36-220 Engineering Stats, Fall 2005
Homework 11 Solutions

Due: December 7, 2005

Regular Exercises

40.

a. $\hat{\mu}_{Y\cdot 10,.5,50,100} = 1.52 + .02(10) - 1.40(.5) + .02(50) - .0006(100) = 1.96$

b. $\hat{\mu}_{Y\cdot 20,.5,50,30} = 1.52 + .02(20) - 1.40(.5) + .02(50) - .0006(30) = 1.40$

c. $\hat{\beta}_4 = -.0006;\ 100\hat{\beta}_4 = -.06$.

d. There are no interaction predictors – e.g., $x_5 = x_1 x_4$ -- in the model. There would be dependence if interaction predictors involving x_4 had been included.

e. $R^2 = 1 - \frac{20.0}{39.2} = .490$. For testing $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ vs. H_a: at least one among $\beta_1, \ldots, \beta_4$ is not zero, the test statistic is $F = \frac{R^2/k}{(1-R^2)/(n-k-1)}$. H_0 will be rejected if $f \ge F_{.05,4,25} = 2.76$. $f = \frac{.490/4}{.510/25} = 6.0$. Because $6.0 \ge 2.76$, H_0 is rejected and the model is judged useful (this even though the value of R^2 is not all that impressive).

41. $H_0 : \beta_1 = \beta_2 = \ldots = \beta_6 = 0$ vs. H_a: at least one among $\beta_1, \ldots, \beta_6$ is not zero. The test statistic is $F = \frac{R^2/k}{(1-R^2)/(n-k-1)}$. H_0 will be rejected if $f \ge F_{.05,6,30} = 2.42$.

$f = \frac{.83/6}{(1-.83)/30} = 24.41$. Because $24.41 \ge 2.42$, H_0 is rejected and the model is judged useful.

42.

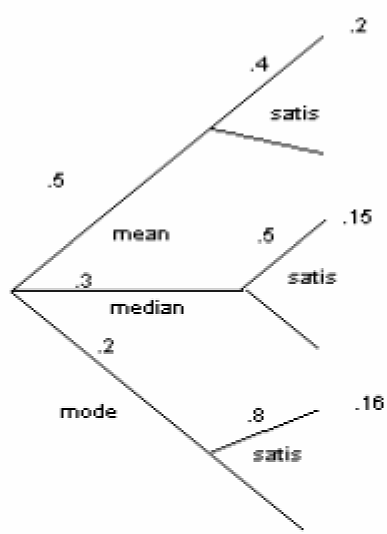
a. To test $H_0 : \beta_1 = \beta_2 = 0$ vs. H_a : at least one $\beta_i \neq 0$, the test statistic is $f = \frac{MSR}{MSE} = 319.31$ (from output). The associated p-value is 0, so at any reasonable level of significance, H_0 should be rejected. There does appear to be a useful linear relationship between temperature difference and at leas one of the two predictors.

b. The degrees of freedom for SSE = n – (k + 1) = 9 – (2 – 1) = 6 (which you could simply read in the DF column of the printout), and $t_{.025,6} = 2.447$, so the desired confidence interval is $3.000 \pm (2.447)(.4321) = 3.000 \pm 1.0573$, or about $(1.943, 4.057)$. Holding furnace temperature fixed, we estimate that the average change in temperature difference on the die surface will be somewhere between 1.943 and 4.057.

c. When $x_1 = 1300$ and $x_2 = 7$, the estimated average temperature difference is $\hat{y} = -199.56 + .2100x_1 + 3.000x_2 = -199.56 + .2100(1300) + 3.000(7) = 94.44$. The desired confidence interval is then $94.44 \pm (2.447)(.353) = 94.44 \pm .864$, or $(93.58, 95.30)$.

d. From the printout, s = 1.058, so the prediction interval is $94.44 \pm (2.447)\sqrt{(1.058)^2 + (.353)^2} = 94.44 \pm 2.729 = (91.71, 97.17)$.

65.



P(satis) = .51

$$P(\text{mean} \mid \text{satis}) = \frac{.2}{.51} = .3922$$

P(median | satis) = .2941

P(mode | satis) = .3137

So Mean (and not Mode!) is the most likely author, while Median is least.

83.

a. Let D_1 = detection on 1st fixation, D_2 = detection on 2nd fixation.

$$P(\text{detection in at most 2 fixations}) = P(D_1) + P(D_1' \cap D_2) = P(D_1) + P(D2 \mid D1')P(D_1) = p + p(1-p) = p(2-p).$$

b. Define $D_1, D_2, \ldots, D_n$ as in a. Then P(at most n fixations)

$$= P(D_1) + P(D_1' \cap D_2) + P(D_1' \cap D_2' \cap D_3) + \ldots + P(D_1' \cap D_2' \cap \ldots \cap D_{n-1}' \cap D_n)$$

$$= p + p(1-p) + p(1-p)^2 + \ldots + p(1-p)^{n-1}$$

$$= p\,[\,1 + (1-p) + (1-p)^2 + \ldots + (1-p)^{n-1}] = p \bullet \frac{1-(1-p)^n}{1-(1-p)} = 1-(1-p)^n$$

Alternatively, P(at most n fixations) = 1 – P(at least n+1 are req'd)

$$= 1 - P(\text{no detection in 1st n fixations}) = 1 - P(D_1' \cap D_2' \cap \ldots \cap D_n') = 1-(1-p)^n$$

c. P(no detection in 3 fixations) = $(1-p)^3$

95. When three experiments are performed, there are 3 different ways in which detection can occur on exactly 2 of the experiments: (i) #1 and #2 and not #3 (ii) #1 and not #2 and #3; (iii) not#1 and #2 and #3. If the impurity is present, the probability of exactly 2 detections in three (independent) experiments is (.8)(.8)(.2) + (.8)(.2)(.8) + (.2)(.8)(.8) = .384. If the impurity is absent, the analogous probability is 3(.1)(.1)(.9) = .027. Thus

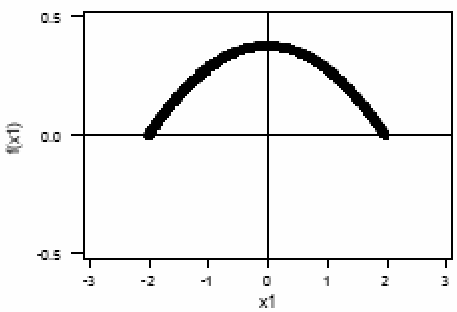
P(present | detected in exactly 2 out of 3) =

$$\frac{P(\det ected.in.exactly.2 \cap present)}{P(\det ected.in.exactly.2)}$$

$$= \frac{(.384)(.4)}{(.384)(.4)+(.027)(.6)} = .905$$

3.

a. Graph of $f(x) = .09375(4 - x^2)$



b. $P(X > 0) = \int_0^2 .09375(4 - x^2)dx = .09375(4x - \frac{x^3}{3})\Big]_0^2 = .5$

c. $P(-1 < X < 1) = \int_{-1}^1 .09375(4 - x^2)dx = .6875$

d. $P(x < -.5 \text{ OR } x > .5) = 1 - P(-.5 \le X \le .5) = 1 - \int_{-.5}^{.5} .09375(4 - x^2)dx$
$= 1 - .3672 = .6328$

12.

a. $P(X < 0) = F(0) = .5$

b. $P(-1 \le X \le 1) = F(1) - F(-1) = \frac{11}{16} = .6875$

c. $P(X > .5) = 1 - P(X \le .5) = 1 - F(.5) = 1 - .6836 = .3164$

d. $F(x) = F'(x) = \frac{d}{dx}\left(\frac{1}{2} + \frac{3}{32}\left(4x - \frac{x^3}{3}\right)\right) = 0 + \frac{3}{32}\left(4 - \frac{3x^2}{3}\right) = .09375\left(4 - x^2\right)$

e. $F(\tilde{\mu}) = .5$ by definition. F(0) = .5 from a above, which is as desired.

64. Let $X_1, \ldots, X_5$ denote morning times and $X_6, \ldots, X_{10}$ denote evening times.

a. $E(X_1 + \ldots + X_{10}) = E(X_1) + \ldots + E(X_{10}) = 5\,E(X_1) + 5\,E(X_6)$
$= 5(4) + 5(5) = 45$

b. $Var(X_1 + \ldots + X_{10}) = Var(X_1) + \ldots + Var(X_{10}) = 5\,Var(X_1) + 5Var(X_6)$
$= 5\left[\frac{64}{12} + \frac{100}{12}\right] = \frac{820}{12} = 68.33$

c. $E(X_1 - X_6) = E(X_1) - E(X_6) = 4 - 5 = -1$

$Var(X_1 - X_6) = Var(X_1) + Var(X_6) = \frac{64}{12} + \frac{100}{12} = \frac{164}{12} = 13.67$

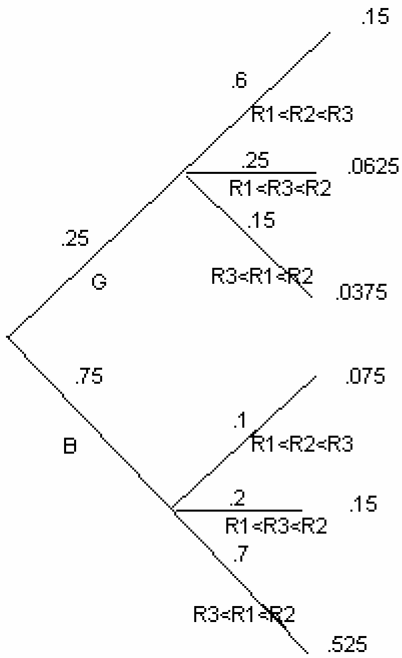
d. $E[(X_1 + \ldots + X_5) - (X_6 + \ldots + X_{10})] = 5(4) - 5(5) = -5$

$Var[(X_1 + \ldots + X_5) - (X_6 + \ldots + X_{10})]$

$= Var(X_1 + \ldots + X_5) + Var(X_6 + \ldots + X_{10})] = 68.33$

Extra Credit

104.



a. $P(G \mid R_1 < R_2 < R_3) = \dfrac{.15}{.15 + .075} = .67$, $P(B \mid R_1 < R_2 < R_3) = .33$, classify as granite.

b. $P(G \mid R_1 < R_3 < R_2) = \dfrac{.0625}{.2125} = .2941 < .05$, so classify as basalt.

$P(G \mid R_3 < R_1 < R_2) = \dfrac{.0375}{.5625} = .0667$, so classify as basalt.

c. P(erroneous classif) = P(B classif as G) + P(G classif as B)
= P(classif as G | B)P(B) + P(classif as B | G)P(G)
= $P(R_1 < R_2 < R_3 \mid B)(.75) + P(R_1 < R_3 < R_2 \text{ or } R_3 < R_1 < R_2 \mid G)(.25)$
= $(.10)(.75) + (.25 + .15)(.25) = .175$

d. For what values of p will $P(G \mid R_1 < R_2 < R_3) > .5$, $P(G \mid R_1 < R_3 < R_2) > .5$, $P(G \mid R_3 < R_1 < R_2) > .5$?

$$P(G \mid R_1 < R_2 < R_3) = \frac{.6p}{.6p + .1(1-p)} = \frac{.6p}{.1 + .5p} > .5 \text{ iff } p > \frac{1}{7}$$

$$P(G \mid R_1 < R_3 < R_2) = \frac{.25p}{.25p + .2(1-p)} > .5 \text{ iff } p > \frac{4}{9}$$

$$P(G \mid R_3 < R_1 < R_2) = \frac{.15p}{.15p + .7(1-p)} > .5 \text{ iff } p > \frac{14}{17} \text{ (most restrictive)}$$

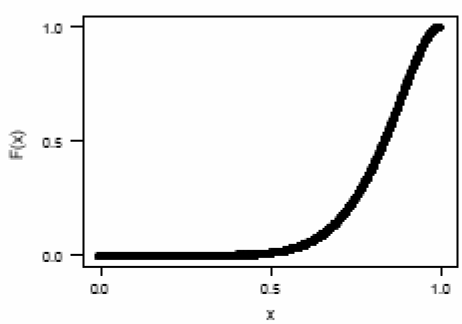
If $p > \frac{14}{17}$ always classify as granite.

15.

a. $F(X) = 0$ for $x \le 0$, $= 1$ for $x \ge 1$, and for $0 < X < 1$,

$$F(X) = \int_{-\infty}^{x} f(y)dy = \int_0^x 90y^8(1-y)dy = 90\int_0^x (y^8 - y^9)dy$$

$$90\left(\tfrac{1}{9}y^9 - \tfrac{1}{10}y^{10}\right)\Big]_0^x = 10x^9 - 9x^{10}$$



b. $F(.5) = 10(.5)^9 - 9(.5)^{10} \approx .0107$

c. $P(.25 \le X \le .5) = F(.5) - F(.25) \approx .0107 - [10(.25)^9 - 9(.25)^{10}]$
$\approx .0107 - .0000 \approx .0107$

d. The 75th percentile is the value of x for which $F(x) = .75$
$\Rightarrow .75 = 10(x)^9 - 9(x)^{10}$ $\Rightarrow x \approx .9036$

e. $E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_0^1 x \cdot 90x^8(1-x)dx = 90\int_0^1 x^9(1-x)dx$

$= 9x^{10} - \frac{90}{11}x^{11}\Big]_0^1 = \frac{9}{11} \approx .8182$

$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx = \int_0^1 x^2 \cdot 90x^8(1-x)dx = 90\int_0^1 x^{10}(1-x)dx$

$= \frac{90}{11}x^{11} - \frac{90}{12}x^{12}\Big]_0^1 \approx .6818$

$V(X) \approx .6818 - (.8182)^2 = .0124$, $\sigma_x = .11134$.

f. $\mu \pm \sigma = (.7068, .9295)$. Thus, $P(\mu - \sigma \le X \le \mu + \sigma) = F(.9295) - F(.7068)$
$= .8465 - .1602 = .6863$

21. $E(\text{area}) = E(\pi R^2) = \int_{-\infty}^{\infty} \pi r^2 f(r)dr = \int_9^{11} \pi r^2 \left(\frac{3}{4}\right)\left(1-(10-r)^2\right)dr$

$= \left(\frac{3}{4}\right)\pi\int_9^{11} r^2\left(1-(100-20r+r^2)\right)dr = \frac{3}{4}\pi\int_9^{11} -99r^2 + 20r^3 - r^4 dr = 100 \cdot 2\pi$

43.

a. $x_1 = 2.6$, $x_2 = 250$, and $x_1x_2 = (2.6)(250) = 650$, so
$\hat{y} = 185.49 - 45.97(2.6) - 0.3015(250) + 0.0888(650) = 48.313$

b. No, it is not legitimate to interpret β_1 in this way. It is not possible to increase by 1 unit the cobalt content, x_1, while keeping the interaction predictor, x_3, fixed. When x_1 changes, so does x_3, since $x_3 = x_1x_2$.

c. Yes, there appears to be a useful linear relationship between y and the predictors. We determine this by observing that the p-value corresponding to the model utility test is < .0001 (F test statistic = 18.924).

d. We wish to test $H_0: \beta_3 = 0$ vs. $H_a: \beta_3 \ne 0$. The test statistic is t=3.496, with a corresponding p-value of .0030. Since the p-value is < alpha = .01, we reject H_0 and conclude that the interaction predictor does provide useful information about y.

e. A 95% C.I. for the mean value of surface area under the stated circumstances requires the following quantities:
$\hat{y} = 185.49 - 45.97(2) - 0.3015(500) + 0.0888(2)(500) = 31.598$. Next,
$t_{.025,16} = 2.120$, so the 95% confidence interval is
$31.598 \pm (2.120)(4.69) = 31.598 \pm 9.9428 = (21.6552, 41.5408)$

47.

a. For a 1% increase in the percentage plastics, we would expect a 28.9 kcal/kg increase in energy content. Also, for a 1% increase in the moisture, we would expect a 37.4 kcal/kg decrease in energy content.

b. The appropriate hypotheses are $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ vs. H_a : at least one $\beta_i \neq 0$. The value of the F test statistic is 167.71, with a corresponding p-value that is extremely small. So, we reject H_0 and conclude that at least one of the four predictors is useful in predicting energy content, using a linear model.

c. $H_0 : \beta_3 = 0$ vs. $H_a : \beta_3 \neq 0$. The value of the t test statistic is t = 2.24, with a corresponding p-value of .034, which is less than the significance level of .05. So we can reject H_0 and conclude that percentage garbage provides useful information about energy consumption, given that the other three predictors remain in the model.

d. $\hat{y} = 2244.9 + 28.925(20) + 7.644(25) + 4.297(40) - 37.354(45) = 1505.5$, and $t_{.025,25} = 2.060$. (Note an error in the text: $s_{\hat{y}} = 12.47$, not 7.46). So a 95% C.I for the true average energy content under these circumstances is $1505.5 \pm (2.060)(12.47) = 1505.5 \pm 25.69 = (1479.8, 1531.1)$. Because the interval is reasonably narrow, we would conclude that the mean energy content has been precisely estimated.

e. A 95% prediction interval for the energy content of a waste sample having the specified characteristics is $1505.5 \pm (2.060)\sqrt{(31.48)^2 + (12.47)^2}$
$= 1505.5 \pm 69.75 = (1435.7, 1575.2)$.

75.

a. $H_0 : \beta_1 = \beta_2 = 0$ will be rejected in favor of H_a : either β_1 or $\beta_2 \neq 0$ if $f = \dfrac{R^2/k}{(1-R^2)/(n-k-1)} \geq F_{\alpha,k,n-k-1} = F_{.01,2,7} = 9.55$. $SST = \Sigma y^2 - \dfrac{(\Sigma y)}{n} = 264.5$, so $R^2 = 1 - \dfrac{26.98}{264.5} = .898$, and $f = \dfrac{.898/2}{(.102)/7} = 30.8$. Because $30.8 \geq 9.55$ H_0 is rejected at significance level .01 and the quadratic model is judged useful.

b. The hypotheses are $H_0 : \beta_2 = 0$ vs. $H_a : \beta_2 \neq 0$. The test statistic value is $t = \dfrac{\hat{\beta}_2}{s_{\hat{\beta}_2}} = \dfrac{-2.3621}{.3073} = -7.69$, and $t_{.0005,7} = 5.408$, so H_0 is rejected at level .001 and p-value < .001. The quadratic predictor should not be eliminated.

c. x = 1 here, and $\hat{\mu}_{Y \cdot 1} = \hat{\beta}_0 + \hat{\beta}_1(1) + \hat{\beta}_2(1)^2 = 45.96$. $t_{.025,7} = 1.895$, giving the C.I. $45.96 \pm (1.895)(1.031) = (44.01, 47.91)$.

76.

a. 80.79

b. Yes, p-value = .007 which is less than .01.

c. No, p-value = .043 which is less than .05.

d. $.14167 \pm (2.447)(.03301) = (.0609, .2224)$

e. $\hat{\mu}_{y \cdot 9,66} = 6.3067$, using $\alpha = .05$, the interval is

$$6.3067 \pm (2.447)\sqrt{(.4851)^2 + (.162)^2} = (5.06, 7.56)$$