Discussion and discovery of logistic regression.

This worksheet uses the code and summary output below. You'll need a calculator.

```
mvmodel <- glm(cases ~ age + income + sex, data=depress, family="binomial")
summary(mvmodel)
##
## Call:
## glm(formula = cases ~ age + income + sex, family = "binomial",
##     data = depress)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0249  -0.6524  -0.5050  -0.3179   2.5305
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.67646    0.57881  -1.169  0.24253
## age         -0.02096    0.00904  -2.318  0.02043 *
## income      -0.03656    0.01409  -2.595  0.00946 **
## sex          0.92945    0.38582   2.409  0.01600 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 268.12  on 293  degrees of freedom
## Residual deviance: 247.54  on 290  degrees of freedom
## AIC: 255.54
##
## Number of Fisher Scoring iterations: 5
```

1. Review the code above that fits a logistic regression model with the presence of symptoms of depression (cases) as the outcome, and age, income and sex as covariates. What specifically is the difference in the model fitting code below and what it would look like if you were fitting a linear regression.

2. What values can the variable `cases` take? What do each of those values represent?

3. Using the summary output showing the numerical estimates for the beta coefficients, write out the right hand side of the mathematical model. E.g. $log(odds) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots$.

4. Calculate the odds of depression for a 1 unit change in age. Do not use any values for age, keep it as a variable x1. Simplify your algebra as much as possible $OR = \frac{Odds\,(Y=1 \mid age+1)}{Odds(Y=1|age)}$

5. Using the simplified version of the equation you figured out above, calculate the odds of depression for a female vs male (male is the reference group for the variable sex).

6. In linear regression, what is the hypothesis (in symbolic form) that we are testing to assess if there is a relationship between x and y?

7. For a logistic regression, what is the value of the odds ratio that we test against in the null hypothesis? That is, what is the value of OR when we assume no relationship exists?

8. Interpret what an OR *above* this testing value means. Interpret what an OR *below* this testing value means.

9. Confidence intervals on βs are calculated as $\hat{\beta} \pm 1.96 * SE$. How would we calculate a confidence interval for the OR?