Take Home Assignment 1

Jan Ryszkiewicz

Task 1

Task 2

subtask 1

The model presented in task can also be written as:

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & X_1 \\ 1 & 0 & 0 & 0 & X_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 & X_{{}_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 & X_{{}_3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 1 & X_{{}_4} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & X_{45} \end{bmatrix}, \quad \beta = \begin{bmatrix} \mu \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ b \end{bmatrix}, \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{45} \end{bmatrix}.$$

Where each column $2, 3, 4$ in X is filled with $0, 1$ depending on corresponding $\alpha_{i(j)}$ And column 1 corresponds to always present μ

Also α_i defined as:

$$\begin{aligned} \alpha_2 &= \text{Temporary -- Research/Academic (relative to Permanent)} \\ \alpha_3 &= \text{Temporary -- Private Consultant (relative to Permanent)} \\ \alpha_4 &= \text{Freelance (relative to Permanent)} \end{aligned}$$

We know that:

$$\begin{aligned} \mathrm{Cov}(\hat{\alpha}_2, \hat{\alpha}_3) &= 22{,}000{,}000 \\ \mathrm{Cov}(\hat{\alpha}_2, \hat{\alpha}_4) &= 20{,}000{,}000 \\ \mathrm{Cov}(\hat{\alpha}_3, \hat{\alpha}_4) &= 21{,}000{,}000 \end{aligned}$$

And:

$$\hat{\text{SE}}(\hat{\mu}) = 20{,}000$$
$$\hat{\text{SE}}(\hat{\alpha}_2) = 24{,}000$$
$$\hat{\text{SE}}(\hat{\alpha}_3) = 23{,}000$$
$$\hat{\text{SE}}(\hat{\alpha}_4) = 22{,}000$$

First we want to calculate the CL for $\alpha_2 - \alpha_3$ with confidence 90%

Which is:

$$\begin{aligned}
c &= [0, 1, -1, 0, 0], \\
\hat{\psi} &= c^T\hat{\beta} = \hat{\alpha}_2 - \hat{\alpha}_3, \\
\hat{\text{SE}}(\hat{\psi}) &= \sqrt{cVar(\hat{\beta})c^T}, \\
&= \sqrt{Var(\hat{\alpha}_2 - \hat{\alpha}_3)} \\
&= \sqrt{Var(\hat{\alpha}_2) + Var(\hat{\alpha}_3) - 2Cov(\hat{\alpha}_2, \hat{\alpha}_3)}, \\
\text{CI}_{90\%} &= \hat{\psi} \pm t_{45-5,\, 1-0.10/2} \cdot \hat{\text{SE}}(\hat{\psi})
\end{aligned}$$

That then by substitution becomes:

$$\begin{aligned}
\hat{\psi} &= -40{,}000 - (-10{,}000) \\
&= -30{,}000 \\
\hat{\text{SE}}(\hat{\psi}) &= \sqrt{(24{,}000)^2 + (23{,}000)^2 - 2 \cdot 22{,}000{,}000} \\
&\approx 32{,}573 \\
t_{40,\,(1-0.10)/2} &\approx 1.68385 \\
\text{CI}_{90\%} &= \hat{\psi} \pm t_{45-5,\, 1-0.10/2} \cdot \hat{\text{SE}}(\hat{\psi}) \\
\text{CI}_{90\%} &= -30{,}000 \pm 1.68385 \cdot 32{,}573 \\
&= -30{,}000 \pm 54848
\end{aligned}$$

Therefore the Confidence interval is:

$$\text{CI}_{90\%} = [-84{,}848\ ,\ 24{,}848]$$

That concludes the first subtask of Task 2.

subtask 2

Next we want to look for statistical evidence for: $\alpha_2 \leq \alpha_3$

Therefore we perform One Sided Hyphotesis Test on:

$$H_0 = \alpha_2 - \alpha_3 > 0$$
$$H_a = \alpha_2 - \alpha_3 \leq 0$$

H_0 - temporary researchers are earning **more** than temporary private consultants

H_a - the opposite

We are using the same $\hat{\psi}$ from previous subtask:

$$\begin{aligned}\hat{\psi} &= \hat{\alpha}_2 - \hat{\alpha}_3 \\ T &= \frac{(\hat{\alpha}_2 - \hat{\alpha}_3) - 0}{\hat{\mathrm{SE}}(\hat{\psi})} \\ &= \frac{-30{,}000}{32{,}573} \\ &\approx -0.921\end{aligned}$$

Then:

```r
qt(0.05, df = 40)
```

```
[1] -1.683851
```

$$-0.921 > -1.683851$$
$$T > t_{40,\,0.05}$$

We cannot reject H_0

Therefore there is not enough statistical evidence that temporary researchers are earning less than temporary private consultants.

Task 3

We have $x_i \in [-2, 2]$ for $i = 1, \dots, n$. The X and β matrces are as follows:

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}.$$

Therefore, the variance of $\hat{\beta}$ is

$$\mathrm{Var}(\hat{\beta}) = \sigma^2 (X^\top X)^{-1} = \sigma^2 \begin{bmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix}^{-1}.$$

Following the inverse this becomes

$$\mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{n \sum_i x_i^2 - (\sum_i x_i)^2} \begin{bmatrix} \sum_i x_i^2 & -\sum_i x_i \\ -\sum_i x_i & n \end{bmatrix}.$$

And then, variance of the slope estimate $(\hat{\beta}_1)$ is:

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2 n}{n \sum_i x_i^2 - (\sum_i x_i)^2}$$

Which further reduces to:

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_i x_i^2 - \frac{1}{n}(\sum_i x_i)^2} = \frac{\sigma^2}{\sum_i x_i^2 - \frac{2}{n}(\sum_i x_i)^2 + \frac{1}{n}(\sum_i x_i)^2} =$$

$$= \frac{\sigma^2}{\sum_i x_i^2 - 2\sum_i x_i \cdot \bar{x} + n\bar{x}^2} = \frac{\sigma^2}{\sum_i (x_i - \bar{x})^2}$$

We want to minimize the $\mathrm{Var}(\hat{\beta}_1)$. Which can be done by maximizing $\sum_i (x_i - \bar{x})^2$ as σ^2 is constant. This can be achieved by spreading $x_1, \dots, x_n$ as much as possible that will ideally set $\bar{x} = \frac{1}{n}\sum_i (x_i) = 0$

We know that:

$$x_i \in [-2, 2], \quad \forall i = 1, \dots, n$$

Then we can choose:

- $x_1, \dots, x_{n/2} = -2$
- $x_{n/2+1}, \dots, x_n = 2$

By that we minimize $\mathrm{Var}(\hat{\beta}_1)$ now equal to:

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_i (x_i - \bar{x})^2} = \frac{\sigma^2}{\sum_i (\pm 2 - 0)^2} = \frac{\sigma^2}{4n}$$

And we achive maximum possible precission

Q.E.D.

Task 4

Only for Master's