BUDT 730 DMD
Practice Final Exam

Question 1

A parking lot company owns a parking lot in Baltimore. One particular attendant works at the parking lot, taking cash parking fare. He hands in all of the revenue he gets to the company at the end of every night. The average income from the parking lot for any given day is $850, with a standard deviation of $150. He has worked 45 days this year, and on average, he has turned in $800 per day to the company. That is, over the course of those 45 days, on average, the company would have expected to receive $38,250 in revenue. However, he has only given us $36,000. He's $2,250 short of what the company would expect.

a) What is the sampling distribution of the average income from the parking lot over 45 days?

A manager in the company should determine if the attendant is cheating him or not. If the manager decides to fire the attendant, he should hire a new attendant using a recruitment agency, which incurs a cost of $1,000.

b) Write out the null and the alternative hypotheses for the manager. Clearly state them in terms of your parameter.

c) What is the risk of Type I and Type II error? Clearly state them in English.

Which one is worse?

How does this influence our significance level we need?

d) **ASSUME** that the p-value is 0.0127. Write out the meaning of this p-value in English.

e) Set up a test statistic for the hypothesis test and compute the value

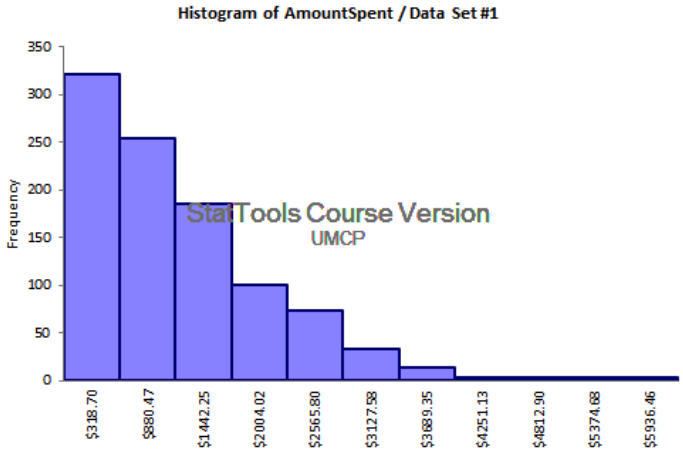
f) Using a 5% level of significance, should the manager fire the attendant? (Use the critical value method)

Question 2

HyTex, a direct marketing company, is trying to understand which factors contribute to the purchasing habits of customers who receive product catalogs in the mail. The company has a data set consisting of 1000 customers, which includes the following variables:

- Age: encoded as 1 for 30 or younger, 2 for 31 to 55, 3 for 56 or older
- Gender: encoded as 0 for females and 1 for males
- Close: encoded as 1 if the customer lives close to stores that sell similar merchandise, 0 otherwise
- Salary (in $): combined annual salary of the customer and spouse (if any)
- History: encoded as NA if the customer had no dealings prior to the current year, 1 if the customer was low-spending in the previous year, 2 if medium-spending, 3 if high-spending
- Catalogs: number of catalogs sent to the customer this year (6, 12, 18, or 24)
- FirstPurchase: date of the customer's first purchase with HyTex
- AmountSpent (in $): total amount of customer purchases this year

The two outputs below provide some descriptive information about the dependent variable of interest, AmountSpent:



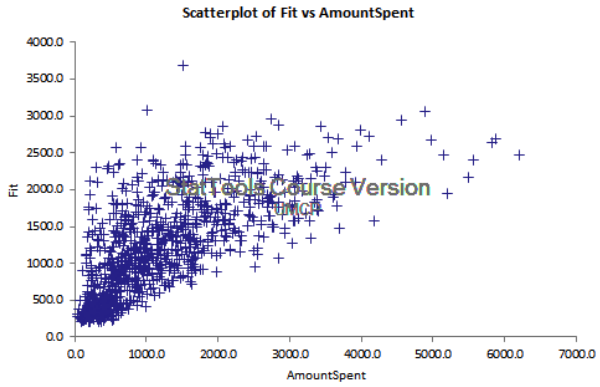
One Variable Summary	**AmountSpent**
Mean	$1216.77
Variance	923665.86
Std. Dev.	$961.08
Median	$961.61
Mode	$584.40
Minimum	$37.81
Maximum	$6217.34
Range	$6179.54
Count	1000

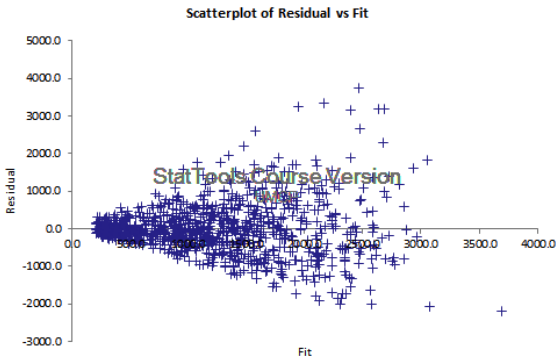
a) Given the variables specified above, name two variables for which it would be reasonable to convert to dummy variables to be used in a regression model.

First, we fit a basic model of Salary vs. AmountSpent and report the following results:

Multiple Regression for AmountSpent Summary	**R-Square**	**Adjusted R-Square**	**StErr of Estimate**
	0.4894	0.4889	687.0676

Regression Table	**Coefficient**	**Standard Error**	**t-Value**	**p-Value**
Constant	-15.3324	45.3743	-0.3379	0.7355
Salary	0.02196	0.00071	30.9307	< 0.0001





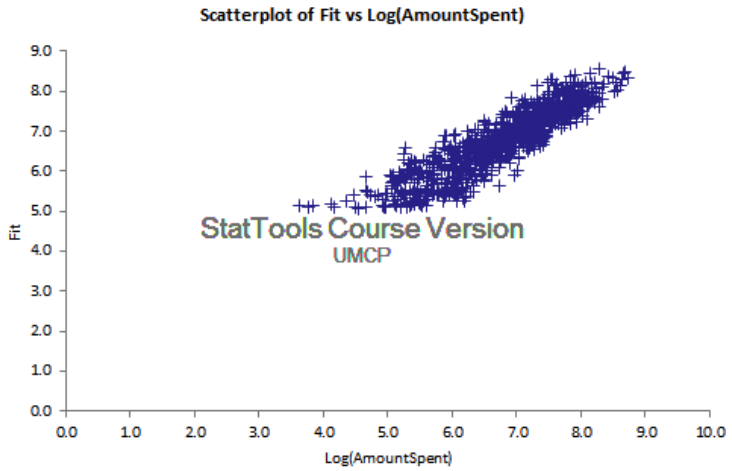
b) Is this model valid as an explanatory model? Why or why not. Explain your answer.

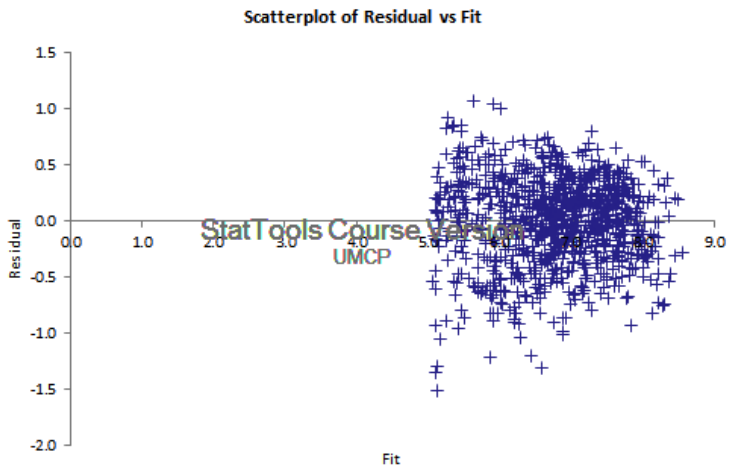
c) Provide an interpretation of the p-value for the Salary.

We attempt to fit a better model by transforming AmountSpent into Log(AmountSpent) and incorporating additional independent variables. The results are summarized below.

Multiple Regression for Log(AmountSpent) Summary	R-Square	Adjusted R-Square	StErr of Estimate
	0.8063	0.8047	0.3869

Regression Table	Coefficient	Standard Error	t-Value	p-Value
Constant	5.99618	0.04239	141.4261	< 0.0001
Close	-0.27409	0.02827	-9.6955	< 0.0001
Salary	1.3796E-05	5.0726E-07	27.1984	< 0.0001
History = 1	-0.7899	0.03554	-22.2249	< 0.0001
History = 2	-0.1441	0.03489	-4.1323	< 0.0001
History = 3	0.08712	0.03825	2.2777	0.0230
Catalogs = 12	0.31396	0.03405	9.2196	< 0.0001
Catalogs = 18	0.57005	0.03629	15.7078	< 0.0001
Catalogs = 24	0.70908	0.03685	19.2415	< 0.0001





d) Specify two pieces of information from these results that support the belief that this model is an improvement over the initial model?

e) Provide an economic interpretation for the coefficient of the Catalogs = 18 variable.

f) Predict AmountSpent for a customer with a household income of $100,000 who lives close to stores that sell similar merchandise, was a high spender in the previous year, and received 6 catalogs. You may round each coefficient to the hundredths place. You must show your work to receive full credit.

Question 3

An analyst is studying airline fares on several routes in the U.S. For each route, she has data on the following variables:

PAX = Number of passengers on that route (demand)
SW = 1 if Southwest is present on that route; 0 otherwise
FARE = average price on that route in dollars

The following regression output was obtained. Note that the dependent variable is *log*(PAX).

Multiple Regression for Log(PAX) ***Summary***	**R-Square**	**Adjusted R-square**	**Std. Err. of Estimate**
	0.0537	0.0492	0.711928389

Regression Table	Coefficient	t-Value	p-Value
Constant	10.5229	22.8820	0.0000
SW	1.4421	2.0148	0.0443
Log(FARE)	-0.2703	-3.0446	0.0024
SW*Log(FARE)	-0.3535	-2.3501	0.0191

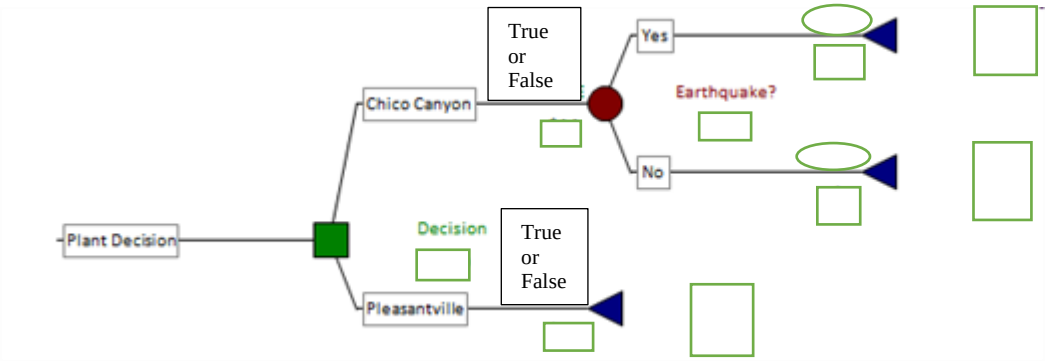
a) What is the price elasticity of demand (the change in demand in relation to a change in its price) on routes on which Southwest is present?

b) Predict the demand for a route on which Southwest does not fly and the fare is $500. You may round each coefficient to the hundredths place. You must show your work to receive full credit.

Question 4

A nuclear power company is deciding whether to build a nuclear plant at Chico Canyon or at Pleasantville. The cost of building the power plant is $14 million at Chico and $20 million at Pleasantville. If the company builds at Chico, however, and an earthquake occurs at Chico during the next 5 years, construction will be terminated and the company will lose $20 million (and will still have to build a power plant at Pleasantville). Without further information, the company believes there is a 20% chance that an earthquake will occur at Chico during the next 5 years.

a) Complete the decision tree below to help the power company decide what to do based on EMV. Make sure to label all decision and chance nodes and include appropriate costs, payoffs and probabilities. Show how to compute the EMVs.



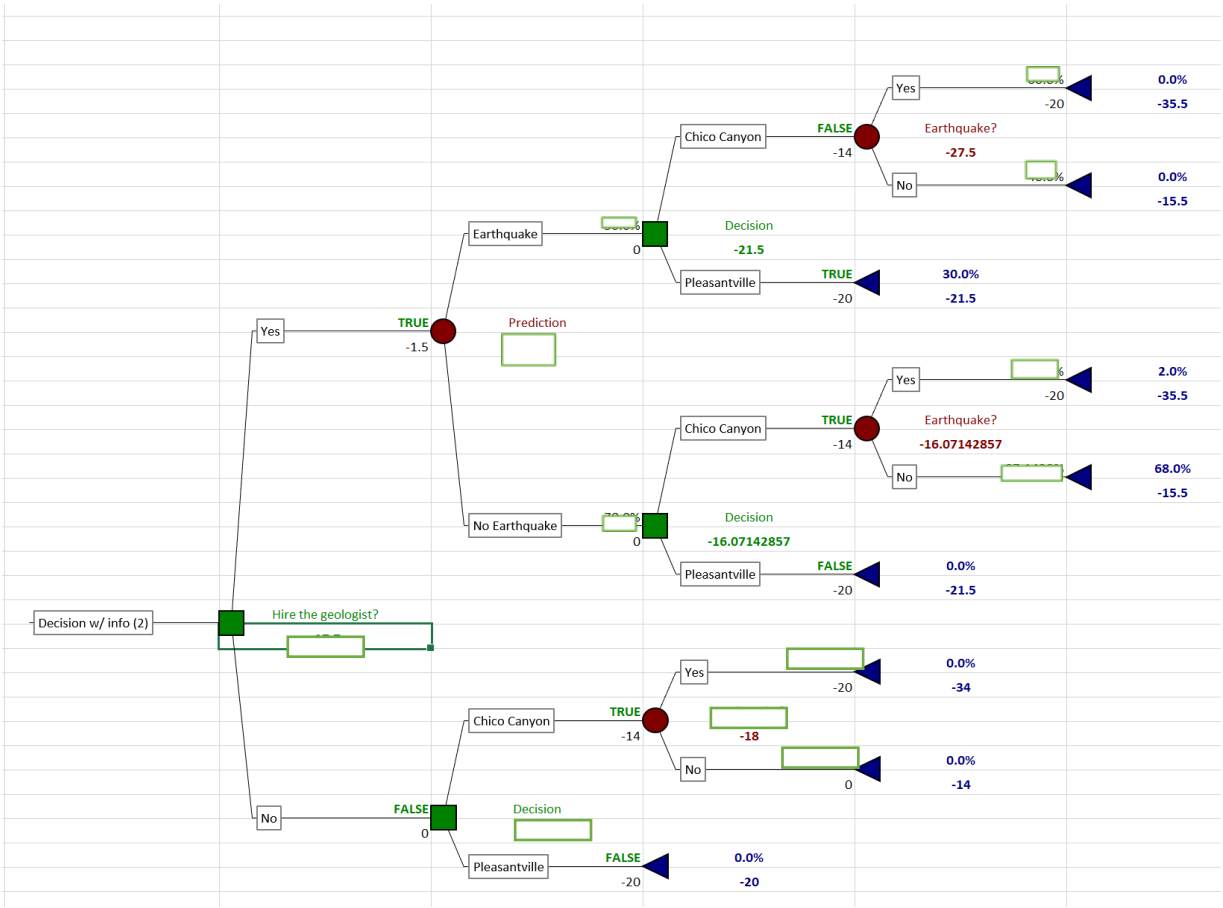
b) Where should the power company build the plant? What is the expected cost?

c) Suppose that an actual (not perfectly reliable) geologist can be hired to analyze the earthquake risk. His fee is $1.5M. The geologist's past record indicates that he will predict an earthquake on 90% of the occasions for which an earthquake will occur and no earthquake on 85% of the occasions for which an earthquake will not occur. Given this information, the posterior probabilities that an earthquake will and will not occur, given the geologists predictions are given as below:

Posterior Prob	
Prob of Earthquake	30%
Prob of No Earthquake	70%

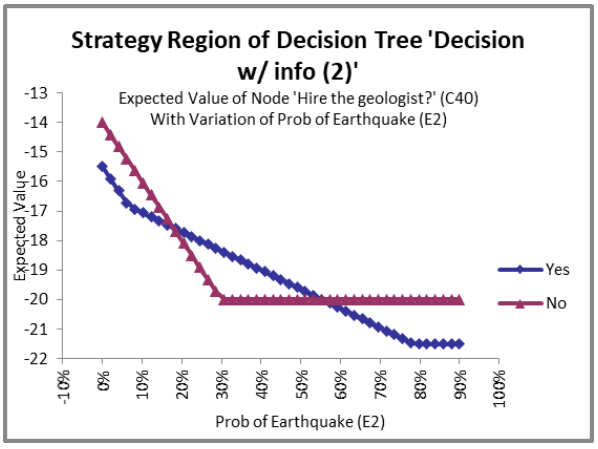
Posterior Prob		
	Prob of Earthquake	Prob of No Earthquake
Test Earthquake	0.6	0.4
Test No Earthquake	0.029	0.971

Complete the decision tree below.



d) Where should the power company build the plant? Should the company hire the geologist? What is the EVI?

e) Suppose that the probability that an earthquake will occur at Chico during the next 5 years is uncertain. The strategy region results on the probability is given below.



Answer the following questions.

i. At what probability, if any, would the **first stage** optimal decision for model (c) change? If it changes, how?

ii. At what probability, if any, would the **second stage** optimal decision for model (c) change? Explain your answers.