1.(a) We have
\[
\ln f(y_i) = -\ln y_i - \pi (\ln y_i - x_i'\beta)^2
\]
\[
\ln L(\beta) = \sum_{i=1}^{N} \{-\ln y_i - \pi (\ln y_i - x_i'\beta)^2\}
\]
\[
\partial \ln L(\beta) / \partial \beta = -\sum_{i=1}^{N} 2\pi (\ln y_i - x_i'\beta)x_i = 0.
\]

(b) Since \(\partial^2 \ln L(\beta) / \partial \beta \partial \beta' = -\sum_{i=1}^{N} 2\pi x_i x_i'\) we have \(\sqrt{N} (\hat{\beta} - \beta_0) \overset{d}{\rightarrow} N[0, -A_0^{-1}]\) by the information matrix equality \(A_0 = \text{plim} N^{-1} \frac{1}{N} \sum_{i=1}^{N} x_i x_i'\).

(c) Need \(E[\ln y_i|x_i] = x_i'\beta_0\) so that \(E[\partial \ln L(\beta) / \partial \beta]_{\beta_0} = -\sum_{i=1}^{N} 2\pi (\ln y_i - x_i'\beta_0)x_i = 0\).

(d) Now \(\sqrt{N} (\hat{\beta} - \beta_0) \overset{d}{\rightarrow} N[0, A_0^{-1}B_0A_0^{-1}]\) where \(A_0\) as above and
\[
B_0 = \text{plim} N^{-1} \frac{1}{N} \sum_{i=1}^{N} (\ln y_i - x_i'\beta_0) x_i \times \frac{1}{N} \sum_{i=1}^{N} (\ln y_i - x_i'\beta_0) x_i'
\]
\[
= 4\pi^2 \text{plim} \frac{1}{N} \sum_{i=1}^{N} (\ln y_i - x_i'\beta_0)^2 x_i x_i'.
\]

(e) The multiples \(2\pi \) and \(4\pi^2\) cancel and we have
\[
\hat{V}[\beta] = \left(\frac{1}{N} \sum_{i=1}^{N} x_i x_i'\right)^{-1} \times \frac{1}{N} \sum_{i=1}^{N} (\ln y_i - x_i'\beta)^2 x_i x_i' \times \left(\frac{1}{N} \sum_{i=1}^{N} x_i x_i'\right)^{-1}.
\]

2.(a) Density for \(i^{th}\) observation is \(p_{1i}^{y_{1i}} \times p_{2i}^{y_{2i}} \times p_{3i}^{y_{3i}}\).
\[
\ln L(\beta) = \sum_i \left[y_{1i} \ln p_{1i} + y_{2i} \ln p_{2i} + y_{3i} \ln p_{3i}\right] = \sum_i \left[y_{1i} + y_{2i} x_i' \beta_2 + y_{3i} x_i' \beta_3 - \ln (1 + e^{x_i' \beta_2} + e^{x_i' \beta_3})\right].
\]

(b) Multinomial logit is ARUM with
\[
U_{1i} = x_i' \beta_1 + \varepsilon_{1i}
\]
\[
U_{2i} = x_i' \beta_2 + \varepsilon_{2i}
\]
\[
U_{3i} = x_i' \beta_3 + \varepsilon_{3i}
\]
where \(\varepsilon_{ij}\) are i.i.d. type 1 extreme value. We choose \(y_i = j\) if \(U_{j,i} \geq U_{k,i}\) for all \(k\).

(c) When a person ages one more year then the probability of health being excellent falls by 0.0076766, when evaluated at the sample mean of age.

(d) An alternative is ordered logit (or ordered probit or ....)
This is likely to be better as the outcomes are ordered.
Compare the log-likelihoods with adjustment for degrees of freedom.
Here multinomial logit has four parameters - see the output.
Ordered logit has three parameters - two cut points and age.
So use AIC or BIC. Choose the model with smallest (given Stata definitions of AIC and BIC).
NOTE: Conditional logit is not an alternative as it is the same as multinomial logit except allows regressors to differ across alternatives which is not the case here.

(e) Logit model is \(\text{Pr}[y_i = 1|x_i] = \Lambda(x_i'\beta) = e^{x_i'\beta} / (1 + e^{x_i'\beta})\).
So \(E[y_i|x_i] = \Lambda(x_i'\beta)\) and \(E[(y_i - \Lambda(x_i'\beta))|x_i] = 0\) and \(E[x_i(y_i - \Lambda(x_i'\beta))]| = 0\).
Assume existence of instrument \(z_i\) such that \(E[z_i(y_i - \Lambda(x_i'\beta))] = 0\).
Then use \(\hat{\beta}\) solves \(\sum_{i=1}^{N} z_i (y_i - \Lambda(x_i'\beta)) = 0\), or in over-identified case \(\hat{\beta}\) minimizes \[
\left[\sum_{i=1}^{N} z_i (y_i - \Lambda(x_i'\beta))\right]' W \left[\sum_{i=1}^{N} z_i (y_i - \Lambda(x_i'\beta))\right]
\] where e.g. \(W = \left[\sum_{i=1}^{N} z_i z_i'\right]^{-1} \).
3. (a) Pr[y_{1i} = 1] = Pr[x'_{1i}\beta_1 + \varepsilon_{1i} > 0] = Pr[-\varepsilon_{1i} < x'_{1i}\beta_1] = \Phi(x'_{1i}\beta_1) as \varepsilon_{1i} \sim N[0,1].

Probit of y_{1i} on x_{1i}.

(b) Now given the general result for (z_1, z_2), \varepsilon_{2i}|\varepsilon_{1i} = \sigma_{12}\varepsilon_{1i} + v_i where v_i is independent of \varepsilon_{1i}.

\[ E[y_{2i}|y_{2i}] = 1 = E[x'_{2i}\beta_2 + \varepsilon_{2i}|x'_{1i}\beta_1 + \varepsilon_{1i} > 0] \]
\[ = x'_{2i}\beta_2 + E[\varepsilon_{2i}|\varepsilon_{1i} > -x'_{1i}\beta_1] \]
\[ = x'_{2i}\beta_2 + E[\sigma_{12}\varepsilon_{1i} + v_i|\varepsilon_{1i} > -x'_{1i}\beta_1] = x'_{2i}\beta_2 + \sigma_{12}E[\varepsilon_{1i}|\varepsilon_{1i} > -x'_{1i}\beta_1] + 0 \]
\[ = x'_{2i}\beta_2 + \sigma_{12}E[\lambda(x'_{1i}\beta_1)]. \]

(1) Probit of y_{1i} on x_{1i} gives \hat{\beta}_1; (2) For y_{1i} = 1, OLS of y_{2i} on x_{2i} and \lambda(x'_{1i}\hat{\beta}_1) gives \hat{\beta}_2.

(c) No. Key is that \varepsilon_{2i}|\varepsilon_{1i} = \delta\varepsilon_{1i} + v_i where v_i is independent of \varepsilon_{1i}.

4. (a) We have \hat{\gamma} = g(\hat{\beta}) = g(\beta_0) + |\partial g(\beta)/\partial \beta| \times (\hat{\beta} - \beta_0) \sim N[g(\beta_0), \sigma^2|\partial g(\beta)/\partial \beta|^2].

Since \beta is unknown we use \tilde{\gamma} = s^2|\partial g(\beta)/\partial \beta| \tilde{\beta}^2 where s^2 \sim \sigma^2.

(b) Now \gamma = g(\beta) = e^{\beta} so \partial g(\beta)/\partial \beta = e^{\beta}, \tilde{\gamma} = \sigma^2(e^{\beta})^2 and \tilde{\gamma} \sim N[e^{\beta_0}, s^2(e^{\beta})^2].

Given s^2 = 1 and \tilde{\beta} = 2 we have \tilde{\gamma} \sim N[e^{\beta_0}, 1 \times (e^2)^2] \sim N[e^{\beta_0}, e^4].

Under H_0: e^{\beta_0} = 1, t = (e^{\beta} - e^{\beta_0})/se(e^{\beta}) = (e^2 - 1)/\sqrt{e^4} = (e^2 - 1)/e^2 = 1 - (1/e^2) = 0.87.

We do not reject H_0 as |t| < 1.96.
[An alternative is to use \tilde{\gamma} = \sigma^2(e^{\beta})^2 = \sigma^2 under H_0: e^{\beta_0} = 1. Then t = (e^2 - 1)/1 = 6.389].

(c) Easiest to do bootstrap without asymptotic refinement.

Resample with replace 999 times, say. Get 999 estimates of \hat{\beta}_* and hence e^{\hat{\beta}*}.

The bootstrap standard error estimate of e^{\hat{\beta}*} is the standard deviation of the 1,000 estimates e^{\hat{\beta}*}. Then reject H_0 if t = (e^{\beta} - 1)/se_{boot}[e^{\hat{\beta}}] exceeds 1.96 in absolute value.

If instead do bootstrap with refinement then form t_* = (e^{\hat{\beta}*} - e^{\beta_0})/se_{boot}[e^{\hat{\beta}_*}] and see whether t_* = 0.87 from part (b) lies within the 25 to 97.5 percentiles of the 999 t_*'s.

5. (a) \hat{\beta} = 0.0179. One more year of job tenure leads to 0.0179 proportionate increase or 1.79% increase in the wage (controlling for the individual specific effect \alpha_i).

(b) Regress (y_{it} - \bar{y}_i) on (x_{it} - \bar{x}_i).

(c) This eliminates a fixed effect \alpha_i in the model y_{it} = \alpha_i + \beta x_{it} + u_{it} where \alpha_i may be correlated with u_{it}.

(d) Hausman test. \tilde{V}[^{\hat{\beta}_{FE} - \hat{\beta}_{RE}}] = \tilde{V}[^{\hat{\beta}_{FE}}] - \tilde{V}[^{\hat{\beta}_{RE}}] = 0.0031^2 - 0.0021^2 = 0.000052.

So se[^{\hat{\beta}_{FE} - \hat{\beta}_{RE}}] = \sqrt{0.000052} = 0.0023.

Hausman t-test: t = (\hat{\beta}_{FE} - \hat{\beta}_{RE})/se[^{\hat{\beta}_{FE} - \hat{\beta}_{RE}}] = (0.0179 - 0.0216)/0.0023 = 0.0037/0.0023 = 1.61 < 1.96.

Hausman \chi^2-test: 1.61^2 = 2.57 < 3.84.

Do not reject H_0 at level 0.05. So RE is okay and is preferred as likely more efficient.

(e) Default standard errors were used. Better to use clustered standard errors.

These are likely to be larger. And the way we did the Hausman test becomes invalid.

The curve for this exam is only a guide. Course grade based on course score.

Scores out of 50
75th percentile 45 (90%) A 42 and above
Median 41 (82%) A- 34 and above
25th percentile 34 (68%) B+ 26 and above