

## Practice final

$$(a) \quad r^2 = \frac{SSR}{SSTO} = \frac{3.01308}{15.44508} = 0.195$$

About 19.5% of the total variation in winning times can be explained by the year.

(b) The winning time decreases, on average, by 0.009 seconds every year (or, by 0.036 seconds every four years).

$$(c) \quad \text{Men: } y_i = \beta_0 + (\beta_1 + \beta_2) x_i + \varepsilon_i$$

$$\text{Women: } y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

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

$$(d) \quad H_0: \beta_2 = 0 \quad H_a: \beta_2 \neq 0$$

$$F = \frac{(13.40617 - 3.01308) / 1}{2.03890 / 37} = 188.6$$

$$F_{0.05}(1, 37) < F_{0.05}(1, 30) = 4.17$$

$$188.6 > 4.17 \Rightarrow \text{reject } H_0.$$

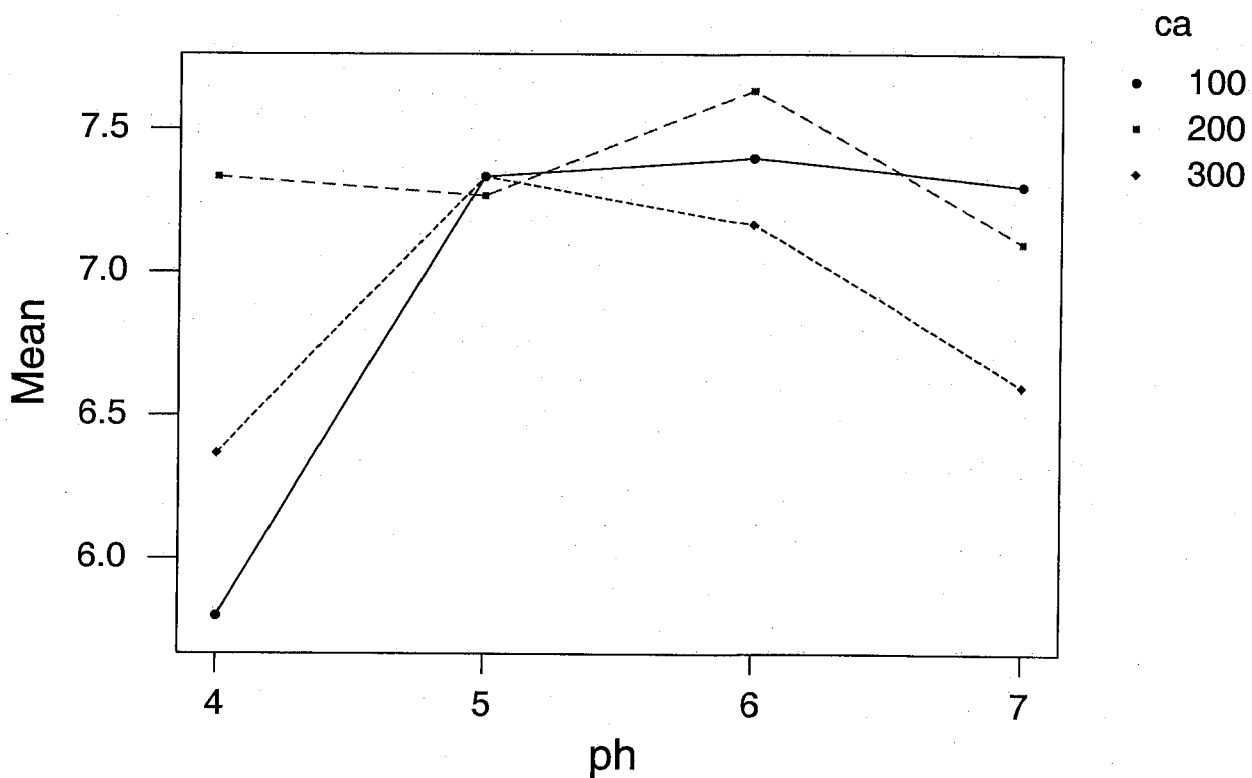
The change in the winning time per year for men is different than that for women.

(e) 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.

2 (a)

Interaction Plot - Data Means for diam



In order to draw this plot by hand, you need first calculate the mean of the three replications in each pH and calcium combination.

It seems that pH and calcium might interact.

2 (b)

This is a completely randomized design with  $4 \times 3$  factorial treatments. There are 3 replications at each factor level combination.

$$y_{ijh} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijh}$$

$$i = 1, 2, 3, 4; \quad j = 1, 2, 3; \quad h = 1, 2, 3$$

$y_{ijh}$ : increase in trunk diameter of the  $h^{\text{th}}$  tree under pH level  $i$  and calcium level  $j$ .

$\mu$ : overall mean

$\alpha_i$ : main effect of the  $i^{\text{th}}$  pH level

$\beta_j$ : main effect of the  $j^{\text{th}}$  calcium level.

$(\alpha\beta)_{ij}$ : interaction effect of the  $i^{\text{th}}$  pH level and  $j^{\text{th}}$  calcium level.

$\varepsilon_{ijh}$ : random error for the  $h^{\text{th}}$  tree under  $i^{\text{th}}$  pH level &  $j^{\text{th}}$  calcium level.

(c)  $H_0: \alpha_i = \beta_j = (\alpha\beta)_{ij} = 0$  for all  $i = 1, 2, 3, 4$  and  $j = 1, 2, 3$ .

$H_a$ : At least one of the terms in  $H_0 \neq 0$ .

$$F = \frac{(4.4608 + 1.4672 + 3.2550) / (3+2+6)}{1.6267 / 24} = 12.32$$

p-value =  $\Pr\{F(11, 24) > 12.32\} < .001 < \alpha \Rightarrow \text{reject } H_0$ .

The pH level, calcium and their interaction together have statistically significant predictive value.

$$2(d) \quad H_0: (\alpha\beta)_{ij} = 0 \quad \text{for all } i, j$$

$$H_a: (\alpha\beta)_{ij} \neq 0 \quad \text{for some } i, j$$

$$F = \frac{3.2550/6}{1.6267/24} = 8.00$$

$$F_{0.05}(6, 24) = 2.51$$

$$8.00 > 2.51 \Rightarrow \text{reject } H_0$$

There is a significant interaction between pH and calcium affecting the trunk diameter increase.

3. (a) Define

$$X_h = \begin{cases} 1 & \text{if treatment } i \\ 0 & \text{else} \end{cases} \quad h=2, 3, 4, 5.$$

$$(*) \quad y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} \\ + \beta_6 X_{1i} * X_{2i} + \beta_7 X_{1i} * X_{3i} + \beta_8 X_{1i} * X_{4i} + \beta_9 X_{1i} * X_{5i} + \epsilon_i$$

$i=1, \dots, 30$

$y$ : response

$\beta_0$ : intercept

$\beta_1, \dots, \beta_9$ : slopes

$X_1$ : covariate

$X_2 - X_5$ : dummy variables for treatment

$X_1 * X_h, h=2, 3, 4, 5$ : interaction between covariate  $X_1$  & treatment

$\epsilon$ : random error

Treatment level 2:

$$y = \beta_0 + \beta_1 x_1 + \varepsilon$$

Treatment level 2:  $y = (\beta_0 + \beta_2) + (\beta_1 + \beta_6) x_1 + \varepsilon$

Treatment level 3:  $y = (\beta_0 + \beta_3) + (\beta_1 + \beta_7) x_1 + \varepsilon$

Treatment level 4:  $y = (\beta_0 + \beta_4) + (\beta_1 + \beta_8) x_1 + \varepsilon$

Treatment level 5:  $y = (\beta_0 + \beta_5) + (\beta_1 + \beta_9) x_1 + \varepsilon$

Hence, model (\*) allows different intercepts & different slopes for the relationship between  $y$  &  $x_1$  for the 5 treatment levels. This is the most flexible model for this situation.

3(b).  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon$  (\*)

Treatment level 1:  $y = \beta_0 + \beta_1 x + \varepsilon$

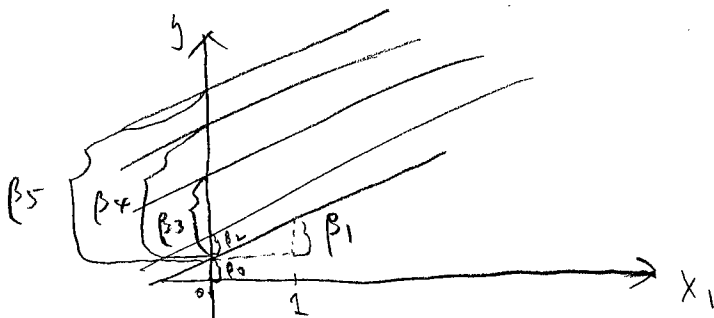
2:  $y = (\beta_0 + \beta_2) + \beta_1 x + \varepsilon$

3:  $y = (\beta_0 + \beta_3) + \beta_1 x + \varepsilon$

4:  $y = (\beta_0 + \beta_4) + \beta_1 x + \varepsilon$

5:  $y = (\beta_0 + \beta_5) + \beta_1 x + \varepsilon$

Hence model (\*) assumes the lines to be parallel (same slope) but not coincident (different intercepts).



Depending on the values of  $\beta_0, \beta_1, \dots, \beta_5$ , the lines may be positioned differently. But they should be parallel.

$$3(c). \quad y = \beta_0 + \beta_1 x_i + \varepsilon$$

Same intercept ( $\beta_0$ ) and same slope ( $\beta_1$ ) for all 5 lines corresponding to the 5 treatment levels.

(d)

Complete model : model (\*)

Reduced model : model (\*),

$$F = \frac{\{SSE(R) - SSE(F)\} / [df\{SSE(R)\} - df\{SSE(F)\}]}{SSE(F) / df\{SSE(F)\}}$$

$$df_1 = df\{SSE(R)\} - df\{SSE(F)\}$$

$$df_2 = df\{SSE(F)\}$$