Problems from Olive, D.J. (2021), Theory for Linear Models: (http://parker.ad.siu.edu/Olive/linmodbk.htm).

There are roughly 9 topics for this set of problems, and some problems use more than one topic. We start with Topic 1).

1) Projection Matrices, Generalized Inverses, and the Column Space

2.32. If $\boldsymbol{P}$ is a projection matrix, prove a) the eigenvalues of $\boldsymbol{P}$ are 0 or 1, b) $rank(\boldsymbol{P}) = tr(\boldsymbol{P})$.

Proof: a) If λ is an eigenvalue of $\boldsymbol{P}$, then for some $\boldsymbol{x} \neq \boldsymbol{0}$, $\lambda \boldsymbol{x} = \boldsymbol{P}\boldsymbol{x} = \boldsymbol{P}^2 \boldsymbol{x} = \lambda^2 \boldsymbol{x}$. So $\lambda(\lambda - 1) = 0$, which only has possible solutions $\lambda = 0$ or $\lambda = 1$.

b) Thus $rank(\boldsymbol{P})$ = number of nonzero eigenvalues of $\boldsymbol{P} = tr(\boldsymbol{P})$ by a).

2.38. a) Define a generalized inverse of a matrix $\boldsymbol{A}$.

b) i) Suppose $\boldsymbol{X}$ is $n \times p$ with rank $r < p$. Give the formula for the projection matrix $\boldsymbol{P}$ onto the column space of $\boldsymbol{X}$.

ii) For

$$\boldsymbol{X} = \begin{bmatrix} 1 & -2 \\ 1 & -2 \\ 1 & -2 \end{bmatrix},$$

calculate $\boldsymbol{P}$.

iii) With $\boldsymbol{X}$ as above and $\boldsymbol{Y} = (1, 2, 3)^T$, calculate the error sum of squares SSE.

Solution: a) $\boldsymbol{A}^-$ is a generalized inverse of $\boldsymbol{A}$ if $\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A} = \boldsymbol{A}$.

b) i) $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T$.

ii) $C(\boldsymbol{X}) = C(\boldsymbol{1})$. Hence $\boldsymbol{P} = \boldsymbol{1}(\boldsymbol{1}^T\boldsymbol{1})^{-1}\boldsymbol{1}^T = \dfrac{1}{3}\boldsymbol{1}\boldsymbol{1}^T$.

iii) $SSE = \boldsymbol{Y}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{Y} = \boldsymbol{Y}^T\boldsymbol{Y} - \frac{1}{3}(\sum Y_i)^2 = 1+4+9-(1+2+3)^2/3 = 14-36/3 = 2.$

2) Quadratic Forms $\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y}$ and terms like $\boldsymbol{A}\boldsymbol{Y}$:

2.21. Let $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ where $\boldsymbol{e} \sim N_n(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$. Assume $\boldsymbol{X}$ has full rank. Let $\boldsymbol{r}$ be the vector of residuals. Then the residual sum of squares RSS $= \boldsymbol{r}^T\boldsymbol{r}$. The sum of squared fitted values is $\hat{\boldsymbol{Y}}^T\hat{\boldsymbol{Y}}$. Prove that $\boldsymbol{r}^T\boldsymbol{r}$ and $\hat{\boldsymbol{Y}}^T\hat{\boldsymbol{Y}}$ independent (or dependent).
(Hint: write each term as a quadratic form.)

2.30. Let $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ where $\boldsymbol{e} \sim N_n(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$. Assume $\boldsymbol{X}$ has full rank and that the first column of $\boldsymbol{X} = \boldsymbol{1}$ so that a constant is in the model. Let $\boldsymbol{r}$ be the vector of residuals. Then the residual sum of squares RSS $= \boldsymbol{r}^T\boldsymbol{r} = \|(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{Y}\|^2$. The sample mean $\overline{\boldsymbol{Y}} = \frac{1}{n}\boldsymbol{1}^T\boldsymbol{Y}$. Prove that $\boldsymbol{r}^T\boldsymbol{r}$ and $\overline{\boldsymbol{Y}}$ independent (or dependent).
(Hint: If $\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\boldsymbol{A}\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{B}\boldsymbol{Y}$ iff $\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}^T = \boldsymbol{0}$.
So prove whether $(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{Y} \perp\!\!\!\perp \dfrac{1}{n}\boldsymbol{1}^T\boldsymbol{Y}$.)

2.33. Suppose that $\boldsymbol{A}\boldsymbol{Y}$ and $\boldsymbol{B}\boldsymbol{Y}$ are independent where $\boldsymbol{A}$ and $\boldsymbol{B}$ are symmetric matrices. Are $\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y}$ and $\boldsymbol{Y}'\boldsymbol{B}\boldsymbol{Y}$ independent? (Hint: show that the quadratic form $\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y}$ is a function of $\boldsymbol{A}\boldsymbol{Y}$ by using the definition of the generalized inverse $\boldsymbol{A}^-$.)

2.35. Let $\boldsymbol{Y}$ be an $n \times 1$ random vector and $\boldsymbol{A}$ an $n \times n$ symmetric matrix. Let $E(\boldsymbol{Y}) = \boldsymbol{\theta}$ and $\text{Cov}(\boldsymbol{Y}) = \boldsymbol{\Sigma} = (\sigma_{ij})$.

a) Prove that $E(\boldsymbol{Y}^T\boldsymbol{A}\boldsymbol{Y}) = tr(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\theta}^T\boldsymbol{A}\boldsymbol{\theta}$.

b) Let $E(Y_i) = \theta$ for all i, $\sigma_{ii} = \sigma^2$ for all i, and $\sigma_{ij} = \rho\sigma^2$ for $i \neq j$ where $-1 < \rho < 1$. Show that $\sum_i (Y_i - \overline{Y})^2$ is an unbiased estimator of $\sigma^2(1-\rho)(n-1)$. Hint: write $\sum_i (Y_i - \overline{Y})^2 = \boldsymbol{Y}^T \boldsymbol{A} \boldsymbol{Y}$ and use a).

c) Show when $\sum_i (Y_i - \overline{Y})^2$ and $\overline{Y}$ are independent if $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$. State the theorems clearly wherever used in your proof.

Solution: a) Note that $E(\boldsymbol{Y}\boldsymbol{Y}^T) = \boldsymbol{\Sigma} + \boldsymbol{\theta}\boldsymbol{\theta}^T$. Since the quadratic form is a scalar and the trace is a linear operator, $E[\boldsymbol{Y}^T\boldsymbol{A}\boldsymbol{Y}] = E[tr(\boldsymbol{Y}^T\boldsymbol{A}\boldsymbol{Y})] = E[tr(\boldsymbol{A}\boldsymbol{Y}\boldsymbol{Y}^T)] = tr(E[\boldsymbol{A}\boldsymbol{Y}\boldsymbol{Y}^T]) = tr(\boldsymbol{A}\boldsymbol{\Sigma} + \boldsymbol{A}\boldsymbol{\theta}\boldsymbol{\theta}^T) = tr(\boldsymbol{A}\boldsymbol{\Sigma}) + tr(\boldsymbol{A}\boldsymbol{\theta}\boldsymbol{\theta}^T) = tr(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\theta}^T\boldsymbol{A}\boldsymbol{\theta}$.

b) Note that $\sum_i (Y_i - \overline{Y})^2$ is the residual sum of squares for the linear model $\boldsymbol{Y} = \boldsymbol{1} + \boldsymbol{e}$. Hence $\displaystyle\sum_i (Y_i - \overline{Y})^2 = \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y} = \boldsymbol{Y}^T(\boldsymbol{I} - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}^T)\boldsymbol{Y}$ where $\boldsymbol{H} = \boldsymbol{1}(\boldsymbol{1}^T\boldsymbol{1})^{-1}\boldsymbol{1}^T$. Now $tr(\boldsymbol{A}\boldsymbol{\Sigma}) = tr(\boldsymbol{\Sigma}) - tr(\dfrac{1}{n}\boldsymbol{1}\boldsymbol{1}^T\boldsymbol{\Sigma})$. Now $\boldsymbol{1}^T\boldsymbol{\Sigma} = (\sigma^2[1+(n-1)\rho], ..., \sigma^2[1+(n-1)\rho]$, $\boldsymbol{1}\boldsymbol{1}^T\boldsymbol{\Sigma} = (\sigma^2[1+(n-1)\rho])$, and $tr(\frac{1}{n}\boldsymbol{1}\boldsymbol{1}^T\boldsymbol{\Sigma}) = \sigma^2[1+(n-1)\rho]$. So $tr(\boldsymbol{A}\boldsymbol{\Sigma}) = n\sigma^2 - \sigma^2[1+(n-1)\rho] = \sigma^2[n-1-(n-1)\rho] = \sigma^2(n-1)(1-\rho)$. Now $\boldsymbol{\theta}^T\boldsymbol{A}\boldsymbol{\theta} = \theta\boldsymbol{1}^T(\boldsymbol{I} - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}^T)\boldsymbol{1} = \theta^2(n - n^2/n) = 0$. Hence the result follows by a).

c) Assume $\boldsymbol{Y} \sim N_n(\boldsymbol{\theta}, \sigma^2\boldsymbol{I})$. Then $\overline{\boldsymbol{Y}} = \boldsymbol{B}\boldsymbol{Y}$ where $\boldsymbol{B} = \dfrac{1}{n}\boldsymbol{1}^T$. Now $\boldsymbol{Y}^T\boldsymbol{A}\boldsymbol{Y} = \boldsymbol{Y}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{Y}$. Hence the two terms are independent if $\boldsymbol{A}\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{B}\boldsymbol{Y}$ iff $\boldsymbol{A}\boldsymbol{B}^T = \boldsymbol{0}$, but $\boldsymbol{A}\boldsymbol{B}^T = \dfrac{1}{n}(\boldsymbol{I} - \dfrac{1}{n}\boldsymbol{1}\boldsymbol{1}^T)\boldsymbol{1} = \dfrac{1}{n}(\boldsymbol{1} - \boldsymbol{1}) = \boldsymbol{0}$.

2.37. a) For an $n \times 1$ vector $\boldsymbol{Y}$ with $E(\boldsymbol{Y}) = \boldsymbol{\mu}$ and $\text{Cov}(\boldsymbol{Y}) = \boldsymbol{\Sigma}$, show $E(\boldsymbol{Y}^T\boldsymbol{A}\boldsymbol{Y}) = trace(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^T\boldsymbol{A}\boldsymbol{\mu}$. Is normality of $\boldsymbol{Y}$ necessary here?

b) Consider the full rank linear model $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ where $\boldsymbol{X}$ is $n \times p$, the first column of $\boldsymbol{X}$ is $\boldsymbol{1}$, $\boldsymbol{\beta}$ is $p \times 1$, and $\boldsymbol{e} \sim N_n(\boldsymbol{0}, \sigma^2\boldsymbol{I})$.

i) Write down an ANOVA table to test $(\beta_2, ..., \beta_p)^T = \boldsymbol{0}$, giving expressions for the regression sum of squares (SSR) and the error sum of squares (SSE).

ii) Find $E(SSR)$ and $E(SSE)$ when H_0 is true.

iii) Derive the distribution of SSE/σ^2 if H_0 is true. State any theorems used.

Solution: a) Note that $E(\boldsymbol{Y}\boldsymbol{Y}^T) = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^T$. Since the quadratic form is a scalar and the trace is a linear operator, $E[\boldsymbol{Y}^T\boldsymbol{A}\boldsymbol{Y}] = E[tr(\boldsymbol{Y}^T\boldsymbol{A}\boldsymbol{Y})] = E[tr(\boldsymbol{A}\boldsymbol{Y}\boldsymbol{Y}^T)] = tr(E[\boldsymbol{A}\boldsymbol{Y}\boldsymbol{Y}^T]) = tr(\boldsymbol{A}\boldsymbol{\Sigma} + \boldsymbol{A}\boldsymbol{\mu}\boldsymbol{\mu}^T) = tr(\boldsymbol{A}\boldsymbol{\Sigma}) + tr(\boldsymbol{A}\boldsymbol{\mu}\boldsymbol{\mu}^T) = tr(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^T\boldsymbol{A}\boldsymbol{\mu}$.

Normality is not needed.

b) i)

Source	df	SS	MS	F	p-value
Regression	p-1	$SSR = \boldsymbol{Y}^T(\boldsymbol{P} - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}^T)\boldsymbol{Y}$	MSR	$F_0 = \frac{MSR}{MSE}$	for H_0:
Residual	n-p	$SSE = \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y}$	MSE		$\beta_2 = \cdots = \beta_p = 0$

ii) $E(MSE) = \sigma^2$, so $E(SSE) = (n-p)\sigma^2$. By a)

$$E(SSR) = \boldsymbol{\beta}^T\boldsymbol{X}^T(\boldsymbol{P} - \frac{\boldsymbol{1}\boldsymbol{1}^T}{n})\boldsymbol{X}\boldsymbol{\beta} + tr[\sigma^2(\boldsymbol{P} - \frac{\boldsymbol{1}\boldsymbol{1}^T}{n})] = \boldsymbol{\beta}^T\boldsymbol{X}^T(\boldsymbol{P} - \frac{\boldsymbol{1}\boldsymbol{1}^T}{n})\boldsymbol{X}\boldsymbol{\beta} + \sigma^2(p-1).$$

When H_0 is true $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{1}\beta_1$ and $E(SSR) = \sigma^2(p-1)$.

iii) By Theorem 2.14 g), if $\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I})$ then $\dfrac{\boldsymbol{Y}^T \boldsymbol{A} \boldsymbol{Y}}{\sigma^2} \sim \chi^2\left(r, \dfrac{\boldsymbol{\mu}^T \boldsymbol{A} \boldsymbol{\mu}}{2\sigma^2}\right)$ iff $\boldsymbol{A}$ is idempotent with rank$(\boldsymbol{A}) = tr(\boldsymbol{A}) = r$.

This theorem applies to SSE/σ^2 with $\boldsymbol{A} = \boldsymbol{I} - \boldsymbol{P}$, $r = n - p$, and $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$. Then $\boldsymbol{\mu}^T(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{\mu} = \boldsymbol{0}$ since $\boldsymbol{P}\boldsymbol{X} = \boldsymbol{X}$. Hence $SSE/\sigma^2 \sim \chi^2(n-p, 0) \sim \chi^2_{n-p}$.

2.40. Let $\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} > 0$, and let $\boldsymbol{A}$ be a symmetric matrix.

a) State the necessary and sufficient condition(s) for $\boldsymbol{Y}^T \boldsymbol{A} \boldsymbol{Y}$ to be a chi-square random variable.

b) Suppose $rank(\boldsymbol{\Sigma}) = n$ and $\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{A} = \boldsymbol{0}$ where $\boldsymbol{B}$ is a $q \times n$ matrix. Prove that $\boldsymbol{Y}^T \boldsymbol{A} \boldsymbol{Y}$ and $\boldsymbol{B}\boldsymbol{Y}$ are independent.

c) If $\boldsymbol{\mu} = \mu \boldsymbol{1}$ and $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$ where $\sigma^2 > 0$, prove that $\overline{Y} = \dfrac{1}{n}\sum_{i=1}^n Y_i$ and $\dfrac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y})^2$ are independent.

Solution: a) $\boldsymbol{Y}^T \boldsymbol{A} \boldsymbol{Y} \sim \chi^2(rank(\boldsymbol{A}))$ iff $\boldsymbol{A}\boldsymbol{\Sigma}$ is idempotent and $\boldsymbol{\mu}^T \boldsymbol{A} \boldsymbol{\mu} = 0$ by Theorem 2.13.

b) This proof similar to the proof of Theorem 2.8. Let $\boldsymbol{u} = \boldsymbol{A}\boldsymbol{Y}$ and $\boldsymbol{w} = \boldsymbol{B}\boldsymbol{Y}$. Then $\boldsymbol{A}\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{B}\boldsymbol{Y}$ iff Cov$(\boldsymbol{w}, \boldsymbol{u}) = \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{A} = \boldsymbol{0}$. Thus $\boldsymbol{A}\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{B}\boldsymbol{Y}$.

Let $g(\boldsymbol{A}\boldsymbol{Y}) = \boldsymbol{Y}^T \boldsymbol{A}^T \boldsymbol{A}^- \boldsymbol{A}\boldsymbol{Y} = \boldsymbol{Y}^T \boldsymbol{A}\boldsymbol{A}^- \boldsymbol{A}\boldsymbol{Y} = \boldsymbol{Y}^T \boldsymbol{A}\boldsymbol{Y}$. Then $g(\boldsymbol{A}\boldsymbol{Y}) = \boldsymbol{Y}^T \boldsymbol{A}\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{B}\boldsymbol{Y}$ since $\boldsymbol{A}\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{B}\boldsymbol{Y}$.

c) $\overline{Y} = \boldsymbol{1}^T \boldsymbol{Y}/n$ and $\sum_{i=1}^n (Y_i - \overline{Y})^2 = \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{Y}$ where $\boldsymbol{P}_1 = \boldsymbol{1}\boldsymbol{1}^T/n$ is the projection matrix on $C(\boldsymbol{1})$ since $\sum_{i=1}^n (Y_i - \overline{Y})^2$ is the residual sum of squares for the model $\boldsymbol{Y} = \boldsymbol{1}\mu + \boldsymbol{e}$ with least squares estimator $\hat{\mu} = \overline{Y}$. Hence the quantities are independent if $\boldsymbol{B}\boldsymbol{Y} = \boldsymbol{1}^T \boldsymbol{Y}$ and $\boldsymbol{Y}^T \boldsymbol{A}\boldsymbol{Y} = \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{Y}$ are independent, or if $\boldsymbol{1}^T \boldsymbol{I}(\boldsymbol{I} - \boldsymbol{P}_1) = 0$ by b). This result holds since $\boldsymbol{1}^T \boldsymbol{P}_1 = \boldsymbol{1}^T$ since $\boldsymbol{P}_1$ is the projection matrix on $C(\boldsymbol{1})$ means $\boldsymbol{P}_1 \boldsymbol{1} = \boldsymbol{1}$.

2.42. a) Suppose $\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $\boldsymbol{A}$ be an $n \times n$ symmetric matrix.

i) Show $E[(\boldsymbol{Y} - \boldsymbol{\mu})^T \boldsymbol{A}(\boldsymbol{Y} - \boldsymbol{\mu})] = tr(\boldsymbol{A}\boldsymbol{\Sigma})$. Is normality of $\boldsymbol{Y}$ necessary here?

ii) State a necessary and sufficient condition for $(\boldsymbol{Y} - \boldsymbol{\mu})^T \boldsymbol{A}(\boldsymbol{Y} - \boldsymbol{\mu})$ to be a chi-square random variable.

iii) State a necessary and sufficient condition for $(\boldsymbol{Y} - \boldsymbol{\mu})^T \boldsymbol{A}(\boldsymbol{Y} - \boldsymbol{\mu})$ and $\boldsymbol{B}\boldsymbol{Y}$ to be independent where $\boldsymbol{B}$ is an $q \times n$ matrix.

b) Suppose $\boldsymbol{Y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I})$ where $\boldsymbol{X}$ is an $n \times p$ matrix of rank p and $\boldsymbol{\beta}$ is $p \times 1$.

i) Derive the distribution of $\dfrac{1}{\sigma}(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$ where $\boldsymbol{H}$ is the projection matrix onto the column space $C(\boldsymbol{X})$.

ii) Derive the distribution of $u = \dfrac{\boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}}{\sigma^2}$.

iii) Show that u and $v = \boldsymbol{H}\boldsymbol{Y}$ are independent.

Solution: Note that $\boldsymbol{H} = \boldsymbol{P}$ and that $\boldsymbol{Z} = \boldsymbol{Y} - \boldsymbol{\mu} \sim N_n(\boldsymbol{0}, \boldsymbol{\Sigma})$.

a) i) $E[(\boldsymbol{Y} - \boldsymbol{\mu})^T \boldsymbol{A}(\boldsymbol{Y} - \boldsymbol{\mu})] = E[\boldsymbol{Z}^T \boldsymbol{A}\boldsymbol{Z}] = tr(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{0}^T \boldsymbol{A}\boldsymbol{0} = tr(\boldsymbol{A}\boldsymbol{\Sigma})$ by Theorem 2.5 using $E(\boldsymbol{Z}) = \boldsymbol{0}$.

Alternatively, $E(\boldsymbol{Z}\boldsymbol{Z}^T) = \boldsymbol{\Sigma}$ since $E(\boldsymbol{Z}) = \boldsymbol{0}$. Since the quadratic form is a scalar and the trace is a linear operator, $E[\boldsymbol{Z}^T \boldsymbol{A}\boldsymbol{Z}] = E[tr(\boldsymbol{Z}^T \boldsymbol{A}\boldsymbol{Z})] = E[tr(\boldsymbol{A}\boldsymbol{Z}\boldsymbol{Z}^T)] = tr(E[\boldsymbol{A}\boldsymbol{Z}\boldsymbol{Z}^T]) = tr(\boldsymbol{A}\boldsymbol{\Sigma})$.

Normality is not needed for this result.

ii) $\boldsymbol{A\Sigma}$ is idempotent by Theorem 2.13.

iii) $\boldsymbol{B\Sigma A} = \boldsymbol{0}$ (or $\boldsymbol{A\Sigma B}^T = \boldsymbol{0}$) by Theorem 2.8.

b) i) $\frac{1}{\sigma}(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y} \sim N_n(\frac{1}{\sigma}(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{X\beta}, \frac{1}{\sigma}(\boldsymbol{I}-\boldsymbol{H})\sigma^2\boldsymbol{I}\frac{1}{\sigma}(\boldsymbol{I}-\boldsymbol{H}) \sim N_n(\boldsymbol{0}, \boldsymbol{I}-\boldsymbol{H})$ since $\boldsymbol{HX} = \boldsymbol{X}$.

ii) By Theorem 2.14 g), if $\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2\boldsymbol{I})$ then $\dfrac{\boldsymbol{Y}^T\boldsymbol{AY}}{\sigma^2} \sim \chi^2\left(r, \dfrac{\boldsymbol{\mu}^T\boldsymbol{A\mu}}{2\sigma^2}\right)$ iff $\boldsymbol{A}$ is idempotent with rank$(\boldsymbol{A}) = tr(\boldsymbol{A}) = r$.

This theorem applies to $u = \dfrac{\boldsymbol{Y}^T(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y}}{\sigma^2} = SSE/\sigma^2$ with $\boldsymbol{A} = \boldsymbol{I}-\boldsymbol{H}$, $r = n-p$, and $\boldsymbol{\mu} = \boldsymbol{X\beta}$. Then $\boldsymbol{\mu}^T(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\mu} = \boldsymbol{0}$ since $\boldsymbol{HX} = \boldsymbol{X}$. Hence $SSE/\sigma^2 \sim \chi^2(n-p, 0) \sim \chi^2_{n-p}$.

iii) By Theorem 2.8 b), independence follows since $\boldsymbol{H}(\boldsymbol{I}-\boldsymbol{H}) = \boldsymbol{0}$.

3) MLE:

Also see 2.36 d) under Topic 4) LS estimators for $p \leq 2$.

2.1. Suppose $Y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$ for $i = 1, ..., n$ where the errors are independent $N(0, \sigma^2)$. Then the likelihood function is

$$L(\boldsymbol{\beta}, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(\frac{-1}{2\sigma^2}\|\boldsymbol{Y} - \boldsymbol{X\beta}\|^2\right).$$

a) Since the least squares estimator $\hat{\boldsymbol{\beta}}$ minimizes $\|\boldsymbol{Y} - \boldsymbol{X\beta}\|^2$, show that $\hat{\boldsymbol{\beta}}$ is the MLE of $\boldsymbol{\beta}$.

b) Then find the MLE $\hat{\sigma}^2$ of σ^2.

Solution: a) For fixed σ^2, maximizing the likelihood is equivalent to maximizing

$$\exp\left(\frac{-1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{X\beta}\|^2\right),$$

which is equivalent to minimizing $\|\boldsymbol{y} - \boldsymbol{X\beta}\|^2$. So $\hat{\boldsymbol{\beta}}$ maximizes $L(\boldsymbol{\beta}, \sigma^2)$ regardless of the value of $\sigma^2 > 0$. Hence $\hat{\boldsymbol{\beta}}$ is the MLE of $\boldsymbol{\beta}$.

b) Let $Q = \|\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\|^2$. Then the MLE of σ^2 can be found by maximizing the log profile likelihood $\log(L_P(\sigma^2))$ where

$$L_P(\sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(\frac{-1}{2\sigma^2}Q\right).$$

Let $\tau = \sigma^2$. Then

$$\log(L_p(\sigma^2)) = c - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}Q,$$

and

$$\log(L_p(\tau)) = c - \frac{n}{2}\log(\tau) - \frac{1}{2\tau}Q.$$

Hence

$$\frac{d\log(L_P(\tau))}{d\tau} = \frac{-n}{2\tau} + \frac{Q}{2\tau^2} \stackrel{set}{=} 0$$

or $-n\tau + Q = 0$ or $n\tau = Q$ or

$$\hat{\tau} = \frac{Q}{n} = \hat{\sigma}^2 = \frac{\sum_{i=1}^n r_i^2}{n} = \frac{n-p}{n} MSE,$$

which is a unique solution.

Now

$$\frac{d^2 \log(L_P(\tau))}{d\tau^2} = \frac{n}{2\tau^2} - \frac{2Q}{2\tau^3}\bigg|_{\tau=\hat{\tau}} = \frac{n}{2\hat{\tau}^2} - \frac{2n\hat{\tau}}{2\hat{\tau}^3} = \frac{-n}{2\hat{\tau}^2} < 0.$$

Thus $\hat{\sigma}^2$ is the MLE of σ^2.

Variant: Recognize that the least squares estimator $\hat{\boldsymbol{\beta}}$ minimizes $\|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$. Then show $\hat{\boldsymbol{\beta}}$ is the MLE of $\boldsymbol{\beta}$.

2.2. Suppose $Y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$ for $i = 1, ..., n$ where the errors are iid double exponential $(0, \sigma)$ with $\sigma > 0$. Then the likelihood function is

$$L(\boldsymbol{\beta}, \sigma) = \frac{1}{2^n}\frac{1}{\sigma^n}\exp(\frac{-1}{\sigma}\sum_{i=1}^n |Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}|).$$

Suppose that $\tilde{\boldsymbol{\beta}}$ is a minimizer of $Q(\boldsymbol{\beta}) = \sum_{i=1}^n |Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}|$.

a) By direct maximization, show that $\tilde{\boldsymbol{\beta}}$ is an MLE of $\boldsymbol{\beta}$ regardless of the value of σ.

b) Find an MLE of σ by maximizing

$$L_P(\sigma) \equiv L(\tilde{\boldsymbol{\beta}}, \sigma) = \frac{1}{2^n}\frac{1}{\sigma^n}\exp(\frac{-1}{\sigma}\sum_{i=1}^n |Y_i - \boldsymbol{x}_i^T\tilde{\boldsymbol{\beta}}|).$$

2.3. Suppose $Y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$ where the errors are independent $N(0, \sigma^2/W_i)$ where $W_i > 0$ are known constants. Then the likelihood function is

$$L(\boldsymbol{\beta}, \sigma^2) = \left(\prod_{i=1}^n \sqrt{W_i}\right)\left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sigma^n}\exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^n W_i(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2\right).$$

a) Suppose that $\hat{\boldsymbol{\beta}}_W$ minimizes $\sum_{i=1}^n W_i(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2$. By direct maximization, show that $\hat{\boldsymbol{\beta}}_W$ is the MLE of $\boldsymbol{\beta}$ regardless of the value of σ.

b) Then find the MLE $\hat{\sigma}^2$ of σ^2.

2.4. Suppose $\boldsymbol{Y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{V})$ for known positive definite $n \times n$ matrix $\boldsymbol{V}$. Then the likelihood function is

$$L(\boldsymbol{\beta}, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{|\boldsymbol{V}|^{1/2}}\frac{1}{\sigma^n}\exp\left(\frac{-1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right).$$

a) Suppose that $\hat{\boldsymbol{\beta}}_G$ minimizes $(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$. Show that $\hat{\boldsymbol{\beta}}_G$ is the MLE of $\boldsymbol{\beta}$.

b) Find the MLE $\hat{\sigma}^2$ of σ^2.

2.14. Suppose $Y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$ with $Q(\boldsymbol{\beta}) \geq 0$. Let c_n be a constant that does not depend on $\boldsymbol{\beta}$ or σ. Suppose the likelihood function is

$$L(\boldsymbol{\beta}, \sigma) = c_n \frac{1}{\sigma^n}\exp\left(\frac{-1}{\sigma}Q(\boldsymbol{\beta})\right).$$

a) Suppose that $\hat{\boldsymbol{\beta}}_Q$ minimizes $Q(\boldsymbol{\beta})$. Show that $\hat{\boldsymbol{\beta}}_Q$ is an MLE of $\boldsymbol{\beta}$.

b) Then find an MLE $\hat{\sigma}$ of σ.

Solution: a) For fixed $\sigma > 0$, $L(\boldsymbol{\beta}, \sigma^2)$ is maximized by minimizing $Q(\boldsymbol{\beta}) \geq 0$. So $\hat{\boldsymbol{\beta}}_Q$ maximizes $L(\boldsymbol{\beta}, \sigma^2)$ regardless of the value of $\sigma^2 > 0$. So $\hat{\boldsymbol{\beta}}_Q$ is the MLE.

b) Let $Q = Q(\hat{\boldsymbol{\beta}}_Q)$. Then the MLE $\hat{\sigma}^2$ is found by maximizing the profile likelihood, $L_p(\sigma^2) = L(\hat{\boldsymbol{\beta}}_Q, \sigma^2) = c_n \, \dfrac{1}{\sigma^n} \exp\left(\dfrac{-1}{2\sigma^2} Q\right)$. Let $\tau = \sigma^2$. The $L_p(\tau) = c_n \, \dfrac{1}{\tau^{n/2}} \exp\left(\dfrac{-1}{2\tau} Q\right)$, and the log profile likelihood $\log L_p(\tau) = d - \dfrac{n}{2}\log(\tau) - \dfrac{Q}{2\tau}$. Thus

$$\frac{d \; \log L_p(\tau)}{d\tau} = \frac{-n}{2\tau} + \frac{Q}{2\tau^2} \overset{set}{=} 0$$

or $-n\tau + Q = 0$ or $\hat{\tau} = \hat{\sigma}^2 = Q/n$, unique. Then

$$\frac{d^2 \; \log L_p(\tau)}{d\tau^2} = \frac{n}{2\tau^2} - \frac{2Q}{2\tau^3}\bigg|_{\hat{\tau}} = \frac{n}{2\hat{\tau}^2} - \frac{2n\hat{\tau}}{2\hat{\tau}^3} = \frac{-n}{2\hat{\tau}^2} < 0$$

which proves that $\hat{\sigma}^2$ is the MLE of σ^2.

2.15. Suppose $Y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + e_i$ with $Q(\boldsymbol{\beta}) \geq 0$. Let c_n be a constant that does not depend on $\boldsymbol{\beta}$ or σ^2. Suppose the likelihood function is

$$L(\boldsymbol{\beta}, \sigma^2) = c_n \, \frac{1}{\sigma^n} \exp\left(\frac{-1}{2\sigma^2} Q(\boldsymbol{\beta})\right).$$

a) Suppose that $\hat{\boldsymbol{\beta}}_Q$ minimizes $Q(\boldsymbol{\beta})$. By direct maximization, show that $\hat{\boldsymbol{\beta}}_Q$ is the MLE of $\boldsymbol{\beta}$ regardless of the value of σ.

b) Then find the MLE $\hat{\tau} = \hat{\sigma}^2$ of $\sigma^2 = \tau$ by maximizing

$$L_P(\tau) \equiv L(\hat{\boldsymbol{\beta}}_Q, \tau) = c_n \frac{1}{\tau^{n/2}} \exp(\frac{-1}{2\tau} Q(\hat{\boldsymbol{\beta}})).$$

2.41. Let $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$ where $\boldsymbol{e} \sim N_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$, $\boldsymbol{X}$ is an $n \times p$ matrix of rank p, and $\boldsymbol{\beta}$ is a $p \times 1$ vector.

a) Write down (do not derive) the MLEs of $\boldsymbol{\beta}$ and σ^2.

b) If $\hat{\sigma}^2$ is the MLE of σ^2, derive the distribution of $(n-p)\hat{\sigma}^2/\sigma^2$.

c) Prove that $\hat{\boldsymbol{\beta}}$ (MLE of $\boldsymbol{\beta}$) and $\hat{\sigma}^2$ are independent.

d) Now suppose $\boldsymbol{e} \sim N_n(\boldsymbol{0}, \sigma^2 \boldsymbol{V})$ where $\boldsymbol{V}$ is a known positive definite matrix. Write down the MLE of $\boldsymbol{\beta}$.

Solution: a) $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{Y}$ and $\hat{\sigma}^2 = \dfrac{1}{n} \sum_{i=1}^n r_i^2 = \dfrac{1}{n} SSE = \dfrac{1}{n} \boldsymbol{Y}^T (\boldsymbol{I} - \boldsymbol{P}) \boldsymbol{Y}$.

b) By Theorem 2.14 g), if $\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I})$ then $\dfrac{\boldsymbol{Y}^T \boldsymbol{A} \boldsymbol{Y}}{\sigma^2} \sim \chi^2\left(r, \dfrac{\boldsymbol{\mu}^T \boldsymbol{A} \boldsymbol{\mu}}{2\sigma^2}\right)$ iff $\boldsymbol{A}$ is idempotent with $\text{rank}(\boldsymbol{A}) = tr(\boldsymbol{A}) = r$.

This theorem applies to SSE/σ^2 with $\boldsymbol{A} = \boldsymbol{I} - \boldsymbol{P}$, $r = n - p$, and $\boldsymbol{\mu} = \boldsymbol{X\beta}$. Then $\boldsymbol{\mu}^T(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{\mu} = \boldsymbol{0}$ since $\boldsymbol{PX} = \boldsymbol{X}$. Hence $SSE/\sigma^2 \sim \chi^2(n-p, 0) \sim \chi^2_{n-p}$. Thus

$$(n-p)\hat{\sigma}^2/\sigma^2 = \frac{n-p}{n}\,\frac{SSE}{\sigma^2} \sim \frac{n-p}{n}\chi^2_{n-p}.$$

c) $\boldsymbol{BY} \perp\!\!\!\perp \boldsymbol{Y}^T\boldsymbol{AY}$ if $\boldsymbol{BA} = \boldsymbol{0}$ by Theorem 2.8 b). Here $\boldsymbol{BA} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{I} - \boldsymbol{P}) = \boldsymbol{0}$ since $\boldsymbol{X}^T\boldsymbol{P} = \boldsymbol{X}^T$. Thus the MLEs are independent.

d) The MLE is the generalized least squares estimator $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{Y}$.

2.43. Consider the regression model $y_i = \beta x_i + e_i$ for $i = 1, ..., n$ where the e_i are iid $N(0, \sigma^2)$.

a) Derive the least squares estimator of β.

b) Write down an unbiased estimator of σ^2.

c) Derive the maximum likelihood estimators of β and σ^2.

Solution: a) $Q(\beta) = \sum_{i=1}^n (y_i - \beta x_i)^2$. By the chain rule,

$$\frac{dQ(\beta)}{d\beta} = -2\sum_{i=1}^n (y_i - \beta x_i)x_i.$$

Setting the derivative equal to 0 and calling the unique solution $\hat{\beta}$ gives $\sum_{i=1}^n x_i y_i = \hat{\beta}\sum_{i=1}^n x_i^2$ or

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}.$$

b) $MSE = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^n r_i^2$ since $p = 1$.

c) Since $y_i \sim N(x_i\beta, \sigma^2)$, the likelihood function

$$L(\beta, \sigma^2) = \prod_{i=1}^n f_{y_i}(y_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}}\,\frac{1}{\sigma}\exp[\frac{-1}{2\sigma^2}(y_i - x_i\beta)^2] = c_n\frac{1}{\sigma^n}\exp[\frac{-1}{2\sigma^2}\sum_{i=1}^n (y_i - x_i\beta)^2] =$$

$$c_n\frac{1}{\sigma^n}\exp[\frac{-1}{2\sigma^2}Q(\beta)]$$

where $Q(\beta)$ is the least squares criterion. For fixed $\sigma > 0$, maiximizing $L(\beta, \sigma)$ is equivalent to minimizing the least squares criterion $Q(\boldsymbol{\beta})$. Thus $\hat{\beta}$ from a) is the MLE of β. To find the MLE of σ^2, use the profile likelihood function

$$L_p(\sigma^2) = L_p(\tau) = c_n\frac{1}{\sigma^n}\exp[\frac{-1}{2\sigma^2}Q] = c_n\frac{1}{\tau^{n/2}}\exp[\frac{-1}{2\tau}Q]$$

where $Q = Q(\hat{\beta})$. Then the log profile likelihood function

$$\log(L_p(\tau)) = d_n - \frac{n}{2}\log(\tau) - \frac{Q}{2\tau},$$

$$\text{and}\quad \frac{\mathrm{d}}{\mathrm{d}\tau}\log(\mathrm{L_p}(\tau)) = \frac{-\mathrm{n}}{2\tau} + \frac{\mathrm{Q}}{2\tau^2} \overset{\text{set}}{=} 0.$$

Thus $n\tau = Q$ or $\hat{\tau} = \hat{\sigma}^2 = Q/n = \sum_{i=1} r_i^2/n$, which is a unique solution. Now

$$\frac{d^2}{d\tau^2}\log(L_p(\tau)) = \frac{n}{2\tau^2} - \frac{2Q}{2\tau^3}\bigg|_{\hat{\tau}} = \frac{n}{2\hat{\tau}^2} - \frac{2n\hat{\tau}}{2\hat{\tau}^3} = \frac{-n}{2\hat{\tau}^2} < 0.$$

Thus $\hat{\sigma}^2$ is the MLE of σ^2.

4) LS Estimators for $p \le 2$:

Also see Problem 2.43.

1.2. Suppose that the regression model is $Y_i = 7 + \beta X_i + e_i$ for $i = 1, ..., n$ where the e_i are iid $N(0, \sigma^2)$ random variables. The least squares criterion is $Q(\eta) = \sum_{i=1}^{n}(Y_i - 7 - \eta X_i)^2$.

a) What is $E(Y_i)$?

b) Find the least squares estimator $\hat{\beta}$ of β by setting the first derivative $\frac{d}{d\eta}Q(\eta)$ equal to zero.

c) Show that your $\hat{\beta}$ is the global minimizer of the least squares criterion Q by showing that the second derivative $\frac{d^2}{d\eta^2}Q(\eta) > 0$ for all values of η.

Final answers: a) $7 + \beta X_i$

b) $\hat{\beta} = \sum(Y_i - 7)X_i / \sum X_i^2$

1.3. The location model is $Y_i = \mu + e_i$ for $i = 1, ..., n$ where the e_i are iid with mean $E(e_i) = 0$ and constant variance $\text{VAR}(e_i) = \sigma^2$. The least squares estimator $\hat{\mu}$ of μ minimizes the least squares criterion $Q(\eta) = \sum_{i=1}^{n}(Y_i - \eta)^2$. To find the least squares estimator, perform the following steps.

a) Find the derivative $\frac{d}{d\eta}Q$, set the derivative equal to zero and solve for η. Call the solution $\hat{\mu}$.

b) To show that the solution was indeed the global minimizer of Q, show that $\frac{d^2}{d\eta^2}Q > 0$ for all real η. (Then the solution $\hat{\mu}$ is a local min and Q is convex, so $\hat{\mu}$ is the global min.)

Solution:

$$a)\frac{dQ(\eta)}{d\eta} = -2\sum_{i=1}^{n}(Y_i - \eta).$$

Setting the derivative equal to 0 and calling the unique solution $\hat{\beta}$ gives $\sum_{i=1}^{n} Y_i = n\hat{\beta}$ or $\hat{\beta} = \overline{Y}$.

b) The second derivative

$$\frac{d^2 Q(\eta)}{d\eta^2} = 2n > 0,$$

hence $\hat{\beta}$ is the global minimizer.

1.4. The normal error model for simple linear regression through the origin is

$$Y_i = \beta X_i + e_i$$

for $i = 1, ..., n$ where $e_1, ..., e_n$ are iid $N(0, \sigma^2)$ random variables.

a) Show that the least squares estimator for β is

$$\hat{\beta} = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}.$$

b) Find $E(\hat{\beta})$.

c) Find VAR$(\hat{\beta})$.

(Hint: Note that $\hat{\beta} = \sum_{i=1}^n k_i Y_i$ where the k_i depend on the X_i which are treated as constants.)

See the solution of 2.36.

1.5. Suppose that the regression model is $Y_i = 10 + 2X_{i2} + \beta_3 X_{i3} + e_i$ for $i = 1, ..., n$ where the e_i are iid $N(0, \sigma^2)$ random variables. The least squares criterion is $Q(\eta_3) = \sum_{i=1}^{n} (Y_i - 10 - 2X_{i2} - \eta_3 X_{i3})^2$. Find the least squares estimator $\hat{\beta}_3$ of β_3 by setting the first derivative $\frac{d}{d\eta_3} Q(\eta_3)$ equal to zero. Show that your $\hat{\beta}_3$ is the global minimizer of the least squares criterion Q by showing that the second derivative $\frac{d^2}{d\eta_3^2} Q(\eta_3) > 0$ for all values of η_3.

Final answers: a) $\hat{\beta}_3 = \sum X_{3i}(Y_i - 10 - 2X_{2i}) / \sum X_{3i}^2$. The second partial derivative $= 2 \sum X_{3i}^2 > 0$.

1.31. For the simple linear regression model, $Y_i = \beta_1 + x_i \beta_2 + e_i$ for $i = 1, ..., n$ or $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$ where $\boldsymbol{X} = [\boldsymbol{1} \;\; \boldsymbol{x}]$ and $\boldsymbol{\beta} = (\beta_1 \; \beta_2)^T$. Find $\hat{\beta}_1$ and $\hat{\beta}_2$ by minimizing the least squares criterion.

Solution: The LS criterion $Q(\beta_1, \beta_2) = \sum_{i=1}^n (Y_i - \beta_1 - x_i \beta_2)^2$. By the **chain rule,**

$$\frac{\partial Q}{\partial \beta_1} = -2 \sum_{i=1}^{n} (Y_i - \beta_1 - \beta_2 X_i)$$

and

$$\frac{\partial^2 Q}{\partial \beta_1^2} = 2n.$$

Similarly,

$$\frac{\partial Q}{\partial \beta_2} = -2 \sum_{i=1}^{n} X_i (Y_i - \beta_1 - \beta_2 X_i)$$

and

$$\frac{\partial^2 Q}{\partial \beta_2^2} = 2 \sum_{i=1}^{n} X_i^2.$$

Setting the first partial derivatives to zero and calling the solutions $\hat{\beta}_1$ and $\hat{\beta}_2$ shows that the OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ satisfy the **normal equations**:

$$\sum_{i=1}^n Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum_{i=1}^n X_i \text{ and}$$

$$\sum_{i=1}^n X_i Y_i = \hat{\beta}_1 \sum_{i=1}^n X_i + \hat{\beta}_2 \sum_{i=1}^n X_i^2.$$

The first equation gives $\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X}$.

There are several equivalent formulas for the slope $\hat{\beta}_2$.

$$\hat{\beta}_2 \equiv \hat{\beta} = \frac{\sum_{i=1}^n X_i Y_i - \frac{1}{n}(\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i)}{\sum_{i=1}^n X_i^2 - \frac{1}{n}(\sum_{i=1}^n X_i)^2} = \frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^n (X_i - \overline{X})^2}$$

$$= \frac{\sum_{i=1}^n (X_i - \overline{X}) Y_i}{\sum_{i=1}^n (X_i - \overline{X})^2} = \frac{\sum_{i=1}^n X_i Y_i - n\overline{X}\ \overline{Y}}{\sum_{i=1}^n X_i^2 - n(\overline{X})^2}.$$

2.36. Consider the regression model $Y_i = \beta x_i + e_i$ for $i = 1, ..., n$ where the e_i are iid $N(0, \sigma^2)$.

a) Show that the least squares estimator of β is

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}.$$

b) Express $\hat{\beta}$ as a linear combination of the responses and derive its mean and variance.

c) Show that $\hat{Y}_i = \hat{\beta} x_i$ is an unbiased estimator of $E(Y_i)$and derive its variance.

d) Derive the maximum likelihood estimators of β and σ^2.

Solution: a) $Q(\beta) = \sum_{i=1}^n (Y_i - \beta x_i)^2$. By the chain rule,

$$\frac{dQ(\beta)}{d\beta} = -2 \sum_{i=1}^n (Y_i - \beta x_i) x_i.$$

Setting the derivative equal to 0 and calling the unique solution $\hat{\beta}$ gives $\sum_{i=1}^n x_i Y_i = \hat{\beta} \sum_{i=1}^n x_i^2$ or

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}.$$

b) $\hat{\beta} = \sum_{i=1}^n k_i Y_i$ where $k_i = x_i / \sum_{j=1}^n x_j^2$. Hence $E(\hat{\beta}) = \sum_{i=1}^n k_i E(Y_i) = \sum_{i=1}^n k_i \beta x_i = \beta \sum_{i=1}^n x_i^2 / \sum_{j=1}^n x_j^2 = \beta$. $V(\hat{\beta}) = \sum_{i=1}^n k_i^2 V(Y_i) = \sigma^2 \sum_{i=1}^n k_i^2$ using $Y_i = Y_i | x_i$ has $V(Y_i) = \sigma^2$. Note that $\sum_{i=1}^n k_i^2 = 1/\sum_{i=1}^n x_i^2$.

c) $E(\hat{Y}_i) = \beta x_i = E(Y_i) = E(Y_i | x_i)$, suppressing the conditioning. $V(\hat{Y}_i) = V(\hat{\beta} x_i) = x_i^2 V(\hat{\beta}) = \sigma^2 x_i^2 / \sum_{j=1}^n x_j^2$ by b).

d) Under this normal model, the MLE of β is $\hat{\beta}$ and the MLE of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} r_i^2 = \frac{n-p}{n}\ MSE$$

with $p = 1$.

2.44. Let Y_1 and Y_2 be iindependent random variables with mean θ and 2θ respectively. Find the least squares estimate of θ and the residual sum of squares.

Solution:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}\theta + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}.$$

Then

$$\hat{\theta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} = \left[(1\ 2)\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right]^{-1}(1\ 2)\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \frac{Y_1 + 2Y_2}{5}.$$

$$\text{Now}\quad \hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\theta} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}\frac{\mathrm{Y_1 + 2Y_2}}{5} = \begin{pmatrix} \frac{Y_1+2Y_2}{5} \\ \\ \frac{2Y_1+4Y_2}{5} \end{pmatrix}.$$

Thus

$$RSS = \left(Y_1 - \frac{Y_1 + 2Y_2}{5}\right)^2 + \left(Y_2 - \frac{2Y_1 + 4Y_2}{5}\right)^2.$$

1.32. Consider the following two simple linear regression models:

Model I: $Y_i = \beta_0 + \beta_1 x_i + e_i$

Model II: $Y_i = \beta_1 x_i + e_i$

with e_i iid with mean 0 and variance σ^2 and $i = 1, ..., n$,

a) State (but do not derive) the least squares estimators of β_1 for both models. Are these estimators "BLUE"? Why or why not. Quote the relevant theorem(s) in support of your assertation.

b) Prove than $V(\hat{\beta}_1) = \sigma^2 / \sum_{i=1}^{n}(x_i - \overline{x})^2$ for model I, and $V(\hat{\beta}_1) = \sigma^2 / \sum_{i=1}^{n}(x_i)^2$ for model II.

c) Referring to b), show that the variance $V(\hat{\beta}_1)$ for Model I is never smaller than the variance $V(\hat{\beta}_1)$ for model II.

Solution: a) Model I:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})Y_i}{\sum_{j=1}^{n}(x_j - \overline{x})^2} = \sum_{i=1}^{n} k_i Y_i \ \ \text{with} \ \ \mathrm{k_i} = \frac{\mathrm{x_i - \overline{x}}}{\sum_{\mathrm{j=1}}^{\mathrm{n}}(\mathrm{x_j - \overline{x}})^2}.$$

Model II:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{j=1}^{n} x_j^2} = \sum_{i=1}^{n} k_i Y_i \ \ \text{with} \ \ \mathrm{k_i} = \frac{\mathrm{x_i}}{\sum_{\mathrm{j=1}}^{\mathrm{n}} \mathrm{x_j^2}}.$$

The models are full rank, so the estimators are BLUE.

b) Model I:

$$V(\hat{\beta}_1) = \sum_{i=1}^n k_i^2 V(Y_i) = \sigma^2 \sum_{i=1}^n k_i^2 = \sigma^2 \frac{\sum_{i=1}^n (x_i - \overline{x})^2}{[\sum_{j=1}^n (x_j - \overline{x})^2]^2} = \sigma^2 / \sum_{i=1}^n (x_i - \overline{x})^2.$$

Model II:

$$V(\hat{\beta}_1) = \sum_{i=1}^n k_i^2 V(Y_i) = \sigma^2 \sum_{i=1}^n k_i^2 = \sigma^2 \frac{\sum_{i=1}^n x_i^2}{[\sum_{j=1}^n x_j^2]^2} = \sigma^2 / \sum_{i=1}^n x_i^2.$$

c) The result follows if $\sum_{i=1} x_i^2 \geq \sum_{i=1} (x_i - \overline{x})^2$, but $\sum_{i=1}^n (x_i - \mu)^2$ is the least squares criterion for the model $x_i = \mu + e_i$, and the criterion is minimized by the least squares estimator $\hat{\mu} = \overline{x}$. Hence using $\tilde{\mu} = 0$ gives a least squares criterion at least as large as that using $\hat{\mu}$, and the result holds.

5) WLS:

See Problem 2.41 d) under Topic 3 if $\boldsymbol{V}$ is diagonal.

6) Non-full rank linear models:

3.12. Consider the linear regression model $Y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i$ or $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$ where $\boldsymbol{Y} \sim N_n(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I})$. Assume $\boldsymbol{X}$ is $n \times p$ with $rank(\boldsymbol{X}) = r \leq p$.

a) Give expressions for SSE and SSR using matrix notation.

b) Find $E(SSE)$ and $E(SSR)$.

c) Find the distribution of i) SSE, ii) SSR, and iii) MSR/MSE under the assumption $\beta_2 = \cdots = \beta_p = 0$.

Solution: a) $SSE = \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y}$ and $SSR = \boldsymbol{Y}^T(\boldsymbol{P} - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}^T)\boldsymbol{Y} = \boldsymbol{Y}^T(\boldsymbol{P} - \boldsymbol{P}_1)\boldsymbol{Y}$ where $\boldsymbol{P}_1 = \frac{1}{n}\boldsymbol{1}\boldsymbol{1}^T = \boldsymbol{1}(\boldsymbol{1}^T\boldsymbol{1})^{-1}\boldsymbol{1}^T$ is the projection matrix on $C(\boldsymbol{1})$.

b) $E(MSE) = \sigma^2$, so $E(SSE) = (n-r)\sigma^2$. By a) and Theorem 2.5,

$$E(SSR) = \boldsymbol{\beta}^T\boldsymbol{X}^T(\boldsymbol{P} - \frac{\boldsymbol{1}\boldsymbol{1}^T}{n})\boldsymbol{X\beta} + tr[\sigma^2(\boldsymbol{P} - \frac{\boldsymbol{1}\boldsymbol{1}^T}{n})] = \boldsymbol{\beta}^T\boldsymbol{X}^T(\boldsymbol{P} - \frac{\boldsymbol{1}\boldsymbol{1}^T}{n})\boldsymbol{X\beta} + \sigma^2(r-1).$$

When H_0 is true $\boldsymbol{X\beta} = \boldsymbol{1}\beta_1$ and $E(SSR) = \sigma^2(r-1)$.

c) By Theorem 2.14 g), if $\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2\boldsymbol{I})$ then $\dfrac{\boldsymbol{Y}^T\boldsymbol{A}\boldsymbol{Y}}{\sigma^2} \sim \chi^2\left(a, \dfrac{\boldsymbol{\mu}^T\boldsymbol{A}\boldsymbol{\mu}}{2\sigma^2}\right)$ iff $\boldsymbol{A}$ is idempotent with rank$(\boldsymbol{A}) = tr(\boldsymbol{A}) = a$.

i) Theorem 2.14 g) applies to SSE/σ^2 with $\boldsymbol{A} = \boldsymbol{I} - \boldsymbol{P}$ and $a = n - r$. Since $\boldsymbol{\mu} = \boldsymbol{X\beta}$, and $\boldsymbol{\mu}^T(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{\mu} = \boldsymbol{0}$ since $\boldsymbol{PX} = \boldsymbol{X}$. Hence $SSE/\sigma^2 \sim \chi^2(n-r, 0) \sim \chi^2_{n-r}$. Thus $SSE \sim \sigma^2\chi^2_{n-r}$ regardless of whether H_0 is true or false.

ii) Theorem 2.14 g) applies to SSR/σ^2 with $\boldsymbol{A} = \boldsymbol{P} - \boldsymbol{P}_1$ and $a = r - 1$. If H_0 is true, then $\boldsymbol{\mu} = \boldsymbol{1}\beta_1$ and and $\boldsymbol{\mu}^T(\boldsymbol{P} - \boldsymbol{P}_1)\boldsymbol{\mu} = \boldsymbol{0}$ since $\boldsymbol{1}$ is the first column of $\boldsymbol{X}$ and $\boldsymbol{P}_1$ is the projection matrix on $C(\boldsymbol{1})$. Thus $\boldsymbol{P1} = \boldsymbol{P}_1\boldsymbol{1} = \boldsymbol{1}$. Hence $SSR/\sigma^2 \sim \chi^2(r-1, 0) \sim \chi^2_{r-1}$. Thus $SSR \sim \sigma^2\chi^2_{r-1}$.

iii) SSE and SSR are independent by Craig's theorem since $(\boldsymbol{I} - \boldsymbol{P})(\boldsymbol{P} - \boldsymbol{P}_1) = \boldsymbol{P} - \boldsymbol{P}_1 - \boldsymbol{P} + \boldsymbol{P}_1 = \boldsymbol{0}$. MSE = SSE/(n-r) and MSR = SSR/(r-1). Thus

$$MSE/MSR = \frac{SSR/[\sigma^2(r-1)]}{SSE/[\sigma^2(n-r)]} \sim F_{r-1,n-r}.$$

7) Estimability and the Gauss Markov Theorem:

3.13. Consider the linear regression model $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ where $\boldsymbol{Y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I})$. Assume $\boldsymbol{X}$ is $n \times p$ with $rank(\boldsymbol{X}) = r \le p$.

a) i) Define what is meant by an estimable linear function of $\boldsymbol{\beta}$.

ii) Write down the least squares estimator of an estimable function of $\boldsymbol{\beta}$.

iii) Write down an unbiased estimator of σ^2.

b) Show the estimators of part a) ii) and iii) are unbiased.

c) State the Gauss Markov Theorem.

d) Give expressions for SSE and SSR using matrix notation.

Solution: a) i) Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be constant vectors. Then $\boldsymbol{a}^T\boldsymbol{\beta}$ is estimable if there exists a linear unbiased estimator $\boldsymbol{b}^T\boldsymbol{Y}$ so $E(\boldsymbol{b}^T\boldsymbol{Y}) = \boldsymbol{a}^T\boldsymbol{\beta}$. Also, the quantity $\boldsymbol{a}^T\boldsymbol{\beta}$ is estimable iff $\boldsymbol{a}^T = \boldsymbol{b}^T\boldsymbol{X}$ iff $\boldsymbol{a} = \boldsymbol{X}^T\boldsymbol{b}$ iff $\boldsymbol{a} \in C(\boldsymbol{X}^T)$.

ii) Let a least squares estimator $\hat{\boldsymbol{\beta}}$ be any solution to the normal equations $\boldsymbol{X}^T\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}^T\boldsymbol{Y}$. Then the least squares estimator of $\boldsymbol{a}^T\boldsymbol{\beta}$ is $\boldsymbol{a}^T\hat{\boldsymbol{\beta}} = \boldsymbol{b}^T\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{b}^T\boldsymbol{P}\boldsymbol{Y}$.

iii) $MSE = \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y}/(n-r) = SSE/(n-r)$.

b) ii) $E(\boldsymbol{b}^T\boldsymbol{P}\boldsymbol{Y}) = \boldsymbol{b}^T\boldsymbol{P}\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{b}^T\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{a}^T\boldsymbol{\beta}$.

iii) $E(SSE) = E(\boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y}) = tr[\sigma^2(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{I}] + \boldsymbol{\mu}^T(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{\mu}$ by Theorem 2.5 where $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$. Hence $E(SSE) = \sigma^2(tr(\boldsymbol{I} - \boldsymbol{P}) = \sigma^2(n-r)$. Hence $E(MSE) = E(SSE)/(n-r) = \sigma^2$.

c) If $\boldsymbol{a}^T\boldsymbol{\beta}$ is estimable and a least squares estimator $\hat{\boldsymbol{\beta}}$ is any solution to the normal equations $\boldsymbol{X}^T\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}^T\boldsymbol{Y}$, then $\boldsymbol{a}^T\hat{\boldsymbol{\beta}}$ is the unique BLUE of $\boldsymbol{a}^T\boldsymbol{\beta}$.

d) $SSE = \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y}$ and $SSR = \boldsymbol{Y}^T(\boldsymbol{P} - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}^T)\boldsymbol{Y} = \boldsymbol{Y}^T(\boldsymbol{P} - \boldsymbol{P}_1)\boldsymbol{Y}$ where $\boldsymbol{P}_1 = \frac{1}{n}\boldsymbol{1}\boldsymbol{1}^T = \boldsymbol{1}(\boldsymbol{1}^T\boldsymbol{1})^{-1}\boldsymbol{1}^T$ is the projection matrix on $C(\boldsymbol{1})$.

3.14. Let $E(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta}$ where $\boldsymbol{Y}$ is 3×1, $\boldsymbol{X}$ is 3×2, and $\boldsymbol{\beta}$ is 2×1. Let

$$i)\ \boldsymbol{X} = \begin{bmatrix} 2 & 0 \\ 1 & 1 \\ 0 & 2 \end{bmatrix} \quad \text{and} \quad \text{ii)}\ \boldsymbol{X} = \begin{bmatrix} 3 & 6 \\ 2 & 4 \\ 1 & 2 \end{bmatrix}.$$

a) In each of cases i) and ii), state whether $\boldsymbol{\beta}$ is estimable and explain your answer.

b) If the answer is "yes," then determine the matrix $\boldsymbol{B}$ in $\hat{\boldsymbol{\beta}} = \boldsymbol{B}\boldsymbol{Y}$.

c) If the answer is "no," then produce one estimable parametric function and its unbiased estimator.

Solution: a) Note that $\boldsymbol{\beta}$ is estimable for i) since $\boldsymbol{X}$ for i) has full rank 2. Note that $\boldsymbol{\beta}$ is not estimable for ii) since $\boldsymbol{X}$ for ii) does not have full rank ($rank(\boldsymbol{X}) = 1$).

b)

$$\boldsymbol{B} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T = \left(\begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 1 & 1 \\ 0 & 2 \end{bmatrix} \right)^{-1} \boldsymbol{X}^T = \begin{bmatrix} 5 & 1 \\ 1 & 5 \end{bmatrix}^{-1} \boldsymbol{X}^T.$$

If

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

and $d = a_{11}a_{22} - a_{21}a_{12} \neq 0$, then

$$\boldsymbol{A}^{-1} = \frac{1}{d}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}.$$

Thus

$$\boldsymbol{B} = \frac{1}{24}\begin{bmatrix} 5 & -1 \\ -1 & 5 \end{bmatrix}\begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix} = \frac{1}{24}\begin{bmatrix} 10 & 4 & -2 \\ -2 & 4 & 10 \end{bmatrix}.$$

c) Note that $\boldsymbol{b}^T\boldsymbol{Y}$ is an unbiased estimator of $\boldsymbol{b}^T\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{a}^T\boldsymbol{\beta}$ with $\boldsymbol{a}^T = \boldsymbol{b}^T\boldsymbol{X}$. If $\boldsymbol{b} = \boldsymbol{1}$, then

$$\boldsymbol{a}^T = \boldsymbol{1}^T\boldsymbol{X} = (1 \;\; 1 \;\; 1)\begin{bmatrix} 3 & 6 \\ 2 & 4 \\ 1 & 2 \end{bmatrix} = (6 \;\; 12).$$

Thus the estimable function $\boldsymbol{a}^T\boldsymbol{\beta} = 6\beta_1 + 12\beta_2$ has unbiased estimator $\boldsymbol{b}^T\boldsymbol{Y} = \boldsymbol{1}^T\boldsymbol{Y} = Y_1 + Y_2 + Y_3$.

The function $\boldsymbol{a}^T\boldsymbol{\beta}$ is estimable iff $\boldsymbol{a} \in C(\boldsymbol{X}^T)$ iff $\boldsymbol{a} = c(1 \;\; 2)^T$ for some nonzero constant c. Thus $(1 \;\; 2)\boldsymbol{\beta} = \beta_1 + 2\beta_2$ is estimable.

The function $\boldsymbol{a}^T\boldsymbol{\beta}$ is estimable iff $\boldsymbol{a}^T = \boldsymbol{b}^T\boldsymbol{X}$ for some constant vector $\boldsymbol{b}$. Let

$$\boldsymbol{a}^T = (1 \;\; 1 \;\; 1)\begin{bmatrix} 3 & 6 \\ 2 & 4 \\ 1 & 2 \end{bmatrix} = (6 \;\; 12).$$

Then $\boldsymbol{a}^T\boldsymbol{\beta} = 6\beta_1 + 12\beta_2$ has unbiased estimator $\boldsymbol{a}^T\hat{\boldsymbol{\beta}} = \boldsymbol{b}^T\boldsymbol{P}\boldsymbol{Y}$ where

$$\boldsymbol{P} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}\left[(3 \;\; 2 \;\; 1)\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}\right]^{-1}(3 \;\; 2 \;\; 1) = \frac{1}{14}\begin{bmatrix} 9 & 6 & 3 \\ 6 & 4 & 2 \\ 3 & 2 & 1 \end{bmatrix}.$$

Since $\boldsymbol{b} = \boldsymbol{1}$, the unbiased estimator is

$$\frac{1}{14}(18 \;\; 12 \;\; 6)\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \frac{18}{14}Y_1 + \frac{12}{14}Y_2 + \frac{6}{14}Y_3.$$

8) Hypothesis Testing:

See Problem 2.37 b) and Problem 3.12 c) iii).

2.39. Consider the usual full rank model $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ where $\boldsymbol{X}$ is $n \times p$ and $\boldsymbol{e} \sim N_n(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T \; \boldsymbol{\beta}_2^T)^T$ where $\boldsymbol{\beta}_i$ is $p_i \times 1$.

a) Write down the complete ANOVA table for the test $H_0 : \boldsymbol{\beta}_2 = \boldsymbol{0}$, including the expected mean squares.

b) Prove that $SSE(R) - SSE$ and MSE are independent.

c) If H_0 is true, show $F_R \sim F_{p_2, n-p}$.

Solution: a)

Source	df	SS	MS	E(MS)	F
Reduced	$n - p_1$	$SSE(R) = \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{Y}$	MSE(R)	E(MSE(R))	$F_R = \frac{SSE(R)-SSE}{p_2 MSE} =$
Full	$n - p$	$SSE = \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y}$	MSE	σ^2	$\dfrac{\boldsymbol{Y}^T(\boldsymbol{P} - \boldsymbol{P}_1)\boldsymbol{Y}/p_2}{\boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y}/(n-p)}$

where

$$E(MSE(R)) = \frac{1}{n-p_1}[\sigma^2 tr(\boldsymbol{I}-\boldsymbol{P}_1)+\boldsymbol{\beta}^T\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{P}_1)\boldsymbol{X}\boldsymbol{\beta}] = \frac{1}{n-p_1}[\sigma^2(n-p_1)+\boldsymbol{\beta}^T\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{P}_1)\boldsymbol{X}\boldsymbol{\beta}].$$

If H_0 is true, then $\boldsymbol{Y} \sim N_n(\boldsymbol{X}_1\boldsymbol{\beta}_1, \sigma^2\boldsymbol{I})$, and $E(MSE(R)) = \sigma^2$.

b) Need to show that $SSE(R) - SSE = \boldsymbol{Y}^T(\boldsymbol{P} - \boldsymbol{P}_1)\boldsymbol{Y}$ and $SSE = \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y}$ are independent. This result follows from Craig's Theorem since $(\boldsymbol{P} - \boldsymbol{P}_1)(\boldsymbol{I} - \boldsymbol{P}) = \boldsymbol{P} - \boldsymbol{P}_1 - \boldsymbol{P} + \boldsymbol{P}_1 = \boldsymbol{0}$.

c) By Theorem 2.14 g), if $\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2\boldsymbol{I})$ then $\dfrac{\boldsymbol{Y}^T\boldsymbol{A}\boldsymbol{Y}}{\sigma^2} \sim \chi^2\left(r, \dfrac{\boldsymbol{\mu}^T\boldsymbol{A}\boldsymbol{\mu}}{2\sigma^2}\right)$ iff $\boldsymbol{A}$ is idempotent with rank$(\boldsymbol{A}) = tr(\boldsymbol{A}) = r$.

This theorem applies to SSE/σ^2 with $\boldsymbol{A} = \boldsymbol{I} - \boldsymbol{P}$ and $r = n - p$. Then $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$, and $\boldsymbol{\mu}^T(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{\mu} = \boldsymbol{0}$ since $\boldsymbol{P}\boldsymbol{X} = \boldsymbol{X}$. Hence $SSE/\sigma^2 \sim \chi^2(n-p, 0) \sim \chi^2_{n-p}$. Similarly, when H_0 is true, the theorem applies to $\boldsymbol{Y}^T(\boldsymbol{P} - \boldsymbol{P}_1)\boldsymbol{Y}/\sigma^2$ with $\boldsymbol{A} = \boldsymbol{P} - \boldsymbol{P}_1$ and $r = p - p_1 = p_2$. Then $\boldsymbol{\mu} = \boldsymbol{X}_1\boldsymbol{\beta}_1$, and $\boldsymbol{\mu}^T(\boldsymbol{P} - \boldsymbol{P}_1)\boldsymbol{\mu} = \boldsymbol{0}$ since $\boldsymbol{P}\boldsymbol{X}_1 = \boldsymbol{P}_1\boldsymbol{X}_1 = \boldsymbol{X}_1$. Hence $\boldsymbol{Y}^T(\boldsymbol{P} - \boldsymbol{P}_1)\boldsymbol{Y}/\sigma^2 \sim \chi^2(p_2, 0) \sim \chi^2_{p_2}$. Thus

$$F_R = \frac{\boldsymbol{Y}^T(\boldsymbol{P} - \boldsymbol{P}_1)\boldsymbol{Y}/p_2}{\boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y}/(n-p)} \sim F_{p_2, n-p}.$$

9) Expected Value, Covariance Matrix and Large Sample Theory for least squares quantities:

See 2.42 b) i).

1.33. Consider the simple linear regression model $\boldsymbol{Y} = \beta_1\boldsymbol{1} + \beta_2\boldsymbol{x} + \boldsymbol{e} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ where the e_i are iid with $E(e_i) = 0$, $V(e_i) = \sigma^2$, and $\boldsymbol{X} = [\boldsymbol{1}\ \ \boldsymbol{x}]$. Let the residual vector $\boldsymbol{r} = \boldsymbol{Y} - \hat{\boldsymbol{Y}}$.

a) Find $E(\boldsymbol{r})$ and $Cov(\boldsymbol{r})$.

b) Find $Cov(\boldsymbol{r}, \boldsymbol{Y})$.

c) Find $Cov(\boldsymbol{r}, \hat{\boldsymbol{Y}})$.

Solution: a) $E(\boldsymbol{r}) = E[(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y}] = (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{0}$. $Cov(\boldsymbol{r}) = Cov[(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y}] = (\boldsymbol{I} - \boldsymbol{P})Cov(\boldsymbol{Y})(\boldsymbol{I} - \boldsymbol{P})^T = \sigma^2(\boldsymbol{I} - \boldsymbol{P})$.

b) Cov$(\boldsymbol{r}, \boldsymbol{Y}) = E([\boldsymbol{r} - E(\boldsymbol{r})][\boldsymbol{Y} - E(\boldsymbol{Y})]^T) =$

$$E([(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y} - (\boldsymbol{I} - \boldsymbol{P})E(\boldsymbol{Y})][\boldsymbol{Y} - E(\boldsymbol{Y})]^T) =$$

$$E[(\boldsymbol{I} - \boldsymbol{P})[\boldsymbol{Y} - E(\boldsymbol{Y})][\boldsymbol{Y} - E(\boldsymbol{Y})]^T] = (\boldsymbol{I} - \boldsymbol{P})Cov(\boldsymbol{Y}) = (\boldsymbol{I} - \boldsymbol{P})\sigma^2\boldsymbol{I} = \sigma^2(\boldsymbol{I} - \boldsymbol{P}).$$

c) Cov$(\boldsymbol{r}, \hat{\boldsymbol{Y}}) = E([\boldsymbol{r} - E(\boldsymbol{r})][\hat{\boldsymbol{Y}} - E(\hat{\boldsymbol{Y}})]^T) =$

$$E([(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y} - (\boldsymbol{I} - \boldsymbol{P})E(\boldsymbol{Y})][\boldsymbol{P}\boldsymbol{Y} - \boldsymbol{P}E(\boldsymbol{Y})]^T) =$$

$$E[(\boldsymbol{I}-\boldsymbol{P})[\boldsymbol{Y}-E(\boldsymbol{Y})][\boldsymbol{Y}-E(\boldsymbol{Y})]^T\boldsymbol{P}] = (\boldsymbol{I}-\boldsymbol{P})\sigma^2\boldsymbol{I}\boldsymbol{P} = \sigma^2(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{P} = \boldsymbol{0}.$$

4.9^Q. Suppose $\boldsymbol{Y}^* = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{r}^W$ where where $E(\boldsymbol{r}^W) = \boldsymbol{0}$ and $Cov(\boldsymbol{r}^W) = Cov(\boldsymbol{Y}^*) = MSE\ \boldsymbol{I}_n$. Then $\hat{\boldsymbol{\beta}}^* = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}^*$. Recall that $\boldsymbol{X}$ is an $n \times p$ constant matrix. Simplify quantities when possible.

a) What is $E(\hat{\boldsymbol{\beta}}^*)$?

b) What is $Cov(\hat{\boldsymbol{\beta}}^*)$?

c) Recall that $\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{P}\boldsymbol{Y}$. What is $E(\hat{\boldsymbol{\beta}}^*_I) = E[(\boldsymbol{X}_I^T\boldsymbol{X}_I)^{-1}\boldsymbol{X}_I^T\boldsymbol{Y}^*]$?

d) What is $Cov(\hat{\boldsymbol{\beta}}^*_I)$?

4.10. Suppose $\boldsymbol{Y}^* \sim N_n(\boldsymbol{X}\hat{\boldsymbol{\beta}}, \sigma_n^2\boldsymbol{I}_n)$. Hence $Y_i^* = \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}} + \epsilon_i^P$ where $E(\epsilon_i^P) = 0$ and $V(\epsilon_i^P) = \sigma_n^2$. Hence $\boldsymbol{A}\boldsymbol{Y}^* \sim N_g(\boldsymbol{A}\boldsymbol{X}\hat{\boldsymbol{\beta}}, \sigma_n^2\boldsymbol{A}\boldsymbol{A}^T)$ if $\boldsymbol{A}$ is a $g \times n$ constant matrix. Recall that $\boldsymbol{X}$ is an $n \times p$ constant matrix. Simplify quantities when possible.

a) What is the distribution of $\hat{\boldsymbol{\beta}}^* = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}^*$?

b) Using a), what is $E(\hat{\boldsymbol{\beta}}^*)$?

c) Recall that $\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{P}\boldsymbol{Y}$. What is the distribution of $\hat{\boldsymbol{\beta}}^*_I = (\boldsymbol{X}_I^T\boldsymbol{X}_I)^{-1}\boldsymbol{X}_I^T\boldsymbol{Y}^*$ if $\hat{\boldsymbol{\beta}}^*_I$ is $k \times 1$?

4.11. Suppose $\boldsymbol{Y}^* = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{r}^W$ where $E(\boldsymbol{r}^W) = \boldsymbol{0}$ and $Cov(\boldsymbol{r}^W) = Cov(\boldsymbol{Y}^*) = diag(r_i^2) = diag(r_1^2, ..., r_n^2)$. Then $\hat{\boldsymbol{\beta}}^* = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}^*$ is the least squares estimator from regressing $\boldsymbol{Y}^*$ on $\boldsymbol{X}$, an $n \times p$ constant matrix. This model is used for the wild bootstrap. Simplify quantities when possible. (Can simplify a) and c), but can't simplify b) and d) much.)

a) What is $E(\hat{\boldsymbol{\beta}}^*)$?

b) What is $Cov(\hat{\boldsymbol{\beta}}^*)$?

c) What is $E(\hat{\boldsymbol{\beta}}^*_I) = E[(\boldsymbol{X}_I^T\boldsymbol{X}_I)^{-1}\boldsymbol{X}_I^T\boldsymbol{Y}^*]$?

d) What is $Cov(\hat{\boldsymbol{\beta}}^*_I)$?

Solution: a) $(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T E(\boldsymbol{Y}^*) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$.

b) $\boldsymbol{A}Cov(\boldsymbol{Y}^*)\boldsymbol{A}^T = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T diag(r_i^2)\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}$.

c) We will use $\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{P}\boldsymbol{Y}$ and $\boldsymbol{P}\boldsymbol{X}_I = \boldsymbol{X}_I$. Then $E(\hat{\boldsymbol{\beta}}^*_I) = (\boldsymbol{X}_I^T\boldsymbol{X}_I)^{-1}\boldsymbol{X}_I^T E(\boldsymbol{Y}^*) = (\boldsymbol{X}_I^T\boldsymbol{X}_I)^{-1}\boldsymbol{X}_I^T\boldsymbol{X}\hat{\boldsymbol{\beta}} = (\boldsymbol{X}_I^T\boldsymbol{X}_I)^{-1}\boldsymbol{X}_I^T\boldsymbol{P}\boldsymbol{Y} = (\boldsymbol{X}_I^T\boldsymbol{X}_I)^{-1}\boldsymbol{X}_I^T\boldsymbol{Y} = \hat{\boldsymbol{\beta}}_I$.

d) $\boldsymbol{A}Cov(\boldsymbol{Y}^*)\boldsymbol{A}^T = (\boldsymbol{X}_I^T\boldsymbol{X}_I)^{-1}\boldsymbol{X}_I^T diag(r_i^2)\boldsymbol{X}_I(\boldsymbol{X}_I^T\boldsymbol{X}_I)^{-1}$.

2.45. a) By the least squares central limit theorem, $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N_p(\boldsymbol{0}, \sigma^2\ \boldsymbol{W})$. Hence the limiting distribution of of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is the $N_p(\boldsymbol{0}, \sigma^2\ \boldsymbol{W})$ distribution. Let $\boldsymbol{A}$ be a constant $r \times p$ matrix. Find the limiting distribution of $\boldsymbol{A}\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$.

b) Suppose $\boldsymbol{Z}_n \xrightarrow{D} N_k(\boldsymbol{\mu}, \boldsymbol{I})$. Let $\boldsymbol{A}$ be a constant $r \times k$ matrix. Find the limiting distribution of $\boldsymbol{A}(\boldsymbol{Z}_n - \boldsymbol{\mu})$.

Solution a) $\sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N_r(\boldsymbol{0}, \sigma^2\boldsymbol{A}\boldsymbol{W}\boldsymbol{A}^T)$.

b) $\boldsymbol{A}(\boldsymbol{Z}_n - \boldsymbol{\mu}) \xrightarrow{D} N_r(\boldsymbol{0}, \boldsymbol{A}\boldsymbol{A}^T)$.