# Math 20340 — Statistics for Life Sciences

Fall 2009 final exam

December 16, 2009, 4.15pm-6.15pm

**Name:** ______________________________ **Instructor:** David Calvin

This examination contains 7 problems on 8 pages (including the front cover). The exam also comes with tables for binomial, standard normal, $t$, chi-squared, and $F$ distributions, and a table of useful formulae. It is closed-book, but you may use up to four single sided pages of handwritten notes. You may use a calculator. **Show all your work** on the paper provided. The honor code is in effect for this examination.

**Scores**

| Question | Score | Out of |
|---|---|---|
| 1 | | 10 |
| 2 | | 10 |
| 3 | | 10 |
| 4 | | 10 |
| 5 | | 10 |
| 6 | | 10 |
| 7 | | 10 |
| Total | | 70 |

**GOOD LUCK !!!**

1. An auto insurance company believes the following, based on historical data: in a given year, the probability that a client makes no claim is .8, and if a client makes a claim, then the probability that it is for 20% of the value of the insured car is .6, the probability that it is for 50% of the value is .3, and the probability that it is for 100% of the value is .1. A client is chosen at random. Her car is valued at $12,000.

   (a) Let $X$ be the amount that the client claims from the insurance company in the year. What down the possible values for $X$, and the probability that it takes each of these values.

   (b) Compute the expected value of $X$.

   (c) What is the probability that the client will make a claim for an amount that is greater than the expected value calculated in part (b)?

2. Adam Vinatieri is known to make 60% of his long (more than 55 yard) field goal attempts. During pre-game warmup on Thursday night, 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:

   (a) What's the expected number of long attempts that Vinatieri will successfully make?

   (b) What's the probability that he successfully makes 3 or fewer long attempts?

   (c) What's the probability that the number of long attempts he successfully makes is within one standard deviation of the expected number?

3. (a) Find a number $z > 0$ such that the probability that a standard normal takes a value between $-1$ and $z$ is .5.

(b) When Robert Hughes rushes, the number of yards he gains is very close to being normally distributed with mean 4 and standard deviation 2. Suppose Hughes makes three rushing plays in a row. What is the probability that his total gain on the three plays is at least 10 yards?

(c) The response time to a certain stimulus has mean 1.2 seconds with variance .04. Forty randomly selected subjects are tested. What is the probability that the average response time of the forty is less than 1.15?

4. 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 attended a political meeting on campus this semester, while 33 of 100 randomly selected seniors said that they did. Let $p_f$ be the proportion of first-years who attended a political meeting on campus this semester, and $p_s$ the proportion of seniors. Does the data provide sufficient evidence, at 5% significance, that $p_f \neq p_s$? Compute the $p$-value for this test.

5. 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 | 9.6 | 9.3 | 9.5 | 9.7 | 9.4 |

(a) Estimate $\sigma^2$, the variance for readings on this sample using this particular test, using a 90% confidence interval.

(b) What assumption(s) do you have to make about the distribution of the readings to make your work in part (a) valid?

6. I want to estimate the average length of time it takes for 1mg of a certain drug to clear out of the body's system. I have a fair idea, before I start my research, that the typical range of times for most people is from 8 hours to 24 hours.

   (a) About how many people will I have to sample, in order to be 98% confident that I have estimated the average to within $\pm 1$ hour?

   (b) The point of my sampling is to perform the hypothesis test $H_0 : \mu = 16$ against $H_1 : \mu \neq 16$ at 2% significance. If the true mean is $\mu = 18$ and I sample 100 people, what is the probability that I will accept $H_0$?

7. (a) What is a Type I error in a hypothesis test?

(b) What is a Type II error in a hypothesis test?

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

(d) 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?

(e) When would you use the pooled estimator for variance?

# Some (possibly) useful formulae

- **Drawing samples 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 $n$:**
$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$
  - **SE of $\bar{x}$ (sample of size $n$):**
$$SE = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$$
  - **SE of $\bar{x_1} - \bar{x_2}$ (sample of size $n_1$ from population $1$, $n_2$ from population $2$):**
$$SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \approx \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$
- **Drawing samples from a binomial population**
  - **Sample proportion of a sample of size $n$:**
$$\hat{x} = \frac{\text{Number of successes}}{n}$$
  - **SE of $\hat{p}$ (sample of size $n$):**
$$SE = \sqrt{\frac{pq}{n}} \approx \sqrt{\frac{\hat{p}\hat{q}}{n}}$$
  - **SE of $\hat{p_1} - \hat{p_2}$ (sample of size $n_1$ from population $1$, $n_2$ from population $2$):**
$$SE = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \approx \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

- **Large sample hypothesis testing for a binomial population**
  - **Pooled estimator for $p$, the common proportion**:
$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$
- **Small sample hypothesis testing for mean of a normal population**
  - **Test statistic**:
$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \text{ is a } t \text{ distribution with } n - 1 \text{ degrees of freedom}$$
- **Small sample hypothesis testing for difference of means of two normal population**
  - **Test statistic**:
$$\frac{\bar{x_1} - \bar{x_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} \text{ is a } t \text{ distribution with } df = n_1 + n_2 - 2$$
  - **Pooled estimator for $\sigma^2$, the common variance**:
$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$
- **Small sample hypothesis testing for variances of normal populations**
  - **Test statistic for variance of a single population**:
$$\frac{(n-1)s^2}{\sigma^2} \text{ is a chi-squared distribution with } df = n - 1$$
  - **Test statistic for variance of a two populations**:
$$\frac{s_1^2}{s_2^2} \text{ is an } F \text{ distribution with } df_1 = n_1 - 1,\ df_2 = n_2 - 1$$
- **Central Limit Theorem**
  - **Version** 1: If $x_1, \ldots, x_n$ is a random sample from a population with mean $\mu$ and standard deviation $\sigma$, then for large enough $n$ the distribution of $\bar{x}$ is approximately normal with mean $\mu$, standard deviation $\sigma/\sqrt{n}$.
  - **Version** 2: If $x_1, \ldots, x_n$ is a random sample from a population with mean $\mu$ and standard deviation $\sigma$, then for large enough $n$ the distribution of $\sum_{i=1}^{n} x_i$ is approximately normal with mean $n\mu$, standard deviation $\sqrt{n}\sigma$.
  - If the population is exactly normal (or very close to it), then the sample doesn't need to be large.