Stat 225 - Final Review

The final exam will be next Monday, December 10. This review sheet consists of problems off of our previous three exams together a couple problems on new material

Exam 1

3. I've got an unfair coin that comes up heads 70% of the time. Let X denote the random variable obtained by flipping the coin once then writing down a 1 if the result is a head and a 2 if the result is a tail. We then let $\{X_i\}_{i=1}^{10000}$ denote a sequence of 10000 independent trials of X and let

$$S = \sum_{i=1}^{10000} X_i.$$

   (a) Show that $E(X) = 0.7$.

   (b) Show that $\sigma(X) \approx 0.458$.

   (c) Find $E(S)$ and $\sigma(S)$.

   (d) Use a normal distribution to estimate $P(S < 7070)$.

   (e) Express $P(S < 7070)$ *exactly* using a sum and binomial coefficients.

4. I've got a twelve-sided die with

   - 6 sides labeled 1

   - 3 sides labeled 2

   - 2 sides labeled 3

   - 1 sides labeled 6

   (a) Show that the expected roll is 2.

   (b) Show that the standard deviation of one roll is $\sqrt{2}$.

   (c) What are the expectation and standard deviation of 100 rolls?

   (d) Use a normal distribution to estimate $P(S < 220)$.

6. Suppose that X is normally distributed with mean 65 and standard deviation 8. Find $P(62 < X < 70)$.

7. Suppose my classes exam scores are normally distributed with a mean of 60 and a standard deviation of 15. What percentage of my students score above 90%?

8. Use a u-substitution to translate the normal integral

$$\frac{1}{\sqrt{200\pi}} \int_{100}^{115} e^{-(x-110)^2/200} dx$$

into a *standard* normal integral.

Exam 2

2. A sample of 123 North Carolina men finds their average weight to be 190 pounds with a standard deviation of 36.75 pounds.

   (a) What is the standard error associated with this sample mean?

   (b) Write down a 96% confidence interval for the weight of a North Carolina man based on this sample.

3. A sample of 12 North Carolinians finds that 5 of them have smoked a significant amount.

   (a) What is the standard error associated with this sample proportion?

   (b) Write down a 95% confidence interval for the height of a North Carolina adult based on this sample.

4. Recall that a sample of 123 North Carolina men finds their average weight to be 190 pounds with a standard deviation of 36.75 pounds. Suppose we would like to use this data to explore the following question:
   Is the average North Carolina man 180 pound or is the average North Carolina man more than 180 pounds.

   (a) Write down the hypothesis test that clearly states the problem.

   (b) Compute the test statistic obtained from this data.

   (c) Compute the p-value obtained from this data.

   (d) What is the conclusion of the hypothesis test?

5. Recall that a sample of 12 North Carolinians found that 5 of them have smoked a significant amount. Suppose we would like to use that data to explore the following question:
   Have 3/4 of North Carolinians smoked a significant amount.

   (a) Write down the hypothesis test that clearly states the problem.

   (b) Compute the test statistic obtained from this data.

   (c) What is the critical cutoff for this problem?

   (d) What is the conclusion of the hypothesis test?

Exam 3

1. In my lifetime, Ohio State and Michigan have played each other 14 times, when both are ranked in the top 10. If, for each of those 14 years, we compute Ohio State's score minus

Michigan's score we get a data set with mean -1.57 and standard deviation 9.43. Using this data, can we reject the null hypothesis that the average difference in scores is zero?

Be sure to

(a) Write down the hypothesis test.

(b) Compute the standard error for the problem and test statistic.

(c) Use a normal table or t-table as appropriate to assess the hypothesis test

2. A recent study surveyed people's attitudes on the following question: Is it or is it not acceptable for social media companies to use users data to send them specific political ads. The results, divided by age category, are summarized in the table below.

	Over 55	Under 55
Acceptable	82	128
Not acceptable	41	23

Perform a hypothesis test to explore the question - do people under 55 genuinely find it acceptable for social media companies to use user's data to send them specific political ads more often than people over 55. Be sure

(a) Compute the observed proportions and their difference,

(b) Compute the standard error and statistic, and

(c) State the conclusion of the hypothesis test with a clear explanation of why.

3. UNCA has six full professors in the math department who make an average of $89,000 a year with a standard deviation of $9400. UNCA also has 3 full professors in the chemistry department who make an average of $105,000 a year with a standard deviation of $24,000. Let's use this sparse data to explore the question of whether math and chemistry professors receive equal pay to a 90% level of confidence. Specifically, viewing these data as samples from a random process,

(a) State the null and alternative hypotheses,

(b) The difference between the two means,

(c) The associated standard error and test statistic.

(d) From the computations, can we conclude with a 90% level of confidence that there is a difference in pay between the two groups?

5. Suppose that I have paired data X and Y that are well correlated with $r = -0.81$, $\bar{x} = 8$, $\bar{y} = 12$, $\sigma_x = 2$, and $\sigma_y = 3$. Find the regression line for this data.

8. I run a regression analysis using the following Python commands:

```
import pandas as pd
from scipy.stats import linregress
df = pd.read_csv('https://marksmath.org/visualization/CFB_Stats/CFB2014.csv')
linregress(df.win_loss_percentage, df.total_points)
```

I obtain the following result:

```
LinregressResult(slope=426.9314687857958, intercept=152.7186211636847,
rvalue=0.8350713150741875, pvalue=1.688515353211888e-34, stderr=25.056701569182948)
```

(a) The linear regression is attempting to understand the relationship between what two variables?

(b) What is a formula for the regression line?

(c) State the null and alternative hypotheses for this regression test.

(d) What is the conclusion of the regression test?

More problems

1. An ultimate frisbee player performed an experiment with three different grips to see what effect it might have on the distance the resulting throw. He tried it with his *normal grip*, with *one finger out*, and with the *frisbee inverted*. The summary and ANOVA tables for the test are shown below.

	Finger out	Inverted grip	Normal grip
$\bar{x}_i$	29	33	34
σ_i	5.5	3.4	5

	Df	Sum Sq	Mean Sq	F value	Pr($>$F)
Grip	2	58.5833	29.2917	2.0453	0.1543
Residuals	21	300.75	14.321		

Conduct a hypothesis test to determine if these data provide convincing evidence that the average score varies across some (or all) groups. Be sure to state both the hypothesis test and the conclusion of the hypothesis test.

2. Let us suppose that four teams (OSU, UM, NU, and MSU) play each other 4 times each resulting in the game graph shown in figure 1. They then play the championship tournament whose results are shown below. We'd like to use the maximum likelihood technique to come up with a final ranking for the teams.

OSU	59
MSU	0

OSU	59
UM	0

NU	10
UM	21
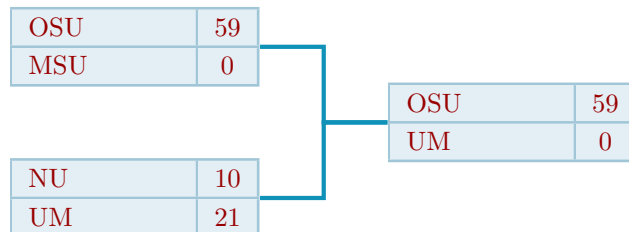
(a) Translate the directed game graph to an adjacency matrix.

(b) Adjust the adjacency matrix from part (a) to account for the results of the tournament.

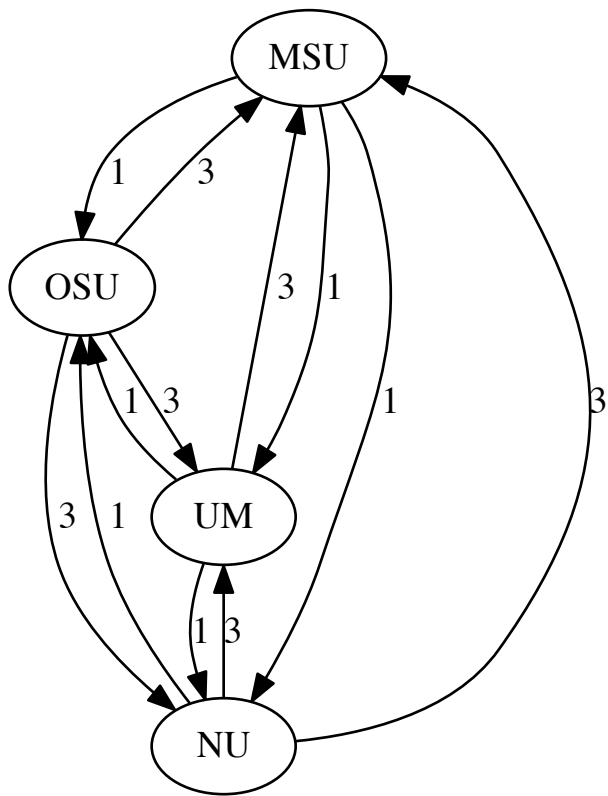(c) Write out the equations that you would need to solve in order to find the Maximum Likelihood ranking of the teams.

Figure 1: A game graph