Problem 1

(a)

i	group	n	Mean	SD
1	Non-smoker	21	3.78	1.79
2	Light smoker	21	3.23	1.86
3	Heavy smoker	21	2.59	1.82

Y_{ij} denotes the lung function measure for the j_{th} subject in the i_{th} smoking group
$K = 3$ total number of smoking groups
$N = 63$ total number of subjects
$H_0 : \mu_1 = \mu_2 = \mu_3$ The lung function measure populations means are equal across the smoking groups.
H_1 : The lung function measure population means are not equal across the smoking groups.

$s_p^2 = \dfrac{\sum_{i=1}^{3}(n_i - 1)s_i^2}{\sum_{i=1}^{3}(n_i - 1)} \approx 3.33$

$\bar{Y} = \dfrac{\sum_{i=1}^{3}\sum_{j=1}^{n_i} Y_{ij}}{N} = 3.2$

Under $H_0 : \hat{\sigma}^2 = \dfrac{\sum_{i=1}^{3} n_i(\bar{Y}_i - \bar{Y})^2}{K-1} \approx 7.45$

Under $H_0 : F \equiv \dfrac{\hat{\sigma}^2}{s_p^2} \sim F_{K-1,N-K}$

F test statistic $= \dfrac{\hat{\sigma}^2}{s_p^2} \approx 2.24$

Critical Region $C_\alpha = \{F : F > F_{2,60;1-\alpha}\}$
Using $\alpha = .05$
Quantile ${F^{-1}}_{2,60;.95} \approx 3.15$
$2.24 < 3.14$ The test statistic is outiside of the critical region thus fail to reject H_0
p-value $= 1 - F_{2.24,2,60} \approx .12 > .05$ The p-value is larger than α
Not enough evidence to reject the null hypothesis that the population means for lung function measure are equal across all smoking groups.

Anova Table

Source of Variation	SS	df	Mean Square	F	P-value
Among groups	14.9	2	$\hat{\sigma}^2 = 7.45$	2.24	.12
Within Groups	199.8	60	$s_p^2 = 3.33$		
Total	214.7	62			

(b) Constraint on model: $\mu_i = \alpha_1 + \alpha_2 i$
Using $\bar{y}_i$ to estimate μ_i
$\bar{y}_1 = \alpha_1 + \alpha_2 * 1$
$\bar{y}_2 = \alpha_1 + \alpha_2 * 2$
$\bar{y}_3 = \alpha_1 + \alpha_2 * 3$
Using collapsed model to obtain $\hat{\alpha}_1$ and $\hat{\alpha}_2$
$\bar{y}_i = [3.78, 3.23, 2.59]^T \; i = [1, 2, 3]$
Essence Matrix $\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$
Grand mean= $\bar{y} = 3.2 \quad \bar{x} = 2$
$SXX = \sum(x_i - \bar{x})^2 = 2$
$SXY = \sum(x_i - \bar{x})y_i = -1.19$
$SYY = \sum(y_i - \bar{y})^2 = .7094$
$\hat{\alpha}_2 = \frac{SXY}{SXX} = \frac{-1.19}{2} = -.595$
$\hat{\alpha}_1 = \bar{y} - \hat{\alpha}_2\bar{x} = 3.2 - 2 * -.595 = 4.39$
$\hat{\alpha}_1 = 4.39 \quad \hat{\alpha}_2 = -.595$
$R\tilde{S}S = SYY - \hat{\alpha}_2^2 SXX = .00135$
Actual Model: $y_{ij} = \alpha_1 + \alpha_2 * i + \epsilon_{ij}$
In order to obtain the SE for the estimates of the actual model we will need to compute SSE of the actual model using $R\tilde{S}S$ from the collapsed model and SS_{Within} from the anova table from part a.
$SSE = SS_{Within} + 63 * R\tilde{S}S$
$SSE = 199.8 + 63 * .00135 = 199.8851$
Then use SSE to compute $\hat{\sigma}^2$
$\hat{\sigma}^2 = SSE/(N-2) \approx 3.28$
Next create an X matrix and calculate $(X'X)^{-1}$

$$X_{63\times 2} = \begin{bmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & 2 \\ \vdots & \vdots \\ 1 & 3 \\ \vdots & \vdots \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} 0.1111 & -0.0476 \\ -0.0476 & 0.0238 \end{bmatrix}$$

Use the diagonal of $(X'X)^{-1}$ matrix to obtain the standard error of the estimates
Standard error of estimates=$\sqrt{\hat{\sigma}^2 * (X'X)^{-1}_{ii}} = [.60, .28]$
$\hat{\sigma}^2 \approx 3.28$

	Estimate	Std Error
$\hat{\alpha_1}$	4.39	.60
$\hat{\alpha_2}$	-.60	.28

(c) $\mu_i = \beta_1 + \beta_2 i + \beta_3 i^2$

$H_0 : \beta_2 = \beta_3 = 0$
since $\beta_2 = \beta_3 = 0$ we have $\mu_i = \beta_1 + 0$
Equivalent to testing $\mu_1 = \mu_2 = \mu_3 (= \beta_1)$
This is the same as the test from part a
Thus under $H_0 : F \approx 2.24 \sim F_{2,60}$
Using $\alpha = .05$
Critical Region $C_\alpha = \{F : F > F_{2,60;.95}\}$
Quantile ${F^{-1}}_{2,60;.95} \approx 3.15$
$2.24 < 3.14$ The test statistic is outiside of the critical region thus fail to reject H_0
p-value $= 1 - F_{2.24,2,60} \approx .12 > .05$ The p-value is larger than α
Not enough evidence to reject the null hypothesis that the population means for lung function measure are equal across all smoking groups.

Problem 2

(a)

$$\bar{x} = 2 \quad SXX = \sum (x_i - \bar{x})^2 = 6$$
$$SXY = \sum (x_i - \bar{x}) y_i = -y_1 - y_2 + 2y_4$$
$$\hat{\beta}_2 = \frac{SXY}{SXX} = (1/6)(-y_1 - y_2 + 2y_4)$$
$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$
$$\hat{\beta}_1 = (1/4)(y_1 + y_2 + y_3 + y_4) - 2 * (1/6)(-y_1 - y_2 + 2y_4)$$
$$\hat{\beta}_1 = (1/4 + 1/3)y_1 + (1/4 + 1/3)y_2 + (1/4)y_3 + (1/4 - 2/3)y_4$$
$$= (7/12)y_1 + (7/12)y_2 + (1/4)y_3 - (5/12)y_4$$
$$\hat{\beta}_1 = .583y_1 + .583y_2 + .25y_3 - .417y_4$$
$$.583 + .583 + .25 - .417 = 1$$

(b)

$$E(\hat{\beta}_1) = (7/12)E(y_1) + (7/12)E(y_2) + (1/4)E(y_3) - (5/12)E(y_4)$$
$$= (7/12)(1) + (7/12)(1) + (1/4)(2) - (5/12)(8)$$
$$= 20/12 - 40/12 = -20/12 \approx -1.667$$
$$E(\hat{\beta}_1) \approx -1.667$$

(c)

$$Var(\hat{\beta}_1) = Var[(7/12)y_1 + (7/12)y_2 + (1/4)y_3 - (5/12)y_4]$$
Since $Cov(Y_i, Y_j) = 0$ we have:
$$Var(\hat{\beta}_1) = (49/144)Var(y_1) + (49/144)Var(y_2) + (1/16)Var(y_3) + (25/144)Var(y_4)$$
$$= (49/144)1 + (49/144)2 + (1/16)4 + (25/144)5 = \frac{77}{36} \approx 2.139$$
$$Var(\hat{\beta}_1) \approx 2.139$$

(d)

$$\hat{\beta}_2 = \frac{SXY}{SXX} = (1/6)(-y_1 - y_2 + 2y_4)$$
$$= (-1/6)y_1 + (-1/6)y_2 + (1/3)y_4$$
$$\hat{\beta}_2 = -.167y_1 - .167y_2 + .333y_4$$
$$E(\hat{\beta}_2) = E[(-1/6)y_1 + (-1/6)y_2 + (1/3)y_4]$$
$$= (-1/6) * 1 + (-1/6) * 1 + (1/3) * 8$$
$$E(\hat{\beta}_2) = 2.333$$
$$Var(\hat{\beta}_2) = Var[(-1/6)y_1 + (-1/6)y_2 + (1/3)y_4]$$
Since $Cov(Y_i, Y_j) = 0$ we have:
$$Var(\hat{\beta}_2) = (1/36)Var(y_1) + (1/36)Var(y_2) + (1/9)Var(y_2)$$
$$= (1/36) * 1 + (1/36) * 2 + (1/9) * 5 = \frac{23}{36} \approx .639$$
$$Var(\hat{\beta}_2) \approx .639$$

(e)

$$Var(\hat{\beta}_1 + \hat{\beta}_2) = Var\left(\frac{1}{12}[7y_1 + 7y_2 + 3y_3 - 5y_4 - 2y_1 - 2y_2 + 4y_4]\right)$$
$$= \frac{1}{144}Var(5y_1 + 5y_2 + 3y_3 - y_4)$$
Since $Cov(Y_i, Y_j) = 0$ we have:
$$Var(\hat{\beta}_1 + \hat{\beta}_2) = \frac{1}{144}(25Var(y_1) + 25Var(y_2) + 9Var(y_3) + Var(y_4))$$

$$Var(\hat{\beta}_1 + \hat{\beta}_2) = \frac{1}{144}(5*1+5*2+3*4-1*5) = \frac{11}{72}$$
$$Cov(\hat{\beta}_1, \hat{\beta}_2) = (1/2)[Var(\hat{\beta}_1) + Var(\hat{\beta}_2) - Var(\hat{\beta}_1 + \hat{\beta}_2)]$$
$$Cov(\hat{\beta}_1, \hat{\beta}_2) = \frac{1}{2}\left(\frac{77}{36} + \frac{23}{36} - \frac{11}{72}\right) = \frac{189}{144} = 1.3125$$

(f)

$$\hat{\mu}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$$
$$\hat{\mu}_3 = \hat{\beta}_1 + \hat{\beta}_2 x_3 = \hat{\beta}_1 + 2\hat{\beta}_2$$
$$\hat{\mu}_3 = (7/12)y_1 + (7/12)y_2 + (1/4)y_3 - (5/12)y_4 + 2[(-1/6)y_1 + (-1/6)y_2 + (1/3)y_4]$$
$$= \frac{7-4}{12}y_1 + \frac{7-4}{12}y_2 + \frac{1}{4}y_3 + \frac{8-5}{12}y_4 = \frac{1}{4}y_1 + \frac{1}{4}y_2 + \frac{1}{4}y_3 + \frac{1}{4}y_4$$
$$\hat{\mu}_3 = \frac{1}{4}(y_1 + y_2 + y_3 + y_4)$$
$$E(\hat{\mu}_3) = (1/4)E\left(y_1 + y_2 + y_3 + y_4\right)$$
$$= (1/4)[E(y_1) + E(y_2) + E(y_3) + E(y_4)] = \frac{1+1+2+8}{4} = 3$$
$$E(\hat{\mu}_3) = 3$$

(g)

$$R_i = Y_i - \hat{\mu}_i$$
$$R_3 = y_3 - \hat{\mu}_3 = y_3 - \frac{1}{4}(y_1 + y_2 + y_3 + y_4)$$
$$R_3 = \frac{-1}{4}(y_1 + y_2 - 3y_3 + y_4)$$
$$E(R_3) = E(y_3) - E(\hat{\mu}_3) = 2 - 3 = -1$$
$$E(R_3) = -1$$
$$Var(R_3) = Var\left(\frac{-1}{4}(y_1 + y_2 - 3y_3 + y_4)\right)$$

Since $Cov(Y_i, Y_j) = 0$ we have:

$$Var(R_3) = \frac{1}{16}[Var(y_1) + y_2 + 9Var(y_3) + Var(y_4)] = \frac{1+2+9*4+5}{16} = \frac{11}{4}$$
$$Var(R_3) = 2.75$$
$$E(R_3^2) = Var(R_3) + E(R_3)^2 = 2.75 + 3^2 = 11.75$$