1.(a) Here

\[ Q_N(\beta) = N^{-1} \sum_i (y_i - \exp(x'_i\beta))^2 \]
\[ = N^{-1} \sum_i \left[ (y_i - \exp(x'_i\beta_0)) + \exp(x'_i\beta_0) - \exp(x'_i\beta) \right]^2 \]
\[ = N^{-1} \sum_i \left[ (y_i - \exp(x'_i\beta_0))^2 + 2(y_i - \exp(x'_i\beta_0))\exp(x'_i\beta_0) - \exp(x'_i\beta) \right] + \exp(x'_i\beta_0) - \exp(x'_i\beta))^2 \].

(b) Now

\[ Q_0(\beta) = \text{plim } Q_N(\beta) \]
\[ = \text{plim } N^{-1} \sum_i (y_i - \exp(x'_i\beta_0))^2 \]
\[ + 2 \text{plim } N^{-1} \sum_i \left[ (y_i - \exp(x'_i\beta_0))\exp(x'_i\beta_0) - \exp(x'_i\beta) \right] \]
\[ + \text{plim } N^{-1} \sum_i \left[ \exp(x'_i\beta_0) - \exp(x'_i\beta) \right]^2 \]
\[ = \lim N^{-1} \sum_i E_{\xi_i}[\sigma_{\eta_0}^2] + 0 + \lim N^{-1} \sum_i E_{\xi_i}\left[\exp(x'_i\beta_0) - \exp(x'_i\beta)\right]^2 \]

where apply a LLN to first term and use \(E[(y_i - \exp(x'_i\beta_0))^2] = E_{\xi_i}[y_i - \exp(x'_i\beta_0)]^2|\xi_i] = E_{\xi_i}[\sigma_{\eta_0}^2]|\xi_i] \) where second last equality uses \(E[y_i|x_i] = \exp(x'_i\beta_0)\) and a LLN to the second term using \(E[y_i - \exp(x'_i\beta_0)|x_i] = 0\) since \(E[y_i|x_i] = \exp(x'_i\beta_0)\).

(c) \(\hat{\beta}\) is consistent for \(\beta_0\) since by inspection \(\text{plim } Q_N(\beta)\) is clearly minimized at \(\beta = \beta_0\), as then the third term (the only term involving \(\beta\)) takes the minimum value of zero.

Or if you want to do more work in this example: differentiate w.r.t. \(\beta\) (not \(\beta_0\))

\[ \frac{\partial Q_0(\beta)}{\partial \beta} = \lim N^{-1} \sum_i -2E_{\xi_i}[\exp(x'_i\beta)|\exp(x'_i\beta_0) - \exp(x'_i\beta)] \]
\[ = 0 \text{ when } \beta = \beta_0. \]

(d) Consider the first term: \(\text{plim } \frac{1}{N} \sum_i (y_i - \exp(x'_i\beta_0))^2\).

This is the average \(X_N = N^{-1} \sum_i x_i\) where \(x_i = (y_i - \exp(x'_i\beta_0))^2\).

Here \(x_i\) is i.i.d. as \((x_i, y_i)\) are i.i.d - best to use Khinchines Theorem (as then minimal assumptions).

This requires \(E[X_i] = E[(y_i - \exp(x'_i\beta_0))^2] = E_{\xi_i}[\sigma_{\eta_0}^2]|\xi_i]\) exists.

Similarly for the second and third terms.

(e) No. In (b) the second term will no longer disappear if \(E[y_i|x_i] \neq \exp(x'_i\beta_0)\), so \(Q_0(\beta)\) will not necessarily be maximized at \(\beta = \beta_0\). (Also the first term will change and now involve \(\beta\).)

So \(\hat{\beta}\) will be consistent if the functional form for \(E[y_i|x_i]\) is misspecified.

2.(a) Differentiating \(Q_N(\beta) = \frac{1}{N} \sum_i (y_i - \exp(x'_i\beta))^2\)

\[ \frac{\partial Q_N}{\partial \beta} = \frac{1}{N} \sum_i -2(y_i - \exp(x'_i\beta))x'_i \]
\[ = \frac{-2}{N} \sum_i (y_i - \exp(x'_i\beta))x'_i \]

Can assume the result that \(\sqrt{N} \frac{\partial Q_N}{\partial \beta} |_{\beta_0} \xrightarrow{d} \mathcal{N}(0, \mathbf{B}_0)\), where

\[ \mathbf{B}_0 = \text{plim } N \left( \frac{\partial^2 Q_N}{\partial \beta \partial \beta} \right) |_{\beta_0} \]
\[ = \text{plim } N \left( \frac{-2}{N} \sum_i (y_i - \exp(x'_i\beta_0))\exp(x'_i\beta_0)x'_i \right) \left( \frac{-2}{N} \sum_i (y_i - \exp(x'_i\beta_0))\exp(x'_i\beta_0)x'_i \right)' \]
\[ = \lim \frac{2}{N} \sum_i E \left[ \left( y_i - \exp(x'_i\beta_0) \right)^2 \exp(x'_i\beta_0)^2 \right] x'_i x'_i \]
\[ = \lim \frac{4}{N} \sum_i E_{\xi_i}[\sigma_{\eta_0}^2(\exp(x'_i\beta_0))^2 x'_i x'_i]. \]
(b) We have differentiating ∂²Qₙ / ∂β∂β by parts

\[
\frac{\partial^2 Q_n}{\partial \beta \partial \beta} = \frac{2}{n} \sum_i (\exp(x_i' \beta))^2 x_i x'_i - \frac{2}{n} \sum_i (y_i - \exp(x'_i \beta))(\exp(x'_i \beta)) x_i x'_i
\]

\[A_0 = \lim \frac{\partial^2 Q_n}{\partial \beta \partial \beta} \bigg|_{\beta_0} = \lim \frac{2}{n} \sum_i (\exp(x'_i \beta_0))^2 x_i x'_i, \quad \text{since } E[y_i - \exp(x'_i \beta_0)|x_i] = 0.\]

(c) Combining and noting that \((\frac{\partial}{\partial \beta})^{-1}(\frac{\partial^2}{\partial \beta^2})^{-1} = (\frac{\partial}{\partial \beta})^{-1}(\frac{\partial^2}{\partial \beta^2})^{-1}\)

\[
\sqrt{N}(\beta_0 - \beta_0) \xrightarrow{d} \mathcal{N}(0, A_0^{-1}B_0A_0^{-1})
\]

(d) Use \(A^{-1}\hat{B}A^{-1}\) where \(A = \frac{1}{n} \sum_i (\exp(x'_i \beta))^2 (\exp(x'_i \beta))^2 x_i x'_i\)

and \(\hat{B} = \frac{1}{n} \sum_i (y_i - \exp(x'_i \beta))^2 (\exp(x'_i \beta))^2 x_i x'_i\).

(e) Reject \(H_0 : \beta_j = 0\) against \(H_a : \beta_j \neq 0\) at level \(a\) if \(|t| > 1.96\)

where \(t = \hat{\beta}_j / s_{\hat{\beta}_j}\) is the \(j\)th diagonal entry in the matrix in (g).

3. (a) A sequence of random variables \(\{b_N\}\) converges in probability to \(b\) if for any \(\varepsilon > 0\) and \(\delta > 0\), there exists \(N^* = N^*(\varepsilon, \delta)\) such that for all \(N > N^*\), \(\Pr[|b_N - b| < \varepsilon] > 1 - \delta\).

(b) A sequence of random variables \(\{b_N\}\) converges in distribution to a random variable \(b\) if \(\lim_{N \to \infty} F_N = F\), at every continuity point of \(F\), where \(F_N\) is the distribution of \(b_N\), \(F\) is the distribution of \(b\), and convergence is in the usual mathematical sense.

(c) In general it will not, unless \(b\) is a constant, rather than a random variable.

(d) Because the statistics we use, such as estimators, involve averages.

(e) Newton Raphson: \(\hat{\theta}_{s+1} - \hat{\theta}_s = -[\partial^2 Q_N(\theta) / \partial \theta \partial \theta]|_{\hat{\theta}_s}^{-1} \times \partial Q_N(\theta) / \partial \theta|_{\hat{\theta}_s}\).

(f) When \(y_i\) equals 1 with probability \(\Lambda(x'_i \beta)\) and 0 with probability \(1 - \Lambda(x'_i \beta)\),

we can write the density as \(f(y_i|x_i) = \Lambda(x'_i \beta)^{y_i}(1 - \Lambda(x'_i \beta))^{1-y_i}\).

Then \(\ln L = \ln \left(\prod_{i=1}^{N} f(y_i|x_i)\right) = \sum_{i=1}^{N} \ln f(y_i|x_i) = \sum_{i=1}^{N} \ln(\Lambda(x'_i \beta)^{y_i}(1 - \Lambda(x'_i \beta))^{1-y_i})\)

\(= \sum_{i=1}^{N} \{y_i \ln(\Lambda(x'_i \beta)) + (1 - y_i) \ln(1 - \Lambda(x'_i \beta))\}\}

4. I expected answer to cover both parameter interpretation (ME’s) and statistical significance.

The coefficients are jointly statistically significant at 5%.

Individually the coefficient of \(x_2\) is statistically significant at 5% but that of \(x_3\) is not.

Here ME \(g' = \beta_1 + \beta_2 x_2 + \beta_3 x_3 \beta_j\) where \(g'(\cdot) < 0\) so sign of ME is the reverse of the sign of \(\beta_j\).

It follows that \(E[y|x_2, x_3]\) decreases when \(x_2\) increases and when \(x_3\) increases.

As .42/.27 ≈ 1.5, a one unit change in \(x_3\) has about 1.5 times the effect of a one unit change in \(x_2\).

Aside: Also since standard errors reported were not labelled “robust” might also want to check this.

Overall - question 1 done well. Questions 2-4 not done as well. Some fundamental errors were made at times - so check for my hand-written comments on your answers.

The course grade will be based on a curve from the combined scores of midterm (35%), final (50%) and assignments (15%).

The curve for this exam is only a guide to give you a rough idea of how you are doing.

<table>
<thead>
<tr>
<th>Scores out of</th>
<th>35</th>
<th>75th percentile</th>
<th>28 (83%)</th>
<th>A</th>
<th>28 and above</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>26.5 (76%)</td>
<td>A-</td>
<td>22 and above</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25th percentile</td>
<td>21.5 (61%)</td>
<td>B+</td>
<td>17 and above</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>