Université d'Ottawa · University of Ottawa
Faculté des sciences
Mathématiques et de statistique
Faculty of Science
Mathematics and Statistics

MAT3375 Regression Analysis
Sample Final with Solutions

Fall 2012 **Professor P.-J. Bergeron**
Duration: 3 hours

Name: _____

Student Number: _____

Read the whole thing attentively, not tentatively.
There are 6 questions.
Were this the actual exam, these answers would get the perfect grade of 90 points out of 90.

Professor's use only

Question	Grade
1	15
2	15
3	15
4	15
5	15
6	15
Total	90

585, av. King-Edward
Ottawa (Ontario) K1N 6N5 Canada
585 King Edward Avenue
Ottawa, Ontario K1N 6N5 Canada
(613) 562-5864 • Téléc./Fax (613) 562-5776
Courriel/Email: uomaths@science.uottawa.ca

[15] 1. Data was gathered from the 2012 NSERC Discovery Grant competition results. We want to know if different departments have different grant sizes. We have five different departments: Computer Science (CS), Mathematics (Math), Physics, Chemistry and Geology.

(a) Explain why the above are categorical covariates and how the design matrix for multiple regression is structured.

Departments are categorical covariates because they are not quantitative measures but discrete categories, which can be included as covariates using indicator functions, i.e.
$X_{ij} = \mathbf{1}$[individual i is in j-th department]. Let n_j be the number of individuals in the j-th department. The design matrix can be constructed as

$$\mathbf{X} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \dots & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \dots & \mathbf{0}_{n_2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{1}_{n_{k-1}} & \mathbf{0}_{n_{k-1}} & \mathbf{0}_{n_{k-1}} & \dots & \mathbf{1}_{n_{k-1}} \\ \mathbf{1}_{n_k} & \mathbf{0}_{n_k} & \mathbf{0}_{n_k} & \dots & \mathbf{0}_{n_k} \end{pmatrix}$$

Solution:

(b) Consider the SAS output below. Which department constitutes the baseline (intercept)? Are any department significantly different from baseline? If so, which? Explain.

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	34929	6393.06342	5.46	<.0001	22131	47726
math	1	-11095	9410.33455	-1.18	0.2432	-29932	7741.59311
CS	1	-10512	9410.33455	-1.12	0.2686	-29349	8324.92644
Physics	1	7178.09524	9410.33455	0.76	0.4487	-11659	26015
Chem	1	34379	9213.38498	3.73	0.0004	15937	52822

Solution:
Geology is the baseline as all the other departments have their own β. Chemistry is the only one significantly different from geology (p-value for $\beta < 0.05$, interval estimate excludes 0). *Solution:*

(c) In particular, we want to compare the "formal sciences" (math and computer science) to the "experimental science" (physics and chemistry). Give the formula for the General Linear Hypothesis Test, including the degrees of freedom and distribution of the test statistic.

Solution: The test statistic is

$$F = \frac{(\mathbf{A}\hat{\beta} - \mathbf{c})^{\mathbf{t}}(\mathbf{A}(\mathbf{X^t X})^{-1}\mathbf{A^t})^{-1}(\mathbf{A}\hat{\beta} - \mathbf{c})}{qs^2}$$

where $\mathbf{A}$ is the constraint matrix (let's just call it that) of dimension $q \times p$, and q is the number of constraints. $s^2 = MSE$ of the unconstrained model. F has an F distribution with q numerator and $n - p$ denominator degrees of freedom under the null hypothesis.

(d) Write down the null and alternative hypotheses from part (c) in terms of the regression coefficient, and in matrix form.

Solution: There's more than one way to write these. Only one set of hypotheses is necessary. First set:
$H_0 : \beta_{Math} = \beta_{CS} = \beta_{Phys} = \beta_{Chem}$
H_1 : at least one of the β's is different from the others.

Using $\mathbf{A}$, we could take all the paired differences to be 0, i.e.

$$H_0 : \mathbf{A}\boldsymbol{\beta} = \mathbf{0}$$

$$H_1 : \mathbf{A}\boldsymbol{\beta} \neq \mathbf{0}$$

where

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \end{pmatrix}$$

Second set: taking the average of math and CS vs average of chemistry and physics
$H_0 : \frac{\beta_{Math}+\beta_{CS}}{2} = \frac{\beta_{Phys}+\beta_{Chem}}{2}$
$H_1 : \frac{\beta_{Math}+\beta_{CS}}{2} \neq \frac{\beta_{Phys}+\beta_{Chem}}{2}$
The matrix version is as above, with $\mathbf{A} = \begin{pmatrix} 0 & 0.5 & 0.5 & -0.5 & -0.5 \end{pmatrix}$. Note that this test is more in line with the way question (c) was formulated, lumping the formal science together and the experimental science together.

(e) There's at least two ways to setup the matrix above. The General Linear Hypothesis test was performed using 2 different A matrices. The results (F statistics and associated p-value) are below. Do these result concur with respect to the hypotheses (i.e. are the decisions the same)? Are the NSERC grants significantly different between the formal and experimental sciences according to the numbers below?

General linear hypothesis for NSERC

F1	pf1	F2	pf2
21.328119	0.000022	10.052973	0.0000198

Solution: The results concur since both have tiny p-values and thus we reject H_0. Thus, we can say that NSERC Discovery grants are significantly different betweeen the formal and experimental sciences. (*N.B.: the following comment is not in any way shape or form part of the solutions* If your prof had used all the public data instead of a systematic sampling of the first 12 or so recipients in each department, and splitting between statisticians and mathematicians, we'd see that statistics profs are getting screwed).

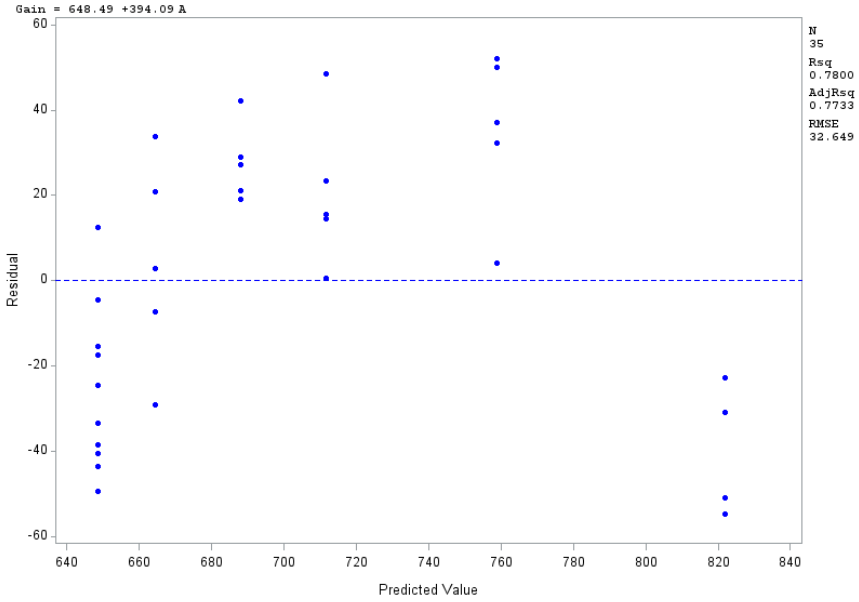
[15] 2. We have data on turkeys grown with identical diets except with an amount A of amino acid methionine as percentage of the diet. We want to model the response weight gain in grams as a function of A.

(a) Show that in simple linear regression $\sum_{i=1}^{n} e_i = 0$.

Solution:

$$\begin{aligned}\sum_{i=1}^{n} e_i &= \sum_{i=1}^{n}(y_i - \hat{y}_i) = \sum_{i=1}^{n}(y_i - b_0 - b_1 x_i) \\ &= \sum_{i=1}^{n} y_i - nb_0 - b_1 \sum_{i=1}^{n} x_i = n\bar{y} - n(\bar{y} - b_1\bar{x}) - b_1 n\bar{x} \\ &= n\bar{y} - n\bar{y} + nb_1\bar{x} - nb_1\bar{x} = 0\end{aligned}$$

(b) A simple regression was done. Consider the following residual plot below. Explain why adding a quadratic term to the model is warranted.



Solution: Under the least-square regression model, the errors (residuals) should be uncorrelated and have the same variance. The clear curved pattern in the residual plot indicates that there is a (nonlinear) correlation between residuals, and suggests in particular, a quadratic relationship between y and x.

(c) Explain the difference between transforming the covariate in simple linear regression and adding a polynomial term in multiple linear regression.

Solution: Transforming the covariate in SLR keeps the *simple* linear regression paradigm, i.e. there is a single, linear effect of the transformed covariate on the response. The number of parameters to estimate and the degrees of freedom are the same, only the "shape" of the covariate space, not its dimensions, is changed. By adding a quadratic term, one adds a parameter to estimate and another dimension to the covariate space. It is a more complex model (the covariate will have both a linear and a quadratic effect) and the number of degrees of freedom is reduced compared to SLR.

(d) Three regressions, with linear, quadratic and cubic terms respectively, were performed. According to the following output, which is the best? Explain.

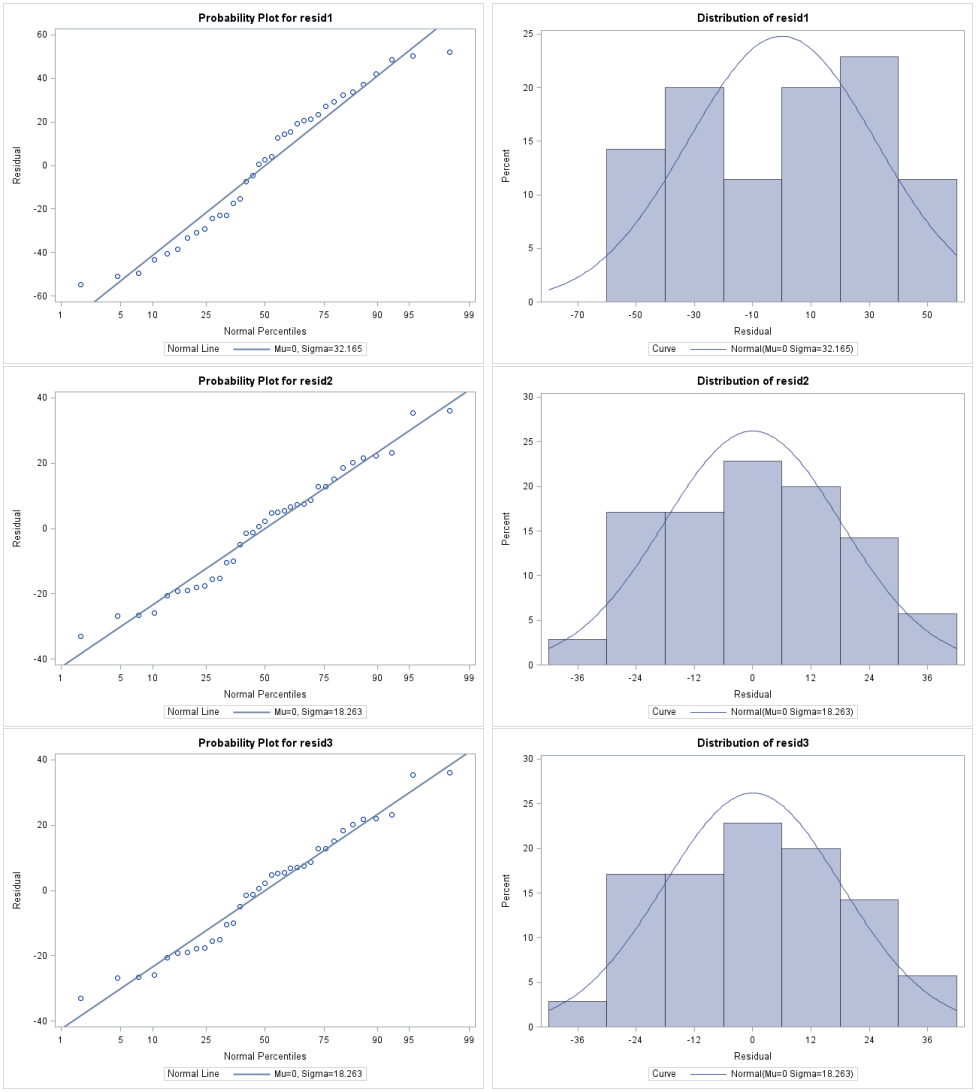
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	648.48989	7.65806	84.68	<.0001
A	1	394.08897	36.43744	10.82	<.0001

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	625.54161	5.22742	119.67	<.0001
A	1	964.47096	72.65057	13.28	<.0001
A2	1	-1362.06864	166.07665	-8.20	<.0001

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	625.57907	5.74006	108.98	<.0001
A	1	961.86794	168.33857	5.71	<.0001
A2	1	-1344.00405	1063.42186	-1.26	0.2157
A3	1	-28.43137	1652.48753	-0.02	0.9864

Solution: Using nested models, the quadratic model is an improvement over the linear one as the estimate for β_2, the quadratic coefficient, is highly significant (small p-value). However, adding the cubic term yields no significance for β_3, the cubic term, accounting for the other two. Thus the quadratic model is the best.

(e) Explain what a Q-Q plot is and what it assesses. Look at the follwing histograms and Q-Q plots of the residuals for the 3 models (linear, quadratic and cubic). Which of the three model, from these plots, best fit the linear model assumption? Explain.



Solution: A Q-Q plot plots the theoretical quantiles from a sample of a given (in this case: normal) distribution against the empirical quantiles (order statistics) of a sample (in this case, the sample residuals). If the plot shows a slightly noisy identity line pattern, then we cannot reject normality of the residuals. If there is a clear pattern of the points away from the diagonal line, then the assumption of normality may be violated. In the above, the Q-Q plots are not so clear, but the histogram for the linear model is clearly bimodal, which conflicts with the symmetric unimodality of the normal density. It's hard to tell apart the residuals of the quadratic vs the cubic model, they both look normal. However, the principle of parsimony would pick the smaller model of the two, thus, like in (d), the quadratic model "wins" over the cubic one.

[15] 3. Data on housing price in all 50 US states and the District of Columbia has been collected, with covariates including median income, population, land area, percentage of African American, population density, population growth and number of registered vehicles per capita. As you may guess, the idea is to do some model selection.

(a) Define R^2 and R_a^2.

$R^2 = 1 - \frac{SSE}{SST}$

$$R_a^2 = 1 - \frac{\frac{SSE}{n-p}}{\frac{SSTO}{n-1}} = 1 - \left(\frac{n-1}{n-p}\right)\frac{SSE}{SSTO} = 1 - \frac{MSE}{MSTO}$$

Solution:

(b) Define variance inflation factor.

Solution:

$$VIF = \frac{1}{1 - R_j^2}$$

where R_j^2 is the R^2 resulting from regressing the j-th covariate on the other $p-1$ covariates

(c) Models using Median Income, Population density and vehicles per capita were fitted.

Model #	Variables
1	Income
2	Density
3	Vehicles
4	Income, Density
5	Income, Vehicles
6	Density, Vehicles
7	Income, Density, Vehicles

Below is the Sum of Squares of press residuals for each of these models. Which one has the best fit according to this criterion? Explain.

The MEANS Procedure

Variable	Sum
p1	465104.11
p2	8805404.56
p3	678710.35
p4	4097833.34
p5	428870.35
p6	8448969.13
p7	3786624.71

Solution: Model 5 has the best fit according to PRESS residuals. PRESS residuals measure the (cross-validated) prediction error at each point, minimizing the sum of square of these thus minimizes the prediction error. Hence Model 5 based on income and vehicles has the best fit.

(d) Give a description of backward selection procedure.

Solution: The backward selection procedure starts with the model with all covariates, if they are not all significant, drops the least significant (the one with the highest p-value) from the model and performs the regression again, rinse and repeat until all the covariates left (id any) are significant.

(e) A backward selection procedure was done on the data, with the following model. Using a Bonferroni correction for multiple testing at $\alpha = 0.05$ significance level, which covariates are related to housing prices?

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-107.00335	71.62598	12119	2.23	0.1422
Medinc	0.01002	0.00150	242577	44.67	<.0001
PopDensmiles	0.03342	0.00862	81597	15.03	0.0003
PopGrowth	-43.49134	21.31571	22606	4.16	0.0472
AreaMiles	-0.23737	0.12474	19662	3.62	0.0635
Pop2011	4.72540	1.53232	51642	9.51	0.0035

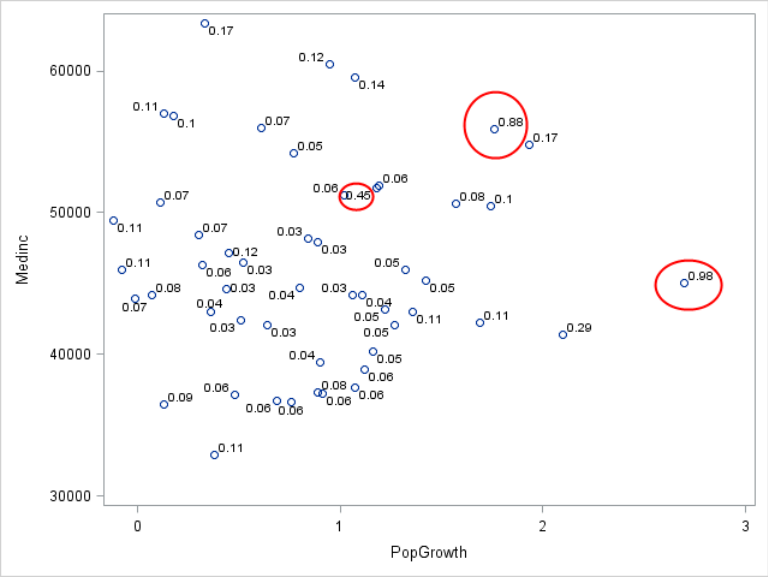
Solution: There are $g = 5$ variables in the model, thus the Bonferroni adjusted significance level is $\alpha/g = 0.05/5 = 0.01$, and we only consider significant covariates with p-value < 0.01. Here, those are Median Income, Population Density and Population in 2011.

[15] 4. Refer to question 3. We are interested in checking the data for outliers and departures from normality.

(a) Define what is meant by leverage.

Solution: Leverage is a measure of the influence of an observation on the prediction (or parameter estimation). For the i-th observation, it is defined to be h_{ii}, the i-th diagonal element of the hat matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

(b) Find two points with high leverage from the graph below. Give the formula for the rule of thumb for the threshold point for high leverage and compute it for the model from question 3 3(e).



Solution: The rule of thumb is any $h_{ii} > \frac{2p}{n}$ where p is the number of parameters in the model. Here the threshold is $\frac{2\times 6}{51} = 0.23529$. High leverage points are usually on the outside boundary of the covariate space in some dimension, thus we'd find them on the outskirts of the cloud of points if said cloud of points is for appropriate covariates. Three high values were circled above (anything close to one is automatically high, unless $p \approx n/2$ which is simply not a good idea).

(c) Describe the added covariate method to check for outlier. The District of Columbia (D.C.) which is a federal district and not a state, may be a potential outlier in the data. Given the SAS output below, confirm whether D.C. is an outlier or not. Explain your decision.

Parameter Estimates

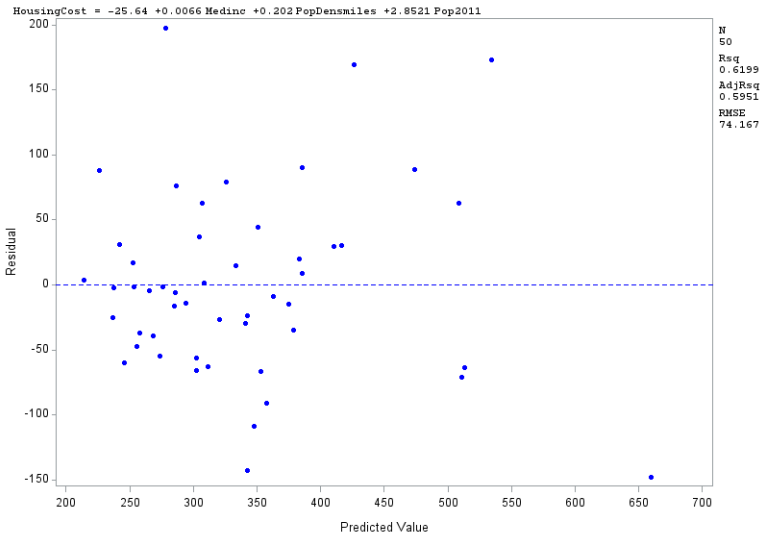
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-44.92473	76.82969	-0.58	0.5617
Medinc	Medinc	1	0.00798	0.00181	4.42	<.0001
PopDensmiles	PopDensmiles	1	0.13709	0.05495	2.49	0.0164
PopGrowth	PopGrowth	1	-27.05576	22.43322	-1.21	0.2342
AreaMiles	AreaMiles	1	-0.18191	0.12466	-1.46	0.1516
Pop2011	Pop2011	1	3.83166	1.56103	2.45	0.0181
dc		1	-1075.04247	563.09942	-1.91	0.0628

Solution: One can test for an outlier by making a categorical (indicator variable) covariate that is 1 for suspected outliers and 0 otherwise. One can then use regular inference on $\beta_{outlier}$ (or multiple outliers using nested models F test) to detect whether the suspected outliers are significantly different from the rest. For a single outlier (or grouped outliers with the same coeffecient), we

need to reject $H_0 : \beta_{outlier} = 0$. In this case β_{DC} has p-value larger than the conventional 0.05, so we cannot say it is an outlier.

(d) Name three measures of influence based on prediction and describe what they measure.
Solution: The three measures are DFFITS, Cook's D and DFBETAS. They measure, respectively, the distance between $\hat{y}_i$ from the model with all the data and the model where the i-th observation was removed, the sum of squares of the distances between all fitted y_i's from the full data and their respectively fitted value when the i-th observation is removed, and the distance between b_k, the estimated β for each p covariates with and without the i-th observation.

(e) D.C. was removed from the data and a regression was run with only Median Income, Population Density and Population as covariates. Looking at the residual vs predicted plot, should a weighted least-square regression be performed? Explain. What formal test would help us decide whether to do WLS?



Solution: It does not appear clear that weighted least-squares should be used as to choose to do so, one would have to notice a clear variation in the dispersion of the residuals against (fitted/covariate). This is not the case here. A properly set-up Brown-Forsythe test, using groups of residuals based on the covariates to separate them according to large and small dispersion, could help us decide whether WLS is warranted or not.

[15] 5. In the winter 1846-1847, about 90 wagon train emigrants in the Donner party were unable to cross the Sierra Nevada Mountains of California before winter, and almost half of them starved to death. Data on members of the party was gathered by historians. The variable of interest is Outcome (1=survided, 0=died of starvation), based on the covariates Age, Sex (genderMale=1 for men, 0 for women) and the categorical variable Status (Hired=1 for hired workers; Single=1 for single, baseline category is Family member).

(a) Define the logistic regression model for this data, clearly identifying the parameters, the link function and the quantities of interest.

Solution:

$$Y_i = E(Y_i) + \epsilon_i$$

where $E(Y_i) = \pi_i$ and $\log\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{x}_i'\boldsymbol{\beta}$ (logit is the link function). $\pi_i = P(Y_i = 1|\mathbf{x}_i)$, $\mathbf{x}_i$ is the vector of covariates. Y_i is the survival status.

(b) Write down the likelihood for the data. Note that $n = 88$.

Solution:

$$\mathcal{L}(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) = \prod_{i=1}^{88} \pi_i^{y_i}(1-\pi_i)^{1-y_i}$$

where π_i is as in (a).

(c) Write down the formula for the deviance based goodness-of-fit statistic and clearly define the null and alternative hypothesis.

Solution:

$$\begin{aligned} G^2 &= -2\left(\log \mathcal{L}(R) - \log \mathcal{L}(F)\right) \\ &= -2\sum_{j=1}^{2}\left[Y_{\cdot j}\log\left(\frac{\hat{\pi}_j}{p_j}\right) + (n_j - Y_{\cdot j})\log\left(\frac{1-\hat{\pi}_j}{1-p_j}\right)\right] \end{aligned}$$

where $Y_{\cdot j} = \sum_{i=1}^{n_j} Y_{ij}$ for each unique $\mathbf{X}_j$ and n_j is the number of replications at $\mathbf{X}_j$. The null hypothesis is $H_0 : p_j = \pi_j$ for $j = 1, \ldots, c$ (c the number of unique $\mathbf{X}_j$), and the alternative is H_1 : at least one $p_j \neq \pi_j$, where $p_j = Y_{\cdot j}/n_j$.

(d) A logistic regression was performed with Age, Age^2 (`Age2`), gender and status. The output is below. Use it to predict the probability of survival of a 14 year old female family member.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Log Likelihood		-46.1815	
Full Log Likelihood		-46.1815	
AIC (smaller is better)		104.3630	
AICC (smaller is better)		105.4001	
BIC (smaller is better)		119.2271	

Analysis Of Maximum Likelihood Parameter Estimates						
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.1986	0.6172	-1.0111 1.4084	0.10	0.7476
Age	1	0.1675	0.0711	0.0282 0.3068	5.55	0.0184
Age2	1	-0.0039	0.0015	-0.0069 -0.0009	6.50	0.0108
genderMale	1	-0.6637	0.5588	-1.7590 0.4315	1.41	0.2349
Hired	1	-1.6254	0.7481	-3.0916 -0.1592	4.72	0.0298
Single	1	-26.3242	143392.4	-281070 281017.6	0.00	0.9999
Scale	0	1.0000	0.0000	1.0000 1.0000		

Solution: Female family member means all the categorical covariates are 0, only Age and Age^2 are needed.

$$\begin{aligned}\pi_i &= \frac{1}{1+\exp(-\mathbf{x}_i'\boldsymbol{\beta})} \\ &= \frac{1}{1+\exp(-(0.1986+0.1675\times 14-0.0039\times 14^2))} \\ &= \frac{1}{1+\exp(-1.7792)} = \frac{1}{1+0.1687731} = 0.8556\end{aligned}$$

(e) A simpler logistic regression model with only Age and gender was performed. Based on the SAS output below, is the model from (d) a significant improvement over this reduced model? Use a significance level of 95%. Note that $\chi^2_{0.05}(2) = 5.991$, $\chi^2_{0.05}(3) = 7.815$, $F_{0.05}(2, 82) = 3.108$, $F_{0.05}(3, 82) = 2.716$

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Log Likelihood		-54.4337	
Full Log Likelihood		-54.4337	
AIC (smaller is better)		114.8674	
AICC (smaller is better)		115.1531	
BIC (smaller is better)		122.2994	

Analysis Of Maximum Likelihood Parameter Estimates						
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.6218	0.5028	0.6363 2.6073	10.40	0.0013
Age	1	-0.0356	0.0152	-0.0655 -0.0057	5.45	0.0195
genderMale	1	-1.0680	0.4823	-2.0132 -0.1227	4.90	0.0268
Scale	0	1.0000	0.0000	1.0000 1.0000		

Solution:

$$G^2 = -2(\log \mathcal{L}(R) - \log \mathcal{L}(F)) = -2 \times (-54.4337 + 46.1815) = 16.5044$$

Under H_0, $G^2 \sim \chi^2(5-2) = \chi^2(3)$, but, conveniently, the statistic is larger than all the given critical values so even if you pick the wrong critical value, the conclusion is still to reject the null that the models are equivalent. Thus, the model from (d) is a significant improvement from the reduced model.

[15] 6. As part of a wildlife reintroduction experiment, 100 tagged family units of wolves (say) were released in a large wildlife preserve. A year later, the size of each group (number of wolves per family unit), their presence in 3 locations of the preserve and the number of sightings of the units are recorded. Scientists want to know if the group size varies by location, and whether it is related to the number of sightings as well.

(a) Explain why this is a Poisson regression problem and give the regression model.

Solution: The group size is not well approximated by the continous, symmetric distribution that is the normal. Since we have counts that go from 1 to (possibly) infinity, with no binomial setting, the appropriate choice would be the Poisson distribution. The model is

$$Y_i = E(Y_i) + \epsilon_i$$

where $E(Y_i) = \lambda_i = \exp(\mathbf{x}_i'\boldsymbol{\beta})$ (or $\log \lambda_i = \mathbf{x}_i'\boldsymbol{\beta}$), and $Y_i \sim \text{Poisson}(\lambda_i)$.

(b) Describe what overdispersion is.

Solution:
Under the Poisson model $E(Y) = Var(Y) = \lambda$, thus the estimated values of the mean and variance of $\hat{\lambda}_i$ should be roughly the same. Overdispersion occurs when $Var(Y) > E(Y)$. Using maximum likelihood or deviance the observed χ^2 statistic should have roughly the same value as the degrees of freedom (the expectation of a χ^2 random variable). Since the χ^2 distribution is right-skewed, a chi^2 statistic significantly larger than the degrees of freedom implies overdispersion.

(c) A Poisson regression with a log link was performed with all the covariates. Explain from the output below why it is overdispersed.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	95	190.1938	2.0020
Scaled Deviance	95	190.1938	2.0020
Pearson Chi-Square	95	209.4047	2.2043
Scaled Pearson X2	95	209.4047	2.2043
Log Likelihood		169.9589	
Full Log Likelihood		-243.2022	
AIC (smaller is better)		496.4043	
AICC (smaller is better)		497.0426	
BIC (smaller is better)		509.4302	

Solution: As per (b), both the Pearson χ^2 and deviance statistics are more than twice the degrees of freedom, which indicates the variance is larger than the mean, thus overdispersion.

(d) A correction for overdispersion was applied. Two models, one with all the covariates and one with only presence in Site 3 as a covariate, were evaluated. Use the output to test which model is better. Show all your work.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	95	190.1938	2.0020
Scaled Deviance	95	95.0000	1.0000
Pearson Chi-Square	95	209.4047	2.2043
Scaled Pearson X2	95	104.5957	1.1010
Log Likelihood		84.8929	
Full Log Likelihood		-243.2022	
AIC (smaller is better)		496.4043	
AICC (smaller is better)		497.0426	
BIC (smaller is better)		509.4302	

Analysis Of Maximum Likelihood Parameter Estimates						
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	1.9403	0.2302	1.4892	2.3914	71.06	<.0001
Sights	1	-0.0071	0.0133	-0.0332	0.0191	0.28	0.5956
Site1	1	0.0207	0.1698	-0.3121	0.3535	0.01	0.9028
Site2	1	-0.1018	0.1599	-0.4151	0.2115	0.41	0.5243
Site3	1	-0.7007	0.1672	-1.0283	-0.3731	17.57	<.0001
Scale	0	1.4149	0.0000	1.4149	1.4149		

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	98	191.4886	1.9540
Scaled Deviance	98	98.0000	1.0000
Pearson Chi-Square	98	212.1865	2.1652
Scaled Pearson X2	98	108.5928	1.1081
Log Likelihood		86.6502	
Full Log Likelihood		-243.8496	
AIC (smaller is better)		491.6992	
AICC (smaller is better)		491.8229	
BIC (smaller is better)		496.9095	

Solution:

$$G^2 = -2(\log \mathcal{L}(R) - \log \mathcal{L}(F)) = -2 \times (-243.8496 + 243.2022) = 1.29$$

Under H_0, this follows a χ^2 with $98 - 95 = 3$ degrees of freedom. Since $E(G^2) = 3$ under H_0 and we only reject H_0 for large values of G^2, we cannot reject that the reduced model is different than the larger one. Thus, the better model is the one with fewer covariates.

(e) Give the formula for a 95% confidence interval for β_{site3} from the reduced model. Compute the interval from the output data below

Parameter	DF	Estimate	Standard Error
Intercept	1	1.8357	0.1095
Site3	1	-0.6930	0.1429

Solution: $\beta_k \in b_k \pm z_{\alpha/2} s(b_k) = -0.693 \pm 1.96 \times 0.1429 = [-0.973, -0.413]$. Note that $-.693 = -\log 2$ thus the group size at site 3 is estimated to be half the group size elsewhere.