1. (a) $X_1 = \begin{cases} 1 & \text{if 1 bedroom} \\ 0 & \text{else} \end{cases}$

$X_2 = \begin{cases} 1 & \text{if 2 bedrooms} \\ 0 & \text{else} \end{cases}$

$X_3 = \begin{cases} 1 & \text{if 3 bedrooms} \\ 0 & \text{else} \end{cases}$

(b) df's are (in order) 4, 17, 21, 7, 14, 21

SSE's are (in order) 6294.5, 6066.7

(c) Full model:

$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_{14} + \beta_6 X_{24} + \beta_7 X_{34} + \varepsilon$

Reduced model: $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$

H_0: $\beta_5 = \beta_6 = \beta_7 = 0$

H_a: At least one of $\beta_5, \beta_6, \beta_7 \neq 0$

$$F = \frac{SSR(X_{14}\ X_{24}\ X_{34} \mid X_1\ X_2\ X_3\ X_4)/3}{SSE(\text{full})/14}$$

$$= \frac{(4800.9 - 4573.1)/3}{6066.7/14} = 0.175 < F_{.05}(3,14) = 3.34$$

$df_1 = 3$, $df_2 = 14$ $\Rightarrow$ fail to reject H_0.

The linear relationship between price and age does not depend on the number of bedrooms.

2.

$y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i * G_i + \varepsilon_i$

(a.) Men: $y_i = \beta_0 + (\beta_1 + \beta_2) X_i + \varepsilon_i$
$(G_i = 1)$

Women: $y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
$(G_i = 0)$

The slopes are $\beta_1 + \beta_2$ and β_1 for men & women, respectively.

(b.) Cook's distance is the only measure among the three diagnostic measures used that is for detecting influential observations.

Compare the largest value 0.579 to $F(3, 37)$, the tail area is about 63%. None of the observations is influential.

3. (a) $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ $\quad i=1,\dots,t;\ j=1,\dots,n_i$

where y_{ij} : jth observation under treatment i

μ : overall mean

α_i : effect due to treatment i

ε_{ij} : random error

$\hat{\alpha}_i = \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}$, $\quad \bar{y}_{\cdot\cdot} = (38 + 32.25 + 24.5)/3 = 31.58$

$\hat{\alpha}_1 = 38 - 31.58 = 6.42$

$\hat{\alpha}_2 = 32.25 - 31.58 = 0.67$

$\hat{\alpha}_3 = 24.5 - 31.58 = -7.08$

It seems that the higher a patient's activity level is before the surgery, the less time is needed for rehab. after surgery.

(b) $\varepsilon_{ij} \overset{\text{indep.}}{\sim} N(0, \sigma^2)$

Normal probability plot or histogram of residuals

(c) $SSTr = \sum_{i=1}^{t} n_i (\hat{\alpha}_i)^2$

$= 8\{6.42^2 + 0.67^2 + (-7.08)^2\} = 734.33$

$SSE = \sum_i (n_i - 1) s_i^2 = 7(30 + 13.93 + 17.43) = 429.52$

Source of variation	SS	df	MS	F
Treatment	734.33	2	367.17	17.95
Error	429.52	21	20.45	
Total	1163.85	23		

3 (d)

$H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0.$

H_a: At least one $\alpha_i \neq 0$, $i = 1, 2, 3$.

$$F = \frac{SSTr/(t-1)}{SSE/(N-t)} = \frac{734.33/2}{429.52/21} = 17.95$$

$df_1 = 2$, $df_2 = 21$.

$$\begin{aligned} p\text{-value} &= Pr\left(F(2,21) \geq 17.95\right) \\ &< .001 \quad \text{(Table 8)} \\ &< \alpha = .01 \end{aligned}$$

$\Rightarrow$ reject H_0

We conclude that the mean number of days required for successful rehabilitation is different for at least one of the three fitness groups.