DS290 Modeling and Simulation. AugDec 2025.
Answer Key to Midterm 2

1.

. Since $\frac{\partial \sum_i e_i^2}{\partial \beta_j}=0$, each $\beta_j=\frac{\sum_i (x_{ji}-\bar{x}_j)(y_i-\bar{y})}{\sum_i (x_{ji}-\bar{x}_j)^2}=\frac{\sum_i (x_{ji}-\bar{x}_j)y_i}{\sum_i (x_{ji}-\bar{x}_j)^2}=:\frac{S_{x_j y}}{S_{x_j x_j}}$, *where* $\bar{x}_j:=\frac{\sum_{i=1}^n x_{ji}}{n}$

Then, $Var(\beta_j)=\frac{\sum_i (x_{ji}-\bar{x}_j)^2 Var(\varepsilon_i)}{\left(\sum_i (x_{ji}-\bar{x}_j)^2\right)^2}=\frac{\sigma^2}{\sum_i (x_{ji}-\bar{x}_j)^2}$; $E(\beta_j)=0$, since

$Var(y_i)=Var(\hat{y}_i)+Var(\varepsilon_i)=Var(\varepsilon_i)=\sigma^2$ and $\hat{y}_i$ is a computed quantity.

for each independent random variable y_i.
For unbiased estimator we need to evaluate the following ,

$$E\left(\sum_{i=1}^n e_i^2\right)=\sum_{i=1}^n E\left(y_i-\frac{\sum_{l=1}^n y_l}{n}-\sum_{j=1}^K \beta_j (x_{ji}-\bar{x}_j)\right)^2$$

$$=\sum_{i=1}^n \left(\frac{1}{n}\right)^2 \left(E(y_1^2)+....+E(y_{i-1}^2)+E(y_{i+1}^2)+....+E(y_n^2)\right)+\left(1-\frac{1}{n}\right)^2 E(y_i^2)-\left(2\left(\frac{n-1}{n}\right)^2+\frac{(n-1)(n-2)}{n^2}\right)E(y_i)^2$$

$$+\sum_{j=1}^K E\left(\beta_j^2 \sum_{i=1}^n (x_{ji}-\bar{x}_j)^2-2\beta_j \sum_{i=1}^n (x_{ji}-\bar{x}_j)(y_i-\bar{y})\right)$$

$$=n\frac{(n-1)}{n}\left(\sigma^2+E(y_i)^2-E(y_i)^2\right)+E\left(\sum_{j=1}^K \beta_j^2 S_{x_j x_j}-2\beta_j^2 S_{x_j x_j}\right)$$

$$=(n-1)\sigma^2-K\sigma^2=(n-1-K)\sigma^2$$

$$Therefore,\ \frac{E\left(\sum_{i=1}^n e_i^2\right)}{(n-1-K)}=Var(\varepsilon_i)=\sigma^2$$

2.

For this problem the true optimal solution can be computed analytically: $x^*=2.611$ years, giving an expected cost of $11,586. This solution is obtained by minimizing the expected cost, which can be written as

$$2000x+\int_0^\infty 20000\, I(y\le 1)\frac{e^{-y/x}}{x}\, dx$$

where I is the indicator function.

3.

Let X represent a process S-T such that S has a distribution that complements the exponential arrival times such that X is normally distributed with zero mean and standard deviation σ . Then:

$$W_{n+1}-w=\phi(W_n-w)+X_n$$

$$W_{n+1}-w=\phi^{n+1}(W_0-w)+\sum_{j=0}^n \phi^{n-i-1} X_j$$

$$E(W_{n+1})=w+\phi^{n+1}(E(W_0)-w)\rightarrow w \quad as \quad n\rightarrow\infty$$

$$Var(W_{n+1})=\phi^{2(n+1)}+\sigma^2\sum_{j=0}^n \phi^{(2j)}\rightarrow \sigma^2/(1-\phi^2) \quad as \quad n\rightarrow\infty$$

The first two lines in the above can be obtained by induction.

4.

. In the calculation below, substitute epsilon = 3 and t = 2.821 for 9 degrees of freedom and at 1% confidence level obtained as 0.04/(5-1). Also use S12, S23, S24, S25 in the calculations below.

$\bar{Y}_{\cdot i}$	1	2	3	4	5
	8.9464	6.5448	7.1387	9.5497	11.1213

S_{ij}^2	1	2	3	4	5
1		0.0251	0.0375	0.0330	0.0620
2			0.0028	0.0104	0.0226
3				0.0051	0.0104
4					0.0046

$t_{0.04/(5-1),(10-1)}=\cancel{2.685}\rightarrow 2.821$

$\bar{Y}_{\cdot 1}=8.9464>\bar{Y}_{\cdot 2}+t\sqrt{S_{12}^2/10}=6.6861$

$\bar{Y}_{\cdot 3}=7.1387>\bar{Y}_{\cdot 2}+t\sqrt{S_{13}^2/10}=6.71755$

$\bar{Y}_{\cdot 4}=9.5497>\bar{Y}_{\cdot 2}+t\sqrt{S_{14}^2/10}=6.70685$

$\bar{Y}_{\cdot 5}=11.1213>\bar{Y}_{\cdot 2}+t\sqrt{S_{15}^2/10}=6.76693$

Similar
Caclulations

Thus, there was adequate data to select the best, policy 2, with 96% confidence.

[illegible] $\widehat{S^2}=\max_{i\neq j} S_{ij}^2=0.0620$. The seconde-stage sample size,

$$R=\max\left(R_0,\left\lceil\frac{t^2\widehat{S^2}}{\epsilon^2}\right\rceil\right)=\max\left(10,\left\lceil\frac{(2.82^2)(0.0620)}{3^2}\right\rceil\right)=10$$

Thus, 10 replication is sufficient to make statistical comparisons.
Since $\min_{i=1}^5\{\bar{Y}_{\cdot i}\}=\bar{Y}_{\cdot 2}$, there was adequate data to conclude that policy 2 has the least expected cost per day with 96% confidence.

However, the conclusion remains as above. The answer does not change.

5. (a)

From the least squares, $\frac{\partial}{\partial \beta_1}\sum_{i=1}^{n} e_i^2=0$, $that\ \ is, \sum_{i=1}^{n} e_i \frac{\partial}{\partial \beta_1} e_i=0$, $that\ \ is, \sum_{i=1}^{n} e_i (x_i-\bar{x})=0.$

$$\sum_{i=1}^{n} (y_i-\bar{y})^2=\sum_{i=1}^{n} (y_i-\hat{y}_i+\hat{y}_i-\bar{y})^2=\sum_{i=1}^{n} e_i^2+r_i^2+2\beta_1 e_i (x_i-\bar{x})$$
$$=\sum_{i=1}^{n} (e_i^2+r_i^2)+2\beta_1 \sum_{i=1}^{n} e_i (x_i-\bar{x})=\sum_{i=1}^{n} e_i^2+\sum_{i=1}^{n} r_i^2$$

5(b) $(n-(1+2\sum_{k=1}^{n-1}\frac{n-k}{n}\rho_k))\sigma^2$

6.

(a) Test for significance ($H_0 : \mu_d = 0$)
Letting $d_i = y_i - z_i$, $\bar{d} = 1.80, S_d = 3.60$

$$t_0 = 1.80/(3.60/\sqrt{4}) = 1.$$

For $\alpha = 0.05, t_{3,0.025} = 3.18$
Since $|t_0| < 3.18$, do not reject the null hypothesis.

(b) Sample size needed for $\beta \leq 0.20$
$\delta = 2/3.60 = 0.556$
For $\alpha = 0.05$, $\beta \leq 0.20$ and $\delta = 0.556$
$n = 30$ observations.