Math 108 Exam 3 Solutions Spring '00

1. An ecologist studying acid rain takes measurements of the pH in 12 randomly selected Adirondack lakes. The results are as follows:

   3.0 6.5 5.0 4.2 5.5 4.7 3.4 6.8 4.6 5.7 6.2 5.2

   Give 90% confidence intervals for
   a) the average
   b) the standard deviation
   of the pH of the Adirondack lakes.

   SOLUTION: We enter the data in the first column of a blank minitab worksheet, and calculate the descriptive statistics for that column. We find that the sample average is approximately 5.067 and the sample S is approximately 1.172 . Thus, the SE for the average is

   $$\text{SE} \approx \frac{1.172}{\sqrt{12}} \approx .3383$$

   Since the sample size is small, we use the t table (with 11 degrees of freedom) to calculate the radius of the 90% confidence interval for the average. The tail size associated to a 90% confidence interval is 5% in each tail, so the table tells us that the radius of the interval is 1.8SE $\approx .609$. Thus, the 90% confidence interval for the average is

   $$(5.067 - .609, 5.067 + .609) = (4.458, 5.676)$$

   For the standard deviation, we use the χ^2 table to find the confidence interval, using the formula $\chi^2 = (n-1)S^2/\sigma^2$. We have 11 degrees of freedom, so the 90% confidence interval is determined by cutting off the 5% tails on both sides of the χ^2 curve with 11 degrees of freedom. We get

   $$4.58 < \chi^2 < 19.68$$
   $$4.58 < \frac{(n-1)S^2}{\sigma^2} < 19.68$$
   $$4.58 < \frac{11 \cdot 1.172^2}{\sigma^2} < 19.68$$
   $$4.58 < \frac{15.11}{\sigma^2} < 19.68$$

$$\frac{1}{19.68} < \frac{\sigma^2}{15.11} < \frac{1}{4.58}$$
$$\frac{15.11}{19.68} < \sigma^2 < \frac{15.11}{4.58}$$
$$\sqrt{\frac{15.11}{19.68}} < \sigma < \sqrt{\frac{15.11}{4.58}}$$
$$.88 < \sigma < 1.82$$

So the 90% confidence interval for σ is $(.88, 1.82)$.

2. The data in the last problem was taken to test the hypothesis that the average pH for an Adirondack lake is 4.4. What are the values of t, df, and P? What do you conclude?

   SOLUTION: The null hypothesis is that the average should be 4.4, and the alternative hypothesis is that the average is more than that. As shown in Problem 1, the standard error is given by SE $\approx .3383$. Thus,

   $$t \approx \frac{5.067 - 4.4}{\text{SE}} \approx \frac{.667}{.3383} \approx 1.97$$

   There are 11 degrees of freedom. From the t-table, P is between 2.5% and 5%, so the test is significant. The average is likely to be more than 4.4.

3. Suppose that 90% of American families have cars and that a sociologist believes the percentage in a particular town is higher. She does a survey to test this. In a random sample of 550 families from the town, 92.5% of the families have cars. What are the values of z and P? What do you conclude?

   SOLUTION: The null hypothesis is given by the box $\boxed{9\boxed{1}\quad\boxed{0}}$. Thus, the expected average is .9 and the standard error is

   $$\text{SE} = \frac{(1-0)\sqrt{.9 \cdot .1}}{\sqrt{550}} \approx .0128$$

   so

   $$z = \frac{.925 - .9}{\text{SE}} \approx 1.95$$

   The normal table gives P $\approx 2.56\%$, so the test is significant. The town is likely different from the country as a whole with respect to car ownership.

4. A new treatment for high blood pressure is tested. The patients are randomly divided into two groups. One group gets the new treatment, and the other gets the old treatment. The results are as follows. Here $\bar{x}$ is the average systolic blood pressure for each group:

	n	$\bar{x}$	S
New treatment group	210	126	18
Old treatment group	200	130	20

What are the values of z and P? What do you conclude?

Solution: This time, we are comparing the averages of two different samples. The null hypothesis is that the averages should be the same, i.e., that the difference of the two averages should be 0. The standard error is given by the Pythagorean rule:

$$\text{SE} = \sqrt{\left(\frac{18}{\sqrt{210}}\right)^2 + \left(\frac{20}{\sqrt{200}}\right)^2} \approx 1.88$$

We have

$$z = \frac{(130 - 126) - 0}{\text{SE}} \approx 2.13$$

The normal table gives P $\approx$ 1.66%, so the test is significant. The new treatment is likely to reduce blood pressure more, on the average, than the old treatment does.

5. A die is tested for fairness. Here are the results:

number on die	1	2	3	4	5	6
observed freq.	50	37	33	48	27	45

What are χ^2, df, and P? What do you conclude?

Solution: The sample size is 240. The null hypothesis is that the die is fair, meaning that the expected frequency for each face should be $240/6 = 40$. Thus,

$$\begin{aligned}\chi^2 = {} & \frac{(50-40)^2}{40} + \frac{(37-40)^2}{40} + \frac{(33-40)^2}{40} + \frac{(48-40)^2}{40} + \frac{(27-40)^2}{40} \\ & + \frac{(45-40)^2}{40} \\ = {} & 10.4\end{aligned}$$

There are 6 categories of data, hence $6-1=5$ degrees of freedom. So the χ^2 table shows that P is between 5% and 10%. The test is inconclusive.

6. A botanist is crossing two strains of peas. Genetic theory says that 9/16 of the offspring should be in group A, 3/16 in group B, 3/16 in group C, and 1/16 in group D, provided the strains are pure.

   He presents data for 320 offspring: 184 are in group A, 58 in group B, 59 in group C, and 19 in group D.

   You run a χ^2 test to see if his data looks to have been fudged. What are the values of χ^2, df, and P? What do you conclude?

   SOLUTION:

	observed	expected
A	184	$9/16 \cdot 320 = 180$
B	58	$3/16 \cdot 320 = 60$
C	59	$3/16 \cdot 320 = 60$
D	19	$1/16 \cdot 320 = 20$

Thus,

$$\chi^2 = \frac{(184-180)^2}{180} + \frac{(58-60)^2}{60} + \frac{(59-60)^2}{60} + \frac{(19-20)^2}{20} \approx .22$$

There are 4 categories of data, hence $4-1=3$ degrees of freedom. The χ^2 table tells us P is between 1% and 5%, so the test is significant. The data is likely to be fudged.

7. A study is made to see if voting and gender are independent. 200 men and 300 women are polled, to find out if they voted in the last election. The data obtained were as follows.

	Men	Women
Voted	95	175
Didn't vote	105	125

What are the values of χ^2, df, and P? What do you conclude?

SOLUTION: We complete the table:

	Men	Women	Total
Voted	95	175	270
Didn't vote	105	125	230
Total	200	300	500

We can then calculate the expected frequencies under the assumption that the variables are independent:

	Men	Women
Voted	$200 \cdot 270/500 = 108$	$300 \cdot 270/500 = 162$
Didn't vote	$200 \cdot 230/500 = 92$	$300 \cdot 230/500 = 138$

Thus,

$$\chi^2 = \frac{(95-108)^2}{108} + \frac{(175-162)^2}{162} + \frac{(105-92)^2}{92} + \frac{(125-138)^2}{138} \approx 5.67$$

The number of degrees of freedom is the product of the numbers of degrees of freedom for the two variables: $\text{df} = (2-1)(2-1) = 1$. So P is between 1% and 5%. The test is signficant. The two variables are not likely to be independent.

8. A new procedure is developed to try to lower the standard deviation for a certain variable. Under the standard procedure, we have $\sigma = 50$. A test of the new procedure gives $S = 35$ with a sample size of 20. What are χ^2, df, and P? What do you conclude?

   SOLUTION: We use the formula for comparing σ and S:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{19 \cdot 35^2}{50^2} = 9.31$$

   There are $n-1 = 19$ degrees of freedom. The χ^2 table shows P is between 1% and 5%, so the test is significant. The new procedure lowers σ.