1. (a) Here \( \ln L(\beta, \alpha) = \sum_i \ln f(y_i) = \sum_i \left\{ (\alpha - 1) \ln y_i - \frac{y_i}{\exp(x_i')\beta} - \alpha x_i' \beta - \ln \Gamma(\alpha) \right\} \)

(b) Differentiation yields
\[
\frac{\partial \ln L}{\partial \beta} = \sum_i \left( \frac{y_i}{\exp(x_i')\beta} x_i - \alpha x_i \right) = \sum_i \left( \frac{y_i - \alpha \exp(x_i')\beta}{\exp(x_i')\beta} x_i \right) = 0.
\]
\[
\frac{\partial \ln L}{\partial \alpha} = \sum_i \left( \ln y_i - x_i' \beta - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \right) \left( \frac{y_i - \alpha \exp(x_i')\beta}{\exp(x_i')\beta} \right) x_i = 0.
\]

(c) Easiest to derive the outer product of the gradient estimate \( \hat{B}^{-1} \). This yields for \( \theta = [\beta' \alpha]' \).
\[
\hat{\theta} = \left[ \sum_i \left( \frac{y_i - \alpha \exp(x_i')\beta}{\exp(x_i')\beta} \right)^2 x_i x_i' \right]^{-1} \sum_i \left( \ln y_i - x_i' \beta - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \right) \left( \frac{y_i - \alpha \exp(x_i')\beta}{\exp(x_i')\beta} \right) x_i.
\]
Or can use Hessian which \( \hat{A}^{-1} \) yields after some algebra yields
\[
\hat{\theta} = \left[ \sum_i \left( \frac{y_i - \alpha \exp(x_i')\beta}{\exp(x_i')\beta} \right) x_i x_i' \right] \sum_s \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - \sum_i \left( \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \right)^2 \right]^{-1}
\]

Note: In general we use \( -E \left[ \frac{\partial^2 \ln L}{\partial \beta \partial \beta} \right] \). Here
\[
\left( \begin{bmatrix} E \left[ \frac{\partial^2 \ln L}{\partial \beta \partial \beta} \right] & E \left[ \frac{\partial^2 \ln L}{\partial \beta \partial \alpha} \right] \\ E \left[ \frac{\partial^2 \ln L}{\partial \alpha \partial \beta} \right] & E \left[ \frac{\partial^2 \ln L}{\partial \alpha \partial \alpha} \right] \end{bmatrix} \right)^{-1} = \left( \begin{bmatrix} E \left[ \frac{\partial^2 \ln L}{\partial \beta \partial \beta} \right] \right)^{-1} \left( \begin{bmatrix} E \left[ \frac{\partial^2 \ln L}{\partial \beta \partial \alpha} \right] \right)^{-1}
\]
except in the special case that \( E \left[ \frac{\partial^2 \ln L}{\partial \beta \partial \alpha} \right] = 0. \)

(d) In general the MLE for both \( \beta \) and \( \alpha \) will be inconsistent.
Here there is some hope that MLE for \( \beta \) may be consistent, since \( E[\partial \ln L / \partial \beta] = 0 \) requires only correct specification of the mean (then \( E \left[ \sum_i \left( \ln y_i - x_i' \beta - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \right) \frac{y_i - \alpha \exp(x_i')\beta}{\exp(x_i')\beta} \right] = 0 \)). [Half credit for saying this].

But \( E[\partial \ln L / \partial \alpha] = 0 \) requires the much stronger assumption that \( E[\ln y_i] = x_i' \beta + \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \)
(then \( E \left[ \sum_i \left( \ln y_i - x_i' \beta - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \right) \right] = 0 \)).

This fails and the two equations jointly estimated will yield inconsistent estimates.

One way to see this is that \( \hat{\alpha} \) inconsistent then contaminates \( \beta \) that solves \( \sum_i \left( \frac{y_i - \alpha \exp(x_i')\beta}{\exp(x_i')\beta} \right)^2 \) = 0.

More formally, the information matrix is not block-diagonal as \( E[\partial^2 \ln L / \partial \beta \partial \alpha] \neq 0 \) and estimation of \( \alpha \) effects estimation of \( \beta \).

(e) Two possible methods are based on \( E[y_i | x_i] = \exp(x_i' \beta) \) are
NLS of \( y_i \) on \( \exp(x_i' \beta) \) which minimizes \( \sum_i (y_i - \exp(x_i' \beta))^2 \).
MM estimation based on \( E[(y_i - \exp(x_i' \beta))x_i] = 0 \) which solves \( \sum_i (y_i - \exp(x_i' \beta))x_i = 0 \).

(f) Here \( E[y] = \alpha \lambda \) and \( V[y] = \alpha \lambda^2 \).
So \( E[x(y - \alpha \exp(x' \beta))] = 0 \) and \( E[(y - \alpha \exp(x' \beta))^2 - 1] = 0 \).
Let \( h(y_i, x_i, \alpha, \beta) = [x_i (y_i - \alpha \exp(x_i' \beta))]' (y_i - \alpha \exp(x_i' \beta))^2 - 1 \)' \( \left( (y_i - \alpha \exp(x_i' \beta))^2 - 1 \right)' \).

The GMM estimator minimizes
\[
Q_N(\alpha, \beta) = \frac{1}{N} \left( \sum_i h(y_i, x_i, \alpha, \beta) \right)' W_N \left( \sum_i h(y_i, x_i, \alpha, \beta) \right),
\]
where any full rank weighting matrix will do since this is just-identified.
2. (a) Here \( \Pr[y = 0] = \Pr[y^* = 0] = e^{-\mu} = \exp(-\exp(\mathbf{x}'\mathbf{\beta})) \). So 
\[
\Pr[y = 1] = 1 - \Pr[y = 0] = 1 - \exp(-\exp(\mathbf{x}'\mathbf{\beta})).
\]
Estimate by binary MLE. \( \hat{\mathbf{\beta}} \) maximizes \( L_N(\mathbf{\beta}) = \sum_i y_i \ln(1 - \exp(-\exp(\mathbf{x}'\mathbf{\beta}))) + (1 - y_i) \ln(\exp(-\exp(\mathbf{x}'\mathbf{\beta}))) \).

(b) This is ordered model
\[
\begin{align*}
p_0 & = \Pr[y = 0] = \Pr[y^* = 0] = e^{-\mu} = \exp(-\exp(\mathbf{x}'\mathbf{\beta})). \\
p_1 & = \Pr[y = 1] = \Pr[y^* = 1] = e^{-\mu - \mu} = \exp(\beta_1 \mathbf{x}_1') \exp(-\exp(\mathbf{x}'\mathbf{\beta})). \\
p_2 & = \Pr[y = 2] = 1 - p_0 - p_1.
\end{align*}
\]
Estimate by multinomial MLE. \( \hat{\mathbf{\beta}} \) maximizes \( L_N(\mathbf{\beta}) = \sum_i (y_i \ln p_{i0} + y_i \ln p_{i1} + y_i \ln p_{i2}) \) where 
\[
y_{i0} = 1 \text{ if } y_i = 0, \ y_{i1} = 1 \text{ if } y_i = 1, \ y_{i2} = 1 \text{ if } y_i = 2.
\]

(c) For notational simplicity initially suppress conditioning on \( \mathbf{x} \)
\[
f(y) = f(y^*|y^* \geq 1) = \frac{f(y^*)}{\Pr[y^* \geq 1]} = \frac{e^{-\mu y^*/y^!}}{(1 - \Pr[y^* = 0])} = \frac{e^{-\mu y/y!}}{(1 - e^{-\mu})}
\]
So
\[
\ln f(y|x) = -\exp(\mathbf{x}_i'\mathbf{\beta}) + y_i \mathbf{x}_i'\mathbf{\beta} - \ln y_i! - \ln(1 - e^{-\exp(\mathbf{x}_i'\mathbf{\beta})}).
\]

(d) Very few got this.
\[
\begin{align*}
\mathbb{E}[y] & = \mathbb{E}[y^*|y^* \geq 1] \\
& = \sum_{y^*=1}^{\infty} y^* \frac{f(y^*)}{\Pr[y^* \geq 1]} = \frac{1}{\Pr[y^* \geq 1]} \sum_{y^*=1}^{\infty} y^* f(y^*) = \frac{1}{\Pr[y^* \geq 1]} \sum_{y^*=0}^{\infty} y^* f(y^*) = \frac{1}{1 - e^{-\mu}} \mu,
\end{align*}
\]
using \( \sum_{y^*=0}^{\infty} y^* f(y^*) \) is \( \mathbb{E}[y^*] \) and we were told that for the Poisson that \( \mathbb{E}[y^*] = \mu. \)

(e) Since
\[
\mathbb{E}[y_i|x_i] = \frac{\exp(\mathbf{x}_i'\mathbf{\beta})}{1 - e^{-\exp(\mathbf{x}_i'\mathbf{\beta})}}
\]
do nonlinear least squares regression of \( y_i \) on \( \exp(\mathbf{x}_i'\mathbf{\beta})/(1 - e^{-\exp(\mathbf{x}_i'\mathbf{\beta})}) \).
Or do MM based on \( \sum_i \mathbf{x}_i (y_i - \exp(\mathbf{x}_i'\mathbf{\beta})/(1 - e^{-\exp(\mathbf{x}_i'\mathbf{\beta})})) = \mathbf{0}. \)

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) Remarkably few got this completely correct. Simplest is Lindeberg-Levy CLT.
Let \( \{X_i\} \) be iid with \( \mathbb{E}[X_i] = \mu \) and \( \mathbb{V}[X_i] = \sigma^2. \) Then \( Z_N = \frac{\sum_{i=1}^{N} X_i - N \mu}{\sqrt{N} \sigma} \) \( \xrightarrow{d} \) \( N[0,1]. \)
[Other CLT's can be given].

(c) \( y^* = 1 + 2x + u \) where \( x \sim \mathcal{N}[0,1] \) and \( u \sim \mathcal{N}[0, x^2] \)
We observe \( y = 1 \) if \( y^* > 0 \) and \( y = 0 \) if \( y^* \leq 0. \)

(d) In (c) I had meant to generate \( y \) from a Tobit model but mistekenly generated a binary variable. So the natural thing would be to try probit estimation. Tobit is inappropriate.
But I gave full vredit if you thought Tobit was still apropiate, but then noted that the Tobit MLE of \( y \) on \( x \) will be inconsistent for \( \mathbf{\beta} \) as the error here is heteroskedastic. It is not enough to say that standard errors will be wrong. Inconsistency is the most serious problem.

(e) I had intended the question to be about the sample selection model, but if you answered correctly for the Tobit model you also got full credit. The sample selection model is
\[
\begin{align*}
y_1^* & = \mathbf{x}_1'\mathbf{\beta}_1 + \varepsilon_1 \\
y_2^* & = \mathbf{x}_2'\mathbf{\beta}_2 + \varepsilon_2.
\end{align*}
\]
and we observe $y_1 = \begin{cases} 1 & \text{if } y_1^* > 0 \\ 0 & \text{if } y_1^* \leq 0 \end{cases}$, and $y_2 = \begin{cases} y_2^* & \text{if } y_1^* > 0 \\ - & \text{if } y_1^* \leq 0 \end{cases}$.

The errors $(\varepsilon_1, \varepsilon_2)$ have means $(0, 0)$, variances $(1, \sigma_2^2)$ and covariance $\rho \sigma_2^2$.

$\varepsilon_1$ is standard normal. If the MLE is used $(\varepsilon_1, \varepsilon_2)$ are joint normal.

(f) $B$ times do the following.

- Completely resample with replacement all the data $\{(y_{1i}, y_{2i}, x_{1i}, x_{2i})$ for $i = 1, \ldots, N\}$
- For each resample get estimate $\hat{\beta}_b$ and form $\hat{\text{ME}}_b = \exp(\hat{\gamma}_b \beta_b)$.

Standard error is the standard deviation of the $B \hat{\text{ME}}_b$.

(g) This is optimal two-step GMM. Minimize

$$Q_N(\theta) = \frac{1}{N} \left( \sum_i h(w_i, \theta) \right)' \hat{S}^{-1} \left( \sum_i h(w_i, \theta) \right),$$

where $\hat{S} = \sum_{i=1}^N h(w_i, \hat{\theta})h(w_i, \hat{\theta})'$ and $\hat{\theta}$ is a consistent initial estimate such as first-step GMM.

4. (a) No. The default se's assume independence of $u_{it}$ and $u_{is}$. But the error $u_{it}$ is likely to be positively correlated with $u_{is}, i \neq s$, decreasing the informational content of the data. Panel robust se's adjust for this.

(b) Yes. The RE-GLS does control for clustering so might expect the two to be similar. The difference is due to the wrong model for clustered errors (equicorrelation) or heteroskedasticity.

(c) $y_{it} = \alpha_i + x_{it} \beta + u_{it} \Rightarrow (y_{it} - \bar{y}_i) = (x_{it} - \bar{x}_i)' \beta + (u_{it} - \bar{u}_i)$.

So do OLS of $(y_{it} - \bar{y}_i)$ on $(x_{it} - \bar{x}_i)$. (Other methods are possible).

(d) `xtreg y x, vce(robust)` or `xtreg y x, vce(Cluster id)`.

(e) That the RE estimator is fully efficient under $H_0$. This requires that the error $y_{it} = \alpha_i + \varepsilon_{it}$ where both $\alpha_i$ and $\varepsilon_{it}$ are i.i.d.

(f) Usual Hausman test is $H = (\hat{\beta}_{\text{FE}} - \hat{\beta}_{\text{RE}})'(\hat{V}(\hat{\beta}_{\text{FE}}) - \hat{V}(\hat{\beta}_{\text{RE}}))^{-1}(\hat{\beta}_{\text{FE}} - \hat{\beta}_{\text{RE}}) \sim \chi^2(q)$.

$\hat{\beta}_{\text{FE}} = 0.17$ with default standard error $0.03$ and $\hat{\beta}_{\text{RE}} = 0.12$ with default standard error $0.02$.

Note that if indeed the RE is fully efficient then the default standard errors are correct and we would use these.

$H = (0.17 - 0.12)^2 / (0.03^2 - 0.02^2) = 0.0025 / 0.0005 = 5 > \chi^2_{0.05}(1) = 3.84$.

Reject $H_0$. Conclude that there is a difference so FE is the model.

(g) Now $H = (\hat{\beta}_{\text{FE}} - \hat{\beta}_{\text{RE}})'(\hat{V}(\hat{\beta}_{\text{FE}}) + \hat{V}(\hat{\beta}_{\text{RE}}) - 2 + \hat{\text{Cov}}[\hat{\beta}_{\text{RE}}, \hat{\beta}_{\text{FE}}])^{-1}(\hat{\beta}_{\text{FE}} - \hat{\beta}_{\text{RE}}) \sim \chi^2(q)$.

$\hat{\beta}_{\text{FE}} = 0.17$ with robust s.e. $0.08$, $\hat{\beta}_{\text{RE}} = 0.12$ with robust s.e. $0.05$, and $\hat{\text{Cov}}[\hat{\beta}_{\text{RE}}, \hat{\beta}_{\text{FE}}] = 0.02^2$.

$H = (0.17 - 0.12)^2 / (0.08^2 + 0.05^2 - 2 \times 0.02^2) = 0.0025 / 0.0081 = 0.31 < \chi^2_{0.05}(1) = 3.84$.

Reject $H_0$. Conclude that there is no difference so RE is the model.

(h) Stacking we have $y_i = X_i \beta + u_i$, where $y_i$ and $u_i$ are $T \times 1$ and $X_i$ is $T \times k$ with $i^{th}$ row $x_i$.

Then $\hat{\beta} = (\sum_i X_i'X_i)^{-1} \sum_i X_i'y_i = \beta + (\sum_i X_i'X_i)^{-1} \sum_i X_i'u_i$.

The asymptotic variance is $\sum_i X_i'X_i)^{-1} \text{Var}(\sum_i X_i'u_i)(\sum_i X_i'X_i)^{-1}$.

Given independence over $i$ and $E[u_i | x_i] = 0$ this becomes $\sum_i X_i'X_i)^{-1} \sum_i E[X_i'u_iu_i'X_i](\sum_i X_i'X_i)^{-1}$.

So use $(\sum_i X_i'X_i)^{-1} (\sum_i X_i'u_iu_i'X_i)(\sum_i X_i'X_i)^{-1}$ where $\hat{u}_i = y_i - X_i\hat{\beta}$.

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

Scores out of 50
75th percentile 38 (76%) A 36 and above
Median 31.5 (63%) A- 30 and above
25th percentile 26 (52%) B+ 24 and above