

Name Jerry

Student Number _____

STA 442/2101 F 2013 Quiz 11

1. (4 Points) In your analysis of the Chick Feed data, Data, you found a 95% confidence interval for the difference between the expected weight for chicks fed horsebean, and the average of the other expected values.

(a) Give the upper and lower confidence limits. Your answer is a pair of numbers.

$(-156.11, -81.33)$

(b) Suppose you wanted to test whether the expected weight for chicks fed horsebean was different from the average of the other expected weights. State the null hypothesis in terms of μ_j values. Use the alphabetical order of the feeds as they appear on your SAS list file.

$$H_0: \mu_2 = \frac{1}{5} (\mu_1 + \mu_3 + \mu_4 + \mu_5 + \mu_6)$$

(c) From your confidence interval, you can tell that you would reject the null hypothesis at $\alpha = 0.05$ with an ordinary t -test or F -test. How can you tell?

The ^{95%} confidence interval does not include zero.

(d) Would you advise a chicken farmer to purchase the Horsebean feed supplement if she wanted big fat chickens? Why?

No. Chicks fed horsebean are lighter than the average of those fed other supplements.

2. (3 Points) In your analysis of the birdkeeping-lung cancer data, the main question was whether, controlling for the other variables in the model, being a bird keeper was a risk factor for lung cancer.

(a) You did a likelihood ratio test to answer this question. Give the likelihood ratio test statistic and the p -value: two numbers.

$$G^2 = 11.67, p = 0.0006352$$

(b) The default output contains another test of the same null hypothesis. Give the test statistic and the p -value: two numbers.

$$Z = 3.313, p = 0.000923$$

(c) How would you state the conclusion in plain, non-statistical language? Begin with "Allowing for other potential risk factors, ..."

Allowing for other potential risk factors, bird keepers are more likely to get cancer.

3. (3 points) U of T administration is very interested in whether the chances of success are different on the three campuses for undergraduate students with similar performance in High School. So, the Statistical Consulting Service carried out a logistic regression analysis in which

$$\log \frac{\pi}{1-\pi} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3,$$

where π is the probability of graduating within five years of admission, x_1 and x_2 are dummy variables for campus, and x_3 is High School Grade Point Average.

- (a) The table below shows how the dummy variables are defined. Write the odds of graduating within 5 years for each campus.

	x_1	x_2	Odds of Graduating
UTM	1	0	$e^{\beta_0 + \beta_1 + \beta_3 x_3}$
UTSC	0	1	$e^{\beta_0 + \beta_2 + \beta_3 x_3}$
St. George	0	0	$e^{\beta_0 + \beta_3 x_3}$

- (b) Controlling for High School Grade Point Average, the odds of graduating within five years are ___ times as great for students on the UTM campus, compared to students on the UTSC campus. Answer in terms of β quantities.

$$\frac{e^{\beta_0 + \beta_1 + \beta_3 x_3}}{e^{\beta_0 + \beta_2 + \beta_3 x_3}} = e^{\beta_1 - \beta_2}$$

- (c) Suppose you concluded $\beta_2 < 0$. How would you express this in plain, non-statistical language? Use the word “chances” instead of “odds” or “probability,” and begin with “Even allowing for High School marks ...”

Even allowing for High School marks, the chances of graduating within 5 years are less on the UTSC campus than on the St. George campus.

You don't need to attach any printout this time.

1 The SAS System
23:01 Friday, November 9, 2012

NOTE: Copyright (c) 2002-2010 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) Proprietary Software 9.3 (TS1M0)
Licensed to UNIVERSITY OF TORONTO/COMPUTING & COMMUNICATIONS, Site 70072784.
NOTE: This session is executing on the SunOS 5.10 (SUN 64) platform.

You are running SAS 9. Some SAS 8 files will be automatically converted by the V9 engine; others are incompatible. Please see <http://support.sas.com/rnd/migration/planning/platform/64bit.html>

PROC MIGRATE will preserve current SAS file attributes and is recommended for converting all your SAS libraries from any SAS 8 release to SAS 9. For details and examples, please see <http://support.sas.com/rnd/migration/index.html>

This message is contained in the SAS news file, and is presented upon initialization. Edit the file "news" in the "misc/base" directory to display site-specific news and information in the program log. The command line option "-nonews" will prevent this display.

NOTE: SAS initialization used:
real time 0.29 seconds
cpu time 0.12 seconds

```
1 /***** 2101f12HW9chickwts.sas *****/
2 options linesize=79 noovp formdlim='_' nodate;
3 title 'STA2101f12 HW9 Check: Chick Weights';
4
5 data cluck;
6 infile 'chickweights.data' firstobs=2; /* Skip the header */
7 input id weight feed $;
8 label weight = 'Weight in grams at 6 weeks';
9
```

NOTE: The infile 'chickweights.data' is:
Filename=/u/brunner/2101f12/hw/hw9chicks/chickweights.data,
Owner Name=brunner,Group Name=dos,
Access Permission=rw-r--r--,
Last Modified=Fri Nov 9 21:59:51 2012,
File Size (bytes)=1440

NOTE: 71 records were read from the infile 'chickweights.data'.
The minimum record length was 19.
The maximum record length was 19.

NOTE: The data set WORK.CLUCK has 71 observations and 3 variables.

NOTE: DATA statement used (Total process time):
real time 0.02 seconds
cpu time 0.03 seconds

```
10 proc freq;
11 tables feed;
12
```

2 The SAS System

NOTE: There were 71 observations read from the data set WORK.CLUCK.
NOTE: The PROCEDURE FREQ printed page 1.
NOTE: PROCEDURE FREQ used (Total process time):
real time 0.12 seconds
cpu time 0.11 seconds

```
13 proc glm;
14     title2 'One-Factor ANOVA: Just the defaults';
15     class feed;
16     model weight=feed / clparm;
17     means feed;
18     lsmeans feed / pdiff adjust=tukey;
19     contrast 'AllButHorsebean' feed 1 0 -1 0 0 0,
20             feed 0 0 1 -1 0 0,
21             feed 0 0 0 1 -1 0,
22             feed 0 0 0 0 1 -1;
23     contrast 'HorseVsOthers' feed 1 -5 1 1 1 1;
24     estimate 'HorseVsOthers' feed -1 5 -1 -1 -1 -1 /
24 ! divisor=5;
25
26     /* Get Scheffe critical value from proc iml */
27
```

NOTE: The PROCEDURE GLM printed pages 2-6.
NOTE: PROCEDURE GLM used (Total process time):
real time 0.39 seconds
cpu time 0.40 seconds

```
28 proc iml;
NOTE: IML Ready
29     title2 'Scheffe critical value for all possible contrasts';
30     numdf = 5;
30 ! /* Numerator degrees of freedom for initial test
30 ! */
31     dendf = 65;
31 ! /* Denominator degrees of freedom for initial test
31 ! */
32     alpha = 0.05;
33     critval = finv(1-alpha,numdf,dendf);
34     scrit = critval * numdf;
35     print "Initial test has" numdf " and " dendf "degrees of
35 ! freedom."
36     "-----"
36 ! --"
37     "Using significance level alpha = " alpha
38     "-----"
39     "Critical value for the initial test is " critval
40     "-----"
41     "Critical value for Scheffe tests is " scrit
42     "-----";
43
44
45
46
47     /* What's what?
```

Saved: Wednesday, November 14, 2012 6:50:49 PM

3	The SAS System			
48	horsebea	160.200000		2
49	linseed	218.750000		3
50	soybean	246.428571		5
51	meatmeal	276.909091		4
52	casein	323.583333		1
53	sunflowe	328.916667		6
54				
55	horsebea	160.200000	A	2
56	linseed	218.750000	A B C D	3
57	soybean	246.428571	B C D	5
58	meatmeal	276.909091	B C D	E4
59	casein	323.583333	C D	E1
60	sunflowe	328.916667	C	6

NOTE: Exiting IML.

NOTE: The PROCEDURE IML printed page 7.

NOTE: PROCEDURE IML used (Total process time):

real time	0.01 seconds
cpu time	0.02 seconds

NOTE: SAS Institute Inc., SAS Campus Drive, Cary, NC USA 27513-2414

NOTE: The SAS System used:

real time	1.10 seconds
cpu time	0.85 seconds

STA2101f12 HW9 Check: Chick Weights 1

The FREQ Procedure

feed	Frequency	Percent	Cumulative Frequency	Cumulative Percent
casein	12	16.90	12	16.90
horsebea	10	14.08	22	30.99
linseed	12	16.90	34	47.89
meatmeal	11	15.49	45	63.38
soybean	14	19.72	59	83.10
sunflowe	12	16.90	71	100.00

STA2101f12 HW9 Check: Chick Weights 2
 One-Factor ANOVA: Just the defaults

The GLM Procedure

Class Level Information

Class	Levels	Values
feed	6	casein horsebea linseed meatmeal soybean sunflowe

Number of Observations Read 71
 Number of Observations Used 71

STA2101f12 HW9 Check: Chick Weights 3
 One-Factor ANOVA: Just the defaults

The GLM Procedure

Dependent Variable: weight Weight in grams at 6 weeks

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	231129.1621	46225.8324	15.36	<.0001
Error	65	195556.0210	3008.5542		
Corrected Total	70	426685.1831			

R-Square 0.541685
 Coeff Var 20.99052
 Root MSE 54.85029
 weight Mean 261.3099

Source	DF	Type I SS	Mean Square	F Value	Pr > F
feed	5	231129.1621	46225.8324	15.36	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
feed	5	231129.1621	46225.8324	15.36	<.0001

STA2101f12 HW9 Check: Chick Weights 4
 One-Factor ANOVA: Just the defaults

The GLM Procedure

Level of feed	N	-----weight----- Mean	Std Dev
casein	12	323.583333	64.4338397
horsebea	10	160.200000	38.6258405
linseed	12	218.750000	52.2356983
meatmeal	11	276.909091	64.9006233
soybean	14	246.428571	54.1290684
sunflowe	12	328.916667	48.8363842

STA2101f12 HW9 Check: Chick Weights 5
 One-Factor ANOVA: Just the defaults

The GLM Procedure
 Least Squares Means
 Adjustment for Multiple Comparisons: Tukey-Kramer

feed	weight LSMEAN	LSMEAN Number
casein	323.583333	1
horsebea	160.200000	2
linseed	218.750000	3
meatmeal	276.909091	4
soybean	246.428571	5
sunflowe	328.916667	6

Least Squares Means for effect feed
 Pr > |t| for H0: LSmean(i)=LSmean(j)

Dependent Variable: weight

i/j	1	2	3	4	5	6
1		<.0001	0.0002	0.3325	0.0084	0.9999
2	<.0001		0.1413	0.0001	0.0042	<.0001
3	0.0002	0.1413		0.1277	0.7933	<.0001
4	0.3325	0.0001	0.1277		0.7391	0.2207
5	0.0084	0.0042	0.7933	0.7391		0.0039
6	0.9999	<.0001	<.0001	0.2207	0.0039	

STA2101f12 HW9 Check: Chick Weights 6
 One-Factor ANOVA: Just the defaults

The GLM Procedure

Dependent Variable: weight Weight in grams at 6 weeks

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
AllButHorsebean	4	112137.7757	28034.4439	9.32	<.0001
HorseVsOthers	1	120984.0182	120984.0182	40.21	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > t
HorseVsOthers	-118.717532	18.7210316	-6.34	<.0001

Parameter	95% Confidence Limits	
HorseVsOthers	-156.106014	-81.329051

 STA2101f12 HW9 Check: Chick Weights 7
 Scheffe critical value for all possible contrasts

numdf dendf
 Initial test has 5 and 65 degrees of freedom.

 alpha

Using significance level alpha = 0.05

 critval

Critical value for the initial test is 2.3560278

 scrit

Critical value for Scheffe tests is 11.780139

R version 2.15.1 (2012-06-22) -- "Roasted Marshmallows"
Copyright (C) 2012 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: i386-apple-darwin9.8.0/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.52 (6188) i386-apple-darwin9.8.0]

[Workspace restored from /Users/brunner/.RData]
[History restored from /Users/brunner/.Rapp.history]

```
> bird = read.table("http://www.utstat.toronto.edu/~brunner/appliedf13/code_n_data/hw/birdlung.data")
> colnames(bird) = c("cancer", "sex", "highses", "birdkeeper", "age", "yrsmoke", "ncigs")
> head(bird)
  cancer sex highses birdkeeper age yrsmoke ncigs
1     1  1  0         0         1  37      19    12
2     1  1  0         0         1  41      22    15
3     1  1  0         1         0  43      19    15
4     1  1  0         0         1  46      24    15
5     1  1  0         0         1  49      31    20
6     1  1  0         1         0  51      24    15
> fullmod = glm(cancer ~ sex + highses + birdkeeper + age + yrsmoke + ncigs, family=binomial,data=bird)
> summary(fullmod)
```

Call:

```
glm(formula = cancer ~ sex + highses + birdkeeper + age + yrsmoke +
     ncigs, family = binomial, data = bird)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.5642	-0.8333	-0.4605	0.9808	2.2460

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.93736	1.80425	-1.074	0.282924
sex	0.56127	0.53116	1.057	0.290653
highses	0.10545	0.46885	0.225	0.822050
birdkeeper	1.36259	0.41128	3.313	0.000923 ***
age	-0.03976	0.03548	-1.120	0.262503
yrsmoke	0.07287	0.02649	2.751	0.005940 **
ncigs	0.02602	0.02552	1.019	0.308055

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 187.14 on 146 degrees of freedom
Residual deviance: 154.20 on 140 degrees of freedom
AIC: 168.2

Number of Fisher Scoring iterations: 5

```
> redmod = update(fullmod, ~ . - birdkeeper)
```

```
> anova(redmod, fullmod, test="Chisq")
```

Analysis of Deviance Table

Model 1: cancer ~ sex + highses + age + yrsmoke + ncigs

Model 2: cancer ~ sex + highses + birdkeeper + age + yrsmoke + ncigs

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
--	-----------	------------	----	----------	----------

1	141	165.87			
---	-----	--------	--	--	--

2	140	154.20	1	11.67	0.0006352 ***
---	-----	--------	---	-------	---------------

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
>
```

```
>
```