STAT 437 - Assignment 2
Due: Friday, February 11 on Crowdmark
Resubmit: Friday, March 4 on Crowdmark

This assignment covers the materials contained in **Lecture 010** through to **Lecture 020**. Reminder that you are permitted to discuss to these problems with classmates, but every student must submit their own solutions which are their own work (including any code, figures, etc.). **Please indicate any students that you discussed solutions with on your submission**. Please ensure that your submissions on Crowdmark are legible, and separated based on the problems included at the submission link. Submissions can be handwritten or typeset.

Part 1: True or False (20 Marks)

For each of the following problems indicate whether the statement is true or false, and give a short justification for your answer. Correct answers without justification will receive only partial credit.

PROBLEM 1. Suppose that a linear mixed effects model is specified for a continuous outcome, Y_{ij}, such that

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 A_i + b_{0,i} + b_{1,i}A_i + \epsilon_{ij}.$$

Here, t_{ij} represents the time, treated as a continuous variable and $A_i \in \{0, 1\}$ is a binary treatment indicator. Assume that $G_i = \sigma^2 I$. **True or false: The within-person correlation structure can be written as $A_i\mathbf{R}(\rho_1)+(1-A_i)\mathbf{R}(\rho_2)$, where $\mathbf{R}(\rho)$ is a correlation matrix which assumes compound symmetry.**

Solution 1: [1] True. [1] Consider that $\text{cov}(Y_{ij}, Y_{i\ell}) = \text{var}(b_{0,i}) + A_i \text{var}(b_{1,i}) + 2A_i \text{cov}(b_{0,i}, b_{1,i}) + I(j = \ell)\sigma^2$. As a result, the correlation will be given by

$$\frac{\sigma_0^2 + A_i\sigma_1^2 + 2A_i\sigma_{01}}{\sigma_0^2 + A_i\sigma_1^2 + 2A_i\sigma_{01} + \sigma^2} = A_i\frac{\sigma_0^2 + \sigma_1^2 + 2\sigma_{01}}{\sigma_0^2 + \sigma_1^2 + 2\sigma_{01} + \sigma^2} + (1 - A_i)\frac{\sigma_0^2}{\sigma_0^2 + \sigma^2} = A_i\rho_1 + (1 - A_i)\rho_2.$$

PROBLEM 2. Suppose that we have observed data $\{W_i, Z_i\}$ for an independent sample, $i = 1, \ldots, n$ where W_i is a binary indicator and Z_i is a discrete random variable, with values $\{-1, 0, 1\}$. Consider the M-estimator given by

$$U(\theta) = \sum_{i=1}^{n} \begin{pmatrix} I(Z_i = -1)(W_i - \theta_1) \\ I(Z_i = 0)(W_i - \theta_2) \\ 1 - \theta_1 - \theta_2 - \theta_3 \end{pmatrix}.$$

Denote $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\theta}_3)$ as the solution to $U(\widehat{\theta}) = 0$. **True or false: $\widehat{\theta}_3$ is consistent for $P(W_i = 1|Z_i = 1)$.**

Solution 2: [1] False. [1] Applying iterated conditioning on Z_i, we can show that the first component is unbiased when $\theta_1 = P(W_i = 1|Z_i = -1)$, the second component when $\theta_2 = P(W_i = 1|Z_i = 0)$, which means the third component will be unbiased when $\theta_3 = 1 - \theta_1 - \theta_2 = 1 - P(W_i = 1|Z_i = -1) - P(W_i = 1|Z_i = 0)$. This is not (in general) going to be $P(W_i = 1|Z_i = 1)$.

PROBLEM 3. You have a friend who is not a particularly strong calculus student. In attempting to implement GEE for a marginal model, they **incorrectly** find that

$$D_i = \frac{\partial}{\partial \beta} g^{-1}(X_i\beta) = \mathbf{1},$$

(the correct sized matrix of all 1's). They also make the **incorrect** assumption of working independence ($V_i = \sigma^2 I$). Despite this, they have correctly specified the mean model and link function. **True or false: the estimators produced using the GEE under their specification will be consistent for the true β.**

Solution 3: [1] True. [1] Under the assumption that $E[Y_i|X_i] = g^{-1}(X_i\beta)$ for the specified g and $X_i\beta$, GEE will produce consistent estimation. This can be argued either by appealing to M-estimation theory (e.g., the estimating equations remain unbiased in spite of the misspecification) or to the GEE theory which states consistency is attained through correct mean specification.

PROBLEM 4. Consider a hypothetical dataset that contains information on the following variables:

- **Smoking Status**: A binary indicator, Y_{ij}, measured for each individual i at each time point j.
- **Age**: A continuous variate denoting time, t_{ij}, recorded for each individual i at each time point j.
- **Baseline Income**: A continuous variate, Inc_i, recorded for each individual i at the baseline.
- **Employment Status**: A binary indicator, E_{ij}, measured for each individual i at each time point j.

Suppose a marginal model is fit to this data, which specifies a logistic link function, binomial variance pattern, and a linear predictor given by

$$\text{logit}(E[Y_{ij}|X_{ij}]) = \beta_0 + \beta_1 t_{ij} + \beta_2 \text{Inc}_i + \beta_3 E_{ij}.$$

True or false: This is a valid marginal model which can be fit using GEE.

Solution 4: [1] False. [1] The specification of the model would be fine except for the inclusions of E_{ij}. Employment status is a time-varying covariate, and one which varies in a way that is stochastic. As a result, it cannot be included in marginal models fit using GEE.

PROBLEM 5. A professional basketball team has hired a new data analyst to try to help guide decision making on the team. As a first project, the analyst fits a generalized linear marginal model to data on the league's players. In particular, the analyst fits a model which takes as the outcome the median number of points scored per game for each player, and controls for various factors (the player's height and weight, the team they play for, whether they were injured, and so forth). The model is fit with a longitudinal trend representing the player's age. The idea is to use this model to consider what the impact of aging is on the quality of play that players manage. **True or false: This model can be used to predict the impact of aging for a specific individual. (For instance, the model can be used to estimate the aging curve for team's star player.)**

Solution 5: [1] False. [1] A marginal model does not predict individual-level effects, and cannot be interpreted as such. The model *could* be used to estimate what the *average* effect of aging would be for players who are similar (in the measured variates) to the specific individual, but it could not be used to make conclusions regarding that individual player, directly.

PROBLEM 6. Suppose that a linear mixed effects model is fit, given by

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0,i} + b_{1,i}t_{ij} + \epsilon_{ij}.$$

True or false: We can interpret β_1 as the expected change (across the whole population) in the outcome for a unit increase in t_{ij}.

Solution 6: [1] True. [1] In a mixed effects model, the fixed effects (β_0 and β_1) have a population-average effect. In this sense, mixed effects models can be used either for marginal inference or individual inference.

PROBLEM 7. Suppose that the following marginal model is fit to data

$$E[Y_{ij}|X_{ij}] = \beta_0 + \beta_1\text{Age}_{ij} + \beta_2\text{Income}_{ij} + \beta_3 Z_i + \beta_4\text{Age}_{i1} + \beta_5\text{Income}_{i1}.$$

True or False: Testing the null hypothesis $H_0 : \beta_4 = \beta_5 = 0$ is equivalent to a test of the hypothesis that the longitudinal effects of age and income are equal to the cross-sectional effects of age and income, respectively.

Solution 7: [1] True. [1] Note that the longitudinal effects can be estimated based on $E[Y_{ij} - Y_{i1}] = \beta_1(\text{Age}_{ij} - \text{Age}_{i1}) + \beta_2(\text{Income}_{ij} - \text{Income}_{i1})$. The cross-sectional effects are based on $E[Y_{i1}] = \beta_0 + (\beta_1 + \beta_4)\text{Age}_{i1} + (\beta_2 + \beta_5)\text{Income}_{i1} + \beta_3 Z_i$. As a result, the longitudinal effects are given by β_1 and β_2, which the cross-sectional effects are given by $\beta_1 + \beta_4$ and $\beta_2 + \beta_5$. If $\beta_4 = \beta_5 = 0$, then these two effects are equal.

PROBLEM 8. A linear mixed effects model is fit which include 5 $(b_0, b_1, b_2, b_3, b_4)$ random effects terms. An analyst wants to test whether 2 of those random effects can be dropped, and so they fit the nested model with the terms removed, including only (b_0, b_1, b_2). They compute a likelihood ratio statistic of $\Lambda = 8.5$. **True or False: At a 5% significance level, the analyst rejects the null hypothesis that $H_0 : \sigma^2_{b_3} = \sigma^2_{b_4} = 0$.**

Solution 8: [1] False. [1] The required distribution to test against is a mixture chi-square, with 3 and 5 degrees of freedom. This results in a test statistic of 9.836774, and so since $\Lambda < 9.836774$, we fail to reject the null hypothesis. Had this test been **incorrectly** compared against a χ^2_2 distribution, we would have used 5.991465 and as such we would have **incorrectly** rejected H_0.

PROBLEM 9. Suppose that a marginal model is fit to the data from Problem 4 (see above) where we take

$$\text{logit}(E[Y_{ij}|X_{ij}]) = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \beta_3 \text{Inc}_i.$$

True or False: $\exp(1000\beta_3)$ **represents the odds ratio associated with propensity to smoke for a** 1000 **increase in baseline income.**

Solution 9: [1] True. [1] Consider two individuals, measured both with $t_{ij} = t$, one with $\text{Inc}_i = x$ and the other with $\text{Inc}_{i'} = 1000 + x$. The difference in their outcomes will give $\text{logit}(\mu_{i'j}) - \text{logit}(\mu_{ij}) = \log\left(\frac{\mu_{i'j}/(1-\mu_{i'j})}{\mu_{ij}/(1-\mu_{ij})}\right) = 1000\beta_3$. As a result, $\exp(1000\beta_3)$ is the odds ratio comparing these two individuals.

PROBLEM 10. Your friend, who is not taking STAT 437, is attempting to make conclusions regarding the impact of aging in their favourite esports league. They know that you're taking STAT 437, and so they tell you the following: "I was interested in determining how aging impacts competitors. As such, I fit a model using GEEs which included the age and character for each of the competitors, and the interaction between these terms. I then fit the model dropping the interaction term. Using a likelihood ratio test, I rejected the null hypothesis that the interaction term was zero. As a result, I concluded that what character a player uses changes their success rate as they age." **True or False: The procedure outlined by your friend, as well as their conclusions, are valid.**

Solution 10: [1] False. [1] Marginal models estimated via GEE are not fit using likelihood methods. As a result, the likelihood ratio test cannot be used to test the significance of parameters. If the friend actually used a Wald test and came to the same conclusion, then their procedure seems valid, however, they used the wrong tests for the models they selected.

Part 2: Conceptual Question (10 Marks)

For the following questions, provide your answers with justification and clear communication. The answers do not need to be long, but correct responses without complete justification will receive only partial credit.

You are asked to analyze a longitudinal dataset where the goal of the study is to compare two treatments (a new experimental drug, with $A_i = 1$ as compared to an existing medication with $A_i = 0$) based on their ability to lower the diastolic blood pressure over time in a group of sick individuals. The individuals in the survey are similar (except for their assigned treatment). Suppose that, in total, there are $n = 500$ patients, with each treatment receiving 250 individuals. Further suppose that blood pressure measurements are taken at times $\{0, 2, 4, 6, 8\}$, measured in days.

PROBLEM 11. (2 Marks) Suppose that you initially decide to fit a linear mixed effects model, with random effects for the intercept, time (which is treated as a continuous variable during this analysis), and treatment terms, and assume $G_i = \sigma^2 I$. We saw in lecture that $\text{var}(Y_i) = Z_i D Z_i' + G_i$. Write down Z_i for patient $i = 1$, who is receiving the experimental drug, and for patient $i = 2$ who is receiving the existing treatment.

Solution 11: [1] Z_1 represents an individual receiving the experimental treatment, [1] Z_2 represents an individual receiving the standard treatment:

$$Z_1 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 2 & 1 \\ 1 & 4 & 1 \\ 1 & 6 & 1 \\ 1 & 8 & 1 \end{bmatrix} \qquad Z_2 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 4 & 0 \\ 1 & 6 & 0 \\ 1 & 8 & 0 \end{bmatrix}.$$

PROBLEM 12. (3 Marks) How many parameters are needed to estimate the variance structure in this model? Under what condition will $\text{var}(Y_i)$ be the same for all individuals in the study?

Solution 12: [1] In order to estimate the correlation structure, we estimate D and σ^2. D is a 3×3 matrix, which is constrained to be diagonal, which results in 6 parameters, plus one variance parameter. That makes 7 total parameters.

[1] G_i is the same for all individuals since they each have the same number of measurements taken. D is assumed to be shared between all members of the population. As a result, if $Z_i = Z_{i'}$ for two individuals, i and i', then they will have the same variance. [1] This occurs when $b_{2,i} = 0$ (or equivalently, $\text{var}(b_{2,i}) = 0$). In words: when the random effect of treatment is not significant, all individuals will have the same variance.

PROBLEM 13. (2 Marks) Suppose that we find that the random effects terms included in the model do not adequately explain the data. Without seeing any data, does it seem reasonable to drop the random effects terms from the model without otherwise modifying the assumed covariance structure?

Solution 13: [1] No. [1] It seems almost certain that for blood pressure measurements taken in the same week there is going to be *some* within-subject correlation. If we drop the random effects terms, and leave $G_i = \sigma^2 I$, this implies independent measurements which will almost certainly be inappropriate for the data.

PROBLEM 14. (3 Marks) Instead of using a linear mixed effects model, you consider fitting a linear marginal model with the same mean structure used for the previous fixed effects, an identity link function, $V(\mu) = \sigma^2$, and an unstructured correlation. Your coworker says that this was a waste of time since you already had fit this model, arguing the linear mixed effects models are also marginal models. Is your coworker correct in this situation? Why?

Solution 14: [1] No, your coworker is not correct. [2] The marginal model you fit required 6 parameters to estimate the correlation structure. Your unstructured correlation matrix will have $\binom{5}{2} = 10$ correlation parameters, which means that the pattern could not be captured by the model you had fit (in general).

While a linear mixed effects model generally has a corresponding marginal model, they are not equivalent to every marginal model. Moreover, marginal models are useful to fit if population-level inference is all that's required, since no distributional assumptions area made.

Part 3: Theoretical or Applied Question (20 Marks)

Please **pick one** of the following two problems and solve it. **If you solve both** only the first problem will be looked at. The first problem corresponds to an **application** problem, while the second is a **theoretical question**. For the application question, please provide the code and relevant output (consider using a software like RMarkdown, or being highly selective with what output you copy to ensure your solution is legible). For the theoretical question, please include enough of your work to justify the steps you have taken.

PROBLEM 15. (Application) On the course website you will find a data file `schoolgirls.csv`. This reports a study of height growth for 20 girls who were followed from age 6 to age 10. The study also records the height of each girl's mother at birth, grouping them into 1 = short, 2 = medium, and 3 = tall.

1. First, consider a model (`model1`) which is a linear mixed effects model with random slope and intercept, and where the marginal mean can vary based on the mother's height, in addition to the girl's age. Fit `model1` using R. Then, write down this model mathematically, identify the key assumptions for it to be valid, and report the estimated parameters from your model. (Note: you should report the parameter values alongside your mathematical notation, perhaps using a table, rather than simply printing the summary output).

2. Provide an estimated 95% confidence interval for: (a) the impact on average height comparing mothers who were tall to those who were short, (b) the variation in the random slope, and (c) the expected height for a 12 year old girl with a medium-height mother.

3. In addition to `model1`, consider fitting `model2` which drops the random slope term, and `model3` which drops the random intercept term. Decide which of these three models is most appropriate for the data, explicitly stating any hypothesis tests that you run.

4. Fit a corresponding linear marginal model, using GEE, deciding whether an unstructured, exchangeable, or autoregressive correlation structure is most appropriate for these data.

5. Using your selected optimal models from (3) and (4): predict the subject level and population level response for `id=7` at `age=10`. Interpret these values.

Solution 15: For complete marks, the solution must display relevant code (and output). The following are example solutions, and yours may differ slightly. However, you should be justifying choices that you make, and fully explaining what you are doing (and why).

1. [5 Marks] The initial model is specified as [1]

$$Y_{ij} = \beta_0 + \beta_1 \text{Age}_{ij} + \beta_2 I(\text{MomHeight}_i = 2) + \beta_3 I(\text{MomHeight} = 3) + b_{0,i} + b_{1,i}\text{Age}_{ij} + \epsilon_{ij}.$$

[1] In this model we assume that $b_i = (b_{0i}, b_{1i})' \sim N(0, D)$, where D is an unstructured covariance matrix. We assume that $\epsilon_i = (\epsilon_{i1}, \epsilon_{i2}, \epsilon_{i3}, \epsilon_{i4}, \epsilon_{i5}) \sim N(0, \sigma^2 I_{5\times 5})$. [1] We assume that $b_i \perp \epsilon_i$. We can fit this model using:

```
model1 <- lme(
  fixed = height ~ age + as.factor(momheight),
  random = ~ age|id,
  data = schoolgirls,
  method = 'ML'
)
```

Based on the model output we get [3]:

Parameter	Description	Estimated Value
β_0	Fixed effect intercept.	79.2623473
β_1	Fixed effect age slope.	5.7165
β_2	Fixed effect for 'medium' height moms.	3.0303304
β_3	Fixed effect for 'tall' height moms.	6.2886773
σ^2	Shared within-subject variance term.	0.4758165
$\text{var}(b_{0,i})$	Variance for random intercept.	9.4631505
$\text{var}(b_{1,i})$	Variance for random slope.	0.2726612
$\text{cov}(b_{0,i}, b_{1,i})$	Covariance between random intercept and slope.	-1.0808107

2. [3 Marks] Using the model we can simply call `intervals`. We are interested in the intervals for β_3 and $\text{var}(b_{1,i})$.

```
intervals(model1, level=0.95)
```

```
## Approximate 95% confidence intervals
##
##  Fixed effects:
##                            lower      est.     upper
## (Intercept)           76.9733510 79.262347 81.551344
## age                    5.4646302  5.716500  5.968370
## as.factor(momheight)2  0.1709973  3.030330  5.889663
## as.factor(momheight)3  3.4293442  6.288677  9.148010
## attr(,"label")
## [1] "Fixed effects:"
##
##  Random Effects:
##   Level: id
##                           lower      est.     upper
## sd((Intercept))       1.8367945 3.0762234 5.1519918
## sd(age)               0.3623740 0.5221697 0.7524304
```

```
## cor((Intercept),age) -0.8927255 -0.6728529 -0.1941870
##
##  Within-group standard error:
##     lower      est.     upper
## 0.5768058 0.6897945 0.8249162
```

[1] From this output we can see that the $\widehat{\beta}_3 = 6.289$ with a 95% CI of $(3.43, 9.15)$. [1] We also see that $\text{var}(b_{1,i}) = 0.5221697^2 = 0.273$, with a 95% CI of $(0.131, 0.566)$. We are also interested in predicting a new observation, at the population level, for a 12 year old girl with a medium height mother.

[1] We can get this via $\widehat{\beta}_0 + \widehat{\beta}_1 \times 12 + \widehat{\beta}_3$ and the corresponding standard errors. This gives a point estimated of 150.891 with a 95% CI of (148.31, 153.471). Note, if you did not use this function and instead computed from the standard errors, your answers may be slightly different.

3. [5 Marks] We can get the other two models simply by calling `update`. [1]

```
model2 <- update(model1, random = ~ 1|id)
model3 <- update(model1, random = ~ age - 1|id)
```

Both of these models are nested within `model1`. We can, as such, test the fit of each model respectively. [2] $H_0 : \sigma^2_{b_1} = 0$. The test statistic can be read off of an `anova()` call as 38.3165526. This will get compared to the critical value from a mixture χ^2 distribution with $df_1 = 1$ and $df_2 = 2$. This results in a p-value of 2.6920638×10^{-9}, which is less than 0.05, so we reject H_0. This model is not acceptable..

We can do exactly the same for comparing `model1` and `model3`. [1] $H_0 : \sigma^2_{b_0} = 0$. The test statistic can be read off of an `anova()` call as 13.1858666. This will get compared to the critical value from a mixture χ^2 distribution with $df_1 = 1$ and $df_2 = 2$. This results in a p-value of 8.2604188×10^{-4}, which is less than 0.05, so we reject H_0. This model is not acceptable..

[1] As a result we conclude that `model1` is the most appropriate.

4. [3 Marks] We next use `geepack::geeglm()` to fit three models. [1] Recall to use GEE correctly the data needs to be sorted by `id` and `age`. [1] We fit these three models.

```
schoolgirls <- schoolgirls[order(schoolgirls$id, schoolgirls$age), ]

unstr <- geeglm(height ~ age + as.factor(momheight),
                corstr = "unstr",
                id = id,
                family = gaussian,
                data = schoolgirls)
exch <- geeglm(height ~ age + as.factor(momheight),
                corstr = "exch",
                id = id,
                family = gaussian,
                data = schoolgirls)
```

```
ar1 <- geeglm(height ~ age + as.factor(momheight),
              corstr = "ar1",
              id = id,
              family = gaussian,
              data = schoolgirls)

cbind(QIC(unstr), QIC(exch), QIC(ar1))
```

```
##                  [,1]        [,2]        [,3]
## QIC        921.084575  878.674449  883.339608
## QICu       919.878499  875.308407  884.098014
## Quasi Lik -455.939249 -433.654204 -438.049007
## CIC          4.603038    5.683021    3.620797
## params       4.000000    4.000000    4.000000
## QICC       926.025751  879.312747  883.977906
```

[1] In order to select models we use the `QIC`. Because we are comparing various models with differing correlations (but the same mean) we consider both `QIC` and `CIC` and want the lowest values. Based on this, we would select either the exchangeable correlation or the AR1, as the two options disagree! Either is acceptable, but the unstructured is not necessary!

5. [4 Marks] [2] Using `model1` we can predict both of these values. We can either do this by extract the random effects for the corresponding row, or by using `predict`. You may run into some issues with the factor levels of `momheight`. One way around this is as follows.

```
which_idx <- which(schoolgirls$id == 7 & schoolgirls$age == 10) # Correct Row
predict(model1, newdata=schoolgirls, level=c(0,1))[which_idx,]
```

```
##    id predict.fixed predict.id
## 87  7      139.4577   137.8236
```

As a result, we can see that the population average is predicted at 139.4577 for girls who are 10 with a medium height mom, while the individual level prediction was 137.8236. As a result, girl 7 in the data has a subject-level effect which is negative.

[1] Using the `exch` (or `ar1`) model we can predict the population average effect for a girl like `id=7` in the data. This takes $L = (1, 10, 1, 0)$ and simply estimates either 138.9444286 using the exchangeable assumption or 138.7530841 using the AR1 model (only need to provide one of these). These are both slightly less than the estimated population effect in the mixed effects model.

[1] We **cannot** estimate a subject specific effect in the marginal model, as marginal models do not estimate subject specific effects!

Problem 16. In class we discussed the best linear unbiased predictor (BLUP) for random effects. We stated that the best predictor (in terms of MSE) for b_i is going to be given by

$$E[b_i|Y_i] = DZ_i'V_i^{-1}(Y_i - X_i\beta).$$

In this question you will prove that this is true!

1. Suppose that $g(W_i)$ is a predictor for Ω_i, where Ω_i and W_i are arbitrary random variables. Prove that the MSE of such a predictor is minimized when $g(W_i) = E[\Omega_i|W_i]$. Recall that the MSE is defined as

$$E\left\{[\Omega_i - g(W_i)]^2\right\}.$$

   Hint: Recall the law of iterated expectation.

2. Demonstrate that, assuming correct specification of a linear mixed effects model as

$$Y_i = X_i\beta + Z_ib_i + \epsilon_i,$$

   we will have that the joint distribution of (Y_i, b_i) is a multivariate normal. Recall that

$$\begin{aligned} Y_i|b_i &\sim N(X_i\beta + Z_ib_i, \sigma^2 I) \\ b_i &\sim N(0, D). \end{aligned}$$

   Here we take Y_i to be $K \times 1$ and b_i to be $q \times 1$.

   Hint: it may help to note that the moment generating function of a multivariate normal, $W \sim N(\mu, \Sigma)$, is

$$M_W(t) = E[e^{t'W}] = \exp\left(\mu't + \frac{1}{2}t'\Sigma t\right).$$

   Further, if (W, Ω) are jointly multivariate normal then the MGF, $M_{(W,\Omega)}(t) = E[\exp(t_1'W + t_2'\Omega)]$, where $t = (t_1', t_2')'$. Finally, if two random quantities have the same MGF, then they have the same distribution.

3. Appealing to parts (1) and (2) of the question, demonstrate that the best predictor of b_i (as a function of Y_i) is given by

$$DZ_i'V_i^{-1}(Y_i - X_i\beta),$$

   where $V_i = Z_iDZ_i' + \sigma^2 I$. Moreover, indicate the variance of the BLUP (e.g. $\text{var}(b_i|Y_i)$).

   Hint: It may be helpful to know that if

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right),$$

   then $Y_1|Y_2$ will follow a normal distribution with mean $\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(Y_2 - \mu_2)$ and variance $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Here the mean and variance are partitioned based on the sizes of Y_1 and Y_2.

Solution 16: Any valid proofs of the claims made will suffice. I am showing those that I think are most approachable, but you will receive full marks for any valid argument.

1. [4 Marks] Note that we want to minimize $E[(\Omega_i - g(W_i)^2]$. If we consider adding and subtracting $E[\Omega_i|W_i]$ in the brackets, we get

$$\begin{aligned}
&E\left[(\Omega_i - E[\Omega_i|W_i] + E[\Omega_i|W_i] - g(W_i))^2\right] \\
&= E\left[(\Omega_i - E[\Omega_i|W_i])^2\right] + 2E[(\Omega_i - E[\Omega_i|W_i])(E[\Omega_i|W_i] - g(W_i))] + E[(E[\Omega_i|W_i] - g(W_i))^2] \\
&= E\left[\text{var}(\Omega_i|W_i)\right] + 0 + E[(E[\Omega_i|W_i] - g(W_i))^2].
\end{aligned}$$

Note that here, $E[\text{var}(\Omega_i|W_i)]$ is independent of our choice of $g(W_i)$. The 0 second term comes by using iterated expectations, conditional on W_i, where the first part of the product becomes 0. That leaves just the third term. We know that $E[(E[\Omega_i|W_i] - g(W_i))^2] \geq 0$, since it's an expectation of a non-negative function, and as a result the MSE is minimized by solving $E[(E[\Omega_i|W_i] - g(W_i))^2] = 0$, which will occur if $g(W_i) = E[\Omega_i|W_i]$, as required.

2.[12 Marks] We will demonstrate that the two are jointly normal, using the MGF. This can be done through densities, but that's more work. Note that, following from the hint, we have

$$\begin{aligned}
M_{Y_i|b_i}(t) &= E[e^{t'Y_i}|b_i] = \exp\left((X_i\beta + Z_ib_i)'t + \frac{1}{2}t'(\sigma^2 I)t\right) \\
M_{b_i}(t) &= \exp\left(\frac{1}{2}t'Dt\right) \\
M_{(Y_i,b_i)}(t) &= E\left[e^{t_1'Y_i + t_2'b_i}\right] \\
&= E\left[E\left[e^{t_1'Y_i+t_2'b_i}\middle| b_i\right]\right] \\
&= E\left[E\left[e^{t_1'Y_i}\middle| b_i\right]e^{t_2'b_i}\right] \\
&= E\left[M_{Y_i|b_i}(t_1)e^{t_2'b_i}\right] \\
&= E\left[\exp\left((X_i\beta + Z_ib_i)'t_1 + \frac{1}{2}t_1'(\sigma^2 I)t_1 + t_2'b_i\right)\right] \\
&= \exp((X_i\beta)'t_1 + \frac{1}{2}t_1'(\sigma^2 I)t_1)E\left[\exp\left\{(Z_ib_i)'t_1 + t_2'b_i\right\}\right] \\
&= \exp((X_i\beta)'t_1 + \frac{1}{2}t_1'(\sigma^2 I)t_1)E\left[\exp\left\{(Z_i't_1 + t_2)'b_i\right\}\right] \\
&= \exp((X_i\beta)'t_1 + \frac{1}{2}t_1'(\sigma^2 I)t_1)M_{b_i}(Z_i't_1 + t_2) \\
&= \exp\left((X_i\beta)'t_1 + \frac{1}{2}t_1'(\sigma^2 I)t_1 + \frac{1}{2}(Z_i't_1 + t_2)'D(Z_i't_1 + t_2)\right) \\
&= \exp\left((X_i\beta)'t_1 + \frac{1}{2}\left[t_1'(\sigma^2 I)t_1 + t_1'(Z_iDZ_i')t_1 + t_1'(Z_iD)t_2 + t_2'DZ_i't_1 + t_2'Dt_2\right]\right) \\
&= \exp\left(\mu't + \frac{1}{2}t'\Sigma t\right).
\end{aligned}$$

Here we are taking

$$\mu = \begin{bmatrix} X_i\beta \\ \mathbf{0}_{q\times 1} \end{bmatrix} \quad \Sigma = \begin{bmatrix} (\sigma^2 I) + Z_i D Z_i' & Z_i D \\ D Z_i' & D \end{bmatrix} \quad t = (t_1', t_2')'.$$

As a result, we conclude that $(Y_i, b_i) \sim N(\mu, \Sigma)$.

3. [4 Marks] Part 1 indicates that $E[b_i|Y_i]$ is the best predictor (as a function of Y_i) of b_i. Using the given property of multivariate normality, we find that $b_i|Y_i \sim N(\mu^*, \Sigma^*)$ with

$$\begin{aligned} \mu^* &= \mathbf{0} + DZ_i' \left(\sigma^2 + Z_i D Z_i'\right)^{-1} (Y_i - X_i\beta) \\ &= DZ_i' V_i^{-1}(Y_i - X_i\beta) \\ \Sigma^* &= D - DZ_i' V_i^{-1} Z_i D. \end{aligned}$$

This gives us the required form for the BLUP, and the variance of the estimator.