# Stat {4601, 6872} Quiz #1 (50 minutes)

- Open book and open note.
- Use a simple calculator if necessary.
- Show your work. (E.g. Next to each numerical answer, you need to put things like $P_{0.01}(B \geq 3)$ or something equivalent)
- Grad students need to do problems with * as well. Undergrad will get extra credit by solving them.

#1. (7 points)
For patients suffering from arthritis, a traditional pain relief treatment exists but has a side effect of problems in the kidney and resulting in 5% risk of heart attacks and strokes after 3 months of treatment. A new medical treatment is proposed that reduces the chances of the kidney problem. But a question has been raised that the new drug may increase the risk of heart attacks and strokes. To test this, a clinical trial data has been collected for 10 patients and it turned out that 1 of them had the problem.

1) What assumptions does one need to check before applying the binomial test? Use the plain language and terms of the subject field that is understandable to general audience. (1 pt)

2) Formulate the null and alternative hypotheses. (1 pt)

3) Give a general definition of the p-value. Then compute the p-value for the above data. (1 pt)

4) What is the significance level of a test that rejects the null hypothesis when the number of patients experiencing a heart attack is greater than or equal to 3? (1 pt)

5) What is the power of the test in 4) at the alternative 'the true chance of a patient having a heart attack or stroke is 10%'? What is the probability of Type II error in such case? (1 pt)

6) A more extensive clinical study has been performed to yield the new data, which consists of 100 patients and 10 cases of heart-attacks and strokes within 3 months. Compute the p-value for the hypotheses same as 2). (1 pt)

7)* Compare the P-value obtained in 6) and 3) and discuss the result. Advocate the 'large sample size' from the results. (Graduate; 1 pt)

#2. (4 points) You are designing a survey asking CSUH students how many of them support change of the name of CSUH to 'CSU East Bay.'

1) You want to make sure you get the true proportion of those who support the name change within 0.05 with 90% confidence. What's the minimum number of responses you need? (1 pt)

2)* you have a strong suspicion that the true proportion of name change supporters is close to 10%. How would you change the answer to the above question given that suspicion? (1 pt)

3) It occurred to you that the total student population may not be large enough to apply the binomial test etc. Which assumption(s) for binomial test is violated? (1 pt)

4) What is the alternative model you can use to model the data? (You won't find it in the textbook. I mentioned it in the class.) (1 pt)

## *Quiz #1 Answers*

#1.
1) Each patient has either heart attack or not after 3 months, patients are independent, the chance of heart attack is constant.
2) Let p be the chance of a patient having a heart attack or stroke. $H_0$: p=0.05, $H_1$: p>0.05
3) P-value is the smallest significance level at which one can reject the null hypothesis. Its value for the current test is $P_{0.05}(B>=1)$ =0.4013
4) $P_{0.05}(B>=3)$ = 0.0115
5) $P_{0.10}(B>=3)$ = 0.0702. P(Type II error) = 1-0.0702 = 0.9298
6) B*= (B-np)/sqrt(np(1-p)) = 2.294, P-value = P(Z>B*) = 0.01089073
7) Both data have the same sample proportion of the heart attack or stroke occurrences but the p-values differ. The p-value is smaller, the result is more significant, and we have more confidence about our conclusion when the sample size is large. That's the point of trying to collect as large samples as possible.

#1.
1) $n=(z_{\alpha/2})^2/4D^2$ = 270.5543. At least 271 samples.
2) $n = (z_{\alpha/2})^2 p(1-p)/D^2$ = 24.34989 = 25
3) Assumptions about independence and constant p are violated.
4) Use hypergeometric distribution