

Quiz Day 20, MAT 128, SPRING 2019

NAME:

You are allowed one 8 1/2" x 11" sheet of paper with hand-written notes on both sides.

- 1) In Lab 10, you wrote code to sample from a dataframe called `movies`, which contained the 1000 highest rated movies on IMDb, and plot a histogram of the mean star ratings of those samples. Code similar to the code from the lab is below:

```
means = []
for i in range(300):
    sample = movies.sample(40)
    means.append(sample["star_rating"].mean())
pd.Series(means).hist()
```

- (a) What is the sample size in the above code?
- (b) How many samples are simulated in the above code?
- (c) Once a mean is computed, where is it stored in the above code?
- (d) What does the `pd.Series()` function do in the last line of code?
- 2) The histograms of two samples of size 1000 from two distributions are shown below. Do you think the distributions are the same? Why or why not?

