STATISTICS PROJECT: Hypothesis Testing

See my comments in red. Scoring last page.

INTRODUCTION

My topic is the average tuition cost of a 4-yr. public college. Since I will soon be transferring to a 4-yr. college, I thought this topic would be perfect. "The College Board" says that the average tuition cost of college is $5836 per year. I will be researching online the costs of different public colleges to test this claim. I will be using the T-test for a mean, since my sample is going to be less than 30 and an unknown population standard deviation. I will also use Chi-Square Test of Independence.

HYPOTHESIS

I think the average cost of tuition is lower than the average stated by "The College Board".

Ho: mu >/= $5836.
H1: mu< $5836 (Claim)

DATA ANALYSIS

I collected my data from various college websites. I looked up the cost of tuition per year and the number of students enrolled. Here is what I came up with:

College	Tuition	Number of Students
Central Washington University	$4392	10,200
University of Washington	$5985	25,469
Washington State University	$5888	18,432
Western Washington University	$4356	13,000
Evergreen State University	$4590	4400
Eastern Washington University	$5904	10,000
Peninsula College	$3639	10,120
University of Oregon	$6174	20,394
Portland State University	$5208	24,284
Oregon State University	$5604	19,362
Southern Oregon University	$5233	5000
Eastern Oregon University	$4500	3000
Western Oregon University	$5763	4500

| University of Idaho | $4410 | 11,739 |
| Idaho State University | $4400 | 13,000 |

There weren't really any large gaps or outliers in the data that I collected. There was a gap between 5,000 – 10,000 students. But the rest was mostly consistent. The lowest tuition was $3639 from Peninsula College and the highest tuition was $6174 from the University of Oregon. Some of the websites were hard to find the information I wanted, but I eventually found it. Some of the websites were specific as to undergraduate or graduate and some probably contain both. I should have done further research to make sure that my numbers only contain undergraduates and not graduates. So, that is one possible mistake in the data collection. You've explained your strengths and weaknesses in collection well.

HYPOTHESIS TESTING

T-Test for a Mean

Step 1: State the hypothesis and identify the claim.

I claim that the average cost of college tuition is less than $5836 per year as concluded from "The College Board". At a=.025, can it be concluded that the average is less than $5836 based on a sample of 15 colleges?

H0: mu>/= $5836
H1: mu<$5836 (claim)

Step 2: Find the critical value

At a=.025 and d.f. = 14, the critical value is -2.145.

Step 3: Compute the sample test value.

m= 5069.73, s=787.80

t= (5069.73-5836)/(787.80/sqrt(15)) = -3.767

Step 4: Make the decision to reject or not reject the null hypothesis.

Reject the null hypotheses since -3.767 falls in the critical region.

Step 5: Summarize the results.

I will reject the null hypotheses since there is enough evidence to support the claim that the average cost of tuition is less than $5836 per year.

Chi-Squared Independence Test

Step 1: State the hypotheses and identify the claim.

I claim that there is a correlation between the number of students at a college and the cost of tuition per year. Here is the data that I collected:

Cost of Tuition			Number of Students		Total
	3000-9,999	10,000-16,999	17,000-23,999	24,000-30,999	
$3500-4500	1	5	0	0	6
$4501-5500	2	0	0	1	3
$5501-6500	1	1	3	1	6
Total	4	6	3	2	15

At .025, can we conclude that the cost of tuition is dependent on the number of students?

Ho: The cost of tuition is independent of the number of students that attend the college. $(x^2=0)$

H1: The cost of tuition is dependent on the number of students that attend the college. (claim) $(x^2>0)$

Step 2: Find the critical value:

The critical value is 14.449 since the degrees of freedom are (3-1)(4-1)=6.

Step 3: Compute the test value.

First we have to find the expected value:

$E1,1 = (6)(4)/15=1.6$ $E2,1 = (3)(4)/15=.8$ $E3,1 = (6)(4)/15=1.6$
$E1,2 = (6)(6)/15=2.4$ $E2,2 = (3)(6)/15=1.2$ $E3,2 = (6)(6)/15=2.4$
$E1,3 = (6)(3)/15=1.2$ $E2,3 = (3)(3)-15=.6$ $E3,3 = (6)(3)/15=1.2$
$E1,4 = (6)(2)/15=.8$ $E2,4 = (3)(2)/15=.4$ $E3,4 = (6)(2)/15=.8$

I did some spot checking and it looks good. I'm going to trust you on chi-square.

The completed table is shown:

Cost of Tuition			Number of Students		Total
	3000-9,999	10,000-16,999	17,000-23,999	24,000-30,999	
$3500-4500	1 (1.6)	5 (2.4)	0 (1.2)	0 (.8)	6
$4501-5500	2 (.8)	0 (1.2)	0 (.6)	1 (.4)	3
$5501-6500	1 (1.6)	1 (2.4)	3 (1.2)	1 (.8)	6
Total	4	6	3	2	15

Then the test value is $x^2 = \sum (O-E)^2/E$

$$= (1-1.6)^2/1.6 + (5-2.4)^2/2.4 + (0-1.2)^2/1.2 + (0-.8)^2/.8 + (2-.8)^2/.8 + (0-1.2)^2/1.2 + (0-.6)^2/.6 + (1-.4)^2/.4 + (1-1.6)^2/1.6 + (1-2.4)^2/2.4 + (3-1.2)^2/1.2 + (1-.8)^2/.8$$
$$= 13.333$$

Step 4: Make the decision to reject or not to reject the null hypothesis.

Do not reject the null hypothesis since 13.333 is less than 14.449.

Step 5: Summarize the results.

There is not enough evidence to support the claim that the cost of tuition is dependent on the number of students that attend the college.

SUMMARY

My first hypothesis test about the tuition cost of 4-year universities being less than the average was correct. The average as stated by "The College Board" said that the tuition was $5836 per year. I thought that was a little high. The average tuition of the fifteen colleges that I researched was $5069.73. Maybe if I would have researched colleges all around the country instead of just our surrounding states I would have come up with different numbers. Another thing that may have caused this test to be a little off was that when I was collecting data, some of the costs of tuition may include other fees and some may not. When I looked them up, some fees were listed separately and some were not. This could have lead to a Type I error where the null hypothesis was true and it was rejected. Good note on the possibility of a Type I error.

My second hypothesis test about whether the cost of tuition is dependant on the number of students that attend the college was rejected. I thought that the fewer the students that attend a specific college, that tuition would be cheaper, but that wasn't the case. One main problem I can see with colleting my data is that on the college websites for the number of students, some said "over" or "approximately". So, these weren't the exact numbers of students enrolled. Also, as stated earlier, some of the students could be undergraduates or graduates. Some of the websites didn't list them separately. Tuition is higher for graduates, so they should not have been included in this study and it would have thrown off the number of students. So, these may have affected the outcome a little, but I don't' think enough for it to change the hypothesis. Small differences, including rounding, can really throw off your test value in this type of test.

It would have also been interesting to test to see whether the tuition is higher in urban areas where more people live verses rural areas where there are not as

many people. I would be inclined to say that this is true, but it would need to be tested further to say for sure. It would also be interesting to do this same testing for private colleges to see if they have the same results. I thought this was fun to come up with our own hypothesis and try to prove ourselves right or wrong using what we have learned all quarter. It was a good test of our skills and it made me get a better understanding of how the formulas really work rather than just doing the homework examples in the book.

Excellent summary/critique and project overall. I've been teaching this class for over 8 years and your project is one of the best. Can I please use this as a sample?

Grading Rubric
Statistics Project, Spring 2007

	Objective Met (C, B-)	**Exceeded (B, B+)**	**Outstanding (A-, A)**	**Score**
Hypothesis and Proposal *10 points*	Statement of hypothesis in words or symbols, type of test you plan to use.	Statement of hypothesis in words and symbols. Description of test you'll use and how you plan to implement it.	Statement of hypothesis in words and symbols, including motivation. Type of hypothesis test and plans for implementation.	**10**
<u>**Written Project:**</u> **Data Analysis** *15 points*	Brief description the data set including outliers, gaps, and other observations.	Description and interpretation of data set (what and *why*).	Description and interpretation of data set. Discussion of potential problems including how data could be improved.	**15**
<u>**Written Project:**</u> **Hypothesis Testing and Confidence Intervals** *30 points*	Conduct a mathematically accurate hypothesis test.	Conduct two mathematically correct hypothesis tests or one test and one confidence interval	Conduct two (or more) mathematically correct hypothesis tests that allow you to draw meaningful conclusions.	**29**
<u>**Written Project:**</u> **Summary and Self-Critique** *25 points*	Re-iterate outcome of hypothesis testing. Identify areas of strength and weakness.	Explain outcome of hypothesis testing, discuss possibility of Type I or Type II errors. Explore areas of strength and weakness by making suggestions for improvement and proposing further research.	Interpret outcome of hypothesis testing and possibility of Type I or II errors. Explain errors. Draw connections between hypothesis tests, data collection. Discuss areas of strength and weakness.	**25**
Overall Subjective Component *10 points*	Project is mathematically and grammatically correct. Displays adequate understanding of hypothesis testing.	Project is accurate, interesting, and well-presented. Displays mastery of hypothesis testing.	Project is accurate, creative, and well-presented. Hypothesis testing done at mastery level with meaningful connections throughout and a thorough summary and recommendation (I was "wowed")	**10**
Draft and Edits *10 points*	Draft in by due date = 5 points	Draft submitted and constructive review of one other classmate's done by due dates = 10 points		**10**

Overall Score			**99**

Written by Angela Redmon. Last updated 07/12/06.