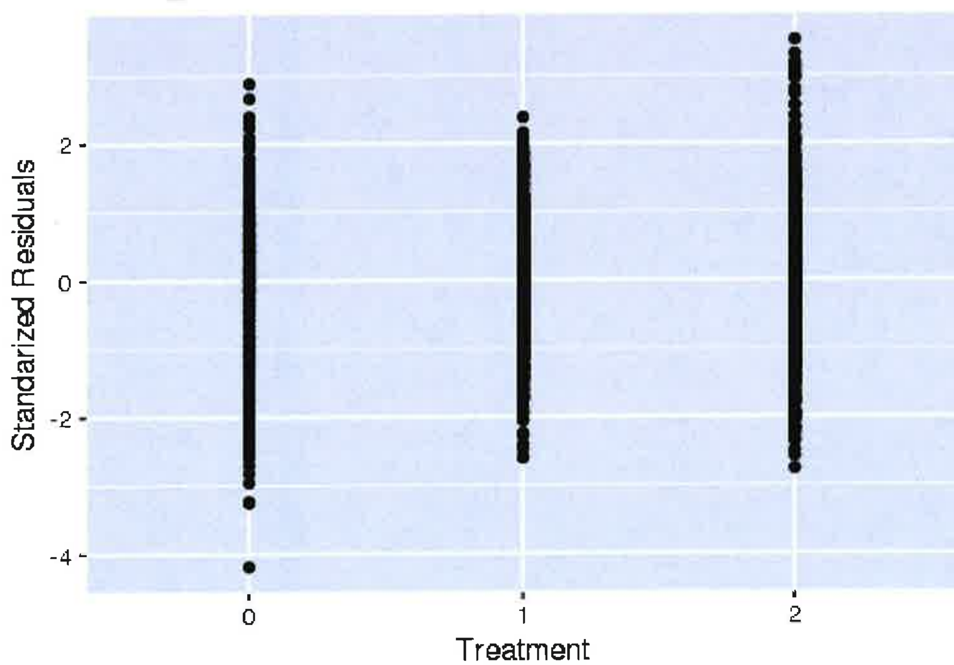
Standarized Residuals vs. Fitted Systolic Change

Model_6



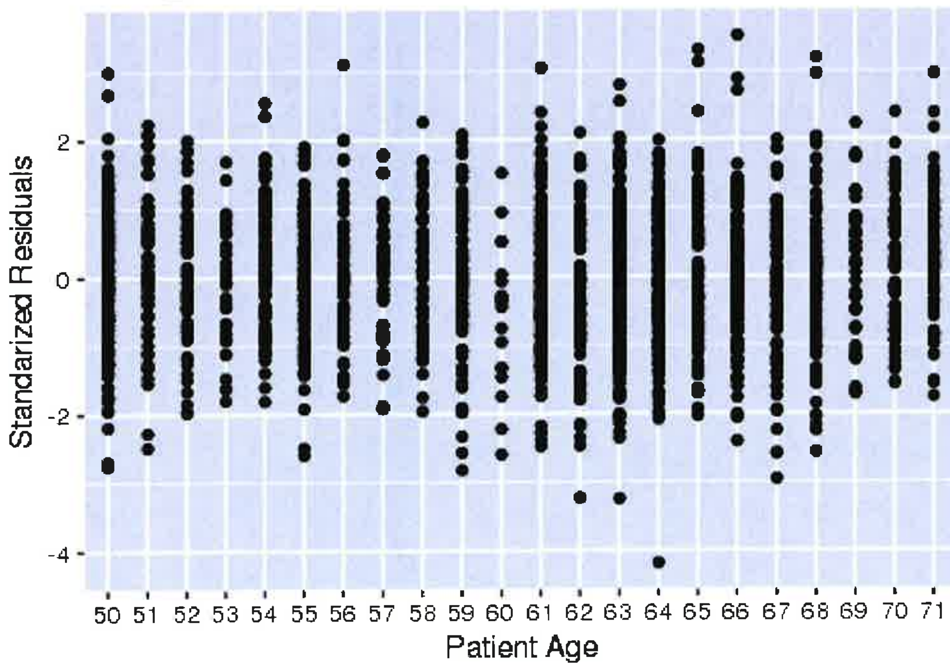
Standarized Residuals vs. Treatment

Model_6



Standarized Residuals vs. Patient Age

Model_6



QQ Plot for Standarized Residuals

Model_6