# Stat 545 Exam

November 7, 2018

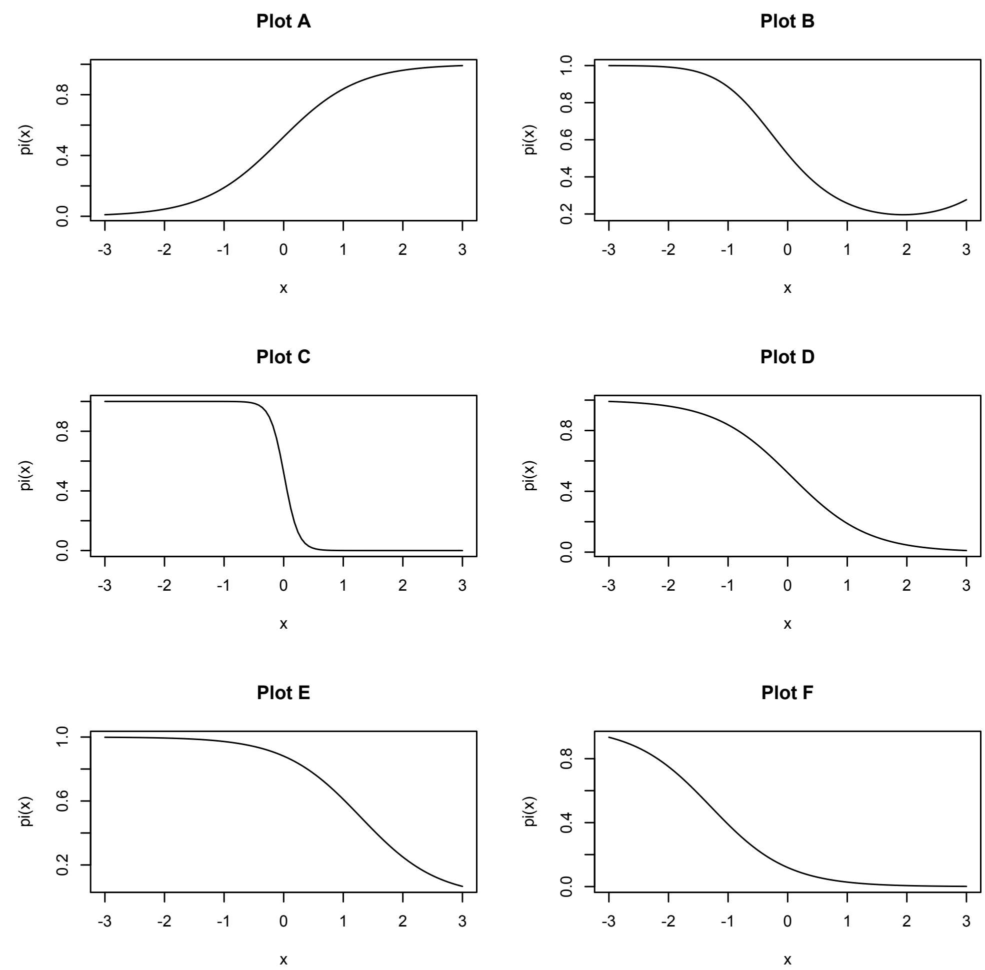
## Directions:

- This is a takehome exam. **YOU ARE ALLOWED 1.5 HOURS (90 MINUTES) TO TAKE THIS EXAM**
- The exam is closed book. You may not use a calculator or a computer or any other electric device. The exam is designed so that answers requiring numerical calculation can be done accurately enough without a calculator.
- You are allowed one 8.5 by 11 inch page of notes with writing or printing on both sides.
- Use your own paper.
- If you are unable to staple your exam sheets, please fold them to keep them together.
- To get full credit, you must justify your answer. For instance, in Problem 1, you should explain why your choice could be correct and none of the others could be.
- **HAND IN YOUR BY 2PM, MONDAY, NOV. 20.**

**1. [25 points]** A logistic model of the form $P[Y = 1|X = x] = \pi(x)$ is fit where $x$ is a 1-dimensional continuous variable. The fitted model is

$$\text{logit}(\pi(x)) \;=\; 0.09 \;-\; 1.55x.$$

Below are plots of $\pi(x)$ for different logistic models. Determine which one could be the plot of the given fitted model.




**2. [20 points] (a)** Define the GLM (Generalized Linear Model) for a response random variable $Y$ with predictor variables $x_1, \ldots, x_p$, including an intercept in your model. Your answer should include the functional form of the conditional distribution of $Y$ given $X = (x_1, \ldots, x_p)$ and the link function.

Give at least three link functions which may be appropriate when $Y$ is a binary outcome (i.e., either $Y = 0$ or $Y = 1$) and discuss their advantages and disadvantages.

**3.** [**25 points**] We want to test if two discrete random variables are independent, say $X$ takes the values $1, \ldots, I$ and $Y$ takes the values $1, \ldots, J$. Suppose we know the marginal probabilities $P[X = i] = \pi_{i+}$ but we don't know the corresponding marginal probabilities for $Y$. Assume we have i.i.d. observations $(X_k, Y_k)$, $k = 1, \ldots, n$. Give the form of the Pearson $\chi^2$ test statistic (be sure to define all the quantities that go into the test statistic) and determine the degrees of freedom of the asymptotic $\chi^2$ null distribution.

**4.** [**20 points**] Below is some output from fitting a logistic regression of a binary outcome $Y$ on a 2-dimensional predictor variable $X$. Use this output to answer the questions below.

```
> fit = glm(y ~ x, family=binomial)

Warning message:
glm.fit: fitted probabilities numerically 0 or 1 occurred

> summary(fit)

Call:
glm(formula = y ~ x, family = binomial)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.40986  -0.00032   0.00000   0.00001   1.51276

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.7932     1.4319   0.554    0.580
x1           24.9077    16.6866   1.493    0.136
x2          -24.0738    15.8765  -1.516    0.129
```

```
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 67.3012  on 49  degrees of freedom
Residual deviance:  5.0176  on 47  degrees of freedom
AIC: 11.018

Number of Fisher Scoring iterations: 11
```

**(a)** Do you think this logistic model is a "good" fit? Give a detailed explanation.

**(b)** What would you predict for $P[Y = 1|X = (-1, 1)]$ based on this fit? You should be able to answer this with minimal calculation. Do you think this is a reasonable predicted value?

**5.** [**10** [**points**] Suppose we have a 2 dimensional parameter $(\theta_1, \theta_2)$ and we want a confidence interval for $\theta_1$. Assume we have a large sample and the likelihood satisfies any needed "regularity" conditions.

**(a)** Describe how to obtain a confidence interval by inverting the likehood ratio test.

**(b)** Describe a simpler approach based on inverting a Wald test.