Name: ______________________________ **Instructor:** David Galvin

Math 20340: Probability and Statistics
Fall Semester 2008
Final Exam
December 16, 4.15pm-6.15pm

This examination contains 10 problems on 11 sheets of paper (including the front cover). It is a closed-book exam. For statistical tables and important formulae, refer to the separate handout. *Show all your work on the paper provided.* Remember that the honor code is in effect for this exam.

Scores

Question	Possible	Actual
1	10	
2	10	
3	10	
4	10	
5	10	
6	10	
7	10	
8	10	
9	10	
10	10	
Total	100	

GOOD LUCK !!!

1. Adam Vinatieri is known to make 60% of his long (more than 55 yard) field goal attempts. During pre-game warmup next weekend, he plans to attempt 8 long field goals. Assuming that the success/failure of long field goal attempts are independent of each other, answer the following questions:

1. What's the expected number of pre-game warmup field goal attempts that Vinatieri will successfully make?

2. What's the probability that he successfully makes 3 or fewer pre-game warmup field goal attempts?

3. A South Bend Tribune journalist claims that Vinatieri has injured his kicking foot, and will not kick long field goals as well this weekend as he usually does. Suppose he kicks 3 of his 8 attempts successfully. Do you think that this provides strong evidence in favour of the journalist's claim? (This does not require calculation.)

2. John hands you three dice, and tells you that

- one of the dice (dice A) is fair (each number comes up with probability $1/6$)
- one (dice B) is loaded towards 6 (6 comes up with probability $1/3$, the remaining numbers are all equally likely) and
- one (dice C) is loaded against 2 (2 comes up with probability $1/11$, the remaining numbers are all equally likely)

You pick one of the dice at random and roll it.

1. What is the probability that you roll a 6?

2. Given that you roll a 6, what is the probability that the dice you picked was the fair dice (dice A)?

3. I have lots of nearly empty cereal boxes in my cupboard. Each one has a amount of cereal in it that is normally distributed with mean 2 ounces, standard deviation .4 ounces. Suppose I choose 5 boxes at random. What is the probability that between the 5 boxes I have at least enough cereal to fill an 8 ounce bowl?

4. It is always sunny in tropical paradise, but the daytime high temperature does vary from day to day. It is normally distributed with average value 78 degrees and standard deviation 8 degrees. You are planning a vacation to tropical paradise, and want to be able to tell your friends that you are 95% sure that the average daytime high temperature for the duration of your visit will be between 74 and 82 degrees.

How long a vacation should you plan?

5. A manufacturer of gunpowder has developed a new powder, which was tested in eight shells. The resulting muzzle velocities, in feet per second, were as follows:

3005	2990	3010	2990
3000	2995	2979	2991

1. Find a 95% confidence interval for the true average velocity μ for shells of this type. State any assumptions that you make.

2. Based on your result to the previous part, would you accept or reject the hypothesis that $\mu = 3015$ (at 5% significance)?

6. The scouting report for high-school basketball player John McJake says that he makes 80% of his free throws, on average. Mike Bray suspects that this is an exaggerated claim, and wants to test it. He watches McJake for a few games one weekend, and observes him making 22 of 32 free throws. Assuming Bray has seen a representative sample of McJake's free throws, is there enough evidence to suggest that McJake's true free throw rate is below 80%?

7. I suspect that students in different years have different levels of politically activity. In a recent *Observer* survey, 32 of 128 randomly selected first-years said that they either attended a political meeting on campus this semester, or actively campaigned for a presidential candidate, while 33 of 100 randomly selected seniors said that they did. Does this provide sufficient evidence, at 5% significance, that there is a difference between the levels of political activity of seniors and first-years?

8. Do women live longer than men? Here's data on the life expectancy (at birth) of 7 randomly selected UN countries, broken down by men and women (the data is from the CIA World Factbook, 2008):

	Yemen	Tanzania	Tonga	Syria	Brunei	Burundi	Andorra
Men	60.61	49.41	67.6	69.27	73.12	50.48	80.35
Women	64.54	52.04	72.76	72.02	77.59	52.12	85.14

Do a paired-difference test on this data to see if it provides evidence for women having a greater life expectancy than men.

9. In a study of the amount of calcium in drinking water undertaken as part of water-quality assessment, the same sample was tested in the laboratory six times at random intervals. The six readings (in parts per million) were

$$9.5 \quad 9.6 \quad 9.3 \quad 9.5 \quad 9.7 \quad 9.4$$

1. Estimate σ^2, the variance for readings on this sample using this particular test, using 90% confidence interval.

2. Based on your answer to part a) would you accept (at 10% significance) the claim that the standard deviation for readings on this sample using this particular test is 2.5?

10. 1. What does it mean to say that two events are *independent*?

2. A certain random variable takes on values $-1, 0, 1$ with probabilities $\frac{1}{4}, \frac{1}{4}, \frac{1}{2}$ respectively. Calculate the expected value and variance of this random variable.

3. What is a Type I error in a hypothesis test?

4. A certain null hypothesis is accepted at 2% significance. With the same data, will it be accepted at 5% significance?

5. John constructs a 90% confidence interval for a certain parameter. Mary uses the same data to construct a 95% confidence interval for the same parameter. Whose interval is shorter?

Some (possibly) useful formulae

- **Binomial distribution**
  - n **trials, probability** p **of success**:
  $$P(X = k) = C_k^n p^k q^{n-k}, k = 0, 1, \ldots, n$$
  where $C_k^n = \frac{n!}{k!(n-k)!}$
- **Central Limit Theorem**
  - **Version** 1: If $x_1, \ldots, x_n$ is a random sample from a population with mean μ and standard deviation σ, then for large enough n the distribution of $\bar{x}$ is approximately normal with mean μ, standard deviation $\sigma/\sqrt{n}$.
  - **Version** 2: If $x_1, \ldots, x_n$ is a random sample from a population with mean μ and standard deviation σ, then for large enough n the distribution of $\sum_{i=1}^n x_i$ is approximately normal with mean $n\mu$, standard deviation $\sqrt{n}\sigma$.
- **Sampling from a general population**
  - **Sample mean of a sample of size** n: $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$
  - **Sample standard deviation of sample of size** n:
  $$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$
  - **Mean and standard deviation of** $\bar{x}$: μ and $\frac{\sigma}{\sqrt{n}}$
  - **Distribution of** $\bar{x}$ ($n \geq 30$):
  $$\frac{\bar{x} - \mu}{s/\sqrt{n}} \approx z, \text{ a standard normal}$$
  - **Distribution of** $\bar{x}$ **(population normal)**:
  $$\frac{\bar{x} - \mu}{s/\sqrt{n}} = t, \text{ a } t \text{ distribution with } n-1 \text{ degrees of freedom}$$
  - **Distribution of** $\bar{x}_1 - \bar{x}_2$ ($n_1, n_2 \geq 30$):
  $$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \approx z$$

- **Pooled estimator for s^2, the common variance:**

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

- **Distribution of $\bar{x}_1 - \bar{x}_2$ (populations normal, variances equal):**

$$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = t \text{ with } n_1 + n_2 - 2 \text{ degrees of freedom}$$

  where s^2 is the pooled estimator for variance

- **Distribution of s^2 (population normal):**

$$\frac{(n-1)s^2}{\sigma^2} = \chi^2, \text{ a } \chi^2 \text{ distribution with } n-1 \text{ degrees of freedom}$$

- **Drawing samples from a binomial population**
  - **Sample proportion of a sample of size n:** $\hat{x} = \frac{\text{Number of successes}}{n}$
  - **Mean and standard deviation of $\hat{p}$:** p and $\sqrt{\frac{pq}{n}}$
  - **Distribution of $\hat{p}$ ($n\hat{p}, n\hat{q} > 5$):**

$$\frac{\hat{p}-p}{\sqrt{pq/n}} \approx \frac{\hat{p}-p}{\sqrt{\hat{p}\hat{q}/n}} \approx z$$

  - **Distribution of $\hat{p}_1 - \hat{p}_2$ ($n_1\hat{p}_1, n_1\hat{q}_1, n_2\hat{p}_2, n_2\hat{q}_2 > 5$):**

$$\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}} \approx z$$

  - **Pooled estimator for p, the common proportion:**

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

  - **Distribution of $\hat{p}_1 - \hat{p}_2$ ($n_1\hat{p}_1, n_1\hat{q}_1, n_2\hat{p}_2, n_2\hat{q}_2 > 5$, proportions equal):**

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}} \approx z$$

  where $\hat{p}$ is pooled estimator for variance