Your Name:

Duration of the Quiz is 50 minutes. There are 20 problems, worth 1 point each. Show all your work for full credit. Books, notes etc. are prohibited.

(1) Every year in the United States, over 120,000 undergraduates submit applications to medical school. Data was gathered on 55 medical school applicants from a liberal arts college in the Midwest. For each applicant, medical school Acceptance status (accepted or denied), GPA, MCAT scores, and Gender were collected. The data is stored in **MedGPA**, which is available from the STAT2 package on R. The following is an R code written considering one explanatory variable, MCAT.

```
> data(MedGPA)
> model<-glm(Acceptance~MCAT,family=binomial,data=MedGPA)
> summary(model)

Call:
glm(formula = Acceptance ~ MCAT, family = binomial, data = MedGPA)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -8.71245    3.23645  -2.692  0.00710 **
MCAT         0.24596    0.08938   2.752  0.00592 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 75.791  on 54  degrees of freedom
Residual deviance: 64.697  on 53  degrees of freedom
AIC: 68.697

Number of Fisher Scoring iterations: 4
```


(a) Write down the logistic regression model based on the R output.

(b) What does β_1 represent in the model?

(c) Calculate the odds scale factor e^{β_1} for the MCAT scores and interpret its value in the context of the problem.

(d) Find a 95% confidence interval for β_1.

(e) Find a 95% confidence interval for the odds ratio comparing an MCAT score of $x + 1$ to a score of x.

(f) Find a 95% confidence interval for the odds ratio comparing an MCAT score of $x + 3$ to a score of x.

(g) Estimate the odds of acceptance when MCAT score is 44 and 38, and compute the odds ratio comparing those two scores. Interpret.

(h) Estimate the probability of acceptance when MCAT score is 44 and 38, and compute the relative risk comparing those two scores.

(i) Test whether the acceptance is independent of MCAT score. Use $\alpha = 0.05$.

Hint: $P(\chi^2 > 0.05) = 3.841459$.

(j) Let's say that you'd like to conduct a likelihood-ratio test about the MCAT effect. Write down the observed value of a chi-squared statistic for the test. You must mention the degrees of freedom as well.

(2) Let's add two more predictor variables to the model. Here is the R output.

```
>
> data(MedGPA)
> model<-glm(Acceptance~MCAT + GPA + Sex,family=binomial,data=MedGPA)
> summary(model)

Call:
glm(formula = Acceptance ~ MCAT + GPA + Sex, family = binomial,
    data = MedGPA)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -23.9851     6.9685  -3.442 0.000578 ***
MCAT          0.1809     0.1080   1.675 0.093946 .
GPA           5.1392     1.8508   2.777 0.005491 **
SexM         -1.2580     0.7303  -1.723 0.084965 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 75.791  on 54  degrees of freedom
Residual deviance: 50.786  on 51  degrees of freedom
AIC: 58.786

Number of Fisher Scoring iterations: 5

>
```

(a) Write down the logistic regression model based on the R output.

(b) After adjusting for a student's gender and GPA, explain what happens to the odds of acceptance when the MCAT score is increased by 1-point?

(c) What does the value $e^{-1.258}$ tell you about the model?

(d) Explain the meaning of the value 5.1392 in the R output.

(3) The Global Longitudinal Study of Osteoporosis in Women (GLOW) is an international study of osteoporosis in women over 55 years of age being coordinated at the Center for Outcomes Research (COR) at the University of Massachusetts/Worcester by its Director, Dr. Frederick Anderson, Jr. The study has enrolled over 60,000 women aged 55 and older in ten countries. The major goals of the study are to use the data to provide insights into the management of fracture risk, patient experience with prevention and treatment of fractures and distribution of risk factors among older women on an international scale over the follow up period. Complete details on the study as well as a list of GLOW publications may be found at the Center for Outcomes Research web site, www.outcomes-umassmed.org/glow. In the GLOW study the covariate self-reported risk is coded at three levels (less, same, and more). The cross tabulation of it with fracture during follow-up (FRACTURE) is shown in the following table. Note that FRACTURE is a polychotomous variable.

Cross-Classification of Fracture During Follow-Up (FRACTURE) by Self-Reported Rate of Risk (RATERISK) from the GLOW Study, n = 500

	RATERISK			
FRACTURE	Less	Same	More	Total
Yes	28	48	49	125
No	139	138	98	375
Total	167	186	147	500

Specification of the Design Variables for RATERISK.

RATERISK (Code)	RATERISK2	RATERISK3
Less (1)	0	0
Same (2)	1	0
More (3)	0	1

Here are the results of fitting the logistic regression model to the data in the first table using the design variables in the second variable.

Variable	Coeff.	Std. Err.	z	p	95% CI
RATERISK2	0.546	0.2664	2.05	0.040	0.024, 1.068
RATERISK3	0.909	0.2711	3.35	0.001	0.378, 1.441
Constant	−1.602	0.2071	−7.74	<0.001	−2.008, −1.196

(a) What is the reference group?

(b) Compute $\log\big(\hat{OR}(\text{Same, Less})\big)$ using the first table.

(c) Compute $\log\big(\hat{OR}(\text{Same, Less})\big)$ using the R output.

(d) Can you explain why the two answers in the previous two problems agree?

(e) Compute $\log\big(\hat{OR}(\text{More, Less})\big)$ using the first table.

(f) Compute $\log\big(\hat{OR}(\text{More, Less})\big)$ using the R output.