Math 222: Exam 2
Spring – 2017
04/12/2017
80 Minutes

Name: *Caleb M^c^Whorter — Solutions*

Write your name on the appropriate line on the exam cover sheet. This exam contains 9 pages (including this cover page) and 4 questions. Check that you have every page of the exam. Answer the questions in the spaces provided on the question sheets. Be sure to answer every part of each question and show all your work. If you run out of room for an answer, continue on the back of the page — being sure to indicate the problem number.

Question	Points	Score
1	10	
2	5	
3	20	
4	15	
Total:	50	

1. (10 points) A group of researchers is trying to determine if there is a relationship between ones education level and whether one has found employment. They survey a group of individuals, asking whether they are employed full–time, part–time, or are unemployed. They also ask the individuals whether they have a high school education, some college education (Associates), a B.A., a Masters, or a Ph.D.. The results are summarized in Table 1 on the next page.

   (a) Complete the missing entires in Table 1.

   (b) Complete the missing entries in Table 2.

   (c) Complete the missing entries in Table 3.

   (d) State H_0, H_a, and the degrees of freedom for this survey.

$$\begin{cases} H_0 : \textit{there is no association between education and employment} \\ \\ H_a : \textit{there is an association between education and employment} \end{cases}$$

   We have degrees of freedom $(\textit{row} - 1)(\textit{column} - 1) = 4 \cdot 2 = 8$.

   (e) Find the p–value and state the conclusion at $\alpha = 0.10$.

   To find χ^2*, we sum the values in Table 3. We find* $\chi^2 = 14.9064$*. With degrees of freedom 8, this gives* $p \approx 0$*. Therefore at the 10% significance level, there is sufficient evidence to reject the null hypothesis: there is sufficient evidence to suggest there is some association between one's education level and employment.*

Table 1: Table of Counts for the Employment Survey.

	High School	Associates	B.A.	Masters	Ph.D.	Total
Full–Time	33	48	59	55	59	254
Part–Time	22	37	36	37	28	160
Unemployed	15	26	12	13	9	75
Total	70	111	107	105	96	489

Table 2: Table of Expected Values for the Employment Survey.

	High School	Associates	B.A.	Masters	Ph.D.
Full–Time	36.36	57.66	55.58	54.54	49.87
Part–Time	22.90	36.32	35.01	34.36	31.41
Unemployed	10.74	17.02	16.41	16.10	14.72

Table 3: Table of Chi–Squared Contributions for the Employment Survey.

	High School	Associates	B.A.	Masters	Ph.D.
Full–Time	0.3105	1.6173	0.2106	0.0039	1.6735
Part–Time	0.0354	0.0128	0.0280	0.2035	0.3704
Unemployed	1.6933	4.7380	1.1856	0.5984	2.2252

2. (5 points) Cornaught University is investigating whether they are admitting under–represented groups 'fairly' or if income may have some influence in admittance. They collect family income data on all African–American students at the University. The breakdown of the students' income levels is given in the table below. Given that 22% of African–Americans make under 15K, 27% make between 15K and 35K, 38% make between 35K and 100K, 11% make between 100K and 200K, and 2% make over 200K, determine whether the students are being admitted 'fairly'. [Use $\alpha = 0.01$.]

Table 4: Breakdown of African–American students on family income.

Income Level	$<$15K	15K–35K	35K–100K	100K–200K	$>$200K
Number of Students	213	312	425	200	100

There were a total of 1,250 students surveyed. Using the percentages given, we find the expected numbers given below (for example, the first expected value is found via $0.22 \cdot 1250 = 275$*):*

Income Level	*$<$15K*	*15K–35K*	*35K–100K*	*100K–200K*	*$>$200K*
Number of Students	*213*	*312*	*425*	*200*	*100*
Expected Number:	*275*	*337.5*	*475*	*137.5*	*25*

We have null and alternative hypotheses given below:

$$\begin{cases} H_0 : \text{there is no association between race, income, and admission} \\ H_A : \text{there is some association between race, income, and admission} \end{cases}$$

Then we have $\chi^2 = 13.9782 + 1.9267 + 5.2632 + 28.4091 + 225 = 274.577$*. Using degrees of freedom* $n - 1 = 5 - 1 = 4$*, we find* $p \approx 0.000$*. Therefore at the 1% significance level, there is sufficient evidence to reject the null hypothesis: there is some association between race, income, and admission.*

3. (20 points) A toy company has hired a group of statisticians to model their costs (in thousands of dollars) based on the number of items they make (in thousands of items). The statisticians use a computer system to create a linear model. The output of the computer program can be found on the next page.

   (a) Complete the missing entries in the computer printout of the model data below.

   Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	29886	29886	93.24	0.000
Items	1	29886	29886	93.24	0.000
Error	23	7372	320.5		
Total	24	37258			

   Model Summary

S	R-sq	R-sq (adj)	R-sq (pred)
17.9025	80.21%	79.35%	75.27%

   Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	113.91	6.95	16.39	0.000	
Items	-9.589	0.993	-9.66	0.000	1.00

   The regression equation is

   production cost = –9.589 Items + 113.91

   (b) What was the total number of data values used to create the model?

   We have dof total $= n - 1 = 24$ *so that* $n = 25$.

   (c) What is the correlation coefficient? What is the coefficient of determination?

   We know that $R^2 = 0.8021$, *which is the coefficient of determination. Then* $R = \pm\sqrt{0.8021} = \pm 0.8956$. *But we know that* $b_1 < 0$ *so that* $R = -0.8956$.

(d) What percentage of the variation in the costs is predicted by the variation in the items for this model?

This is the coefficient of determination, R^2*; therefore, the percentage is 80.21%.*

(e) Construct a 95% confidence interval for β_1.

The variable β_1 *corresponds to items. We are constructing a 95% confidence interval with degrees of freedom* $n-2=23$ *(DFE), then* $t^*=2.069$. *Therefore using the values from the table, we find*

$$\begin{array}{rcl} b_1 & \pm & t^* SE_{b_1} \\ -9.589 & \pm & 2.069(0.993) \\ -9.589 & \pm & 2.05452 \end{array}$$

which gives 95% confidence interval $(-11.6435, -7.53448)$.

(f) Test $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$. Be sure to give the t–value, p–value, and the degrees of freedom. State the conclusion. Is the model linear?

$$\begin{cases} H_0 : \beta_1 = 0 \\ \\ H_a : \beta_1 \neq 0 \end{cases}$$

We have degrees of freedom $n-2=23$ *(DFE), and using the table we find* $t=-9.66$ *and* p*-value* 0.000. *Therefore, we reject the null hypothesis that* $\beta_1 = 0$*, i.e. there is some association between production cost and the number of items produced. This does not imply that the relationship is linear. However, we have* $R^2 = 0.8021$ *so that a good percentage of the variation in the data is explained by a linear model.*

4. (15 points) Concrete is a commonly used material in Civil Engineering. Compressive strength measures the ability of concrete materials to endure various strains. Researchers attempt to try to predict the Compressive strength of various mixtures of concrete using the cement amount, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, and fine aggregate used in the construction as well as the age of the concrete.[1] The model summary is given on the next page.

   (a) Fill in the missing entries in the model on the next page.

   (b) How many concrete mixtures were used to create the model?

   We have degrees of freedom of the total $= n - 1 = 1029$ *so that* $n = 1030$.

   (c) Test $H_0 : \beta_6 = 0$ versus $H_a : \beta_6 < 0$. Be sure to state the degrees of freedom, t–value, and p–value. [Use $\alpha = 0.05$.]

   The variable β_6 *corresponds to the variable CoarseAg.*

$$H_0 : \beta_6 = 0$$
$$H_a : \beta_6 < 0$$

   We have degrees of freedom 1021*, and using the table we find* $t = 1.92$ *and* $p = 0.055/2 = 0.0275$. *Therefore, there is sufficient evidence to reject the null hypothesis; there is some association between CoarseAg and the compression strength of cement.*

   (d) If one re–ran the model using only the variables "FlyAsh", "FineAg", and "Age", would these variables p–values change or remain the same? Explain.

   They would most likely change—whether variables are good predictors or not, their utility as predictors may only be useful in the presence or absence of other variables.

[1] I–Cheng, Yeh, "Modeling of strength of high performance concrete using artificial neural networks.", *Cement and Concrete Research*, Vol. 28, No. 12, pp.1797–1808 (1998).

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	8	176745	22093.1	204.27	0.000
Cement	1	21533	21533.3	199.09	0.000
BF	1	11353	11352.5	104.96	0.000
FlyAsh	1	5281	5281.3	48.83	0.000
Water	1	1513	1513.4	13.99	0.000
SP	1	1046	1046.3	9.67	0.000
CoarseAg	1	398	398.4	3.68	0.000
FineAg	1	384	383.5	3.55	0.000
Age	1	47905	47905.2	442.92	0.000
Error	1021	110428	108.2		
Total	1029	287173			

Model Summary

S	R-sq	R-sq (adj)	R-sq (pred)
10.4019	61.55 %	61.55 %	60.68%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-23.2	26.6	-0.87	0.384	
Cement	0.11979	0.00849	14.11	0.000	7.49
BF	0.1038	0.0101	10.28	0.000	7.28
FlyAsh	0.0879	0.0126	6.99	0.000	6.17
Water	-0.1503	0.0402	-3.74	0.000	7.00
SP	0.2907	0.0935	3.11	0.002	2.97
CoarseAg	0.01803	0.00939	1.92	0.055	5.08
FineAg	0.0201	0.0107	1.88	0.060	7.01
Age	0.11423	0.00543	21.05	0.000	1.12

The regression equation is

Compression Strength = $-23.2 + 0.11979$ Cement $+ 0.1038$ BF $+ 0.0879$ FlyAsh

-0.1503 Water $+ 0.2907$ SP $+ 0.01803$ CoarseAg $+ 0.0201$ FineAg $+ 0.11423$ Age

Bonus (5 points): Below is a partial ANOVA table for a linear regression model.

ANOVA				
	DF	SS	MS	F
Regression	5	833	166.6	45.15
Residual	61	225	3.69	
Total	66	1058		

Find the degrees of freedom for the regression and the residual. You must show all the steps involved in your calculation.

We know that

$$F = \frac{MSM}{MSE} = \frac{SSM/DFM}{SSE/DFE} = \frac{SSM}{DFM} \cdot \frac{DFE}{SSE}$$

From this, we have $\dfrac{DFE}{DFM} = \dfrac{SSE}{SSM}\, F$. *But then we have*

$$DFE = \frac{SSE}{SSM} \cdot F \cdot DFM = \alpha\, DFM$$

where have defined $\alpha := SSE/SSM \cdot F = 225/833 \cdot 45.15 = 12.1954$. *But we know also that* $DFM + DFE = DFT$. *However,*

$$\frac{DFE}{DFM} = \alpha$$
$$DFE = \alpha\, DFM$$

Therefore, using substitution

$$\begin{aligned} DFM + DFE &= DFT \\ DFM + \alpha\, DFM &= 66 \\ DFM\,(1+\alpha) &= 66 \\ DFM &= \frac{66}{1+\alpha} \\ DFM &= \frac{66}{1+12.1954} \\ DFM &= 5.00174 \end{aligned}$$

Then $DFM = 5$, *so that* $DFE = 61$. *Using MS– = SS–/DF–, we easily fill in the remaining two entries.*