1. Questão 1

a) Temos que:

$$
\begin{aligned}
f_Y(y) &= \exp\left\{-\frac{\phi y}{2\mu^2} + \frac{\phi}{\mu} - \frac{\phi}{2y} + \frac{1}{2}\left(\ln\phi - \ln(2\pi y^3)\right)\right\} \\
&= \exp\left\{\phi\left(-\frac{y}{2\mu^2} + \frac{1}{\mu}\right) - \frac{\phi}{2y} + \frac{1}{2}\left(\ln\phi - \ln(2\pi y^3)\right)\right\} \\
&= \exp\left\{\phi\left[y\theta - b(\theta)\right] + c(y,\phi)\right\} I_A(y), \qquad (1)
\end{aligned}
$$

em que $\theta = -\frac{1}{2\mu^2}$, $\left(\mu = \sqrt{\frac{1}{-2\theta}}\right)$, $\phi = \phi$, $b(\theta) = -\sqrt{-2\theta}$, $c(y_i,\phi) = -\frac{\phi}{2y} + \frac{1}{2}\left(\ln\phi - \ln(2\pi y^3)\right)$, $A = (0,\infty)$.

b) Temos que:

$$
\begin{aligned}
b'(\theta) &= (-2\theta)^{-1/2} \to \mathcal{E}(Y) = \mu \\
b''(\theta) &= \frac{2}{2}(-2\theta)^{-3/2} = (-2\theta)^{-3/2} \to \mathcal{V}(Y) = \phi^{-1}\mu^3
\end{aligned}
$$

c) Temos que (do item a)):

$$
\begin{aligned}
L(\beta) &\propto \exp\left\{-\frac{1}{\beta^2}\sum_{i=1}^{n}\frac{y_i}{2X_i^2} + \frac{1}{\beta}\sum_{i=1}^{n}\frac{1}{X_i}\right\} \\
\to l(\beta) &= -\frac{1}{\beta^2}\sum_{i=1}^{n}\frac{y_i}{2X_i^2} + \frac{1}{\beta}\sum_{i=1}^{n}\frac{1}{X_i} + c \\
S(\beta) &= \frac{1}{\beta^3}\sum_{i=1}^{n}\frac{y_i}{X_i^2} - \frac{1}{\beta^2}\sum_{i=1}^{n}\frac{1}{X_i} \\
H(\beta) &= -\frac{3}{\beta^4}\sum_{i=1}^{n}\frac{y_i}{X_i^2} + \frac{2}{\beta^3}\sum_{i=1}^{n}\frac{1}{X_i} \\
I(\beta) &= \frac{3}{\beta^3}\sum_{i=1}^{n}\frac{1}{X_i} - \frac{2}{\beta^3}\sum_{i=1}^{n}\frac{1}{X_i} = \frac{1}{\beta^3}\sum_{i=1}^{n}\frac{1}{X_i}
\end{aligned}
$$

d) ANULADA, teria de ser $H_0 : \beta = 1$ vs $H_1 : \beta_1 \neq 1$.

2. Questão 2

a) Temos que:

$$f_Y(y) = \exp\left\{y\ln p + r\ln(1-p) + \ln\left[\frac{\Gamma(y+r)}{\Gamma(y+1)\Gamma(r)}\right]\right\},$$

em que $\theta = \ln p$, $b(\theta) = -r\ln(1-p)$, $\phi = 1$, $c(y,\theta) = \ln\left[\frac{\Gamma(y+r)}{\Gamma(y+1)\Gamma(r)}\right]$, $A = \{0,1,2,...\}$. Ademais $(p = e^\theta)$

$$\begin{aligned} b(\theta) &= -r\ln\left(1-e^\theta\right) \to b'(\theta) = r\frac{e^\theta}{1-e^\theta} \to \mathcal{E}(Y) = \mu = \frac{rp}{1-p} \to \frac{r}{\mu} = \frac{1}{p} - 1 \\ &\to p = \frac{\mu}{\mu+r} \end{aligned}$$

b) Temos que:

$$\begin{aligned} f_Y(y) &= \exp\left\{y\ln\left(\frac{\mu}{\mu+r}\right) - r\ln(r+\mu) + r\ln r + \ln\left[\frac{\Gamma(y+r)}{\Gamma(y+1)\Gamma(r)}\right]\right\}\mathbb{1}_A(y) \\ &= \exp\left\{\phi\left(y\theta - b(\theta)\right) + c(y,\theta)\right\}\mathbb{1}_A(y), \end{aligned}$$

em que $\theta = \ln\left(\frac{\mu}{\mu+r}\right) \to e^{-\theta} = 1 + \frac{r}{\mu} \to \mu = \frac{r}{e^{-\theta}-1}$, $b(\theta) = r\ln(r+\mu) = r\ln\left(\frac{re^{-\theta}}{e^{-\theta}-1}\right) = r\left(\ln r - \theta - \ln(e^{-\theta}-1)\right)$, $\phi = 1$, $c(y,\phi) = r\ln r + \exp\left[\ln\frac{\Gamma(y+r)}{\Gamma(y+1)\Gamma(r)}\right]$, $A = \{0,1,2,...\}$.

c) Temos que:

$$\begin{aligned} Y_i &\overset{ind.}{\sim} \text{BN}(\mu_i) \\ \ln\left(\frac{\mu_i}{\mu_i+r}\right) &= \sum_{j=1}^{p} x_{ji}\beta_j \end{aligned}$$

d) Temos que:

$$\begin{aligned}
\ln\left(\frac{\mu_i}{\mu_i + r}\right) &= \beta \rightarrow 1 + \frac{r}{\mu_i} = e^{-\beta} \rightarrow \mu_i = \frac{r}{e^{-\beta} - 1} \\
\frac{\partial \mu_i}{\partial \boldsymbol{\beta}} &= \frac{-re^{-\beta}}{(e^{-\beta} - 1)^2} = \mu_i(\mu_i + r), \mu_i + r = \frac{re^{-\beta}}{e^{-\beta} - 1}
\end{aligned}$$

$$\begin{aligned}
l(\beta) &= \sum_{i=1}^{n} y_i \ln \mu_i - \sum_{i=1}^{n} (y_i + r) \ln (\mu_i + r) + c \\
S(\beta) &= \sum_{i=1}^{n} \frac{y_i}{\mu_i} \mu_i(\mu_i + r) - \sum_{i=1}^{n} \frac{y_i + r}{\mu_i + r} \mu_i(\mu_i + r) \\
&= \sum_{i=1}^{n} y_i(\mu_i + r) - \sum_{i=1}^{n} (y_i + r)\mu_i \\
S(\beta) &= \mu \sum_{i=1}^{n} y_i + rn\overline{y} - \mu n\overline{y} - r\mu n = rn\left(\overline{y} - \mu\right) \\
H(\beta) &= -rn\mu(r + \mu); I(\beta) = rn\mu(r + \mu).
\end{aligned}$$

Sob as condições de regularidade, temos que $\widehat{\boldsymbol{\beta}} \approx N\left(\beta, (rn\mu(r+\mu))^{-1}\right)$, para n suficientemente grande.

3. Questão 3

   a) A estrutura linear para cada covariável parece ser apropriada, além do que utilizar os valores padronizados permite comparar, de forma mais direta, as estimativas dos parâmetros de regressão. Ademais, o modelo gama tem uma certa vantagem, pois garante que as médias preditas estejam dentro do espaço paramétrico.

   b) Os dois modelos aparentem ajustes razoáveis, nos quais percebemos, através dos respectivos resíduos, ausência de heterocedasticidade e do correlação e bom comportamento dos preditores escolhidos. Além disso, os praticamente todos os resíduos estão dentro das bandas de confiança (QQ plots e worm plots). No entanto, o modelo M1 apresenta um ajuste modicamente melhor.

   c) Pelo exposto no item b) e pelos resultados da Tabela 1) não restam dúvidas do que o M2 é preferível ao M1. Com efeito, todas as medidas indicam que o M2 é melhor que o M1. Adicionalmente, os valores preditos e observados são satisfatoriamente próximos entre si, em ambos os modelos e, apesar de assumir uma função de ligação identidade, o M1 fornece valores preditos que respeitam o espaço amostral da resposta.

   d) Pelas Tabelas 2 e 3 podemos ver que ambas as covariáveis impactam, significativamente, a resposta. Os anos impactam de forma negativa enquanto que a posição na empresa impacta de forma positiva, sendo o impacto desta bem maior do que daquela.

4. Questão 4

   a) Como se trata de contagem, a resposta é positiva e todas as covariáveis são qualitativas (fatores), os modelo foram boas escolhas. Ademais, devido ao pequeno (as vezes unitário) número de observações ao longo das combinações dos níveis dos fatores, desconsiderar todas as interação também fora uma boa opção. Com as informações apresentadas, não é possível conjecturar a respeito de uma possível (sub/super) dispersão.

   b) Nenhum dos dois modelos apresentou um ajuste satisfatório. Para ambos, os resíduos indicam uma heterocedasticidade não captada e indícios de superdispersão que, mesmo no M2, não foi acomodada de forma apropriada. Aparentemente, o M2 apresentou, apesar dos resultados do CI's serem favoráveis ao M1, um ajuste um pouco menos pior.

   c) Como no modelo não há interações de nenhuma ordem, a verificação dos efeitos principais (de cada fator) ficam mais simples. Há efeito de todas as 3 covariáveis. No entanto, enquanto que para a covariável período, os dois únicos níveis diferem entre si, há uma equivalência entre os níveis A, D e E, e uma aparente equivalência entre os níveis B e C, sendo estes inferiores aos tipos anteriores (tipos de navio). Adicionalmente, há uma diferença entre todos os anos de fabricação, sendo que os anos intermediários parecem ser os piores. No modelo reduzido a ser ajustado, eu excluiria os parâmetros $(\beta_4, \beta_5,)$ e imporia que $\beta_2 = \beta_3$.

   d) Interpretações:

      - e^{α} :número médio de avarias para navios do tipo A, fabricados entre 1960-64, com o período de operação entre 196-1974.
      - e^{β_2}: razão entre o número médio de avarias entre os tipos B e A de navios, independentemente dos outros dois fatores.
      - e^{γ_2}: razão entre o número médio de avarias entre os anos de fabricação 1965-69 e 1960-64, independentemente dos outros dois fatores.
      - e^{δ_2}: razão entre o número médio de avarias entre os períodos de fabricação 1975-79 e 1960-74, independentemente dos outros dois fatores.