UNIVERSITY OF TORONTO
Faculty of Arts and Science

APRIL 2011 EXAMINATIONS

STA 303 H1S / STA 1002 HS

Duration - 3 hours

Examination Aids: Calculator

LAST NAME:________________________ **FIRST NAME:**________________________

STUDENT NUMBER: ____

- There are 23 pages including this page.
- Pages 14 to 21 contain SAS output.
- The last page (page 23) is a table of formulae that may be useful. For all questions you can assume that the results on the formula page are known.
- Some quantiles from the standard normal distribution and a table of the chi-square distribution can be found on page 22.
- Total marks: 90

1abc	1def	2abcd	2efgh	2ij	3abc	3de

4abc	4d	5	6a(i,ii)	6a(iii)bc

1. A study was carried out on mice to see how their diet affects their lifetime, with particular focus on the effect of restricting caloric intake. Three hundred and forty-nine female mice were randomly assigned to one of the following six diet groups:

   (1) `N/N85` – Mice in this group were fed normally before weaning and then afterwards they were restricted to 85 kilocalories per week.
   (2) `N/R40` – Mice in this group were fed normally before weaning and then afterwards they were restricted to 40 kilocalories per week.
   (3) `N/R50` – Mice in this group were fed normally before weaning and then afterwards they were restricted to 50 kilocalories per week.
   (4) `NP` – Mice in this group ate as much as they pleased of a standard diet.
   (5) `R/R50` – Mice in this group were fed a diet restricted to 50 kilocalories per week both before and after weaning.
   (6) `lopro` – This group had a similar diet to `N/R50` but the protein content was restricted.

   Lifetimes, in months, for the mice were recorded. Some output from SAS is given on pages 14 to 15. The questions below relate to this output.

   (a) (1 mark) Why is the `Model DF` equal to 5?

   (b) (1 mark) The least squares mean for diet `N/N85` has been replaced by `X`'s. What is it?

   (c) (2 marks) Explain, in practical terms, what you can conclude from the 6 t-tests on the top of page 15. (The 6 tests you should be considering are the tests with test statistics: 44.47, -5.57, 4.38, 2.19, -9.40, and 2.54.)

(d) (2 marks) Suppose that we are particularly interested in the comparison in mean lifetime between diets `N/R40` and `N/R50`. Using the first formula on the formula sheet, we could construct a pooled two-sample t-test for this comparison with $\overline{y}_1 = 45.12$, $\overline{y}_2 = 42.30$, $n_1 = 60$, $n_2 = 71$, and $s_p = \sqrt{((60-1)6.70^2 + (71-1)7.77^2)\,/((60-1)+(71-1))}$. Will the resulting p-value for this pooled two-sample t-test be 0.0166? (0.0166 is taken from the matrix of p-values on page 15.) Explain why or why not.

(e) (2 marks) On page 15, there is the following note in the SAS output:

```
NOTE: To ensure overall protection level, only probabilities associated with
      pre-planned comparisons should be used.
```

What is the purpose of this note from SAS? What should you do and why?

(f) (4 marks) On page 15 you are given a plot of the standardized residuals versus the predicted values and a normal quantile plot of the standardized residuals. What are you looking for in each plot? What do you conclude? How do your conclusions affect your answers to the previous questions?

2. For this question, we will consider the data from assignment 2. For children born in 1990 in South Africa, their race (black or white) and whether or not their mother had medical aid was recorded. Attempts were made for follow-up medical evaluations in 1995 and the data includes whether or not the children participated in the follow-up, recorded as yes or no in the variable `Traced`. We are interested in the relationship among race, medical aid status, and whether or not a child had a follow-up. Output from SAS is given on pages 16, 17, and 18 for 3 models fit to these data. The variable `Count` is the number of children in each category.

   (a) (3 marks) In the output for model 1, a few numbers have been replaced by `X`'s. Find the values of the following:

   BIC = __________________________

   Lower limit for the missing Wald 95% Confidence Interval = __________________________

   The missing Wald Chi-Square = __________________________

   (b) (2 marks) Write the **model** that was fit for model 1, defining all terms.

   (c) (2 marks) For model 1, give a practical interpretation of the coefficient whose estimate is 1.7223 (assuming the model is appropriate).

   (d) (3 marks) From model 2, what are the estimated odds of being traced for a child with medical aid (assuming the model is appropriate)?

(e) (3 marks) From model 3, what is the odds ratio of being traced, comparing a black child to a white child (assuming the model is appropriate)?

(f) (3 marks) For model 1, the deviance is large. Ignoring what you learn from the other models, give at least three reasons why the deviance might be large when fitting a model of this type to data.

(g) (1 mark) For model 3, under the `Criteria for Assessing Goodness of Fit`, why is `DF` equal to 2?

(h) (4 marks) Is it possible to carry out a Likelihood Ratio Test comparing the fits of models 1 and 3? If not, explain why not. If yes, carry it out, giving each of the following: (I) the test statistic, (II) the distribution of the test statistic under the null hypothesis, (III) the p-value, (IV) the conclusion.

(i) (2 marks) Wald tests for the model parameters for each of these models use chi-square distributions to calculate the p-values. Explain why chi-square is the appropriate distribution.

(j) Choosing from the 3 models for which you are given SAS output, pick the model that you think is most appropriate for these data.

i. (2 marks) Which of the 3 models did you choose? Why?

ii. (2 marks) For the model that you chose in part i., characterize in practical terms what you conclude about the relationship among race, medical aid status, and whether or not a child had a follow-up.

iii. (2 marks) When you analysed these data in assignment 2, one of the analyses treated `Traced` as a response variable and fit a logistic regression model with `Race` and `MedicalAid` as explanatory variables. Explain how the model you chose in part i. can tell you which variables were statistically significant predictors in the logistic regression.

3. In this question, we will consider the data from assignment 1. The data were weights collected on 72 girls suffering from anorexia. The girls were randomly assigned to receive one of three therapies: cognitive behavioural (coded `b`), family (coded `f`), or the control therapy (coded `c`). The girls' weights were measured at the beginning of the study and after following the therapy for a period of time. Therapies are considered successful if girls gain weight on the therapy.

   For this question, our interest is whether a girl gained or lost weight (and not how much). A new variable `gained` is defined as 1 if a girl gained weight and 0 otherwise.

   Some edited output from SAS for an analysis of these data is on page 19. Some numbers have been replaced by `X`'s.

   (a) (2 marks) From what you are given, do you have any concerns about the appropriateness of the inferences from the logistic regression model that was fit? What else would you like to see?

   (b) (1 mark) What is the estimated probability that a girl on therapy `c` gains weight?

   (c) (4 marks) Carry out an hypothesis test with null hypothesis that the log-odds of gaining weight are the same for all three therapies; include: (I) the test statistic, (II) the distribution of the test statistic under the null hypothesis, (III) the p-value, (IV) the conclusion.

(d) In the SAS output, you are given odds ratio estimates for therapy `b` versus therapy `f` and for therapy `c` versus therapy `f`.

i. (1 mark) What is the odds ratio estimate for therapy `b` versus therapy `c`?

ii. (2 marks) Calculate the missing 95% Wald Confidence Interval for the odds ratio of therapy `b` versus therapy `f`.

iii. (2 marks) Explain how the confidence interval in part ii. is consistent with one of the p-values in the output.

(e) (2 marks) The table below gives the counts of the numbers of girls who did or did not gain weight for each therapy.

	Therapy		
	b	c	f
Gained weight	18	11	13
Did not gain weight	11	15	4

An alternative analysis for these data would test whether the row and column variables in this table are independent. Do you prefer this proposed analysis or the analysis that you are given in the SAS output for this question? Why?

4. In this question, we will again consider the data from assignment 1. The data were weights collected on 72 girls suffering from anorexia. The girls were randomly assigned to receive one of three therapies: cognitive behavioural (coded `b`), family (coded `f`), or the control therapy (coded `c`). The girls' weights were measured at the beginning of the study and after following the therapy for a period of time. Therapies are considered successful if girls gain weight on the therapy.

   For this question, we will use the weight of the girls as the response variable, with two measurements on each girl. The variable `when` is equal to `baseline` if the weight was measured at the beginning of the study and is equal to `end` if the weight was measured after the therapy period.

   Some edited output from SAS for an analysis of these data is given on pages 20 and 21. The fitted model assumes variances and covariances are the same for all subjects and includes a random effect for subject.

   (a) (3 marks) From the output that you are given, what can you conclude about the relative effectiveness of the therapies? Support your answer with appropriate numbers from the SAS output.

   (b) (3 marks) Write the model being fit; define all terms. State clearly which parts of the model are random and which are not random.

   (c) (3 marks) What is the estimated variance-covariance matrix of the 144 observed weights?

(d) The table of the standard deviations suggests that it may be worth considering a model that has different variances for baseline and end measurements, and that estimates a different variance-covariance matrix for each therapy group.

i. (1 mark) How many variance-covariance parameters would need to be estimated to accommodate this structure?

ii. (2 marks) How could you compare whether this proposed model fits the data better than the model fit in the SAS output?

5. Suppose people are categorized by three variables. Variable 1 has I categories, variable 2 has J categories, and variable 3 has K categories. Thus there are $I \times J \times K$ categories in total. We observe y_{ijk}, the count of the number of people for whom variable 1 is i, variable 2 is j, and variable 3 is k. We will assume that the y_{ijk} can be considered observations from Poisson distributions with means μ_{ijk} and use Poisson regression. We will fit a model that assumes that variables 1, 2, and 3 are independent.

   (a) (4 marks) Show that the deviance is

   $$2 \sum_{k=1}^{K} \sum_{j=1}^{J} \sum_{i=1}^{I} y_{ijk} \log \left(\frac{y_{ijk}}{\hat{\mu}_{ijk}} \right)$$

   where $\hat{\mu}_{ijk}$ are the estimated values of μ_{ijk} from the fitted model.

   (b) (2 marks) What are the estimated values of μ_{ijk} from the fitted model? Give how they can be calculated from the observed counts; you do not need to derive them.

6. In this course, we have studied the following (generalized) linear models:
(1) one-way analysis of variance, (2) two-way analysis of variance, (3) binary logistic regression, (4) binomial logistic regression, (5) Poisson regression, and (6) mixed models.

   (a) (12 marks (4 each)) Three scenarios (below and on the next page) relate to a study of 73 breakfast cereals sold at a large grocery store. In marketing a cereal, a consideration is whether or not it is displayed at eye level on the grocery store shelf. For each of the cereals in the study, it was recorded whether the cereal was on the lower, middle, or upper shelf. For each scenario indicate:
   (I) which of the 6 types of generalized linear model is appropriate
   (II) the model you would use for the analysis, defining all terms
   (III) the null and alternative hypotheses for the test that addresses the question of interest.

      i. The cereals were examined for their content of various vitamins and minerals. The researcher believes that stores may tend to put healthier cereals on the upper shelf since they are more likely to appeal to adults. We are interested in whether the content (in grams per serving) of three specific vitamins in the cereals are useful in predicting whether a cereal is displayed on the upper shelf.

      ii. We are interested in learning whether there are differences in the average sugar content (in grams per serving) of the cereals depending on their placement on the lower, middle, or upper shelves.

iii. Many of the cereals come with an incentive to buy them such as a free toy in the box or a chance to win a prize. We count the number of cereals with and without an incentive on each of the lower, middle, and upper shelves. We are interested in learning if shelf placement and whether or not a cereal has an incentive are related.

(b) (3 marks) Of the 6 models we have studied (as identified at the beginning of this question), which have random error terms in their models? Why do some models need the random error term and some models do not?

(c) (2 marks) In order to carry out inference about the coefficients of the explanatory variables, which of the 6 models we have studied require a large sample size? Why is a large sample size necessary for these models?

The SAS output on pages 14 to 15 is relevant to question 1.

SAS output for QUESTION 1

diet	lifetime Mean	lifetime Std Dev	lifetime N
N/N85	XXXXX	5.13	57.00
N/R40	45.12	6.70	60.00
N/R50	42.30	7.77	71.00
NP	27.40	6.13	49.00
R/R50	42.89	6.68	56.00
lopro	39.69	6.99	56.00

```
                   The GLM Procedure
                Class Level Information
 Class         Levels    Values
 diet               6    N/N85 N/R40 N/R50 NP R/R50 lopro
          Number of Observations Read         349
          Number of Observations Used         349

Dependent Variable: lifetime

                                     Sum of
Source                      DF      Squares    Mean Square   F Value   Pr > F
Model                        5  12733.94181     2546.78836     57.10   <.0001
Error                      343  15297.41532       44.59888
Corrected Total            348  28031.35713

             R-Square     Coeff Var      Root MSE    lifetime Mean
             0.454275      17.21323      6.678239         38.79713

Source                      DF  Type III SS    Mean Square   F Value   Pr > F
diet                         5  12733.94181     2546.78836     57.10   <.0001
```

(SAS output for question 1 continues on the next page.)

(SAS output for question 1 continued)

Parameter		Estimate		Standard Error	t Value	Pr > |t|
Intercept		39.68571429	B	0.89241725	44.47	<.0001
diet	N/N85	-6.99448622	B	1.25652099	-5.57	<.0001
diet	N/R40	5.43095238	B	1.24085583	4.38	<.0001
diet	N/R50	2.61146881	B	1.19355007	2.19	0.0293
diet	NP	-12.28367347	B	1.30636509	-9.40	<.0001
diet	R/R50	3.20000000	B	1.26206858	2.54	0.0117
diet	lopro	0.00000000	B	.	.	.

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

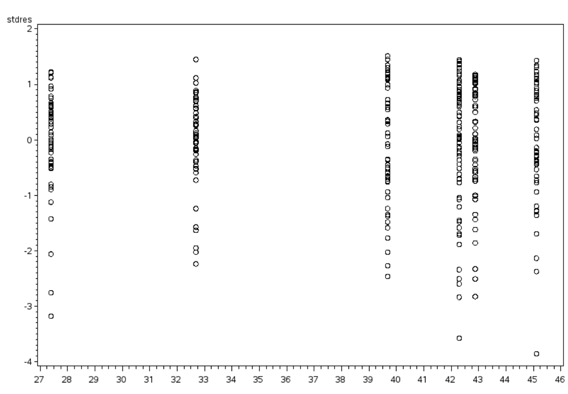
Least Squares Means

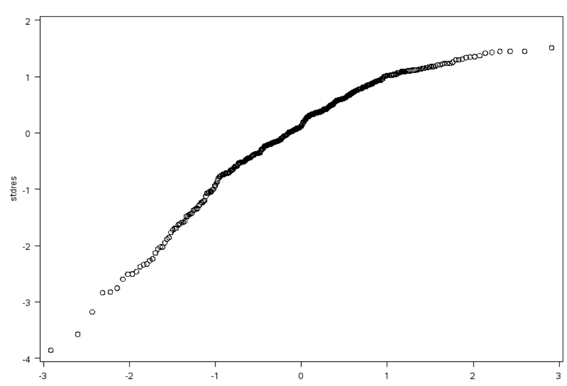
diet	lifetime LSMEAN	LSMEAN Number
N/N85	XXXXXXXXXX	1
N/R40	45.1166667	2
N/R50	42.2971831	3
NP	27.4020408	4
R/R50	42.8857143	5
lopro	39.6857143	6

Least Squares Means for effect diet
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: lifetime

i/j	1	2	3	4	5	6
1		<.0001	<.0001	<.0001	<.0001	<.0001
2	<.0001		0.0166	<.0001	0.0731	<.0001
3	<.0001	0.0166		<.0001	0.6223	0.0293
4	<.0001	<.0001	<.0001		<.0001	<.0001
5	<.0001	0.0731	0.6223	<.0001		0.0117
6	<.0001	<.0001	0.0293	<.0001	0.0117	

NOTE: To ensure overall protection level, only probabilities associated with pre-planned comparisons should be used.





Continued

The SAS output on pages 16 to 18 is relevant to question 2.

SAS output for QUESTION 2
MODEL 1

The GENMOD Procedure

Model Information

Data Set	WORK.FORTABLES
Distribution	Poisson
Link Function	Log
Dependent Variable	Count

Number of Observations Read	8
Number of Observations Used	8

Class Level Information

Class	Levels	Values
Traced	2	No Yes
MedicalAid	2	No Yes
Race	2	Black White

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	4	393.0504	98.2626
Scaled Deviance	4	393.0504	98.2626
Pearson Chi-Square	4	634.8177	158.7044
Scaled Pearson X2	4	634.8177	158.7044
Log Likelihood		8071.2358	
Full Log Likelihood		-219.7333	
AIC (smaller is better)		447.4665	
AICC (smaller is better)		460.7998	
BIC (smaller is better)		XXXXXXXX	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	1.6998	0.1120	XXXXXX	XXXXXX	XXXXXX	<.0001
Traced	No	1	1.0375	0.0571	0.9257	1.1493	330.62	<.0001
Traced	Yes	0	0.0000	0.0000	0.0000	0.0000	.	.
MedicalAid	No	1	1.7223	0.0699	1.5853	1.8594	606.54	<.0001
MedicalAid	Yes	0	0.0000	0.0000	0.0000	0.0000	.	.
Race	Black	1	2.3534	0.0891	2.1789	2.5280	698.00	<.0001
Race	White	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

(SAS output for question 2 continues on the next page.)

SAS output for QUESTION 2
MODEL 2
(The first part of the output that is the same as for MODEL 1 has been omitted.)

The GENMOD Procedure

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1	0.0011	0.0011
Scaled Deviance	1	0.0011	0.0011
Pearson Chi-Square	1	0.0011	0.0011
Scaled Pearson X2	1	0.0011	0.0011
Log Likelihood		8267.7604	
Full Log Likelihood		-23.2086	
AIC (smaller is better)		60.4172	
AICC (smaller is better)		.	
BIC (smaller is better)		60.9733	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept			1	2.2986	0.2928	1.7246	2.8725	61.61	<.0001
Traced	No		1	2.3462	0.3040	1.7503	2.9421	59.56	<.0001
Traced	Yes		0	0.0000	0.0000	0.0000	0.0000	.	.
MedicalAid	No		1	-1.5856	0.2895	-2.1530	-1.0183	30.00	<.0001
MedicalAid	Yes		0	0.0000	0.0000	0.0000	0.0000	.	.
Race	Black		1	1.2861	0.3229	0.6531	1.9190	15.86	<.0001
Race	White		0	0.0000	0.0000	0.0000	0.0000	.	.
Traced*MedicalAid	No	No	1	0.0301	0.1997	-0.3613	0.4215	0.02	0.8803
Traced*MedicalAid	No	Yes	0	0.0000	0.0000	0.0000	0.0000	.	.
Traced*MedicalAid	Yes	No	0	0.0000	0.0000	0.0000	0.0000	.	.
Traced*MedicalAid	Yes	Yes	0	0.0000	0.0000	0.0000	0.0000	.	.
Traced*Race	No	Black	1	-1.4204	0.3415	-2.0897	-0.7511	17.30	<.0001
Traced*Race	No	White	0	0.0000	0.0000	0.0000	0.0000	.	.
Traced*Race	Yes	Black	0	0.0000	0.0000	0.0000	0.0000	.	.
Traced*Race	Yes	White	0	0.0000	0.0000	0.0000	0.0000	.	.
MedicalAid*Race	No	Black	1	3.9090	0.2462	3.4264	4.3916	252.03	<.0001
MedicalAid*Race	No	White	0	0.0000	0.0000	0.0000	0.0000	.	.
MedicalAid*Race	Yes	Black	0	0.0000	0.0000	0.0000	0.0000	.	.
MedicalAid*Race	Yes	White	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale			0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

(SAS output for question 2 continues on the next page.)

SAS output for QUESTION 2
MODEL 3
(The first part of the output that is the same as for MODEL 1 has been omitted.)

The GENMOD Procedure

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	2	0.0237	0.0119
Scaled Deviance	2	0.0237	0.0119
Pearson Chi-Square	2	0.0237	0.0119
Scaled Pearson X2	2	0.0237	0.0119
Log Likelihood		8267.7491	
Full Log Likelihood		-23.2199	
AIC (smaller is better)		58.4398	
AICC (smaller is better)		142.4398	
BIC (smaller is better)		58.9164	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept			1	2.2939	0.2913	1.7229	2.8648	62.01	<.0001
Traced	No		1	2.3514	0.3021	1.7593	2.9435	60.58	<.0001
Traced	Yes		0	0.0000	0.0000	0.0000	0.0000	.	.
MedicalAid	No		1	-1.5581	0.2246	-1.9983	-1.1180	48.13	<.0001
MedicalAid	Yes		0	0.0000	0.0000	0.0000	0.0000	.	.
Race	Black		1	1.2711	0.3074	0.6685	1.8736	17.09	<.0001
Race	White		0	0.0000	0.0000	0.0000	0.0000	.	.
Traced*Race	No	Black	1	-1.3982	0.3077	-2.0013	-0.7950	20.64	<.0001
Traced*Race	No	White	0	0.0000	0.0000	0.0000	0.0000	.	.
Traced*Race	Yes	Black	0	0.0000	0.0000	0.0000	0.0000	.	.
Traced*Race	Yes	White	0	0.0000	0.0000	0.0000	0.0000	.	.
MedicalAid*Race	No	Black	1	3.9031	0.2430	3.4268	4.3795	257.92	<.0001
MedicalAid*Race	No	White	0	0.0000	0.0000	0.0000	0.0000	.	.
MedicalAid*Race	Yes	Black	0	0.0000	0.0000	0.0000	0.0000	.	.
MedicalAid*Race	Yes	White	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale			0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Continued

The SAS output on page 19 is relevant to question 3.

SAS output for QUESTION 3

```
              The LOGISTIC Procedure

                 Model Information
        Data Set                     WORK.MULTI
        Response Variable            gained
        Number of Response Levels    2
        Model                        binary logit
        Optimization Technique       Fisher's scoring
            Number of Observations Read          72
            Number of Observations Used          72

                   Response Profile
           Ordered                     Total
             Value       gained     Frequency
                 1            1            42
                 2            0            30
           Probability modeled is gained=1.

               Class Level Information
                                     Design
            Class       Value     Variables
            therapy     b          1      0
                        c          0      1
                        f          0      0

                 Model Fit Statistics
                                      Intercept
                       Intercept            and
            Criterion         Only     Covariates
            AIC             99.804         98.472
            SC             102.080        105.302
            -2 Log L        97.804         92.472

           Analysis of Maximum Likelihood Estimates
                                     Standard          Wald
Parameter      DF    Estimate        Error    Chi-Square    Pr > ChiSq
Intercept       1      1.1787       0.5718        4.2494        0.0393
therapy   b     1     -0.6862       0.6880        0.9946        0.3186
therapy   c     1     -1.4888       0.6961        4.5749        0.0324

                Odds Ratio Estimates
                          Point          95% Wald
         Effect          Estimate      Confidence Limits
         therapy b vs f       0.503       XXXXX       XXXXX
         therapy c vs f       0.226       0.058       0.883
```

The SAS output on pages 20 to 21 is relevant to question 4.

SAS output for QUESTION 4

```
--------------------------------------------
|                |         weight          |
|                |-------------------------|
|                |  Std  Dev  |     N      |
|----------------+------------+------------|
|therapy |when   |            |            |
|--------+-------|            |            |
|b       |baseline|       4.85|       29.00|
|        |--------+------------+------------|
|        |end     |       8.35|       29.00|
|--------+--------+------------+------------|
|c       |baseline|       5.71|       26.00|
|        |--------+------------+------------|
|        |end     |       4.74|       26.00|
|--------+--------+------------+------------|
|f       |baseline|       5.02|       17.00|
|        |--------+------------+------------|
|        |end     |       8.48|       17.00|
--------------------------------------------

                    The Mixed Procedure

          Model Information

Data Set                     WORK.UNIV
Dependent Variable           weight
Covariance Structure         Variance Components
Estimation Method            REML
Residual Variance Method     Profile
Fixed Effects SE Method      Model-Based
Degrees of Freedom Method    Containment

          Class Level Information

  Class      Levels    Values
  when            2    baseline end
  subject        72    1 2 3 4 5 6 7 8 9 10 11 12 13
                       14 15 16 17 18 19 20 21 22 23
                       24 25 26 27 28 29 30 31 32 33
                       34 35 36 37 38 39 40 41 42 43
                       44 45 46 47 48 49 50 51 52 53
                       54 55 56 57 58 59 60 61 62 63
                       64 65 66 67 68 69 70 71 72
  therapy         3    b c f
```

(SAS output for question 4 continues on the next page.)

Dimensions

Covariance Parameters	2
Columns in X	12
Columns in Z	72
Subjects	1
Max Obs Per Subject	144

Number of Observations

Number of Observations Read	144
Number of Observations Used	144
Number of Observations Not Used	0

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	920.09384022	
1	1	913.85550195	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Estimate
subject	11.8019
Residual	28.3387

Fit Statistics

-2 Res Log Likelihood	913.9
AIC (smaller is better)	917.9
AICC (smaller is better)	917.9
BIC (smaller is better)	922.4

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
when	1	69	12.92	0.0006
therapy	2	69	6.20	0.0033
when*therapy	2	69	5.42	0.0065

Solution for Fixed Effects

Effect	when	therapy	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			90.4941	1.5366	69	58.89	<.0001
when	baseline		-7.2647	1.8259	69	-3.98	0.0002
when	end		0	.	.	.	.
therapy		b	-4.7976	1.9353	69	-2.48	0.0156
therapy		c	-9.3864	1.9761	69	-4.75	<.0001
therapy		f	0	.	.	.	.
when*therapy	baseline	b	4.2578	2.2996	69	1.85	0.0684
when*therapy	baseline	c	7.7147	2.3482	69	3.29	0.0016
when*therapy	baseline	f	0	.	.	.	.
when*therapy	end	b	0	.	.	.	.
when*therapy	end	c	0	.	.	.	.
when*therapy	end	f	0	.	.	.	.

Continued

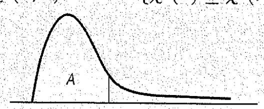
Percentiles of the standard normal distribution

Probability to left of quantile	0.95	0.975	0.99	0.995
Quantile	1.645	1.960	2.326	2.576

Percentiles of the chi-square distribution

TABLE B.3 **Percentiles of the χ^2 Distribution.**

Entry is $\chi^2(A;\nu)$ where $P\{\chi^2(\nu) \le \chi^2(A;\nu)\} = A$



ν	A: .005	.010	.025	.050	.100	.900	.950	.975	.990	.995
1	0.0^4393	0.0^3157	0.0^3982	0.0^2393	0.0158	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	4.61	5.99	7.38	9.21	10.60
3	0.072	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	7.78	9.49	11.14	13.28	14.86
5	0.412	0.554	0.831	1.145	1.61	9.24	11.07	12.83	15.09	16.75
6	0.676	0.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	0.989	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.4	104.2
80	51.17	53.54	57.15	60.39	64.28	96.58	101.9	106.6	112.3	116.3
90	59.20	61.75	65.65	69.13	73.29	107.6	113.1	118.1	124.1	128.3
100	67.33	70.06	74.22	77.93	82.36	118.5	124.3	129.6	135.8	140.2

Continued

Some formulae:

Pooled t-test

$$t_{obs} = \frac{\overline{y}_1 - \overline{y}_2}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

Test for two proportions

$$z_{obs} = (\hat{\pi}_1 - \hat{\pi}_2)\Big/\sqrt{\hat{\pi}_p(1-\hat{\pi}_p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

Linear Regression

$$b_1 = \frac{\sum(x_i-\overline{x})(y_i-\overline{y})}{\sum(x_i-\overline{x})^2} = \frac{\sum x_i y_i - n\overline{x}\overline{y}}{\sum x_i^2 - n\overline{x}^2} \qquad b_0 = \overline{y} - b_1\overline{x}$$

One-way analysis of variance

$$\text{SSTO} = \sum_{i=1}^{N}(y_i-\overline{y})^2 \qquad \text{SSE} = \sum_{g=1}^{G}\sum_{(g)}(y_i-\overline{y}_g)^2 \qquad \text{SSR} = \sum_{g=1}^{G} n_g(\overline{y}_g-\overline{y})^2$$

Bernoulli and Binomial distributions

If $Y \sim \text{Bernoulli}(\pi)$
$\text{E}(Y) = \pi$, $\text{Var}(Y) = \pi(1-\pi)$

If $Y \sim \text{Binomial}(m, \pi)$
$\text{E}(Y) = m\pi$, $\text{Var}(Y) = m\pi(1-\pi)$

Logistic Regression with Binomial Response formulae

$$\text{Deviance} = 2\sum_{i=1}^{n}\{y_i\log(y_i) + (m_i-y_i)\log(m_i-y_1) - y_i\log(\hat{y}_i) + (m_i-y_i)\log(m_i-\hat{y}_1)\}$$

$$D_{res,i} = \text{sign}(y_i - m_i\hat{\pi}_i)\sqrt{2\left\{y_i\log\left(\frac{y_i}{m_i\hat{\pi}_i}\right) + (m_i-y_i)\log\left(\frac{m_i-y_i}{m_i-m_i\hat{\pi}_i}\right)\right\}}$$

$$P_{res,i} = \frac{y_i - m_i\hat{\pi}_i}{\sqrt{m_i\hat{\pi}_i(1-\hat{\pi}_i)}}$$

Multinomial distribution for 2×2 table

$$\Pr(\mathbf{Y}=\mathbf{y}) = \frac{n!}{y_{11}!y_{12}!y_{21}!y_{22}!}\pi_{11}^{y_{11}}\pi_{12}^{y_{12}}\pi_{21}^{y_{21}}\pi_{22}^{y_{22}}$$

Poisson distribution

$$\Pr(Y=y) = \frac{\mu^y e^{-\mu}}{y!},\ y = 0, 1, 2, \ldots$$
$$\text{E}(Y) = \mu,\ \text{Var}(Y) = \mu$$

Two-way contingency tables (easily generalizable to three-way tables)

$$X^2 = \sum_{j=1}^{J}\sum_{i=1}^{I}\frac{(y_{ij}-\hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} \qquad G^2 = 2\sum_{j=1}^{J}\sum_{i=1}^{I} y_{ij}\log\left(\frac{y_{ij}}{\hat{\mu}_{ij}}\right)$$

$$D_{res,ij} = \text{sign}(y_{ij}-\hat{\mu}_{ij})\sqrt{2\left\{y_{ij}\log\left(\frac{y_{ij}}{\hat{\mu}_{ij}}\right) - y_{ij} + \hat{\mu}_{ij}\right\}}$$

$$P_{res,ij} = \frac{y_{ij}-\hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}}}$$

Model Fitting Criteria

$$\text{AIC} = -2\log(L) + 2(p+1) \qquad \text{SC} = \text{BIC} = -2\log(L) + (p+1)\log(N)$$

Total pages: 23
Total marks: 90