Practice problems

Note that there are also many practice problems and solutions on the QR class website, under the “Archive” section!

1. A group of researchers were interested in the effects of repeated exposure to television advertising. They believed that perhaps the effects depend on both the length of the advertising and how often it is repeated. They investigated this topic using four groups of undergraduate students. All groups viewed a 40-minute television program that included ads for a digital camera. Students in Group 1, however, saw one 30-second commercial one time during the program. Group 2 saw two 30-second commercials three times during the program. Group 3 saw one 90-second commercial one time during the program. Group 4 saw two 90-second commercials three times during the program. After the program, students in all groups were asked a series of questions about their recall of the ad, their attitude toward the camera advertised, and whether they thought they might buy this or any other digital camera in the next 12 months.

What is the null and alternative hypothesis of the researchers? What type of research design did this study employ? Explain this in words, and also draw a diagram that summarizes the group(s) of observational units (e.g., people, schools, states) and the treatment(s) involved. Evaluate this research design with respect to internal validity and external validity.

It’s not clear what the null and alternative hypotheses are. We could reasonably expect that there would be 2 null hypotheses, 1 of which deals with the impact of television ad length individuals who see the ad, and the other which deals with the impact of ad repetition on individuals. There may be up to three variables involved in testing such hypotheses—we could call these “recall,” which measures how well individuals remember what they’ve seen in the ads; “attitude,” which measures how individuals feel about the product advertised; and “purchase intention,” which measures an individual’s intent to purchase the product advertised in the future. Without additional information, however, we don’t know specifically what kinds of hypothesis tests might be employed with the data collected in this research.

The study employed a static group comparison; that is, measurements were taken for each group only after they were exposed to a treatment. What’s interesting here is that we really have four different treatments, which would allow us to make lots of different kinds of comparisons. For example, comparing the measurements we obtain for groups 1E and 2E would allow us to test for the effects of repeated viewing of a commercial of a particular length; comparing groups 1E and 3E would allow us to test for the effects of viewing a single ad of differing length.

\[
\begin{align*}
X_1 & \quad O_{1E} \\
X_2 & \quad O_{2E} \\
X_3 & \quad O_{3E} \\
X_4 & \quad O_{4E}
\end{align*}
\]
This type of design has a relatively high degree of internal validity, because we can feel pretty confident that factors external to the study have been well-controlled for. The criteria of time order, covariation, and no better rival explanation seem well satisfied, and although we don’t know much about the theory involved, it seems plausible that length and repetition of television advertising should impact one’s recall and attitude toward the product advertised.

The study is conducted in a laboratory setting, which might give us pause about its external validity. The students know that the study has something to do with television viewing, and might thus focus more on the program and/or advertising content than they would if watching television in their homes. We might question whether the results obtained in this study can be generalized to “real world” TV viewing.

2. You read a study based on sample data that says the average length of service for home health care among people over the age of 65 who have used this type of service is 96 days, with a standard error of 5.1 (note: not standard deviation!). Assuming a large number of degrees of freedom, calculate a 90% confidence interval for the true (population) mean length of service. Interpret your result in plain language that someone with no training in statistics would understand.

This is a standard confidence interval approach, and we can use the z values from a standard normal table because we are told that we have a large number of degrees of freedom. For a 90% confidence level, this value is 1.645. Thus, we have:

Given \( \bar{x} = 96.0 \) and \( \frac{s}{\sqrt{n}} = 5.1 \),

\[
\bar{x} - E < \mu < \bar{x} + E
\]

\[ 96.0 - (1.645 \times 5.1) < \mu < 96.0 + (1.645 \times 5.1) \]

\[ 96.0 - 8.39 < \mu < 96.0 + 8.39 \]

\[ 87.6 < \mu < 104.4 \]

We interpret by saying that we are 90% confident that the average length of service for home health care among people over the age of 65 who have used this type of service is between 87.6 and 104.4. (It’s also true that the probability that the average value of this variable in the population is not between 87.6 and 104.4 is .10.)

3. In an effort to test how accurate radon detectors produced for purchase by homeowners are, a researcher placed 12 such detectors in a chamber and exposed them to exactly 105 picocuries per liter of radon. Among these detectors, the mean reading was 104.13 and the standard deviation was 9.40.

Is there convincing evidence that these detectors provide readings that are, on average, different from the known concentration of 105 PPL? Write out your null and alternative
hypotheses, set up your formula and compute your test statistic, find the associated p-value and interpret your findings in clear language.

We use a two-tailed test because we are interested in significant differences from the mean in either direction. We know that we must use the t distribution because (1) \( \sigma \) is unknown and our sample size is only 12. Thus:

\[
H_0 : \mu = 105 \quad H_A : \mu \neq 105
\]

\[
t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{104.13 - 105}{9.4 / \sqrt{12}} = -0.87 = -0.32
\]

First, we are asked whether there is “convincing evidence” that the detectors are defective. This means that we must choose a significance level ourselves. Let’s say that, because we’re dealing with an issue that involves significant health risks to individuals, we want a .01 significance level in our test. (If you chose another level, this is fine—just make sure you set up the test statistic correctly and found the correct critical and/or p-values.)

We can proceed in two different ways here (the question asks for the p-value, but let’s go through both approaches to remind you of them). We can look for the critical values associated with a significance level of .01, a test statistic value of –0.32, and \((n-1=11)\) degrees of freedom in the t-table (be careful—this is a two-tailed test!). These values are 1.796 and –1.796. Our value does not exceed 1.796; thus we fail to reject our null hypothesis that \(\mu=105\). (Draw a picture if this helps you in such problems.) In practical terms, we do not have evidence that the radon detectors are giving readings that are significantly different from the truth.

The second approach is to find the p-value for our test statistic. Looking in the row of the t table for 11 degrees of freedom, we see that we cannot find our value anywhere along the row—it’s too small. This suggests that the p-value associated with our test statistic is greater than the largest p-value in the table—.10. In fact, because this is a two-tailed test, we know that the p-value for our test statistic is at least \(2 \times .10 = .20\) (check your handout on using the t-table if this does not make sense). We can report this as \(p > .20\). Remember that we set our significance level for this test at .01—thus, something greater than .20 is clearly not statistically significant. Again, we fail to reject our null hypothesis that \(\mu=105\).

4. An education policy researcher believes that “directed reading” activities are useful in helping elementary school students improve some aspects of their reading ability. She chooses a class of 21 third-graders and introduces a directed reading curriculum to them for an 8-week period. A control classroom of 23 third-graders follows a normal curriculum with no directed reading activities. At the end of the 8 weeks, all students are given a reading ability test. The results were as follows:

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Mean score</th>
<th>Stand. Dev.</th>
</tr>
</thead>
</table>
What would be the researcher’s null and alternative hypotheses? Write these both in words and using formulas. Then use the appropriate technique to test the null hypothesis. Compute the test statistic, find the relevant p-value, and interpret your results in simple language.

The researcher has a hypothesis about which of these groups is expected to do better on the reading test. Thus, she does not think that the Thus, her hypotheses are:

\[ H_0: \mu_D \leq \mu_N \quad \quad H_A: \mu_D > \mu_N \]

(Remember that the null hypothesis could also be expressed with an = sign; this is technically more correct but either will give you the same result.)

We need to decide on a significance level and an approach for this test. Let’s use a .05 significance level. Also, because we have no information about whether the variance of reading scores is equivalent in the two populations, and because the sample standard deviations seem fairly different, let’s use the approach for different population variance values. (Note that on any exam you will not have to guess which approach to use; we will make it clear.) Thus:

\[ s.e._D = \frac{s_D}{\sqrt{n_D}} = \frac{11.01}{\sqrt{21}} = 2.40 \]
\[ s.e._N = \frac{s_N}{\sqrt{n_N}} = \frac{17.15}{\sqrt{23}} = 3.58 \]
\[ s.e._{overall} = \sqrt{s.e.^2_D + s.e.^2_N} = 4.31 \]
\[ t = \frac{\bar{x}_D - \bar{x}_N}{s.e._{overall}} = \frac{51.48 - 41.52}{4.31} = 2.31 \]

For a one-tailed test, the critical value for a .05 significance level and a one-tailed test is 1.645. Our test value exceeds this, thus providing evidence that our null hypothesis \( H_0: \mu_D \leq \mu_N \) is not correct. We reject the hypothesis that directed reading students have, on average, reading scores that are equal to or less than students who followed the normal curriculum. Although we cannot conclude definitively that the directed reading program is improving reading test grades, we have evidence that suggests that students following this program are scoring better, on average, than their counterparts in the normal curriculum.

5. A survey of 17,096 college students was designed to collect information on how serious a problem alcohol abuse is on US campuses. Researchers found that 19.4% of those surveyed reported “frequent binge drinking” behavior. Using this information, construct a
95% confidence interval for the proportion of all college students who engage in binge drinking frequently. Interpret your findings in one clear sentence.

This is a straightforward confidence interval for proportions. We have:

\[ \hat{p} - E < p < \hat{p} + E \]

\[ .194 - E < p < .194 + E \]

\[ E = z\left(\frac{\alpha}{2}\right) \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = 1.96 \sqrt{\frac{.194(1-.194)}{17096}} = .006 \]

\[ .194 - .0006 < p < .194 + .0006 \]

\[ .193 < p < .194 \]

What we learn here is that a sample of 17,096 students results in a 95% confidence interval for the entire student population that is very narrow! We are 95% confident that the proportion of all college students who engage in binge drinking is between .193 and .194.

6. Suppose you are working for Hillary Clinton in her bid to become a NY senator. You supervise an SRS of 1785 NY registered voters and ask whether they are currently planning to vote for Mrs. Clinton in the upcoming election. Of these, 750 answer affirmatively. Do these results provide good evidence that less than half of registered NY voters support Mrs. Clinton? There are two ways to answer this question—in words identify each of the approaches, then choose one and employ it to find an answer. Finally, determine how large a sample would be required to obtain a margin of error of +/- 1% in a 99% confidence interval for the proportion of registered NY voters who favor Mrs. Clinton.

Our sample proportion is 750/1785=0.42. We can either test the null hypothesis that \( p \leq 0.50 \), or we can create a confidence interval for \( p \) and determine whether it includes values below 0.50. Since we’ve used a confidence interval approach elsewhere in the problem set, let’s use a hypothesis testing approach here. Let’s set our significance level at .05.

\[ H_0: p \geq .50 \quad H_a: p < .50 \]

\[ np_o = 0.50 \times 1785 = 892.5 \quad n(1-p_o) = 0.50 \times 1785 = 892.5 \]

\[ z = \frac{\hat{p} - p_o}{\sqrt{p_o(1 - p_o)/n}} = \frac{0.42 - 0.50}{\sqrt{(0.50 \times 0.50)/1785}} = -6.76 \]

We have a one-tailed test, so the critical value (from the normal table) is –1.645 (why negative? Look at your hypotheses). Our value –6.76 is far beyond the critical value, so we reject the null hypothesis of \( p \geq .50 \) and conclude that we have evidence to suggest that, in fact, it appears that a minority of registered NY voters support Mrs. Clinton. (Again, draw a picture if this helps you understand the problem better!)
To determine the sample size required to reduce the margin of error in our confidence interval for \( p \) to +/- 1\%, we find the \( z \) value for a 99\% confidence level to be 2.575. We know that we want our interval to range from only .01 below to .01 above the true population proportion. Thus we substitute in the values for \( z \) and \( E \) into the formula:

\[
n = \left( \frac{z \times \frac{1}{2}}{E} \right)^2 = \left( \frac{2.575 \times \frac{1}{2}}{.01} \right)^2 = 16,577
\]

7. In question 5 above, the researchers also split out the information on binge drinking among students by gender. Here are the data:

<table>
<thead>
<tr>
<th>Group</th>
<th>( n )</th>
<th>Proportion “binge drinkers”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men (1)</td>
<td>7180</td>
<td>0.227</td>
</tr>
<tr>
<td>Women (2)</td>
<td>9916</td>
<td>0.170</td>
</tr>
<tr>
<td>TOTAL</td>
<td>17096</td>
<td>0.194</td>
</tr>
</tbody>
</table>

Suppose we want to determine whether the proportion of binge drinkers among females is significantly different than that of males. First, determine what criterion you want to use for statistical significance. Next, state your null and alternative hypotheses. Finally, use two different approaches to test your null hypothesis (hint: one of these employs confidence intervals). In each case, state clearly what you conclude.

Let’s use a .01 significance level. We can either test the equivalence of these proportions directly, or we can develop two 99\% confidence intervals for the proportion of male and female students who are binge drinkers and see whether they overlap.

For males:

\[
\hat{p} - E < p < \hat{p} + E
\]

\[
.227 - E < p < .227 + E
\]

\[
E = z \left( \frac{\alpha}{2} \right) \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = 2.575 \sqrt{\frac{.227(1 - .227)}{7180}} = .01
\]

\[
.227 - .01 < p < .227 + .01
\]

\[
.217 < p < .237
\]

For females:

\[
\hat{p} - E < p < \hat{p} + E
\]

\[
.170 - E < p < .170 + E
\]
Suppose you wanted to explore the relationship between smoking and socioeconomic status (SES). You have the following data from a sample of 356 people, 211 of whom are considered to have high SES status, 52 Middle SES status, and 93 low SES status. Devise a statistical test of the null hypothesis that there is no association between socioeconomic status and smoking behavior. Show your work and interpret your findings in plain language. (Be careful!)

<table>
<thead>
<tr>
<th></th>
<th>High SES</th>
<th>Middle SES</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current smoker</td>
<td>24.2%</td>
<td>42.3%</td>
<td>46.2%</td>
</tr>
<tr>
<td>Former smoker</td>
<td>43.6%</td>
<td>40.4%</td>
<td>30.1%</td>
</tr>
<tr>
<td>Never smoked</td>
<td>32.2%</td>
<td>17.3%</td>
<td>23.7%</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

We said “be careful” to remind you that the correct approach here, a $X^2$ test, requires count data. Thus, we start by generating a new table:

<table>
<thead>
<tr>
<th></th>
<th>High SES</th>
<th>Middle SES</th>
<th>Low</th>
<th>Column totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current smoker</td>
<td>51</td>
<td>22</td>
<td>43</td>
<td>116</td>
</tr>
<tr>
<td>Former smoker</td>
<td>92</td>
<td>21</td>
<td>28</td>
<td>141</td>
</tr>
<tr>
<td>Never smoked</td>
<td>68</td>
<td>9</td>
<td>22</td>
<td>99</td>
</tr>
<tr>
<td>Row totals</td>
<td>211</td>
<td>52</td>
<td>93</td>
<td>356</td>
</tr>
</tbody>
</table>

Our null hypothesis is that there is no association between socioeconomic status and smoking behavior. Assuming for the moment that this were true, we need to compute the cell count values we would expect to observe. This would mean that smoking behavior would occur with the same frequency in each socioeconomic category. To find these values for any given cell, we multiply the row total for the cell by the column total for the cell, then divide by the total sample size. This gives us the following values:

<table>
<thead>
<tr>
<th></th>
<th>High SES</th>
<th>Middle SES</th>
<th>Low</th>
<th>Column totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current smoker</td>
<td>69</td>
<td>17</td>
<td>30</td>
<td>116</td>
</tr>
<tr>
<td>Former smoker</td>
<td>84</td>
<td>21</td>
<td>37</td>
<td>141</td>
</tr>
<tr>
<td>Never smoked</td>
<td>59</td>
<td>14</td>
<td>26</td>
<td>99</td>
</tr>
<tr>
<td>Row totals</td>
<td>211</td>
<td>52</td>
<td>93</td>
<td>356</td>
</tr>
</tbody>
</table>

We compute the test statistic as follows:
\[ X^2 = \sum \frac{(O-E)^2}{E} \]

\[ X^2 = \frac{(51-69)^2}{69} + \frac{(22-17)^2}{17} + \ldots \text{(do this for all cells)} + \frac{(22-26)^2}{26} \]

\[ X^2 = 18.51 \]

We need to compute our degrees of freedom for this test. This is done with the formula \((r-1) \times (c-1)\), where \(r\) is the number of rows in your table and \(c\) is the number of columns. In our case \((3-1) \times (3-1) = 4\) degrees of freedom. Looking in that row of the table, we see that the value 18.51 falls between 18.47 and 20.00, which have p-values of .001 and .0005, respectively. We would thus reject our null hypothesis of no association between smoking behavior and SES, and report a p-value between .001 and .0005 (.0005 < p < .001). Note that we have not learned which SES category members are more likely to be smokers! You would need to do additional analysis to determine this.

9. A multiple linear regression model was developed in which a student’s college GPA in his/her 6th semester is the dependent variable, and the independent (explanatory) variables are the student’s SAT verbal score (SATV); the student’s SAT math score (SATM); the student’s average high school math course grade (HSM); the student’s average high school science course grade (HSS); and the student’s average high school English course grade (HSE). Each of the high school grade variables were coded as follows: A=10, A-=9, B+=8, B=7, etc. GPA is coded on a traditional 4-point scale, i.e., A=4, B=3, C=2, D=1, and F=0. The SAT scores are as they appear on score reports (i.e., a maximum of 800 is possible for each part of the SAT).

Here are the results of this analysis:

| Variable | DF | Parameter estimate | Standard error | T for HO: Parameter=0 | Prob > |T| |
|----------|----|--------------------|----------------|-----------------------|--------|---|
| INTERCEP | 1  | 0.326719           | 0.3999643      | 0.817                 | 0.4149 |
| SATM     | 1  | 0.000944           | 0.00068566     | 1.376                 | 0.1702 |
| SATV     | 1  | -0.000408          | 0.00059189     | -0.689                | 0.4915 |
| HSM      | 1  | 0.145961           | 0.03926097     | 3.718                 | 0.0003 |
| HSS      | 1  | 0.035905           | 0.03779841     | 0.950                 | 0.3432 |
| HSE      | 1  | 0.055293           | 0.03956869     | 1.397                 | 0.1637 |
|          |    | R-square           | 0.21151        |                       |        |

a. Write out the formula for this regression model in the population, with each variable and parameter identified.

---

\(^1\) Yes, we could take issue with this form of coding—why?
\[ Y(GPA) = \alpha \text{(intercept)} + \beta_1 X_1 \text{(SAT math score)} + \beta_2 X_2 \text{(SAT verbal score)} + \beta_3 X_3 \text{(HS math grade)} + \beta_4 X_4 \text{(HS science grade)} + \beta_5 X_5 \text{(HS English grade)} + e \]

b. Interpret the intercept value for the model results above. Does this make sense intuitively? Why or why not?

The intercept value 0.33 can be interpreted as the expected GPA for a student who had 0 on his/her SAT scores and a failing average for his/her high school math, science and English grades. This is likely a case of generating predicted values for a set of independent variables that don’t really exist in the dataset.

c. Interpret in one clear sentence what the coefficient value for the variable HSM means.

For each one-unit increase in a student’s average high school math grade, we predict a 0.146 increase in his/her 6th semester college GPA, all else held constant.

d. What does the column heading “T for HO: Parameter=0” mean in plain language? What is being tested in this column?

For each independent variable (the Xs), we are testing the null hypothesis that \( \beta = 0 \), or that the relevant independent variable has no significant impact on the dependent variable.

e. Test the hypothesis that average high school science grades have no significant effect on a student’s 6th semester college GPA. Write out your null and alternative hypotheses in formulas; compute a test statistic; and find the relevant p-value. State clearly the outcome of your test. Then describe what information in the table above indicates the same conclusion to this test.

As mentioned above, we are testing the null hypothesis that \( \beta_4 = 0 \). We use the following approach:

\[
 t = \frac{b - \beta_4}{\text{s.e.}(b)} = \frac{0.036}{0.038} = 0.95
\]

We cannot use a critical value approach, because we do not have the sample size (and thus the degrees of freedom) for the model. We find the relevant p-value from the computer output (the final column). In this case, it is 0.34, which is clearly higher than any commonly accepted significance level (.10, .05, .01). We thus do not reject the null hypothesis that \( \beta_4 = 0 \). In practical terms, we do not have compelling evidence to believe that high school science grades have a significant effect on a student’s 6th semester college GPA.

f. To what do the values in the last column of the table refer? Interpret the value in this column for the variable SATM.

This column presents the p-values associated with the hypothesis tests for each population parameter \( \beta \). Each one refers to a test of the null hypothesis of \( \beta = 0 \), that is, no significant effect of the particular independent variable on the dependent variable. The probability being presented can be interpreted as the probability that we would obtain the \( b \) (sample
statistic) value that our model generated, assuming that the null hypothesis of no effect is true, by chance alone.

We cannot use a critical value approach, because we do not have the sample size (and thus the degrees of freedom) for the model. We find the relevant p-value from the computer output (the final column). In this case, it is 0.34, which is clearly higher than any commonly accepted significance level (.10, .05, .01). We thus do not reject the null hypothesis that $\beta_4 = 0$. Thus, the p-value 0.17 for the variable SATM indicates that, if $\beta_1 = 0$ in the population, the probability that we would obtain our value for $b_1$ of .0009 by chance alone is 0.17. (It would also be correct to say that the chances of obtaining such a value are 1 in $1/0.17=5.8$.) Thus, we don’t have compelling evidence to reject the null hypothesis that, in the population, SAT math scores have no significant impact on a student’s 6th semester college GPA.

g. What is the predicted GPA for a student with a 600 math SAT score, a 680 SAT verbal score, a B average in high school math, an A- average in high school science, and a B+ average in high school English?

Just plug in the values here, along with the $b$ values from the computer output, into the equation you wrote for part A:

$$\hat{Y} = 0.327 + (.0009 \times 600) + (-.0004 \times 680) + (0.146 \times 7) + (0.036 \times 9) + (0.055 \times 8)$$

$$\hat{Y} = 2.381$$

h. Suppose that two students have identical values on all the explanatory variables except SATV. Student A obtained a 540 score on his verbal SAT; Student B obtained a 700 on hers. What is the predicted impact of this difference in SATV scores on the students’ 6th semester college GPAs?

You could do this the long way—use the approach in part g above for both students, and then subtract the predicted value for one student from that of the other. It is much easier, however, to remember what the $b$’s mean—the predicted impact of a one-unit change in the independent variable on the value of the dependent variable, all else held constant. Thus, we can simply use the following approach:

$$(700-540) = 160$$

$$160 \times .000408 = .065$$

Thus, the student with the higher SAT verbal score (B) is predicted to have a GPA that is .065 lower than that of the student with the lower verbal score (A).

i. Explain in one clear sentence what the $r^2$ value in the table means. What meaning does $r=0.4599$ (the square root of $r^2$) have?

We can say that 21% of the variation in 6th semester college GPA values is explained by the set of independent variables in the model. In a multiple regression setting, the value of $r$ tells us nothing.
j. The sample analyzed for this research was comprised of 224 randomly selected students. Do you have any concerns about the model results given this piece of information?

*The point here was to get you to look at your notes about how sample size and model size are related. At a bare minimum, most textbooks and analysts will tell you that you need 10-30 observations in your sample for every independent variable in your model. We have five independent variables, so we’d like a minimum of 50-150 observations. We are OK with this sample size.*