Problem 1

a) We have

$$\ln f(y|n,p) = y\ln\frac{p}{1-p} + n\ln(1-p) - \ln(1-(1-p)^n) + \ln\binom{n}{y} \quad (1)$$
$$= y\theta - b(\theta) + c(y) \quad (2)$$

where the canonical parameter

$$\theta = \ln\frac{p}{1-p}$$

and

$$b(\theta) = -n\ln(1-p) + \ln(1-(1-p)^n).$$

b) Differentiation $b(\theta)$, keeping in mind that $p = \frac{e^\theta}{1+e^\theta}$ such that $dp/d\theta = p(1-p)$, we find that

$$EY = b'(\theta) = \left(\frac{n}{1-p} + \frac{n(1-p)^{n-1}}{1-(1-p)^n}\right)p(1-p)$$
$$= np\left(1 + \frac{(1-p)^{n-1}}{1-(1-p)^n}\right)$$
$$= \frac{np}{1-(1-p)^n}.$$

Asymptotically, as $n \to \infty$, $P(X=0) \to 0$ and hence EY should be asymptotically equal to np which indeed is the case since

$$\lim_{n\to\infty} EY/(np) = \lim_{n\to\infty}\frac{1}{1-(1-p)^n} = 1.$$

Problem 2

a) Each observation

$$y_i \sim \text{Poisson}\,\mu_i$$

and independent, and

$$\ln\mu_i = \mu + \alpha_{j(i)} + \beta_{k(i)}$$

$i = 1, 2, \ldots, 112$, where $j(i) \in \{1, 2\}$ is the location at which observation i was made and $k(i) \in \{1, 2, \ldots, 7\}$ is the weekday. The unknown parameters are μ, α_1, α_2 and $\beta_1, \beta_2, \ldots, \beta_7$ but to make the model identifiable the constraints $\alpha_1 = 0$ and $\beta_1 = 0$ have been imposed.

The expected number of cyclists passing road point B on a monday becomes

$$e^{2.9767-0.65555} = 10.187$$

The estimated coefficient for `locationB`, that is, $\hat{\alpha}_2 = -0.65555$ means that the expected number of cyclist passing point B differ from the expected number at point A (the reference level of location) by a factor of $e^{-0.65555} = 0.519$, that is, the expected number of 48% lower at location B compared to A.

b) A model H_0 is nested in H_1 if the distribution of the data under H_0 is identical to that under H_1 for a particular subset of possible parameter values under H_1. In terms of their designmatrices $\mathbf{X}_0$ and $\mathbf{X}_1$, the models are nested if the columnspace of $\mathbf{X}_0$ is a subset of the columnspace of $\mathbf{X}_1$. In the present case, `mod0` corresponds to `mod1` with $\beta_1 = \beta_2 = \cdots = \beta_5$ and $\beta_6 = \beta_7$.

The likelihood ratio statistic $LRT = D_0 - D_1$ is chi-square with $p_1 - p_0 = 8 - 3 = 5$ degrees of freedom so the critical value of the test is $\chi^2_{0.05,5} = 11.07$. Given the observed value $LRT = 6.27$ we can thus not reject the null hypothesis that there is a difference between the weekdays beyond the weekend effect.

c) Under the null hypothesis that the model is correct (including the hypothesis of no overdispersion), the deviance D is chi-square with $n - p = 109$ degrees of freedom. This gives a critical value of 134.36 and given the observed value of $D = 141.31$ we can reject the null hypothesis.

Possible mechanisms that could generate overdispersion could be cyclists passing in clusters (such the the Poisson process assumption of indendence between disjoint time intervals is violated), missing covariates and wrong choice of link functions.

An estimate of the dispersion parameter φ is $\hat{\varphi} = D/(n-p) = 141.31/109 = 1.29$. The adjusted estimated standard error for $\hat{\beta}_2$ becomes $\widehat{SE(\hat{\beta}_2)} = 0.05294 \cdot \sqrt{1.29} = 0.06013$.

d) For day number $i = 1, 2, \ldots, 56$ and for observation $j = 1, 2$, conditional on γ_i, the number of cyclists passing $y_{ij}|\gamma_i \sim \text{Poisson}\, \mu_{ij}$ and conditionally independent with $\ln \mu_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \gamma_i$. In addition, the random effects $\gamma_i \sim N(0, \tau^2)$ and independent. The unknow parameters are $\boldsymbol{\beta}$ and τ^2.

e) Conditional on γ_i, $E(y_{ij}|\gamma_i) = e^{\mathbf{x}_{ij}^T \boldsymbol{\beta} + \gamma_i}$.

The conditional covariance $\text{Cov}(y_{i1}, y_{i2}|\gamma_i) = 0$ because the observations are conditionally independent.

Hence, using the law of total covariance, and that $e_i^{\gamma} \sim \text{lognormal}(0, \tau^2)$,

$$\begin{aligned}
\text{Cov}(y_{i1}, y_{i2}) &= E\,\text{Cov}(y_{i1}, y_{i2}|\gamma_i) + \text{Cov}(E(y_{i1}|\gamma_i), E(y_{i2}|\gamma_i)) \\
&= E0 + \text{Cov}(e^{\mathbf{x}_{i1}^T\boldsymbol{\beta}+\gamma_i}, e^{\mathbf{x}_{i2}^T\boldsymbol{\beta}+\gamma_i}) \\
&= e^{(\mathbf{x}_{i1}+\mathbf{x}_{i2})^T\boldsymbol{\beta}}\,\text{Var}(e^{\gamma_i}) \\
&= e^{(\mathbf{x}_{i1}+\mathbf{x}_{i2})^T\boldsymbol{\beta}} e^{\tau^2}(e^{\tau^2} - 1)
\end{aligned}$$

f) Testing the GLM without a random intercept against the random intercept GLMM amounts to testing the null hypothesis $H_0 : \tau^2 = 0$ against $H_1 : \tau^2 > 0$. Under this null hypothesis the likelihood ratio statistic is approximately a 50-50% mixture of chi-squares with 0 and 1 degrees of freedom. Thus, the critical value c satisfy

$$P(\text{LRT} > c|H_0) = \frac{1}{2}P(\chi_0^2 > c) + \frac{1}{2}P(\chi_1^2 > c) = \alpha.$$

Since $\chi_0^2 = 0$, $P(\chi_1^2 > c) = 2\alpha$ so the critical value $c = \chi^2_{2\alpha,1} = \chi^2_{0.1,1} = 2.7055$.

The observed value is $LRT = 2(-312.6 - (-314.26)) = 3.32$ so we can reject H_0 in factor of the random intercept GLMM.

g) The GLM `mod` assumes independent observations (wich don't agree with point e) and f)), no overdispersion (which don't agree with point c)). Neglecting this will lead to underestimation of the standard errors. Even when adjusting for overdispersion for `mod0`, the standard errors are still most likely underestimated because the quasi-Poisson model still incorrectly assumes independent observations.

Thus, we can't trust the standard errors for `mod0`.

It can also noted that while the over-dispersion corrected standard error in point c) is very close corresponding standard error of the GLMM, it is still most likely underestimated as the MLEs obtained using the GLMM are likely more efficient.

h) Letting $\mathbf{y}_i$ denote all observations on day i, the likelihood of the GLM can be expressed as

$$
\begin{aligned}
L(\boldsymbol{\beta}, \tau^2) &= \prod_{i=1}^{56} f(\mathbf{y}_i|\boldsymbol{\beta}, \tau^2) && \text{(independence between days)} \\
&= \prod_{i=1}^{56} \int f(\mathbf{y}_i, \gamma_i|\boldsymbol{\beta}, \tau^2) d\gamma_i && \text{(law of total probability)} \\
&= \prod_{i=1}^{56} \int f(\mathbf{y}_i|\gamma_i, \boldsymbol{\beta}) f(\gamma_i|\tau^2) d\gamma_i && \text{(the product rule)} \\
&= \prod_{i=1}^{56} \int \left(\prod_{j=1}^{2} f(y_{ij}|\gamma_i, \boldsymbol{\beta}) \right) f(\gamma_i|\tau^2) d\gamma_i && \text{(conditional independence)}
\end{aligned}
$$

Since the integrands are typically quite well approximated by Gaussian functions, good methods for computing the integrals numerically are adaptive Gauss-Hermite quadrature and the Laplace approximation (which corresponds to adaptive Guass-Hermite quadrature with a single quadrature point).