Prof. Bret Larget
Math 125
Final Examination

Name: ______________________________

Please show work! The score you earn on each problem is based on your complete solution, not only on the final answer.

Problem 1: (16 points)
Repeated measurements of dust in the air in a coal mine are normally distributed with a mean of 123.917 mg and a standard deviation of 0.085 mg.

(a) What is the probability that a single measurement is larger than 124 mg?

(b) What is the probability that the mean of three measurements is larger than 124 mg?

(c) There is only a 1% chance that the mean of three sampled measurements will be larger than what value?

Problem 2: (18 points)
In an experiment on computer-assisted learning of pictographs sometimes used by learning-impaired children, two groups of 24 children each worked through computer lessons with the same content and examples and were then tested on their recall of the meaning of the symbols. In the Active group, the children were required to interact with the material during the lesson while children in the Passive group merely controlled the pace of the lesson. The correct number of symbols identified out of 56 by each child is displayed in back-to-back stemplots below.

```
      Active | | Passive
----------------------------
            |1|2234
         765|1|555566777889
444433211100|2|00111
       98887|2|566
           1|3|
           5|3|
           4|4|
```

(a) Find the median and first and third quartile for each group.

(b) Display these distributions with side-by-side boxplots. If any individual observations are more than 1.5 IQR units from the box, display them individually as potential outliers.

(c) Is the Active group skewed somewhat to the left, skewed somewhat to the right, or fairly symmetric?

(d) Informally compare the two groups. Would you say that the Active group did better than, about the same as, or worse than the Passive group did overall?

(e) In a one-sided t test comparing the two population means with $H_0 : \mu_{\text{active}} = \mu_{\text{passive}}$ and $H_a : \mu_{\text{active}} > \mu_{\text{passive}}$, the test statistic is $t = 4.28$ and there are 39 degrees of freedom. Is this test significant at the 1% level?

Problem 3: (18 points)
Consider the 24 observations in the Active group from the previous problem to be a simple random sample from a population of similar children.

(a) Use your calculator to find the mean and standard deviation of the scores from the Active group.

(b) Find a 95% confidence interval for the population mean score on this test for children who complete the active computer lesson.

(c) Is there evidence that the population mean test score for children who complete the active lesson would exceed 20? Test this hypothesis with a t test and give a numerical range for the p-value.

(d) Use plain language to summarize the results of the test in the context of the problem. Do not use words such as hypothesis, significance, or p-value.

Problem 4: (18 points)
A laboratory chemist measures the nitrate level in a water sample by passing light through the solution and measuring the absorbance. In a sequence of calibration measurements for nitrate level (mg per liter water), the chemist measures this data.

Nitrate Level	50	50	100	200	400	800	1200	1600	2000	2000
Absorbance	7.0	7.5	12.8	24.0	47.0	93.0	138.0	183.0	230.0	226.0

The mean and standard deviation of the nitrate variable are $\bar{x} = 840$ and $s_x = 802.7$. The mean and standard deviation of the absorbance variable are $\bar{y} = 96.83$ and $s_y = 90.95$. The correlation coefficient is $r = 0.99994$.

(a) Is there a positive or negative association between nitrate level and absorbance.

(b) Is the relationship between these variables strong or weak?

(c) Find the regression equation for predicting absorbance from nitrate level.

(d) What is the value of the residual at the point where $x = 800$?

(e) Estimate the absorbance if the nitrate level is 500. Do you expect this estimate to be accurate? Explain.

Problem 5: (12 points)
Of the nearly six million votes cast in the past election in Florida, 1.5% of all votes made on punch cards did not register a vote for president when counted by machine. A simple random sample of 600 Florida punch card ballots from this election is taken. Let X be the number of sampled ballots that do not register a vote for president.

(a) Explain why X may be modeled as a binomial random variable.

(b) Find the parameters n and p as well as the mean and standard deviation of the distribution of X.

(c) Use the normal approximation to the binomial distribution to find $P(X > 20)$.

Circle **True** or **False** for Problems 6 through 15. If you answer **False**, explain why. Each problem is worth two points.

Problem 6:
True or **False**:

If a distribution is perfectly symmetric, the mean and median will be equal.

Problem 7:
True or **False**:

Because height is a trait that is determined in part by genetic inheritance, a correlation coefficient of $r = 0.50$ inches might be a reasonable value for the measure of correlation between the heights of adult males and their fathers.

Problem 8:
True or **False**:

Random assignment to treatment groups in an experiment helps to prevent bias.

Problem 9:
True or **False**:

A large population has a mean μ and a standard deviation σ. A simple random sample of size n is taken from the population. The central limit theorem implies that there is a 95% chance that the sample mean will be within $1.96\sigma/\sqrt{n}$ of μ regardless of the sample size, even if the population is not normal.

Problem 10:
True or **False**:

The probability is $\frac{1000!}{500!} \times (0.5)^{500}(1-0.5)^{500}$ that a fair coin tossed 1000 times lands heads exactly 500 times.

Problem 11:
True or **False**:

If every individual has the same chance of being selected, it is a simple random sample.

Problem 12:
True or **False**:

A statistician on the witness stand during the trials in Florida did a hypothesis test to compare the probabilities that a ballot would contain an "undervote" if it were made on a punch card ballot or on a ballot read with an optical scanner. The p-value for this test was extremely small, less than the probability of winning the lottery. The following day the newspaper reported that the statistician had claimed that the chances that the undervote probability was the same for both types of ballots was less than the chance of winning the lottery. This interpretation of the p-value is correct.

Problem 13:
True or **False**:

If a 95% confidence interval for the population mean number of times a ballerina can pirouette to the right is 3.2 ± 0.7, for a population of advanced teen-age ballerinas, this means that 95% of all such ballerinas can pirouette to the right three times.

Problem 14:
True or **False**:

If a 95% confidence interval for a comparison between two population means contains 0, this means that a two-sided hypothesis test with null hypothesis that the population means are equal would be significant at the 5% level.

Problem 15:
True or **False**:

If a hypothesis test for a population proportion with $H_0 : p = 0.05$ and $H_a : p < 0.05$ is not significant at the 5% level, then the sample proportion $\hat{p} = 0.05$.