Question #1 (40 points)

For the \(i\)th individual, in the \(j\)th household and \(k\)th zip code, let \(X_{ijk1}, X_{ijk2}, X_{ijk3}\) be income measured in 1000s of dollars at the individual, the household and the community levels respectively. The dependent variable, \(Y_{ijk}\) is a continuous measure of health status.

a) [5] All regressors have the \(ijk\) subscript, but the value of the regressor changes only with certain subscripts. For each of \(X_{ijk1}, X_{ijk2}\) and \(X_{ijk3}\), indicate dependency on subscripts by re-writing them with only the relevant subscripts:

\[ X_{ijk1} = X_{ijk1} \quad X_{ijk2} = X_{jk2} \quad X_{ijk3} = X_{k3} \]

b) [5] Write a model for the response \(Y_{ijk}\) using \(\beta_0\) for the intercept, \(\beta_1\) for the slope on \(X_{ijk1}\) etc. Include an error term (residual), denoted by \(e_{ijk}\):

\[
Y_{ijk} = \beta_0 + \beta_1 X_{ijk1} + \beta_2 X_{ijk2} + \beta_3 X_{ijk3} + e_{ijk}
\]

\[
Y_{ijk} = \beta_0 + \beta_1 X_{ijk1} + \beta_2 X_{jk2} + \beta_3 X_{k3} + e_{ijk}
\]

c) [6] Fill in the expectation and variance of \(Y_{ijk}\):

\[
E[Y_{ijk}] = \beta_0 + \beta_1 X_{ijk1} + \beta_2 X_{ijk2} + \beta_3 X_{ijk3}
\]

\[
E[Y_{ijk}] = \beta_0 + \beta_1 X_{ijk1} + \beta_2 X_{jk2} + \beta_3 X_{k3}
\]

\[
V[Y_{ijk}] = V(e_{ijk}) = \sigma^2
\]

Note that \(\sigma^2\) may be the sum of other variances and covariance, but it is still just one number, the overall variance for an observation.

d) [7] Provide a brief interpretation for each of \((\beta_0, \beta_1, \beta_2, \beta_3)\).

\(\beta_0\) : Expectation of \(Y_{ijk}\) when \(X_{ijk1} = X_{jk2} = X_{k3} = 0\). If simultaneously having each \(X = 0\) is not reasonable (as it likely is not for income), then then intercept should not be interpreted other than as the value needed to produce a good fit to the data.

\(\beta_1\) : Expected change in \(Y_{ijk}\) when individual income \((X_{ijk1})\) changes by one unit and the other \(X\)s remain constant.

\(\beta_2\) : Expected change in \(Y_{ijk}\) when household income \((X_{jk2})\) changes by one unit and the other \(X\)s remain constant.

\(\beta_3\) : Expected change in \(Y_{ijk}\) when zip code income \((X_{k3})\) changes by one unit and the other \(X\)s remain constant.

The foregoing interpretation of \(\beta_1, \beta_2, \beta_3\) depends on whether the other \(X\)s can/do remain constant in “real life.” This premise is very unlikely for individual, household and community code income, so the coefficients must be interpreted with care.

Question #1 continued ⇒
Question #1 (continued)

f) [5] Does your interpretation depend on whether the $e_{ijk}$ are independent or correlated? Why or why not?

No, for the linear model interpretation of the fixed effects coefficients doesn’t depend on the correlation structure of the $e_{ijk}$. The expectation of a sum is the sum of the expectations and is not affected by the correlation structure. However, expectations conditional on some $Y$’s can be affected.

g) [5] Now, assume that the slope on family income ($\beta_2$) can vary from zip code to zip code (denote it by $\beta_{2k}$). Augment your model in (b) with a random effect that represents this variation, decomposing $\beta_{2k}$ into its population-level and zip code specific random components. Use $\tau^2$ for the between-zipcode variance of the $\beta_{2k}$ ($V(\beta_{2k}) = \tau^2$).

\[
Y_{ijk} = \beta_0 + \beta_1 X_{ijk1} + \beta_{2k} X_{jk2} + \beta_3 X_{k3} + e_{ijk}
\]

$\beta_{2k} \sim (\beta_2, \tau^2)$

or

\[
Y_{ijk} = \beta_0 + \beta_1 X_{ijk1} + (\beta_2 + b_{2k}) X_{jk2} + \beta_3 X_{k3} + e_{ijk}
\]

$b_{2k} \sim (0, \tau^2)$

Note that the residuals ($e_{ijk}$) will be different from the $e_{ijk}$.

h) [7] For the model in (g), fill in the expectation and variance of $Y_{ijk}$ and the covariance of $Y_{ijk}$ with $Y_{i'jk}, (i \neq i')$.

\[
E[Y_{ijk}] = \beta_0 + \beta_1 X_{ijk1} + \beta_2 X_{jk2} + \beta_3 X_{k3} \quad (\text{as in part (c)})
\]

\[
V[Y_{ijk}] = \tau^2 X_{ijk2}^2 + \sigma^2 = \tau^2 X_{jk2}^2 + \sigma^2
\]

\[
\text{cov}(Y_{ijk}, Y_{i'jk}) = \tau^2 X_{ijk2} X_{i'jk2} = \tau^2 X_{jk2}^2 \quad (i \neq i')
\]

Note that for the covariance we need $(i \neq i')$ so that the covariance of the residuals is 0 and the only term is that listed.
**Question #2 (30 points)**

Consider the following logistic regressions, all with a 0 intercept.

For the random effects models, \( E(b_0) = 0, \ V(b_0) = \tau^2 \).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>( \text{logit}{ pr(Y = 1 \mid X) } = \beta X )</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEX</td>
<td>Fixed effects, with X:</td>
<td>( \text{logit}{ pr(Y = 1 \mid b_0, X) } = b_0 + \beta X )</td>
</tr>
<tr>
<td>REX</td>
<td>Random effects, with X:</td>
<td>( \text{logit}{ pr(Y = 1 \mid b_0, X) } = b_0 + \beta X )</td>
</tr>
<tr>
<td>FEXW</td>
<td>Fixed effects, with X and W:</td>
<td>( \text{logit}{ pr(Y = 1 \mid b_0, X, W) } = b_0 + \beta X + \gamma W )</td>
</tr>
<tr>
<td>REXW</td>
<td>Random effects, with X and W:</td>
<td>( \text{logit}{ pr(Y = 1 \mid b_0, X, W) } = b_0 + \beta X + \gamma W )</td>
</tr>
</tbody>
</table>

Data analysis produces the following results (* indicates that \( \tau \) is forced to = 0):

<table>
<thead>
<tr>
<th>Model</th>
<th>( \hat{\beta} )</th>
<th>( \hat{\tau} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEX</td>
<td>1.0</td>
<td>*</td>
</tr>
<tr>
<td>REX</td>
<td>1.8</td>
<td>1.6</td>
</tr>
<tr>
<td>FEXW</td>
<td>1.0</td>
<td>*</td>
</tr>
<tr>
<td>REXW</td>
<td>1.1</td>
<td>0.7</td>
</tr>
</tbody>
</table>

a) [8] Explain the relation between the \( \hat{\beta} \) for the FEX and REX models. Specifically, why is \( \hat{\beta} \) in REX bigger than in FEX? What does this difference in estimates imply about population-level predictions?

\( \hat{\beta} \) is bigger in REX compared to FEX due to the “flattening” of the odds ratio when producing the marginal (population level) odds ratio. That is, averaging probabilities conditional on the random effect to produce the population probability attenuates the constant slope relation in the logit-linear model because, unlike logits, probabilities are constrained to the interval \([0, 1]\) and the averages are “squished” near the boundaries, producing the attenuated OR. So, for REX to end up with a population level \( \log(\text{OR}) \) of about 1.0, it needs to use a cluster-specific \( \log(\text{OR}) = 1.8 \). This 1.8 gets “flattened” to 1.0. The foregoing indicates that population-level predictions are about the same for FEX and REX.
Question #2 (continued)

b) [8] Explain why including $W$ reduces $\hat{\tau}$ from 1.6 in REX to 0.7 in REXW.

Including the additional covariate accounts for some of the unexplained variation in REX, thus reducing the estimated standard deviation of the latent, random effect ($\hat{\tau}$).

c) [8] Explain why including $W$ moves $\hat{\beta}$ from 1.8 in REX to 1.1 in REXW.

Because $\hat{\tau}$ went from 1.6 to 0.7, there is less cluster level variance in the intercept and therefore less “flattening” of the conditional OR when computing the population OR. Therefore, less expansion away from the population-level log(OR) is needed to compensate for the flattening. For this smaller $\tau$ case, $\hat{\beta}$ needs to be only 1.1 rather than 1.8 to compensate and produce the population level log(OR) = 1.0.

Though it may be that $W$ is a confounder, the movement of $\hat{\beta}$ from 1.8 to 1.1 can happen even in the absence of confounding and needs to be listed as the primary explanation.

d) [6] Does including $W$ change population-level predictions that are averaged over the distribution of $W$? Why or why not?

Including $W$ (using REXW) will produce very little change in population level predictions that are averaged over $W$ compared to population level predictions computed from REX. To produce the population level prediction ($pr(Y = 1 \mid X)$), you need to average the following over the distribution of $b_0$ (with it’s $\hat{\tau} = 0.7$) and over the values of $W$:

$$pr(Y = 1 \mid b_0, X, W)pr = \frac{e^{b_0 + X + W}}{1 + e^{b_0 + X + W}}$$

This combined averaging produces a very similar result to averaging the following over the distribution of $b_0$ (with it’s $\hat{\tau} = 1.8$):

$$pr(Y = 1 \mid b_0, X)pr = \frac{e^{b_0 + X}}{1 + e^{b_0 + X}}.$$ 

That the estimated coefficient on $X$ ($\hat{\beta}$) will be different in REX and REXW allows for additional adjustment to make the predictions similar.
Question #3 (30 points) Dependent variable $Y$ takes on the values 0 or 1, with the $pr(Y = 1 | X, RE)$ displayed in the following table for two levels of a predictor $X$ and two levels of an underlying random effect. The random effect takes on the "Low" and "High" values, each with probability 0.5.

<table>
<thead>
<tr>
<th>RE Value</th>
<th>Population Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>$X = 0$</td>
<td>1/4</td>
</tr>
<tr>
<td>$X = 1$</td>
<td>2/3</td>
</tr>
</tbody>
</table>

$\text{OR} = 6.0 \quad 6.0 \quad 16/3 \quad (5\frac{1}{3})$

a) [8] Fill in the 2 missing probabilities (located at the two ".") and the 3 missing ORs (located at the three ","). Keep entries as fractions.

\[ \frac{3}{8} = \frac{1}{2} \left( \frac{1}{4} \right) + \frac{1}{2} (.)) = \Rightarrow (.)) = \frac{1}{2} \]

\[ \frac{16}{21} = \frac{1}{2} \left( \frac{2}{3} \right) + \frac{1}{2} \left( \frac{6}{7} \right) \]

\[ OR_{Low} = \left( \frac{2}{1} \right) / \left( \frac{1}{3} \right) = 6.0 \]

\[ OR_{High} = \left( \frac{6}{1} \right) / \left( \frac{1}{1} \right) = 6.0 \]

\[ OR_{pop} = \left( \frac{16}{5} \right) / \left( \frac{16}{3} \right) = 5\frac{1}{3} \]

b) [5] Briefly discuss the relation between the conditional odds ratios for the "Low" and "High" levels of the random effect and the population level odds ratio.

As we have seen in many examples, the conditional odds ratio is “flattened” when computing the population level OR, in our case going from 6.0 to 5\frac{1}{3}. This change is due to the non-linear transform taking us from the logit-linear model to probabilities. Averaging probabilities conditional on the random effect to produce the population probability attenuates the constant slope relation in the logit-linear model because, unlike logits, probabilities are constrained to the interval [0, 1] and the averages are “squished” near the boundaries, producing the attenuated OR.
Question #3 (continued)

c) [7] For $X = 0$ only, consider a pair of responses $(Y_1, Y_2)$ coming from the same cluster (therefore, having the same value for the random effect). In the following table, fill in the 8 probabilities (located at the eight “.”). Keep entries as fractions.

Hint: Compute the joint probability for one of the interior cells. Then, use this and the “Population Level” probability in the previous table to fill in the remaining cells.

<table>
<thead>
<tr>
<th>For $X = 0$</th>
<th>$Y_2 = 0$</th>
<th>$Y_2 = 1$</th>
<th>$Y_1$ marginal</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_1 = 0$</td>
<td>13/32</td>
<td>7/32</td>
<td>5/8</td>
</tr>
<tr>
<td>$Y_1 = 1$</td>
<td>7/32</td>
<td>5/32</td>
<td>3/8</td>
</tr>
<tr>
<td>$Y_2$ marginal</td>
<td>5/8</td>
<td>3/8</td>
<td>1</td>
</tr>
</tbody>
</table>

The interior of the table is the joint $pr(Y_1, Y_2)$. For example, the upper-left cell is $pr(Y_1 = 0, Y_2 = 0)$.

Example computation using the values in part (a):

$$pr(Y_1 = 1, Y_2 = 1) = \frac{5}{32} = \frac{1}{2} \left( \frac{1}{4} \times \frac{1}{4} + \frac{1}{2} \times \frac{1}{2} \right) = \frac{1}{2} \left( \frac{1}{16} + \frac{1}{4} \right)$$


d) [6] Compute,

$$pr(Y_2 = 1 \mid Y_1 = 0) = \left( \frac{7}{32} \right) / \left( \frac{5}{8} \right) = \frac{7}{20} < \frac{3}{8} = pr(Y_1 = 1)$$

$$pr(Y_2 = 1 \mid Y_1 = 1) = \left( \frac{5}{32} \right) / \left( \frac{3}{8} \right) = \frac{5}{12} > \frac{3}{8} = pr(Y_1 = 1)$$


e) [4] Discuss the relation between these two probabilities and their relation to the population-level $pr(Y = 1)$.

We have that,

$$pr(Y_2 = 1 \mid Y_1 = 0) < pr(Y_2 = 1) < pr(Y_2 = 1 \mid Y_1 = 1)$$

and that the weighted average of the two conditional probabilities is exactly the marginal probability,

$$pr(Y_2 = 1) = \frac{3}{8} = \frac{5}{8} \times \frac{7}{20} + \frac{3}{8} \times \frac{5}{12}$$

So, conditioning on the value of $Y_1$ changes the distribution of $Y_2$; they are statistically dependent. Knowing that $Y_1 = 1$ raises the probability that $Y_1 = 1$; knowing that $Y_1 = 0$ lowers it, showing positive dependence.

Though not requested, it is also true that knowledge of the value of $Y_1$ changes the distribution of the random effect from 50/50 to some other distribution (you can work it out). Applying this new, random effect distribution also gives the conditional probabilities in part (d).