The Swift meatpacking plant in Marshalltown, Iowa, employs thousands of Latino men and women, many of whom may not be US citizens. The plant has been raided at least twice by US government agencies, most notably in 1996 by the Immigration and Naturalization Service (after which 148 undocumented workers were deported) and in 2006 by Immigration and Customs Enforcement (after which 90 workers were arrested). Whereas many Marshalltown residents have interpreted these raids as just punishment for the “criminal aliens” in their midst, many others have expressed their appreciation for the “color” Latino and Latina residents add to their otherwise mundane Midwestern existence.

You believe that after the mid-1990s attitudes toward Latino/a people improved among Marshalltown residents. To evaluate this, you decide to do a content analysis of editorials and “letters to the editor” (LtEs) published in the Marshalltown newspaper, the Times-Republican between 1996 and 2010. You begin by randomly sampling one editorial or LtE per month for each of the 15 years of your study. This yields a sample size of 180 (=15x12) editorials. Next, you count the number of times that Latino/a workers were referred to as “illegal” or “criminal” in each editorial. Among the 180 editorials, you find an average of 3.4 mentions of one of these two words. The standard deviation among these word-mentions is 2.1 mentions.

a. What is the unit of analysis in your study? [weight 1]

   The editorial or “letter to the editor” (LtE)

b. Calculate a 95% confidence interval for the average number of mentions of Latino/a workers as “illegal” or “criminal” among editorials or LtEs that appeared in the Times-Republican between 1996 and 2010. [weight 4]

   The Central Limit Theorem holds, because \( n = 180 > 30 \). From the z-table, we find that \( z_{0.025} = 1.96 \).

   \[
   \hat{\sigma}_x = \frac{2.1}{\sqrt{180}} = \frac{2.1}{13.42} = .157
   \]

   \[
   \Delta = \frac{z}{\sqrt{\hat{\sigma}_x}} = 1.96 \times .157 = .307
   \]

   And so the confidence interval, \( \bar{x} \pm \Delta \), is \( 3.4 \pm .307 \), or \( (3.1, 3.7) \).
For parts c through h (i.e., the rest of the exam): Since many of the editorials and LtEs in your sample contain no mentions of Latino/a workers as either “illegal” or “criminal,” you decide to assign to each editorial a “1” if at least one such mention was made in it, and a “0” if no such mention was made. Below are the results of this operationalization of the variable, MENTIONS, among three 5-year periods.

Table 1: Numbers of times that Latino/a workers were referred to as “illegal” or “criminal” in the Times-Republican, 1996-2010*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) one or more mentions</td>
<td>12</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>(0) no mention</td>
<td>48</td>
<td>50</td>
<td>57</td>
</tr>
</tbody>
</table>

* n=180

c. Obtain the expected frequency (i.e., $f_e$) of editorials that mentioned Latino/a workers once or more as “illegal” or “criminal” and that were published between 2006 and 2010.

\[
f_e = \hat{\Pr}(\text{MENTIONS} = 1) \times \hat{\Pr}(\text{YEAR} = 2006 - 2010) \times n
\]

\[
= \frac{25}{180} \times \frac{60}{180} \times 180 = 8.33
\]

d. Estimate the joint probability that a Times-Republican editorial mentioned Latino/a workers as “illegal” or “criminal” and was published between 2006 and 2010.

\[
\hat{\Pr}(\text{MENTIONS} = 1 \text{ and } \text{YEAR} = 2006 - 2010) = \frac{3}{180} = .017
\]
e. Given that the chi-square statistic for Table 1 equals 6.22, do you have statistically significant evidence at the .05 level of significance in support of your theory that after the mid-1990s attitudes toward Latino/a people improved among Marshalltown residents? (Hints: Explain your answer. Also, since you are already given the value of chi-square for this problem, do not waste your time calculating chi-square yourself.) [weight 6]

The degrees of freedom in Table 1 are \((r-1)(c-1) = (2-1)(3-1) = 1 \times 2 = 2\).

At the .05 significance level, the critical value of chi-square from probability Table C is 5.991, which is smaller than the empirical value of chi-square for the table. Thus I conclude that there is evidence of a statistical dependence between YEAR and MENTIONS.

To evaluate whether or not this evidence supports my theory, I need to compare a few conditional probabilities. In particular, if attitudes have improved over time, the probability of mentions in later years should be smaller than the probability of mentions in earlier years. Since \(\Pr(MENTIONS=1|\text{YEAR}=1996-2000)=12/60=.2 > \Pr(MENTIONS=1|\text{YEAR}=2006-2010)=3/60=.05\), it seems that the dependence in Table 1 is in the direction suggested by (thereby supporting) my theory.
f. Calculate a point estimate of the proportion of all editorials published in the Times-Republican between 1996 and 2010 within which Latino/a workers were referred to as “illegal” or “criminal.” [weight 2]

\[
\hat{\pi} = \frac{12 + 10 + 3}{180} = \frac{25}{180} = .139
\]

g. How large a sample would be needed at the .01 significance level to ensure that a point estimate like that calculated in part f (namely, a point estimate of the proportion of all editorials published in the Times-Republican between 1996 and 2010 within which Latino/a workers were referred to as “illegal” or “criminal”) would take a value within .03 of the true value of this proportion? (Hint: Be sure to justify your choice of a variance estimate.) [weight 5]

\[
\Delta = .03 \quad \alpha = .01 \quad z_{\frac{\alpha}{2}} = z_{.005} = 2.576
\]

Justification: I have a “reasonable guess” based on my sample of 180, making a larger sample unnecessary. Another justification might be for the most conservative estimate in which .5 instead of .139 is used.

From the Big Picture: \[ \Delta = z_{\frac{\alpha}{2}} \hat{\sigma}_{\hat{\pi}} \]

Thus \[ .03 = 2.576 \sqrt{\frac{.12}{\sqrt{n}}} \]

or \[ \sqrt{n} = \frac{2.576 \times .346}{.03} = 29.7 \]

or \[ n = 881.8 \]

Rounding up, the answer is \[ n = 882 \]. If \( \pi_0 = .5 \) is assumed, the answer is \[ n=1844 \].
h. Based on the 25 editorials listed in the first row of Table 1, what is the value of the variable, YEAR, at the 50\textsuperscript{th} percentile among the editorials in which Latino/a workers were mentioned as “illegal” or “criminal” one or more times? ( Hint: Your answer should NOT take into account the 155 editorials in which no such mentions were made. Only consider the 25 other editorials when answering this question.) [weight 3]

What is needed here is the value at the \[\left(\frac{50 \times 25}{100} + .5\right)\textsuperscript{th}\] (i.e., the 13\textsuperscript{th}) observation. The 13\textsuperscript{th} observation among the editorials listed in Table 1’s first row is the first of 10 editorials in the year-range of 2001-2005. If we consider the 10 editorials as if they were uniformly distributed across this range, the value of YEAR would be 2001 at the 50\textsuperscript{th} percentile. (No credit lost if “2001-2005” is the answer given.)