Review for Quiz 3 answers

1) In this study[1] mice were raised in a strong magnetic field to see if this affected their weight. There are 20 cages, each containing the same number of mice. 10 cages were placed in a strong magnetic field. The other 10 were not in a magnetic field. After 12 days the mice were weighed and the weight gains (in grams) were recorded (giving 20 numbers). The standard deviations of these samples was $S_M = 5.67$, $S_N = 3.18$ where M, N stand for magnetic and non-magnetic.

Are these standard deviations significantly different?

Use the F-distribution:

$$F_{9,9} = \frac{S_M^2}{S_N^2} = 3.18$$

The critical value is 4.026. Since $3.18 < 4.026$ there is no significant difference between S_M and S_N.

Assuming that the population standard deviations are the same ($\sigma_M = \sigma_N$), (This is worded like this so that, in case you do the first part of the problem incorrectly, you can still do the second part.) how would you proceed to determine whether the magnetic field affected the mice? (Calculate S_p^2, find the critical value of t and explain how you would get a conclusion if you were given the numbers $\overline{X}_M, \overline{X}_N$.)

The pooled variance is

$$S_p^2 = \frac{9S_M^2 + 9S_N^2}{18} = 21.13$$

Then, under the null hypothesis that $\mu_M = \mu_N$ we get

$$t_{18} = \frac{\overline{X_M} - \overline{X_N}}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{\overline{X_M} - \overline{X_N}}{2.055}$$

Since the critical value of t_{18} is $t_{18,.975} = 2.101$, the criterion is: Accept H_0 if

$$|\overline{X_M} - \overline{X_N}| < 4.319$$

The difference is significant only if it is at least 4.319 in absolute value.

2) Find a multiple of the sample variance S_2 (I probably meant S_M^2 or S_N^2 or S_p^2) which is χ^2_ν distributed. What is ν?

The theorem is that

$$\frac{9}{\sigma_M^2} S_M^2 \sim \chi^2_9 \sim \frac{9}{\sigma_N^2} S_N^2$$

and

$$\frac{18}{\sigma^2} S_p^2 \sim \chi^2_{18}$$

A company makes precision metal parts. To keep control of the production process samples are regularly taken.

[1] Data from *Introduction to Mathematical Statistics and Its Applications* by Larsen and Marx.

a) On a small sample of 10 metal parts the diameter of the hole drilled into each part has a sample mean of 2.25 mm and a sample variance of .0004 mm^2. Using the t-distribution and χ^2-distribution find the 99% confidence interval for

(1) the actual mean of the diameters and
(2) an upper bound for the standard error (σ) of the diameter.

The t-distribution is:

$$t_9 = \frac{\overline{X} - \mu}{S/\sqrt{10}} = \frac{2.25 - \mu}{.00632}$$

The critical value of t_9 is $t_{9,.995} = 3.25$. So, the 99% confidence interval for μ is

$$\mu = 2.25 \pm .021mm$$

or

$$2.229mm < \mu < 2.271mm$$

The χ^2 distribution is

$$\chi_9^2 \sim \frac{9S^2}{\sigma^2}$$

To get an 99% confidence upper bound for σ we need the left tail of χ_9^2:

$$\sigma^2 < \frac{9S^2}{\chi_{9,.01}^2} = \frac{9 * .0004}{2.0879} = 0.001724$$

Taking the square root we get

$$\sigma < 0.0415mm$$

with 99% confidence. Note that this number is not the same as 0.021

b) If we assume that the actual mean is 2.24 mm then find, with 99% confidence, an upper bound for the standard error.

This is a theoretical question which comes from the proof that $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$. The answer in general is:

$$\frac{(n-1)S^2}{\sigma^2} + \frac{n(\overline{X} - \mu)^2}{\sigma^2} \sim \chi_n^2$$

In this case:

$$\frac{9S^2 + 10(.01)^2}{\sigma^2} = \frac{.0046}{\sigma^2} \sim \chi_{10}^2$$

Using the left tail $\chi_{10,.01}^2 = 2.558$ we get

$$\sigma < 0.0424mm$$

with 99% confidence. This is larger than 0.0415 ! The reason might be that $\overline{X} - \mu = 0.01$ is more than one standard deviation of $\overline{X}$ which is approximately

$$S/\sqrt{10} = \sqrt{.00004} = 0.0063$$

For this problem, since I already gave you all the numerical answers, the only question is where did the numbers come from and what is the conclusion.

A	*B*	*C*	*D*	*E*	*F*	*G*
No of cells	Concen. 1	Observed frequency	$A*C$	Poisson	Expected	Chi squared
0	213	0.5325	0	0.505352032	202.1408127	0.583365364
1	128	0.32	0.32	0.344902762	137.9611047	0.71921435
2	37	0.0925	0.185	0.117698067	47.07922697	2.157869252
3	18	0.045	0.135			
4	3	0.0075	0.03			
5	1	0.0025	0.0125			
6	0	0	0			
7	0	0	0			
8	0	0	0			
9	0	0	0			
10	0	0	0			
11	0	0	0			
12	0	0	0			
> 2	22	0.055	0.1775	0.032047139	12.81885567	6.575736041
Total	400	1	0.6825	1	400	10.03618501

G) Finally, we do the chi-squared test. Note that the data for $x \geq 3$ is added together in the subtotal so that the expected numbers are at least 5. (The rule is that all E_i must be at least 5. If they are < 5 you are supposed to combine categories in a suitable way.) $df = 4 - 1 - 1 = 2$. Does the data support the hypothesis?

First of all the number of degrees of freedom is the number of categories (4) minus 1 (as always) minus the number of parameters estimated (1). There are 4 categories: 0,1,2 and ≥ 3.

The critical value is the right tail of χ^2_2 which is $\chi^2_{2,.95} = 5.99$ The test statistic is 10.036 which is much larger. (The p-value is 0.0066) So, we reject the null hypothesis. The survival rate of the bacteria is significantly different from the Poisson distribution. One possible explanation is that the bacteria clump together and protect each other. (The ones in the middle of a cluster would be protected.) Notice that the "at least 5" rule prevents small denominators which would give rise to large deviations in the χ^2 sum.

Question: What happens if we ignore the "at least 5" rule and make 5 categories: 0,1,2,3 and ≥ 4 ??