(1, 3.5pt) For `age` and `iq`, a Gaussian glm (i.e., ordinary linear model) could be used. For `sleep`, `sex` and `life`, a binomial glm could be used. For `anxiety` and `depression`, a multinomial GLM with a special treatment for ordinal responses (e.g., proportional odds model, ordered probit model, or proportional hazard model) could be used.

(2, 1pt) Because 0+8+1+5+55+30+4+7=110, the number of cases with missing values is at least 8, calculated by $118 - 110 = 8$.

(3, 2pts) Sampling scheme 1. One possible scenario is to collect data for a certain period of time, meaning that 118 is produced by a random variable, not a fixwd value set prior to data collection.

(4, 1pt) The value is zero because it (the last line in the output) corresponds to the satured model in which $y = \hat{y}$.

(5, 2pts) Since only the `life:sex` interaction is insignficant and `sleep` is the common factor appearing in the 2 significant interactions `sleep:life` and `sleep:sex`, we can say that probability of having considered suicide and having lost interest in sex are independent given knowledge of the sleeping status (i.e., `sex` and `life` are conditionally independent given `sleep`).

(6, 2pts) Conditional independence between `sex` and `life` given `sleep` does not necessarily imply marginal independence between `sex` and `life` without conditioning on `sleep`. From the 3-way table, we can see that when `sleep` is fixed, the indpendence status of `sex` and `life` varies. For instance, when `sleep`=1, the probability of `life`=2 is higher across all sex categories. However, when `sleep`=2, the probabilities of `life`=1 and `life`=2 are more similar across all sex categories. This would lead to a lack of independence in the 2-way table of `sex` and `life`.

(7, 1pt) There are roughly about as many patients who have considered suicide as those who have not (i.e., the marginal distribution of `life` is uniform).

(8, 4pts) When the marginal totals for the four combinations of (`sleep`, `sex`) are considered fixed, `life` follows a binomial distribution where the probability of `life`=1 is a function of `sleep` and `sex`. In this case, a binomial GLM can be used, with the response being the 4 counts of `life`=1 (or `life`=2) in the 3-way table, denoted by $y_{\texttt{life}}$, and the predictors being `sleep` and `sex`, each with two levels (resulting in 4 level combinations). When `sex` and `life` are conditionally independent given `sleep` (i.e., the 2-fis in the fitted log-linear model are `sleep:sex` and `sleep:life`), the binomial GLM $y_{\texttt{life}} \sim$ `sleep` would produce a good fit.

(9, 2pts) The null hypothesis of the MH test is that X1 and X2 are independent given X3 (i.e., conditional independence). According to the answer to question (5), we know that conditional independence exists only when X3 is `sleep`, meaning when X3 is `sex` or `life`, the other two variables do not exhibit conditional independence. When X3=`sex` or `life`, the null hypothesis of MH test would be rejected, resulting in significant results. Thus, X3=`sleep` would produce the most insignificant p-value.

(10, 2pts) The MH test statistic can utilize an *exact* null distribution based on the hypergeometric distribution, which is particularly useful when a significant number of cells have low counts (≤ 5), as opposed to the *asymptotic* null distribution used in the analysis of deviance table. Here, half of the cells have low counts.

(11, 2pts) The `unclass` command converts an ordinal variable with k levels into an interger variable, assigning the levels the *numbers* 1, 2, ..., k according to their original order. The scores assigned to levels 1, 2, 3, 4 of `anxiety` are 1, 2, 3, 4. Notice that in R program, these two sets of 1234 have different meanings and are handled differently. The scores for levels 1, 2, 3 of `depress` are 1, 2, 3. The evidence for this being an appropriate assignment includes: (1) the `Oanx:Odep` interaction is extremely significant, and (2) the deviance of 3.3504 on 5 dfs indicates a good fit.

(12, 1pt) The positive sign in the extremely significant coefficient of `Oanx:Odep` gives the evidence.

(13, 1.5pt) $\exp(1.8514 \times (4-1) \times (3-1)) = \exp(11.1084) = 66729.34$, a positive value much larger than 1, indicating a strong positive dependence between anxiety and depression.

(14, 2pts) The coefficient of `sleep` is positive, indicating that patients who sleep normally (`sleep`=1) have a reduced probability of ending up in the higher levels of the depression variable than patients who have sleeping problems (`sleep`=2). Given the way that this variable is coded, this means that those who sleep well are also less likely to be depressed.

(15, 2pts) After adjusting their means appropriately with the predictors, these latent variables follow the same logistic distribution.

(16, 3pts) Notice that there is a negative sign in the front of the coefficient of `sleep` in the model. Therefore, $P(\texttt{depress}{=}1) = P(\texttt{depress}{\leq}1) = \text{logistic}(6.930 - (4.36215) \times 2) = \text{logistic}(-1.7943) = \frac{\exp(-1.7943)}{1+\exp(-1.7943)} = 0.1425463$.

(17, 2pts) The difference in deviance is $158.29 - 153.10 = 5.19$, which is not particularly large on 3 dfs. We prefer the smaller model using just `sleep`.

~~(18, 3pts) Because~~

$$~~f(y|\lambda) = (1/\lambda)\exp(-y/\lambda) = \exp\{y(-1/\lambda) - [-\log(-(-1/\lambda))]\},~~$$

~~we have~~

- ~~$\theta = -1/\lambda$ (i.e., $\lambda = -1/\theta$)~~
- ~~$b(\theta) = -\log(-\theta)$~~
- ~~$\phi = 1$~~
- ~~$a(\phi) = 1$~~
- ~~$c(y, \phi) = 0$~~

~~(19, 2pts) Denote $\mu = E(y)$. Then, because $\mu = b'(\theta) = -\theta^{-1}$, the canonical link is~~

$$~~\theta = \eta = g(\mu) = -\frac{1}{\mu}.~~$$

~~Because $Var(y) = b''(\theta)a(\phi) = \theta^{-2}$, the variance function is~~

$$~~V(\mu) = \mu^2.~~$$

(20, 1pt) Because $-\infty < \eta < \infty$, some choice of the values of the predictors could result in a negative μ, which is undesirable because the mean of an exponential distribution is always positive.