Oh boy, where to begin. We covered **alot** of material since the last exam. Certainly enough that this test won't be comprehensive, but instead focus on the following topics:

1. Confidence intervals for μ. The key formula

$$\bar{x} \pm a\mathsf{SE}(\bar{x}), \quad \text{where} \quad \mathsf{SE}(\bar{x}) = s/\sqrt{n}$$

   The value a comes from the t distribution with $n-1$ degrees of freedom.

2. Confidence interval for $p_1 - p_2$. The key formula is

$$(\widehat{p}_1 - \widehat{p}_2) \pm a\mathsf{SE}(\widehat{p}_1 - \widehat{p}_2), \qquad \text{where} \quad \mathsf{SE}(\widehat{p}_1 - \widehat{p}_2) = \sqrt{\widehat{p}_1(1-\widehat{p}_1)/n_1 + \widehat{p}_2(1-\widehat{p}_2)/n_2}$$

3. Signficance tests. We studied 4 different tests of the type

$$T = \frac{\text{observed} - \text{expected}}{\mathsf{SE}}:$$

$$\begin{aligned}
H_0: \widehat{p} = p_0 & \qquad & Z = (\widehat{p} - p_0)/\sqrt{p_0(1-p_0)/n} \\
H_0: \mu = \mu_0 & \qquad & T = (\bar{x} - \mu)/(s/\sqrt{n}) \\
H_0: \widehat{p}_1 = \widehat{p}_2 & \qquad & Z = (\widehat{p}_1 - \widehat{p}_2)/\sqrt{\widehat{p}(1-\widehat{p})(1/n_1 + 1/n_2)} \\
H_0: \mu_1 = \mu_2 & \qquad & T = (\bar{x}_1 - \bar{x}_2)/\sqrt{s_1^2/n_1 + s_2^2/n_2}
\end{aligned}$$

4. Chi-squared tests. We also did two tests which use the χ^2 statistic. The first tests to see if a set of categorical values comes from a specified set of probabilities. For example M&Ms where we considered 3 different specifications for the probability of brown, blue, yellow, etc. For this, the chi-squared statistic has $k-1$ degrees of freedom (not $n-1$).

   The other test is a test of homogeneity. This checks if two or more distributions have the same probabilities describing it. The example we did in class was is there a gender difference in junk food preference? For this test, we use the expected number in cell ij is the sum of the ith row times the sum of the jth column divided by n the sample size. The chi-squared statistic in this case has $(n_r - 1)(n_c - 1)$ degrees of freedom (not $k-1$).

Some sample problems follow:

1. Is coin tossing truly random? Not according to a 2004 article in *Science Times*. The article claims that a coin is more likely to land on the same face it started on. To test, suppose a coin is tossed 1,000 times, each time starting on heads. A summary of the data is

n	no.heads
1000	512

Is there statistically significant evidence that the proportion of heads is more than $p = 0.5$?

2. Have gas prices gone up from last month? Suppose it was well known that average gas prices last month were \$3.05. This month a random sample of 10 gas stations found a mean of \$3.10 with a standard deviation of \$0.08. Is the difference statistically significant at the $\alpha = 0.05$ level?

3. According to a new report from Israel, taking aspirin up to the day of coronary bypass grafting, as opposed to stopping one week prior to surgery, seems to speed lung function recovery after the surgery without increasing the risk of bleeding significantly.

   Suppose the data summarizing the study are

Group	n	xbar	s
took aspirin	14	5	3
stopped aspirin	18	7	2.5

   The variable x records the time spent on a ventilator after surgery.

   Is there statistically significant evidence that the mean time spent on the ventilator has decreased?

4. Has support for the Iraq War among Americans declined since 2005? The data from US Today surveys is

date	n	x
2005	492	246
2006	514	180

   (The number of people who "*think the war was worth waging*" is recorded in x, the number questioned is n.)

   Find a 95% confidence interval for the unknown change in proportion $p_{2005} - p_{2006}$.

5. Is Weight Watchers a better diet than the Atkins diet for long-term weight loss? According to a May 2005 *Consumer Reports* this is the case. Suppose the data used to support the claim came from a study of two groups of dieters for one year. The amount of weight lost over that one-year period is measured, and summarized below.

Group	n	xbar	s
Weight Watchers	25	5.4	3.5
Atkins Diet	20	4.2	5.0

Is the difference in weight loss statistically significant at the $\alpha = 0.05$ level?

6. Boy do I love skittles, even more so than M&Ms. We saw that the distribution of colors for Milk Chocolate M&Ms varies with color. Is it the same for Skittles?

   Suppose a 16-ounce bag of 400 skittles had the following color distribution. Is the data consistent with the hypothesis that all five colors are equally likely, or is there a statistically significant difference between the proportions?

Color	Flavor	no.	percent
Purple	Grape	65	16.2%
Orange	Orange	81	20.2%
Red	Strawberry	75	18.8%
Green	Lime	87	21.8%
Yellow	Lemon	92	23.0%
		400	100.0%

7. Is there a difference in color distribution of M&Ms between milk chocolate and Peanut? Two bags were opened and sorted by color. The following distributions were found.

Type	Brown	Yellow	Red	Blue	Orange	Green
Milk Chocolate	9	10	20	22	22	16
Peanut	12	15	12	23	23	15

Is there a statistically significant difference between the two distributions?