

NOTE: These are just Practice Problems. This is NOT meant to look just like the test, and it is NOT the only thing that you should study. Make sure you know all the material from the notes, quizzes, suggested homework and the corresponding chapters in the book.

Questions 1 – 7 Former kicker for the Gator football team, Chris Hetland, was very good at making field goals in the 2005 season, but in the 2006 regular season had only made 3 out of 12. The following is the Logistic Regression Output to predict the probability of making a field goal (yes/no), based on how far the kick is (in yards) and the year (2005 or 2006).

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P
Constant	8312.97	3073.50	2.70	0.007
yards	-0.173760	0.0901421	-1.93	0.054
year	-4.14174	1.53141	-2.70	0.007

- What kind of variables do we have here?
 - a quantitative predictor and a quantitative response
 - two quantitative predictors and a quantitative response
 - a quantitative predictor and a categorical response
 - two quantitative predictors and a categorical response
- Write down the fitted logistic regression equation:
 - $\hat{y} = 8312.97 - 0.173760 - 4.14174$
 - $\hat{y} = 8312.97 - 0.173760 \text{ yards} - 4.14174 \text{ year}$
 - $\hat{p} = e^{8312.97 - 0.173760 - 4.14174} / 1 + e^{8312.97 - 0.173760 - 4.14174}$
 - $\hat{p} = e^{8312.97 - 0.173760 \text{ yards} - 4.14174 \text{ year}} / 1 + e^{8312.97 - 0.173760 \text{ yards} - 4.14174 \text{ year}}$
- The coefficients of yards and years are both negative. This means:
 - that neither variable is a good predictor of whether the kicker will make or not the field goal
 - that simple linear regression would have been more appropriate than logistic regression
 - that there was a mistake in the way the data was entered into the computer
 - that the chances of making the field goal go down as the yardage increases, and as the years increase

Find the probability of making a field goal:

- from the 30 yd line in 2006
- from the 30 yd line in 2005
- from the 40 yd line in 2006
- from the 40 yd line in 2005

Question 8 - 13 As part of a project for their Intro Stat course, two students compared two brands of chips, Frito Lays and Golden Flakes, to see which company gives you more for your money. Five bags of each brand (which, according to the label, each contained 35.4 grams) were measured with a very accurate scale. Use the Wilcoxon Rank-Sum test to see if there are any significant differences between the two brands in the amount of product they put in their bags.

Frito Lays: 35.3 35.4 35.8 35.9 35.9

Golden Flake: 35.3 37.8 38.8 38.1 42.5

8. The null hypothesis is about:

- a) the mean contents of the bags for Frito Lays and Golden Flakes brands
- b) the mode of the contents of the bags for Frito Lays and Golden Flakes brands
- c) the distribution of the contents of the bags for the two brands
- d) the number of bags with contents below the label weight for the two brands

9. The alternative hypothesis, according to the problem stated above, is that:

- a) Frito Lays gives you more chips than Golden Flakes
- b) Frito Lays gives you less chips than Golden Flakes
- c) Frito Lays gives you either more or less chips than Golden Flakes
- d) Golden Flakes gives you more chips than the amount stated on the label

10. The bags that contained 35.9 grams will receive a rank of:

- a) 4
- b) 4.5
- c) 5
- d) 5.5
- e) 6

11. The p-value for the test was .1164. We conclude that:

- a) Frito Lays gives you more chips.
- b) Golden Flakes gives you more chips.
- c) There is not enough evidence to prove a difference between the two brands.
- d) There is enough evidence to prove a difference between the two brands.

12. If the assumptions for the Normal based procedure were satisfied, we could analyze the data with a confidence interval for:

- a) μ
- b) $\mu_1 - \mu_2$
- c) μ_d
- d) $\eta_1 - \eta_2$

13. Why is it not a good idea to use the Normal-based procedure here?

- a) the data was not randomly selected
- b) the data does not have a continuous distribution
- c) the outlier violates the assumption of Normality
- d) the nonparametric method is always better

Questions 14 - 18 Do plain and peanut m&m's have the same distribution of colors? Several bags of each variety (plain and peanut) were randomly selected, and the number of candies of each color were counted before eating any of them. The data appears below.

	brown	yellow	red	blue	green	orange	
plain	81	84	41	17	30	41	294
peanut	17	7	27	13	14	16	94
	98	91	68	30	44	57	388

14. The null hypothesis is that:
- plain and peanut varieties are independent
 - the colors are independent of each other
 - color and variety are independent of each other
 - all of the above
15. The expected number of blue, peanut m&m's (under independence) is:
- 15.0
 - 15.67
 - 7.27
 - 32.33
16. The sampling distribution is χ^2 with degrees of freedom equal to:
- 5
 - 10
 - 11
 - 12
17. The test statistic was 32.67. Use the table to approximate the p-value for this test:
- smaller than .001
 - equal to .05
 - between .05 and .10
 - between .950 and .975
18. What conclusions can you reach from this analysis, based on the data and the test statistic given?
- There is something wrong with the data, maybe the m&m's were not randomly selected.
 - The distribution of colors is not significantly different for plain and peanut m&m's.
 - Peanut m&m's are significantly more colorful than plain m&m's.
 - There are significantly more brown m&m's than orange m&m's.