Model Response 1

a) Suppose six of the students in this class had been released late from their previous class and so decided to run to class. If all the students in the class measured their sitting pulse rates first, then the sitting pulse rates of these six students would be higher than usual. By the time they measure their standing pulse rates, the effect of the previous jog to class could have worn off. As a result, the difference between their standing and sitting pulse rates might be smaller than if they had walked to class.

By having the students flip a coin to determine the order in which they will measure their pulse rates, it is likely that about half of the students who ran to class would measure sitting pulse rates first and that about half would measure standing pulse rates first. Any increase in pulse rate due to the recent exercise will be split about evenly between standing and sitting pulse rate measurements. This will help prevent the kind of systematic inflation of sitting pulse rates that was described as a problem before.

If we observe a difference in students' sitting and standing pulse rates, we want to ensure that the difference is a direct result of whether they are sitting or standing and not some other factor. Randomly assigning the order of the two measurements (with a coin flip, in this case) should help avoid systematic effects on either the standing or sitting pulse rates.

b) The data are paired by individual student, so a matched-pairs *t*-test should be used. A two-sample *t*-test should only be used when the data come from two independent samples.

c) Inference problems should follow the four-step outline described earlier.

Step 1: Identify the population(s) and parameter(s) of interest. State hypotheses if you are performing a hypothesis test.

Our population of interest, according to the problem statement, is "people." Because the data were collected only from high school students, however, it may be problematical to generalize beyond this group. So we'll restrict our attention to the population of high school students. We are interested in performing inference about the mean difference, $\mu_{Diff} = \mu_{\text{stand}-\text{sit}}$, between the standing and sitting pulse rates of high school students. The appropriate hypotheses are:

$$H_0 : \mu_{Diff} = 0$$

$$H_a : \mu_{Diff} > 0$$

The alternative hypothesis is one-sided because we want to determine whether students' standing pulse rates are higher than their sitting pulse rates, in which case μ_{Diff} would be positive.

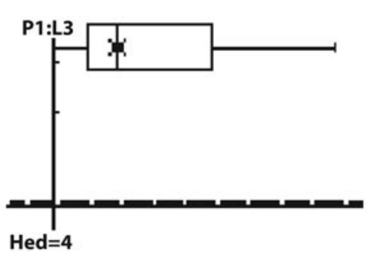
Step 2: Select the appropriate inference procedure and verify conditions for using it.
In part (b), we determined that a matched-pairs *t*-test would be the correct procedure in this case. There are two conditions we should examine before carrying out the procedure:

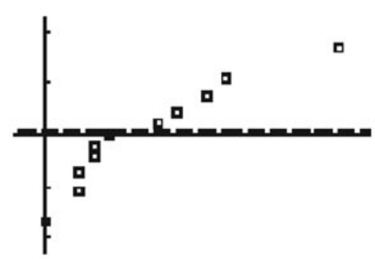
- Data obtained using an SRS from the population of interest

  In this problem, the data come from 12 members of an AP Statistics class. This is clearly a convenience sample. As a result, we may not be able to generalize our findings to the population of all students. However, assignment to treatments was random. So we should be able to determine whether there is evidence of a treatment effect based on sitting or standing.

- Population distribution is normally distributed.

  We do not know that the population distribution of differences in students' standing and sitting pulse rates is normally distributed. A boxplot and a normal probability plot of the differences in our reduced sample of 11 students are shown below. We are concerned by the obvious right-skewed shape of the boxplot, because this may suggest that the data did not come from a normal distribution. The normal probability plot shows some evidence of a nonlinear pattern, which again makes us question the normality of the population distribution. The sample size is small, and the *t*-procedures are not very robust against nonnormality in the population distribution. On the positive side, there are no outliers. We will proceed with the analysis, but exercise appropriate caution when we interpret the results.





Step 3: Carry out the procedure.

- Test statistic: $t = \dfrac{\bar{x}_{diff} - \mu_0}{\dfrac{s_{diff}}{\sqrt{n}}} = \dfrac{6.182 - 0}{\dfrac{5.288}{\sqrt{11}}} = 3.877$

- p-value: For a t-distribution with df = 11 – 1 = 10, the p-value is between 0.0025 and 0.001. (The TI-83 gives a p-value of 0.00154.)

Step 4: Interpret your results in the context of the problem.
Since the p-value is so small (less than 0.0025), it is highly unlikely that we would observe a value of $\bar{x}$ as large as we did in our sample (6.182) if $H_0 : \mu_{Diff} = 0$ were true. So we reject H_0 and conclude that the population mean difference in students' standing and sitting pulse rates is likely to be positive. As pointed out earlier, this gives us statistically significant evidence of a treatment effect. However, we are cautious about generalizing these findings to the population of all students based on data from a convenience sample.

d) Assuming the teacher is correct, we can define the random variable as X = the number of individuals with a higher standing pulse rate. Considering each student's taking of both pulse measurements as a trial, we claim that X is a binomial random variable. When checking the conditions for a binomial setting, we use the "BINS" method:

Binary: "Success" = standing pulse rate higher, "failure" = sitting pulse rate higher.
Independent trials: We hope that one student's pulse rate readings will have no connection to another student's.
Number of trials fixed: n = 12.
Success probability constant: p = 0.5.

So we can calculate the probability of observing 11 or more of the 12 students with higher sitting than standing pulse rates:

$$P(X \geq 11) = \binom{12}{11}(0.5)^{11}(0.5)^{1} + \binom{12}{12}(0.5)^{12}(0.5)^{0} = 0.0029 + 0.00024 = 0.00314$$

Model Problem 2: Do dogs resemble their owners?[1]

Researchers at the University of California, San Diego, designed an experiment to see whether undergraduate psychology students could determine which of two dogs belonged to a dog owner. A total of 45 dogs and their owners were photographed. Of these 45 dogs, 25 were purebreds, and 20 were not purebreds. A dog was classified as resembling its owner if more than half of the 28 undergraduate students matched dog to owner. Here are some results. For the purebred dogs, 16 resembled their owners. For the nonpurebred dogs, only seven resembled their owners.

The researchers believe that people who own purebred dogs would be more likely to resemble their dogs than owners of nonpurebred dogs. Unfortunately, the researchers disagree about what statistical method they should use in this situation.

Researcher 1 begins by constructing the following table to summarize the identifications made by the students.

	Correct	**Not Correct**
Purebred dogs	16	7
Nonpurebred dogs	9	13

Then, she performs a chi-square test of association/independence using computer software. The results are shown below.

```
Expected counts are printed below observed counts
             Correct   Not Corr  Total
Purebred       16         9        25
              12.78     12.22
Nonpurebred     7         13       20
              10.22      9.78
Total          23         22       45
Chi-sq = 0.813 + 0.849 + 1.016 + 1.062 = 3.740
DF = 1, p-value = 0.053
```

a) State hypotheses *and* draw a conclusion based on the chi-square test that was performed.

b) Researcher 2 argues that the chi-square test is not appropriate since they believe that the proportion of correct identifications should be *higher* for purebred dogs than for nonpurebred dogs. He recommends carrying out a two-proportion hypothesis test. Carry out such a test and explain your results.

c) Researcher 3 constructs a 95 percent confidence interval for the difference in the proportion of correct identifications for purebred and nonpurebred dogs and obtains (0.00875, 0.57125). Interpret this interval in the context of the problem.

Model Response 2

a) If a chi-square test of association/independence was performed, then the hypotheses would be H_0 : There is no association between dogs' breeding status and whether they resemble their owners versus H_a : There is an association between dogs' breeding status and whether they resemble their owners, *or* H_0 : Dogs' breeding status is independent of resemblance to their owners versus H_a : Dogs' breeding status is not independent of resemblance to their owners.

With a p-value of 0.053, we have borderline evidence against the null hypothesis. We could legitimately decide either to reject or fail to reject the null hypothesis. If we reject the null hypothesis, then we could conclude that there is evidence of an association between dogs' breeding status and their resemblance to owners. However, if we reject the null hypothesis, we could conclude that there is insufficient evidence of an association between dogs' breeding status and their resemblance to owners.

b) Following the four-step inference problem outline described earlier:

Step 1: Identify the population(s) and parameter(s) of interest. State hypotheses if you are performing a hypothesis test.
We want to determine whether the proportion of purebred dogs that resemble their owners is greater than the proportion of nonpurebred dogs that resemble their owners. Let $p_1 =$ the proportion of all purebred dogs that resemble their owners and $p_2 =$ the proportion of all nonpurebred dogs that resemble their owners. We want to test the hypotheses:

$$H_0 : p_1 = p_2$$
$$H_a : p_1 > p_2$$

We will use a 0.05 significance level.

Step 2: Select the appropriate inference procedure and verify conditions for using it.
Since we are comparing two population proportions, we should use a two-proportion z-test. The two primary conditions for using this procedure are:

- Data obtained using two independent SRSs from the respective populations of interest

We doubt that the researchers obtained any kind of random sample of either purebred or nonpurebred dogs. Therefore the samples of dogs being used in this study may not be representative of their corresponding populations.

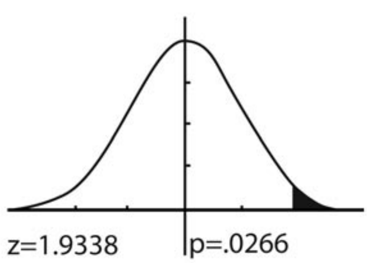
This would severely limit our ability to generalize. If the data come from a randomized comparative experiment, then we could test for evidence of a treatment effect. Insufficient information is given in the stem of the problem to determine whether this is the case here. We proceed with caution.

- It is reasonable to use a normal approximation.

  When performing a two proportion z-test, we begin by assuming that the null hypothesis is true. In that case, $p_1 = p_2$. We estimate the common value of these two unknown population parameters using $\hat{p} = \dfrac{X_1 + X_2}{n_1 + n_2} = \dfrac{16+7}{25+20} = 0.511$. We need only check that $n_1\hat{p}$, $n_1(1-\hat{p})$, $n_2\hat{p}$ and $n_2(1-\hat{p})$ are all at least 5. Since
  25(0.511) = 12.775, 25(0.489) = 12.225,
  20(0.511) = 10.22, and 20(0.489) = 9.78, this condition is satisfied.

Step 3: Carry out the procedure.

- Test statistic: $z = \dfrac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}} = \dfrac{0.64 - 0.35}{\sqrt{(0.511)(0.489)\left(\dfrac{1}{25}+\dfrac{1}{20}\right)}} = 1.93$.
- p-value: A standard normal table yields a p-value of 0.0268. (Running a two proportion z-test on our calculator, we obtain a p-value of 0.0266.)

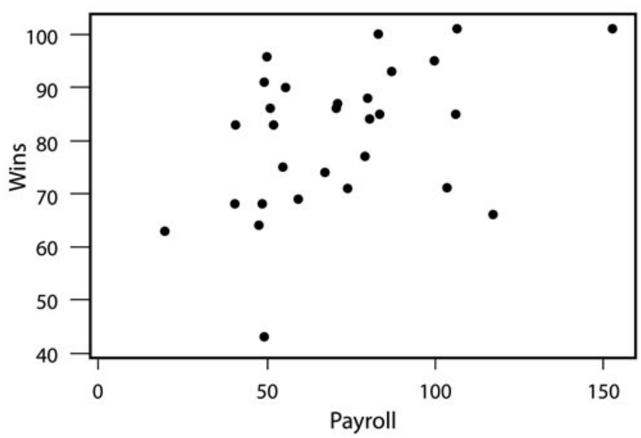


Step 4: Interpret your results in the context of the problem.
Since the p-value is less than our stated cutoff of 0.05, we reject the null hypothesis in favor of the alternative. We conclude that the proportion of purebred dogs that

resemble their owners is significantly higher than the proportion of nonpurebred dogs that resemble their owners, provided we can view our samples as representative of their corresponding populations.

c) We are 95 percent confident that the actual difference in the proportion of purebred dogs that resemble their owners and the proportion of nonpurebred dogs that resemble their owners is between 0.00875 and 0.57125. (Interpretation of *confidence level*: 95 percent confident means that we are using a method to produce intervals that capture the true difference in population proportions in 95 percent of all possible samples of the same size from the respective populations.)

Model Problem 3: Major League Baseball payrolls and winning percentage[2]
Data were collected on the opening day payrolls (in millions of dollars) and the number of wins during the season for a random sample of Major League Baseball teams in the early 2000s. A scatterplot of the data is displayed below, together with partial computer output from a linear regression calculation.



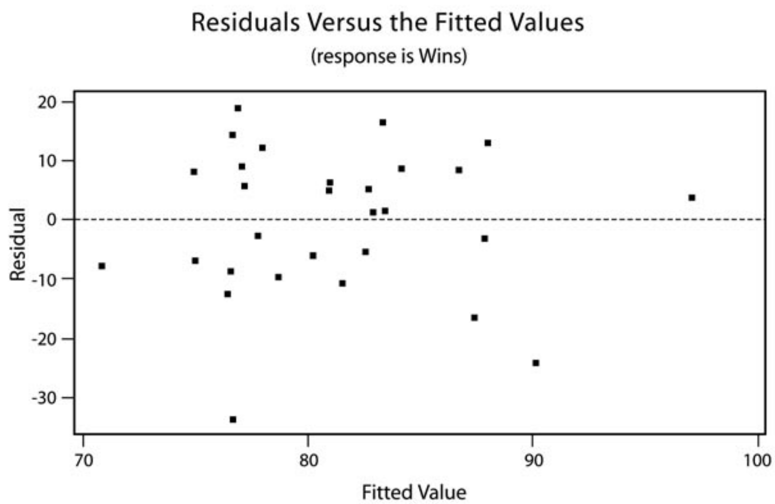
Predictor	Coef	SE Coef	T	P
Constant	66.901	6.261	10.68	0.000
Payroll	0.19803	0.08221	2.41	0.023

S = 12.38 R-Sq = 17.2% R-Sq(adj) = 14.2%

a) Determine the value of the correlation coefficient and interpret its meaning in the context of the problem.

b) Give the equation of the least squares regression line. Be sure to define any variables that you use in your equation.

c) A residual plot resulting from the linear regression is shown below. Estimate the payroll *and* the number of wins for the team with the most negative residual.



d) Is there statistically significant evidence of a positive linear relationship between payroll and games won for Major League Baseball teams? Justify your answer.

Model Response 3

a) Since $r^2 = 0.172$, and the scatterplot shows that the correlation should be positive, $r = +\sqrt{r^2} = +\sqrt{0.172} = 0.414$. This tells us that there is a moderate, positive, linear relationship between team payroll and number of wins in a season for these Major League Baseball teams from the early 2000s.

b) The regression equation is: Predicted Wins = 66.9 + 0.198 Payroll.

c) The team with the largest negative residual appears to have a fitted value (predicted number of wins) of around 77. Substituting this value into the regression equation yields

77 = 66.9 + 0.198 Payroll → 10.1 = 0.198 Payroll → Payroll = 51.01 million.

Since residual = actual y-value – predicted y-value,

–33 = actual y-value – 77 → actual y-value = 44 wins.

d) Following the four-step inference problem outline:

Step 1: Identify the population(s) and parameter(s) of interest. State hypotheses if you are performing a hypothesis test.
In this problem, we are interested in determining whether there is significant evidence of a positive linear relationship between opening day payroll and number of wins in a season for Major League Baseball teams in the early 2000s. Our hypotheses would be $\begin{matrix} H_0 : \beta = 0 \\ H_a : \beta > 0 \end{matrix}$, where β represents the slope of the population regression line relating payroll to wins.

Step 2: Select the appropriate inference procedure and verify conditions for using it.
Here, we should perform a linear regression *t*-test on the slope of the population regression line. The required conditions for using this procedure are:

- Data obtained using an SRS from the population of interest.
  We are told that the data come from a random sample of Major League Baseball teams in the early 2000s.
- There is an underlying linear relationship between the variables.
  The shape of the scatterplot suggests that the variables may be linearly related.
- The standard deviation of the response variable is the same for all *x*-values in the data set.
  From the residual plot, we see a few unusual observations, but for the most part, the spread of the residuals around the zero error line is somewhat similar.
- The response varies normally around the true regression line.
  This is not really possible for us to verify with the information provided.

Step 3: Carry out the procedure.
From the computer output, we see that:

- Test statistic: $t = 2.41$
- *p*-value: The two-sided *p*-value is 0.023, so the one-tailed *p*-value for this problem is approximately 0.015.

Step 4: Interpret your results in the context of the problem.
Due to the low *p*-value (0.015), it is highly unlikely that we would have obtained a sample regression line with a slope as large as or larger than 0.198 if the null hypothesis, $\beta = 0$, was true. So we would decide to reject the null hypothesis and conclude that there is statistically significant evidence of a positive linear relationship between payroll and number of wins for Major League Baseball teams in the early 2000s.