1. **True/False** The *p*-value of a test statistic is the probability that the null hypothesis is true.

2. **True/False** An article reports that a 95% confidence interval for the average weight of adult males in the US is $180 \pm 15$. This means that 95% of all males weight between 165 and 195 pounds.

3. There were 14 American League baseball teams that played in 2003. You don’t have any reason to believe that the number of games that such a team won was related in any way to either the average temperature or total rainfall in their home city. In other words you don’t believe the model $W \sim 1 + \text{TEMP} + \text{RAIN}$ is any better than using two random variables to predict wins. Suppose in fact that you fit the model and find that $R^2 = 0.15$. Someone used to reading economics papers thinks that this value proves that the model is useful for predicting wins. However you know that even if we would use two random variables instead of TEMP and RAIN, we would expect to see a value of $R^2$ that is on average about

4. A study was done on a new filter that was supposed to reduce noise pollution in automobiles. Different sizes of cars were measured (size), two different types of filter (Type) and measurements were taken on each size of the car (Side). The model was $\text{Noise} \sim 1 + \text{Size} + \text{Type} + \text{Side}$ (with Size converted to a factor). The ANOVA table is below with some of the quantities replaced by letters.

**Analysis of Variance Table**

<table>
<thead>
<tr>
<th>Response: Noise</th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>factor(Size)</td>
<td>2</td>
<td>26051.4</td>
<td>13025.7</td>
<td>145.987</td>
<td>C</td>
</tr>
<tr>
<td>Type</td>
<td>1</td>
<td>1056.2</td>
<td>1056.2</td>
<td>0.0017</td>
<td>**</td>
</tr>
<tr>
<td>Side</td>
<td>1</td>
<td>0.7</td>
<td>0.7</td>
<td>0.0078</td>
<td>0.9505</td>
</tr>
<tr>
<td>Residuals</td>
<td>31</td>
<td>2766.0</td>
<td>89.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) Compute A.
(b) Compute B.
(c) We haven’t learned enough to compute C exactly. But you can say approximately what it is.
(d) How many cars were tested?
(e) Write a sentence that interprets correctly the value .930268 in the lower right of the table. Your sentence should interpret the number in the particular context of the problem.
These questions concern data from the National Education Longitudinal Study conducted in 1988. (You actually first met this data on the first test.) A nationally representative sample of eighth graders was taken and many variables were recorded. A sample from that sample is represented in the nels88 data frame, the first few rows of which are below. Only a few variables from the original study were retained. There are 260 cases in the data frame. The variables are

- `sex`: sex of student
- `race`: race of student
- `ses`: quantitative variable computed to measure socioeconomic status with higher numbers meaning higher status
- `paredu`: highest level of education of parents
- `math`: score on a standardized math test

```r
head(nels88)
```

<table>
<thead>
<tr>
<th>sex</th>
<th>race</th>
<th>ses</th>
<th>paredu</th>
<th>math</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>White</td>
<td>-0.13</td>
<td>hs</td>
<td>48</td>
</tr>
<tr>
<td>Male</td>
<td>White</td>
<td>-0.39</td>
<td>hs</td>
<td>48</td>
</tr>
<tr>
<td>Male</td>
<td>White</td>
<td>-0.80</td>
<td>hs</td>
<td>53</td>
</tr>
<tr>
<td>Male</td>
<td>White</td>
<td>-0.72</td>
<td>hs</td>
<td>42</td>
</tr>
<tr>
<td>Female</td>
<td>White</td>
<td>-0.74</td>
<td>hs</td>
<td>43</td>
</tr>
<tr>
<td>Female</td>
<td>White</td>
<td>-0.58</td>
<td>hs</td>
<td>57</td>
</tr>
</tbody>
</table>

A model is fit as follows

```r
> lfull = lm(math ~ sex + ses + race + paredu, nels88)
> anova(lfull)
```

Analysis of Variance Table

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>sex</td>
<td>1</td>
<td>9</td>
<td>9</td>
<td>0.13</td>
<td>0.72092</td>
</tr>
<tr>
<td>ses</td>
<td>3</td>
<td>12383</td>
<td>12383</td>
<td>172.87</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>race</td>
<td>5</td>
<td>1570</td>
<td>314</td>
<td>4.38</td>
<td>0.00077</td>
</tr>
<tr>
<td>paredu</td>
<td>249</td>
<td>17836</td>
<td>72</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. $R^2$ for this model is **less than 50%** **greater than 50%**. (Choose one.)

2. How many different levels of parents education are in these data? __________

3. On the basis of the ANOVA analysis, a natural model to fit next would be

   (a) `lm(math~sex+ses,nels88)`
   (b) `lm(math~sex+race,nels88)`
   (c) `lm(math~sex+paredu,nels88)`
   (d) `lm(math~ses+race,nels88)`
   (e) `lm(math~ses+paredu,nels88)`
   (f) `lm(math~race+paredu,nels88)`
The next two questions are about this story. In Mathematics 143 a few years ago, I had 65 students. There were two different forms of the first test, one on yellow paper and the other on red. The tests were handed out to the 65 students at random. The average score of those taking the yellow test was higher than that of those taking the red test. This led to complaints that the red version of the test was harder. Here are some of the data and what I think is an appropriate analysis of the data. The test was a 40 point test.

> head(M143T1)

Scores Color
1 40 Y
2 39 R
3 21 R
4 35 Y
5 37 Y
6 35 R

> lm143 = lm(Scores ~ Color, M143T1)
> summary(lm143)

Call:
  lm(formula = Scores ~ Color, data = M143T1)

Residuals:
     Min      1Q  Median      3Q     Max
-28.471  -2.613   0.529   4.529   7.387

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)       32.613       1.055   30.91  <2e-16
ColorY             0.858       1.459    0.59    0.56

Residual standard error: 5.87 on 63 degrees of freedom
Multiple R-squared:  0.00546, Adjusted R-squared:  -0.0103
F-statistic: 0.346 on 1 and 63 DF,  p-value: 0.559

1. The average score on the yellow version of test was how much higher than that on the red version of the test?

2. What is the best statistical conclusion from this analysis?
   (a) The complainers have a case. The red test was harder than the yellow test.
   (b) The complainers have a case. The $p$-value is so high that it gives strong evidence that the red test is harder than the yellow test.
   (c) The complainers are wrong. The difference in these two averages is too small to worry about.
   (d) The complainers are wrong. The $p$-value is so high that there is not strong evidence that the red test is harder than the yellow test.
   (e) This statistical test is not appropriate. We would have to give every student both versions and see whether they did better on the red or yellow version.
The dataframe `TenMileRace` has data on a ten mile race that is run in Washington D.C. each year. The variable `net` gives the time necessary to finish the race (from the starting line and not from the gun). Variables `age` and `sex` give the age and sex of the runner.

1. What is the relationship between net running time and the runner’s age? Is the relationship significant? Is it substantial?
2. What is the relationship between net running time and the runner’s sex? Is the relationship significant? Is it substantial?
3. Is there an interaction between sex and age? Is the relationship significant? Is it substantial?

The dataframe `cathedral` in the `faraway` package has data on the cathedral nave heights and lengths of several cathedrals in England.

- `x`: the length of the nave
- `y`: the height of the nave
- `style`: the style of the cathedral, romanesque or gothic

Fit a model to predict the height of the nave from the length of the nave and the style. \( y \sim 1 + x + \text{style} \)

1. Write the equation of the fitted model.
2. Use R to estimate the P-value for the null hypothesis that `style` does not matter in this model.
3. Instead, use a simulation method (using `shuffle`) to compute the same P-value.
4. Should `style` be included in this model?