Math 10B, Quiz 8 Solutions

1. (12 points) Suppose you roll two 4-sided dice and each time you record the sum of the two rolls. You repeat this 49 times and obtain the following data:

Value	Observed frequency
2	7
3	7
4	7
5	7
6	7
7	7
8	7
Total	49

Perform a χ^2 test on the hypothesis that both dice are fair.

Solution: First we need to calculate the expected frequencies given the null hypothesis is true. To do this, we calculate the probability of each outcome occurring in a single trial and then multiply by the total number of trials to get the corresponding frequency. The probability that two fair four-sided dice sum to 2 is 1/16: there are 16 possible outcomes when your roll two four-sided dice and in exactly one of those outcomes (both dice are 1) the sum is 2. Likewise the probability of getting a sum of 3 is 2/16 since there are two ways for this to happen (first die is 1 and second die is 2 or first die is 2 and second die is 1). Continuing with this reasoning gives us the following chart of expected frequencies:

Value	Observed frequency	Probability in a single trial (assuming null hypothesis)	Expected frequency (assuming null hypothesis)
2	7	1/16	$(1/16)\cdot 49$
3	7	2/16	$(2/16)\cdot 49$
4	7	3/16	$(3/16)\cdot 49$
5	7	4/16	$(4/16)\cdot 49$
6	7	3/16	$(3/16)\cdot 49$
7	7	2/16	$(2/16)\cdot 49$
8	7	1/16	$(1/16)\cdot 49$
Total	49	1	49

Now to find the value of the χ^2 statistic on this data, we calculate $(\text{expected} - \text{observed})^2/\text{expected}$ for each outcome and take the sum. In other words, the value of the χ^2 statistic for this data is

$$\frac{(49/16-7)^2}{49/16}+\frac{(49\cdot 2/16-7)^2}{49\cdot 2/16}+\frac{(49\cdot 3/16-7)^2}{49\cdot 3/16}+\frac{(49\cdot 4/16-7)^2}{49\cdot 4/16}+\frac{(49\cdot 3/16-7)^2}{49\cdot 3/16}$$
$$+\frac{(49\cdot 2/16-7)^2}{49\cdot 3/16}+\frac{(49/16-7)^2}{49/16}=41/3\approx 13.7.$$

Since this is a χ^2 goodness of fit test (rather than a test for independence), the degrees of freedom is one less than the number of outcomes. In this case there are 7 outcomes, so the degrees of freedom is 6. With 6 degrees of freedom, a p-value of 0.05 corresponds to a χ^2 statistic of 12.59. Since our χ^2 value is greater than that, **we have enough evidence to reject the null hypothesis at the 5% significance level.**

2. (1 point) Suppose you perform some experiment 5 times and collect the following results: 4, 8, -2, 2, 3. Then the sample mean is 3 and the sample variance is 13.

√ **True** ◯ False

Solution: Recall that the sample mean is literally the average of the samples and the sample variance is

$$\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \widehat{\mu})^2$$

where n is the number of samples, $X_1, \ldots, X_n$ are the samples and $\widehat{\mu}$ is the sample mean.

3. (1 point) Suppose you have perform a χ^2 test on same data and get a χ^2 value of 100 with 9 degrees of freedom. You do not have enough evidence to reject the null hypothesis at the 5% significance level.

   ○ True √ **False**

Solution: When there are 9 degrees of freedom, the χ^2 statistic just needs to be greater than 16.92 to reject the null hypothesis. (Remember that when the null hypothesis is true, you are unlikely to see large values for the χ^2 statistic. So when the χ^2 statistic is large enough, it is considered sufficient evidence to believe the null hypothesis is not true.)

4. (1 point) A student performs a χ^2 test for independence for the random variables X and Y on the following data:

	X = 0	X = 1
Y = 0	300	100
Y = 1	200	400

   The student claims that the degrees of freedom is 3 since there are 4 possible outcomes and the degrees of freedom is always number of outcomes − 1. The student's answer is:

   ○ Too low
   ○ Correct
   √ **Too high**

Solution: The degrees of freedom is only one less than the number of outcomes when you are doing a χ^2 test for goodness of fit (i.e. checking that some distribution matches the one implied by the null hypothesis). But here we are doing a χ^2 test for *independence* in which case the degrees of freedom is (number of outcomes for first variable − 1)(number of outcomes for second variable − 1) which here gives us $(2-1)(2-1) = 1$. So the student's answer is too high.