STAT 502 - Homework 4

Due date: Thursday, Nov 11, 11:59PM. Submit your homework solutions to the course Canvas page. Total points: 20. **Late homework will not be accepted.**

1. **(5pt)** Consider again the zinc data set from homework 3. Suppose that the true treatment effect vector τ is $\tau = (\tau_{back}, \tau_{low}, \tau_{high}, \tau_{med}) = (1, -5, 3, 0)$

   (a) **(1pt)** What is the power of the 0.05-level F-test, for H_0: $\tau = (0, 0, 0, 0)$ if the sample size per zinc level is $n = 5$ and $\sigma = 20$?

   (b) **(2pt)** What is the minimal sample size per zinc level for which the power of the 0.05-level F-test, for H_0: $\tau = (0, 0, 0, 0)$ is $\geq 80\%$ if $\sigma = 20$?

   (c) **(2pt)** Given a sample size $n = 10$ per zinc level, what is the largest value σ can take from $\{1, 2, \cdots, 20\}$, so that the power of the 0.05-level F-test, for H_0: $\tau = (0, 0, 0, 0)$ is $\geq 80\%$?

2. **(6pt)** Consider the treatment effects model for data from a completely randomized design:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad i = 1, 2, 3, 4, \quad j = 1, 2$$

   where μ is the grand mean, index i labels the treatments, j labels the replicates and ϵ_{ij} are i.i.d. $\mathcal{N}(0, \sigma^2)$ random errors.

   Let $\boldsymbol{\beta} = (\mu, \tau_1, \tau_2, \tau_3, \tau_4)^T$ be the vector of unknown mean parameters.

   (a) **(2pt)** Let the vector of unknown parameters $\boldsymbol{\beta'}$ be $\boldsymbol{\beta'} = (\mu, \tau_1, \tau_2, \tau_3)^T$. Write the design matrix $\mathbf{X}$ such that $\mathbf{X}$ is of full rank, $\mathbf{X}$ and $\boldsymbol{\beta'}$ incorporate constraints $4 * \tau_1 - 4 * \tau_2 - 4 * \tau_3 - 4 * \tau_4 = 0$, and $E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta'}$.

   (b) **(2pt)** Calculate the hat matrix $\mathbf{H}$ using the design matrix from 2a.

   (c) **(2pt)** Consider hypothesis $H_0 : \tau_2 = 2\tau_3,\ \tau_1 = \tau_2$. Write out a design matrix $\mathbf{X}'$ that has full column rank and whose column span is equal to the space $\{\mathbf{X}\boldsymbol{\beta} : \boldsymbol{\beta} \in \mathbb{R}^4, \tau_1 + \tau_2 + \tau_3 + \tau_4 = 0,\ \tau_2 = 2\tau_3,\ \tau_1 = \tau_2\}$.

3. **(9pt)** We wish to compare how three diets affect weight gain. The diets are differentiated by protein source (beef, pork, or grain). Nine test animals were randomly assigned to the diets, three animal per diet. This gave rise to the following data:

```
  Weight.Gain Source
1        78.0   Beef
2        86.8   Beef
3       103.8   Beef
4        83.7   Pork
5        89.0   Pork
6        99.2   Pork
7        83.8  Grain
8        81.5  Grain
9        86.2  Grain
```

   (a) **(1pt)** Consider model $Y_{ij} = \mu_i + \epsilon_{ij}$, $i = 1, 2, 3$, $j = 1, 2, 3$, with ϵ_{ij} are i.i.d. $\mathcal{N}(0, \sigma^2)$ random variables. Let $\boldsymbol{\beta} = (\mu_1, \mu_2, \mu_3)$ correspond to $(\mu_{beef}, \mu_{pork}, \mu_{grain})$. What is the design matrix $\mathbf{X}$ for which $\mathbb{E}(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$?

(b) **(2pt)** What is the least squares estimate $\hat{\beta}_1$ of β_1?

(c) **(2pt)** Calculate the standard error of $\hat{\beta}_1$.

(d) **(2pt)** What is the 90% confidence interval for β_1?

(e) **(2pt)** Are $\hat{\mu}_{beef}$ and $\hat{\mu}_{pork}$ independent?