GDBA 502
Fall 2010

Review Problems: DMorin

#1) ① $H_0: \pi_1 - \pi_2 = 0$ π_1 = proportion of blue-collars in favor
$H_1: \pi_1 - \pi_2 \neq 0$ π_2 = " of white-collars "

② $$Z = \frac{(p_1 - p_2) - 0}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

③ reject Ho if $|Z| > Z_{\alpha/2} = Z_{.025} = 1.96$

④ calculation:

$$\bar{p} = \frac{67+33}{130+50} = 0.556, \quad p_1 = \frac{67}{130} = 0.515, \quad p_2 = \frac{33}{50} = 0.66$$

$$Z = \frac{(0.515 - 0.66) - 0}{\sqrt{0.556(1-.556)\left(\frac{1}{130}+\frac{1}{50}\right)}} = \frac{-0.145}{0.0827} = -1.754$$

⑤ $|Z| = |-1.754| \overset{?}{>} 1.96$ no → cannot reject Ho

⑥ Not enough evidence to conclude that the responses differ among the two groups

⑦ $p\text{-value} = 2P(Z < -1.754) = 0.0794$

Note: another way to solve this problem

Ho: responses and groups are independent
H_1: " " not "

reject Ho if $\chi^2 > \chi^2_{(2-1)(2-1),.05} = \chi^2_{1,.05} = 3.8415$

	Blue	White	
For	67 (72.2)	33 (27.8)	100
against	63 (57.8)	17 (22.2)	80
	130	50	180

$$\chi^2 = \frac{(67-72.2)^2}{72.2} + \dots + \frac{(17-22.2)^2}{22.2}$$

$= 3.033$

we cannot reject Ho ⇒ not enough evidence to reject the independence of responses & groups

$p\text{-value} = P(\chi^2_1 > 3.033)$ between .05 and .10.

#2) a) $b_1 = \dfrac{\Sigma x_i y_i - n\bar{x}\bar{y}}{(n-1)S_x^2} = \dfrac{7897 - 5(20)(75.4)}{(5-1)(5.7)^2} = 2.75$

$b_0 = \bar{y} - b_0\bar{x} = 75.4 - 2.75(20) = 20.4$

$y = 20.4 + 2.75x$

For every additional hour of study, we can expect an additional 2.75 points on the Exam score

b). $SSE = \Sigma y_i^2 - n\bar{y}b_0 - b_1 \Sigma x_i y_i$
$= 29747 - 5(75.4)(20.4) - 2.75(7897)$
$= 339.45$

c) $b_1 \pm t_{n-2, \alpha/2} \, s_{b_1}$ $t_{5-2, \frac{.10}{2}} = t_{3,.05} = 2.353$

$s_{b_1} = \dfrac{\sqrt{MSE}}{\sqrt{(n-1) s_x^2}}$ $MSE = \dfrac{SSE}{n-2} = \dfrac{339.45}{5-2}$

$= 113.15$

$s_{b_1} = \dfrac{\sqrt{113.15}}{\sqrt{(5-1)(5.7)^2}} = 0.933$

$2.75 \pm 2.353\,(0.933)$

2.75 ± 2.195

$\Rightarrow 0.555 < \beta_1 < 4.945$

d) Yes since ϕ is not in the 90% C.I for β_1

e) $R^2 = \dfrac{SSR}{SST}$ where $SST = (n-1) s_y^2 =$

$= (5-1)(18.17)^2 = 1320.6$

$SSR = SST - SSE = 1320.6 - 339.45$

$= 981.15$

$R^2 = \dfrac{981.15}{1320.6} = 0.743$

74.3% of the variation in Exam score can be explained by the reg of score on hours of study.

f) $r = \pm\sqrt{R^2} = \sqrt{.743} = 0.862$ (positive since $b_1 > 0$)

g) $\hat{y} \pm t_{n-2, \frac{.10}{2}} \sqrt{MSE} \sqrt{1 + h_i}$

$t_{3,.05} = 2.353$, $h_i = \dfrac{1}{n} + \dfrac{(x_i - \bar{x})^2}{(n-1)s_x^2} = \dfrac{1}{5} + \dfrac{(21-20)^2}{(5-1)(5.7)^2}$

$\hat{y} = 20.4 + 2.75(21) = 78.15$ $= 0.2077$

$\Rightarrow 78.15 \pm 2.353 \sqrt{113.15} \sqrt{1 + .2077} \Rightarrow 78.15 \pm 27.51 \Rightarrow 50.64 < \hat{y}_{21} < 105.66$

#3 a) ① $H_0: \mu \geq 20$ ② $t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$ where $\mu_0 = 20$

$H_1: \mu < 20$

③ reject Ho if $t < -t_{n-1,\alpha} = -t_{9-1,.025} = -2.3060$

④ $t = \frac{18-20}{2/\sqrt{9}} = -3.0$ ⇒ ⑤ reject Ho ⇒ ⑥ we have enough evidence to conclude shelftime is shorter

⑦ pvalue = $P(t_8 < -3) = P(t_8 > 3) = 0.0085$

b) $n = 7$ X = # beginners that moves to intermediate

π = P(beginner moves to intermediate) = 0.4

$P(X \leq 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$

$= .0280 + .1306 + .2613 + .2903$

$= 0.7102$

#4) ① $H_0: \pi \leq .14$ ③ $Z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$ where $\pi_0 = 0.14$

$H_1: \pi > .14$

② $\alpha = .10$

④ reject Ho if $Z > Z_{.10} = 1.28$

⑤ $p = \frac{20}{100} = 0.20$ → $Z = \frac{0.20 - 0.14}{\sqrt{\frac{.14(1-.14)}{100}}} = 1.7292$

⑥ reject Ho since $Z = 1.7292 > 1.28$

⑦ there is enough evidence to conclude that more than 14% would pay over $800 for a system

⑧ p-value = $P(Z > 1.7292) = 0.0419$

#5) a) CI for $\mu_1 - \mu_2$: $(\bar{X}_1 - \bar{X}_2) \pm t_{n_1+n_2-2,\frac{\alpha}{2}} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

assuming $\sigma_1 = \sigma_2$

Group 1: formula Group 2: breast $S_p^2 = \frac{(9-1)(1.424)^2 + (9-1)(1.364)^2}{9+9-2}$

$\bar{X}_1 = 7.56$ $\bar{X}_2 = 8.89$

$S_1 = 1.424$ $S_2 = 1.364$ $= 1.944$

$t_{16,.05} = 1.746$

$$(7.56 - 8.89) \pm 1.746\sqrt{1.944}\sqrt{\tfrac{1}{9}+\tfrac{1}{9}}$$
$$-1.33 \pm 1.1476$$

$$-2.4776 < \mu_1 - \mu_2 < -0.1824$$

b) ① $H_0: \mu_2 - \mu_1 \le D$ ② $t = \dfrac{(\bar{X}_2 - \bar{X}_1) - D}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{16}$

$H_1: \mu_2 - \mu_1 > D$

③ reject Ho if $t > t_{16,.05} = 1.7459$

④ $t = \dfrac{(8.89 - 7.56) - 1}{\sqrt{1.944}\sqrt{\frac{1}{9}+\frac{1}{9}}} = 2.0235$

⑤ $t = 2.0235 > 1.7459 \Rightarrow$ we reject Ho

⑥ Enough evidence to conclude that the average weight gain of breastfed baby is more than that of formula fed babies.

#6) ANOVA

source	df	SS	MS	F
Reg	3	4649.4	1549.8	85.8
Residuals	16	288.9	18.06	
total	19	4938.3		

Parameter	Estimate	St Er	t ratio
Intercept	184.28	113.05	1.63
X_1	8.44	1.51	5.59
X_2	10.13	2.31	4.39
X_3	7.02	10.81	0.65

b) $\hat{y} = 184.28 + 8.44X_1 + 10.13X_2 + 7.02X_3$

c) $H_0: \beta_1 = \beta_2 = \beta_3 = 0$

H_1: at least one $\beta_i \ne 0$

reject Ho if $F > F_{3,16,.05} = 3.2389$

$F = 85.8 > 3.2389 \Rightarrow$ reject Ho $\Rightarrow$ model is sign.

d) most significant predictor is the predictor with largest |t ratio| . therefore X_1 (age) is most significant

e) $b_1 \pm t_{16,.025}\, S_{b_1}$

$8.44 \pm 2.1199\,(1.51) \Rightarrow 8.44 \pm 3.2011$

$5.2389 < \beta_1 < 11.6411$

f) we should eliminate x_3 (previous attendance) which has the lowest t ratio and is not significant

g) $R^2 = \frac{SSR}{SST} = \frac{4649.4}{4938.3} = 94.15\%$

h) X_1 age $= 100$, $X_2 = 50$, $X_3 = 4$ (4 previous attendance)

$\hat{y} = 184.28 + 8.44(100) + 10.13(50) + 7.02(4)$

$= \$\,1562.86$

#7)

	pass test	Fail Test	
perform satisfactorily	.72	.08	0.80
do not perform well	.07	.13	0.20
	.81	.21	1.00

a) P(satisfactorily | pass test) = $\frac{.72}{.81} = 0.91$

b) P(do not perform | fail test) = $\frac{.13}{.21} = 0.62$

#8) $S = \sqrt{MSE} = 11.1 \Rightarrow MSE = 123.21$

a)

Source	df	SS	MS	F
Regr	3	4807.54	1602.51	13.0064
Residual	26	3203.46	123.21	
Total	29	8011		

$H_0: \beta_1 = \beta_2 = \beta_3 = 0$

H_1: at least one $\beta_i \neq 0$

reject H_0 if $F > F_{3,26,.05} = 2.98$

$F = 13.0064 > 2.98 \Rightarrow$ reject H_0, model is significant

b) For every additional km from the coast the average annual precipitation is reduced by 0.1429 cm

c) $R^2 = \frac{SSR}{SST} = \frac{4807.54}{8011} = 0.60$

d)

	coefficient	stdev	t ratio
Constant	-102.36		
X_1	0.00409	0.0012	3.4083
X_2	3.4511	0.7948	4.3421
X_3	-0.1429	0.03634	3.9323

the variable with the largest t-ratio contributes the most $\Rightarrow X_2$ = latitude

e) $\hat{y} = -102.36 + .00409(5000) + 3.4511(37.3) - 0.1429(65)$

$= 44.43$ cm

#9)

obs

	meet deadline	do not meet	
little	20	12	32
moderate	19	19	38
high	21	9	30
	60	40	

Ho: independence

H1: dependence

reject Ho if $\chi^2 > \chi^2_{(3-1)(2-1),\alpha}$

if $\chi^2 > \chi^2_{2,.05} = 5.9915$

exp:

Exp

	MEET	NOT MEET	
little	19.2	12.8	32
mod	22.8	15.2	38
high	18	12	30
	60	40	100

$$\chi = \Sigma \frac{(o_i - e_i)^2}{e_i} = 2.9167$$

$\Rightarrow$ $\chi^2 = 2.9167 \not> 5.9915 \Rightarrow$

Cannot reject independence

no, we cannot conclude that the completion of a project on time and the level of cross-functionality are dependent

#10)

Ho: $\mu_A = \mu_B = \mu_C$

H1: at least one μ_i is different

$SSW = (n_1-1)S_1^2 + (n_2-1)S_2^2 + (n_3-1)S_3^2$

$= (5-1)(39)^2 + (5-1)(44.4)^2 + (5-1)(57.6)^2 = 27247.2$

$SSA = n_1(\bar{X}_1 - \bar{\bar{X}})^2 + n_2(\bar{X}_2 - \bar{\bar{X}})^2 + n_3(\bar{X}_3 - \bar{\bar{X}})^2$ $\bar{\bar{X}} = 649.87$

$= 5(697 - 649.87)^2 + 5(662 - 649.87)^2 + 5(590.6 - 649.87)^2$

8 $= 29406.5$

ANOVA

Source	SS	df	MS	F
Among	29406.5	3-1=2	14703.25	6.4755
Within	27247.2	12	2270.58	
total.	56653.7	15-1=14		

Reject Ho if $F > F_{2,12,.05} = 3.8853$

$F = 6.4755 > 3.8853 \Rightarrow$ reject Ho
at least one μ_i is different

b) assumptions: 3 indep normal popn's with $\sigma_1 = \sigma_2 = \sigma_3$

not required

c)

$\mu_1 \neq \mu_2$? $\quad |\bar{X}_1 - \bar{X}_2| \overset{?}{>} Q_U \sqrt{\frac{MSW}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

Q_U at $3, 12, .05 = 3.77$

$|697 - 662| = 35 \overset{?}{>} 3.77 \sqrt{\frac{2270.6}{2}\left(\frac{1}{5} + \frac{1}{5}\right)} = 80.34$ No

$\mu_1 \neq \mu_3$? $\quad |697 - 590.6| = 106.40 > 80.34$

$\Rightarrow \mu_1 \neq \mu_3$

$\mu_2 \neq \mu_3$? $\quad |662 - 590.6| = 71.4 \not> 80.34$ -

only $\mu_1 \neq \mu_3$

#11

a) $\hat{y} = 12.77 + 0.069X_1 + 9.828X_2 + 2.560X_3 + 80.23X_4$

for every increase of 1 sq feet the sales price ↑ \$69
for every additional bedroom " " \$9828
for every additional bathroom " " \$2560
a house downtown sells of \$80230 more

b) $\hat{y} = 12.77 + 0.069(2000) + 9.828(2) + 2.560(2) + 80.23(1)$
$= 255.776 \times \$1000$
$= \$255,776$

c) $H_0: \beta_i = 0$
$H_1: \beta_i \neq 0$

reject Ho if $|t| = \left|\frac{b_i - 0}{S_{b_i}}\right| > t_{319-5, .025} = 1.96$

	Coefficients	Standard Error	t Stat
Intercept	12.77	6.88	1.856
Sq. Feet	0.069	0.0035	19.714
Bedrooms	9.828	2.804	3.505
Bathrooms	2.560	1.489	1.719
Area	80.23	3.993	20.09

$> 1.96 \Rightarrow \beta_1 \neq 0$
$> 1.96 \Rightarrow \beta_2 \neq 0$
$\not> 1.96 \Rightarrow \beta_3$ not sign different from 0
$> 1.96 \Rightarrow \beta_4 \neq 0$

∴ Sq feetage, # bedrooms and area make a sign contribution at 5%

d) $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$
H_1: at least one $\beta_i \neq 0$

reject Ho if $F > F_{4,314,.05} = 2.37$

Standard error $S = 25.7862 = \sqrt{MSE} \Rightarrow MSE = 664.928 = \frac{SSE}{314} \Rightarrow SSE = 208787.4$

ANOVA

	df	SS	MS
Regression	4	1064245.2	266061.3
Residual	314	208 787.4	664.928
Total	318	1273 032.6	

F
$400.13 > 2.37 \Rightarrow$ reject Ho

at least one $\beta_i \neq 0$.

this multiple regression model is valid

e) Since $R^2 = \frac{SSR}{SST} = \frac{1064245.2}{1273032.6} = 0.8360$

we can find $R^2_{adj} = 1 - \left[(1-R^2)\frac{(n-1)}{(n-k-1)}\right]$

$= 1 - \left[(1-.8360)\frac{(319-1)}{(319-4-1)}\right]$

$= 1 - 0.1661$

$= 0.8339$

83.39% of the variation in sales is explained by the multiple regression model adjusted for the number of predictors and sample size

f) downtown $X_4 = 1$

$\hat{y} = 12.77 + 0.069X_1 + 9.828X_2 + 2.560X_3 + 80.23(1)$

$\hat{y} = 93 + 0.069X_1 + 9.828X_2 + 2.560X_3$

outside downtown $X_4 = 0$

$\hat{y} = 12.77 + 0.069X_1 + 9.828X_2 + 2.560X_3$