LOGISTIC REGRESSION
Joseph M Hilbe

Assignment 2
(20 points)

Q1 (2 points)
What is the primary difference in probabilities produced by a logit model compared to the probabilities produced by a complementary loglog model?

The following data and model relate to questions 2-8.
The following binary logistic regression output uses the **affairs** data set, with *affair* (1 = had affair; 0 = never had affair) as the response and predictors:

male : 1 = male; 0 = female
kids : 1 = children in family; 0 = no children in family.
and a 5-level categorical predictor, *religious*, which has been factored into indicator variables or levels
a*nti-religious:* 1- anitreligious --- the reference.
notrel : 2- not religious
slghtrel : 3- slightly religious
smerel : 4- somewhat religious
vryrel : 5- very religious.

```
Logistic regression                             Number of obs =    601
                                                LR chi2 (6)   =  34.36
                                                Prob > chi2   = 0.0000
Log likelihood = -320.50848                     Pseudo R2     = 0.0509
------------------------------------------------------------------------------
affair  |  Odds Ratio    Std. Err.    z      P>|z|   [95% Conf.Interval]
------------------------------------------------------------------------------
male    |    1.2409   .2424715     1.10    0.269   .8460812    1.819958
kids    |  2.346751   .5681122     3.52    0.000   1.460172    3.771636
notrel  |   .470281   .1651811    -2.15    0.032   .2362555    .9361227
slghtrel|  .6688562   .2381866    -1.13    0.259   .3328214    1.344171
smerel  |  .2633311   .0942904    -3.73    0.000   .1305312    .5312389
vryrel  |  .2792907   .1209288    -2.95    0.003    .119536    .6525508
------------------------------------------------------------------------------
```

Q2 (2 points)
How can the significance of a predictor be determined from the confidence intervals?

Q3 (2 points)
Interpret the odds ratio of *kids*.

Q4 (2 points)
What is the odds ratio of *vryrel* for having an affair. Interpret it in full.

Q5 (2 points)
What is the odds ratio of *vryrel* for not having an affair. Interpret it in full.

Q6 (2 points)
What is the value of the parameter estimate of *male* to the nearest hundredths?

Q7 (2 points)
If the model were re-estimated excluding the predictor *male*, what would you expect to happen to the odds ratio of *kids*. Why?

The following data and model relate to questions 8-9
Below is the **kyp** data set, which is from a study of kyphosis, a disease of the spine. The data comes from Bell et al (1989) as found in Hastie and Tibshirani (1990). *kyp* is the binary response with 1 indicating that the patient has kyphosis, 0 that they do not. Patients in the study underwent corrective spine surgery. Predictors include

start: continuous - indicating the first disk or vertebrae level of the surgery (1-18)
numb: continuous - indicating the number of disks involved in the surgery (2-14)
age : continuous - age in months

Logistic regression **Number of obs = 83**
LR chi2(3) = 21.79
Prob > chi2 = 0.0001
Log likelihood = -32.508134 **Pseudo R2 = 0.2510**

kyp	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
start	-.1981754	.0657199	-3.02	0.003	-.3269841	-.0693667
numb	.2981585	.1778791	1.68	0.094	-.0504781	.6467951
age	.0059798	.0055195	1.08	0.279	-.0048383	.0167979
_cons	-1.213598	1.234168	-0.98	0.325	-3.632523	1.205327

Q8 (2 points)
Interpret the coefficient *start*.

Q9 (2 points)
What does the intercept, *_cons*, indicate?

Q10 (2 points) More Challenging

Given the table of counts below, with the reponse y on the vertical axis and predictor, x, on the horizontal. Calculate the logistic coefficient of x. Then prove using a logistic regression.

	x		
1=y	0	1	Total
0	825	157	982
1	431	82	513
Total	1,256	239	1,495