Module 22: Proportions: One Sample

This module presents confidence intervals and tests of hypotheses for proportions for the situation with one random sample from a population.
Proportions

\[ P = \text{Population parameter, the proportion of population with characteristic} \]

\[ x = \text{Number in sample with the characteristic} \]

\[ n = \text{Total number in sample} \]

\[ p = \frac{x}{n}, \text{the sample estimate of the proportion with the characteristic} \]
Hypothesis tests and confidence intervals are based on the normal approximation to the binomial distribution. For the hypothesis:

\[ H_0: P = P_0 \text{ vs } H_1: P \neq P_0 \]

The options for the test statistic are:

\[
\begin{align*}
    z &= \frac{p - P_0}{\sqrt{p(1-p)/n}} \\
    z &= \frac{p - P_0}{\sqrt{P_0(1-P_0)/n}} \\
    z &= \frac{p - P_0}{\sqrt{P_0 Q_0/n}}
\end{align*}
\]

where \( q = 1 - p \)
Appointment-Keeping Behavior Re-Evaluated

PHILIP HEITZ, M.D., M.S.P.H., AND PAULA L. STAMPS, M.S., PH.D.

Abstract: Many of the traditional approaches to the problem of appointment-keeping behavior have ignored the organizational factors that may be implicated in differentially high broken appointment rates leading to an implicit assumption that low-income and ethnic minority patients will be more likely to break appointments. A case study at a Model Cities Health Center which maintains a kept appointment rate of 83 per cent examined the relationship of broken appointments to age, sex, ethnic background, and payment mechanisms. The results suggest alternative explanations for differentially high broken appointment rates centering on the role of the institution in reinforcing appointment-keeping behavior. (Am. J. Public Health 67:1033-1036, 1977)

Broken appointment rates in ambulatory care facilities range from 16 per cent to 44 per cent. It is important to decrease the broken appointment rate as much as possible in order to provide for the best utilization of the clinic's staff members. Explorations of broken appointment rates have concentrated primarily upon factors related to the patients, with emphasis placed upon demographic characteristics, such as socioeconomic status, race, age, sex, and educational levels, with other factors such as attitudes toward health care and personality factors also noted. Occasionally, additional factors have also been investigated, including the effect of weather and distance, and variables associated with the organization itself such as the appointment system, staffing patterns, and information flow in the organization. Most studies have concluded with acceptance of a primary myth which is that low-income patients do not keep appointments as well as people in middle class socioeconomic groups. This has been consistently repeated by several investigators over the last ten years.

The conclusion of the relationship between low income and broken appointments has been propounded for ambulatory care in fee-for-service settings, psychiatric care, and ambulatory care in a prepaid setting. The magnitude of the difference varies, but in one of the better designed studies, Greenfield compared appointment-keeping rates of a prepaid group and an OEO Comprehensive Neighborhood Health Care Program that was part of the same medical system. The broken appointment rate for the pre-paid health plan sample was 8 per cent and for the OEO health plan sample 25 per cent.

Confounding variables with low social class include age, ethnic identification, and educational levels. In general, the conclusion of most studies is that the older the patients the more likely they are to keep an appointment; and the more education of ethnic minorities, especially Blacks and Spanish-speaking are more likely to break appointments.

Patients with lower educational levels are generally considered to be poorer keepers of appointments, although recent studies did not confirm this relationship for psychiatric patients. Social disorganization of urban families, especially lower income families, has also been related to poor appointment keeping behavior.

All of these studies have in common the emphasis upon the failure of the patient to keep an appointment, without analyzing the organizational factors which might be responsible for reinforcing poor appointment keeping behavior. In those studies where efforts were made to increase the kept appointment rate, some interesting observations have been made with respect to the role of the health facility in broken appointments. Three major factors can be identified as being related to appointment-keeping behavior: physician continuity, communication systems, and type of appointment system. Becker found that physician continuity, i.e., seeing the same physician each time, was positively correlated to appointment-keeping among a group of low-income, mostly Black patients. Increased efforts at communication, including various reminders for missed appointments, also seems to be effective in reducing the broken appointment.
Methodology and Data Collection

Data were collected for the adult medical clinics for a 12-month period of time between April 1, 1974 and March 31, 1975, on number of appointments kept and broken, sex, age, ethnic group, and payment mechanism. Chi-square was utilized at a significance level of 5 per cent to test for the existence of relationships between these variables. The mobility data for each variable were analyzed. Additionally, a chi-square test on the combined 12 months was calculated. There were a total of 3,172 visits recorded to the adult medical clinic during this time period.

Results

The overall kept appointment rate in the adult medical clinics during the study period was 85 per cent. Out of the 3,172 adult patient visits, 34 per cent were male and 66 per cent were female. As can be seen from Table 1, there were no significant differences in appointment-keeping behavior between males and females. The age distribution of the population was broken into two categories using age 45 as the cut-off point. As with other studies, the data in Table 1 reveal a tendency for older patients to keep appointments better than younger patients. It should be added that the differences with respect to age are significant for only three months (June, October, and November), but these differences are large enough to contribute to a statistical difference over the entire 12-month period.

The majority of the clinic’s adult patients are Puerto Ricans (69 per cent), with 27 per cent of the patients classified as “Other” and 4 per cent classified as Black. Table 1 also shows that there is a statistical relationship between ethnic group membership and appointment-keeping behavior, with Puerto Ricans having a higher broken appointment rate than Whites and Blacks. When these data were analyzed for specific months, May, June, August, September, and November, February, and March show significant differences. Among the months of June and August, neither the appointment-keeping rates of the different ethnic groups nor the bilingual health assistant were present at the Health Center. Both of these persons play key roles in scheduling and keeping clinic patients. In November, February, and March the adult clinics were closed due to their regular closing hours. These changes were not adequately conveyed to the non-English speaking patients. The 18.5 per cent broken appointment rate among this group as compared to the 8.5 per cent broken appointment rate among Whites and the 7 per cent broken appointment rate among Blacks appears to be closely associated with these three events. The clinic’s failure to fully communicate with the non-English speaking patient population, the closing of the clinics for medical reasons, and the changes in the clinic schedules are likely to have contributed to the higher rates of broken appointments for the Spanish-speaking population. As can be seen from Table 1, these factors did not affect the English-speaking clients.

Sixty-five per cent of the Health Center’s patient population is an Medicaid. As can be seen from Table 1, there is no relationship between Medicaid status and appointment-keeping behavior.

In order to further clarify the relationship of ethnic status and payment mechanism, the three ethnic groups were classified by their payment mechanism. Table 2 shows that there is an association between ethnic group and payments, but this is stronger among those not on Medicaid. In general, the two groups share the same characteristics as the total clinic sample, and thus reflects the same unmeasured problems that have been previously cited with respect to language problems.

Summary and Conclusions

The data presented suggest that the difference in the kept appointment rates of different ethnic groups does not reflect a true ethnic difference in health behavior. For the months when there was no significant difference in appointment-keeping behavior between the White and Puerto Rican population, the appointment staff was present, and there were no changes in clinic days or medical staff. During five out of the eight months when such differences in appointment rates were found, it was because of the absence of the bilingual appointment staff, the changes made in clinic days, and the physician
### TABLE 1—Distribution of Percentage of Kept and Broken Appointments According to Selected Patient Characteristics

<table>
<thead>
<tr>
<th>Patient Characteristic</th>
<th>N</th>
<th>% Kept (N = 2,603)</th>
<th>% Not Kept (N = 469)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1,109</td>
<td>85.39</td>
<td>14.61</td>
</tr>
<tr>
<td>Female</td>
<td>2,063</td>
<td>64.15</td>
<td>15.88</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 45</td>
<td>2,603</td>
<td>83.00</td>
<td>17.00</td>
</tr>
<tr>
<td>45+</td>
<td>489</td>
<td>90.37</td>
<td>9.63</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Puerto Rican</td>
<td>2,206</td>
<td>81.46</td>
<td>18.54</td>
</tr>
<tr>
<td>White</td>
<td>837</td>
<td>91.52</td>
<td>8.48</td>
</tr>
<tr>
<td>Black</td>
<td>129</td>
<td>93.02</td>
<td>6.98</td>
</tr>
<tr>
<td><strong>Payment Mechanism</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Welfare</td>
<td>2,229</td>
<td>63.40</td>
<td>16.60</td>
</tr>
<tr>
<td>Non-Welfare</td>
<td>943</td>
<td>87.36</td>
<td>12.62</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>3,172</td>
<td>84.58</td>
<td>15.42</td>
</tr>
</tbody>
</table>

\[X^2 = 8.54; df = 1; p > .05\]
\[X^2 = 24.22; df = 1; p < .05\]
\[X^2 = 53.94; df = 2; p < .05\]
\[X^2 = 8.05; df = 1; p > .05\]
Example of one sample hypothesis test and confidence interval

Table 1 on the following page indicates that 85.39% of the 1,109 males kept their appointments. Suppose that you had worked in this area previously and the percent of appointments kept before had always been 90% or higher. A question this raises is whether this observed rate is sufficiently different from 90% for you to think your population was different. That is, you would like to test the hypothesis:

\[ H_0: P_M = 0.90 \] vs \[ H_1: P_M \neq 0.90 \]
Hypothesis test for males keeping appointments

1. **The hypothesis:**
   \[ H_0: P_M = 0.90 \text{ vs } H_1: P_M \neq 0.90 \]

2. **The assumptions:**
   Independence

3. **The \( \alpha \) level:**
   \( \alpha = 0.05 \)

4. **The test statistic:**
   \[ z = \frac{p - P_M}{\sqrt{\frac{p(1-p)}{n}}} \]

5. **The rejection region:**
   Reject \( H_0: P_M = 0.90 \) if \( z \) is not between \( \pm 1.96 \)
6. The result:

\[ p - \frac{x}{n} = \frac{947}{1,109} = 0.85 \]

\[ P_M = 0.90 \]

\[ z = \frac{p - P_M}{\sqrt{\frac{p(1-p)}{n}}} \]

\[ = \frac{0.85 - 0.90}{\sqrt{(0.85)(1-0.85)} / 1,109} \]

\[ = \frac{0.85 - 0.90}{\sqrt{(0.85)(0.15) / 1,109}} = \frac{-0.05}{\sqrt{0.000115}} = \frac{-0.05}{0.0107} = -4.66 \]

7. The conclusion: Reject H₀: \( P_M = 0.90 \) since \( z \) is not between \( \pm 1.96 \).
Confidence Interval for $P$, with $\alpha = 0.05$

\[ C = \left[ p - 1.96 \sqrt{\frac{p(1-p)}{n}} , p + 1.96 \sqrt{\frac{p(1-p)}{n}} \right] = 0.95 \]
Appointment keeping behavior for males

\[ p = \frac{x}{n} = \frac{947}{1,109} = 0.85 \]

\[
C = \left[ p - 1.96 \sqrt{\frac{p(1-p)}{n}} < P < p + 1.96 \sqrt{\frac{p(1-p)}{n}} \right] = 0.95
\]

\[
C = \left[ 0.85 - 1.96 \sqrt{\frac{(0.85)(0.15)}{1,109}} < P < 0.85 + 1.96 \sqrt{\frac{(0.85)(0.15)}{1,109}} \right] = 0.95
\]

Note: \[ \sqrt{\frac{(0.85)(0.15)}{1,109}} = \sqrt{0.000115} = 0.0107 \]

\[
C[0.85 - 1.96(0.0107) < P < 0.85 + 1.96(0.0107)] = 0.95
\]

\[
C[0.85 - 0.021 < P < 0.85 + 0.021] = 0.95
\]

\[
C[0.829 < P < 0.871] = 0.95
\]
Atopy in Children and Parental Social Class

ABSTRACT

Objectives. This analysis was conducted to determine whether atopic disorders were related to social class in a pediatric population of a former socialist country.

Methods. A cross-sectional study of 2,471 schoolchildren was carried out in 1992 and 1993 in 5 states in the former East Germany. Parents completed a standardized questionnaire regarding health events and lifestyle factors. In addition, skin-prick tests were performed and total serum immunoglobulin E (IgE) was determined.

Results. Lifetime prevalence rates for atopic diseases and rates of atopic sensitization were highest in children from social class IIIA (in which parents had more than 10 years of formal education) and lowest in social class III (less than 10 years of parental education), while rates in social class IIIIB (10 years of parental education were completed at an intermediate level).

Conclusion. The data confirmed the association that in formerly socialist countries social stratification remains under the social system, which were influenced by a social gradient in health outcomes. The findings support the hypothesis that access to modern health care could be seen reasons for the increasing rate of atopic disorders during the last 3 decades. (Am J Public Health. 1999;89(11):19-25)

Joschka Heinrich, DrMed, Manel A. Pongor, DrMed, MPH, Markus Weis, DrMed, Inge F. Goldstein, MSc, MBA, DrMed, and H. Erich Wichmann, DrMed, DrDentMed

In industrialized Western countries, atopic diseases and allergic sensitization in both children and adults have been reported to occur more frequently in higher than in lower socioeconomic groups. Little is known about the distribution of health parameters in general and atopic disease in particular in the different social categories in populations of the formerly communist countries of Eastern Europe. Studies performed mostly after the collapse of the communist system revealed that prevalence rates for atopic diseases and allergic sensitization in children and young adults in Eastern European countries were substantially lower than those observed in Western Europe. The fact that differences in the prevalence of allergic diseases parallel differences in resource distribution within Western societies and between Western and Eastern Europe leads to the speculation that increased access to modern health care might be associated with a rise in allergic disorders. Defining a similar social gradient for atopy in an Eastern European population would provide additional evidence in favor of this theory.

In the past, East German society was regarded as characterized by a relatively uniform distribution of income and living conditions. In the German Democratic Republic, as in all former communist countries, income differences between social groups were relatively small. Education and social status were more equally accessible to all groups in society. Health prevention programs represented an important aspect of the health care system and were implemented by a network of medical facilities in kindergartens, schools, and places of employment. Immunization programs were highly successful and day care facilities employing well-trained personnel were readily available. Nevertheless, in spite of a general lifestyle uniformity, social stratification in former socialist countries remains. This study was designed to determine whether the occurrence of atopic disorders in schoolchildren in the former East Germany was related to parental social class.

Subjects and Methods

Study Population

The cross-sectional health survey was conducted in 1992 and 1993 in 3 communities of the state of Saxony-Anhalt. All eligible schoolchildren in Bernau and Zerbst were invited to participate. One third of the schools from each administrative subdistrict of the former East Germany were arbitrarily selected to represent the total regional area of Bernau. All eligible children in Bernau selected schools were included in the study. There were 2,472 valid subjects, masturbation were distributed to 2,444 children and returned by 2,471. Information on social status was obtained for 2,422 subjects (98.7%). Participation rates in the different phases of the study and age and sex.

<table>
<thead>
<tr>
<th>Region</th>
<th>Social class</th>
<th>Prevalence</th>
<th>%</th>
<th>Prevalence</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1</td>
<td>4.7 (4.0, 5.0)</td>
<td>1132</td>
<td>9.8 (6.5, 25.4)</td>
<td>1091</td>
</tr>
<tr>
<td>II</td>
<td></td>
<td>4.2 (3.0, 3.5)</td>
<td>1117</td>
<td>20.3 (16.5, 32.5)</td>
<td>905</td>
</tr>
<tr>
<td>III</td>
<td></td>
<td>5.5 (4.4, 6.6)</td>
<td>1117</td>
<td>14.1 (16.5, 21.8)</td>
<td>1091</td>
</tr>
<tr>
<td>IV</td>
<td></td>
<td>5.5 (4.4, 6.6)</td>
<td>1117</td>
<td>4.1 (3.0, 5.4)</td>
<td>1091</td>
</tr>
<tr>
<td>V</td>
<td></td>
<td>4.2 (3.0, 3.5)</td>
<td>1117</td>
<td>14.1 (16.5, 21.8)</td>
<td>1091</td>
</tr>
<tr>
<td>VI</td>
<td></td>
<td>5.5 (4.4, 6.6)</td>
<td>1117</td>
<td>4.1 (3.0, 5.4)</td>
<td>1091</td>
</tr>
</tbody>
</table>

Note: CI confidence interval

The results of this study indicate that the prevalence of allergic diseases was higher in children from social class I compared to classes II, III, IV, and V. The prevalence rate for allergic diseases was highest in social class I, followed by classes II and III, and lowest in classes IV and V. The differences were statistically significant, with a p-value of <0.05.

The study also found that the prevalence of serum IgE was highest in social class I, followed by classes II and III, and lowest in classes IV and V. The differences were statistically significant, with a p-value of <0.05.

The study concludes that the prevalence of allergic diseases and serum IgE levels are higher in children from social class I compared to children from social classes II, III, IV, and V. The differences were statistically significant, with a p-value of <0.05.
TABLE 2—Lifetime Prevalence of Atopic Diseases and Characteristic Symptoms as Reported by Parents and Point Prevalence of Physician-Diagnosed Atopic Dermatitis, Allergic Sensitization, and Elevated Total Serum IgE in Children Aged 5 to 14 Years (n = 2402), by Social Class: Sachsen-Anhalt, Germany, 1992–1993

<table>
<thead>
<tr>
<th>Atopic disease or symptom (questionnaire information)</th>
<th>Social class&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Social class&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Social class&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
<td>III</td>
</tr>
<tr>
<td></td>
<td>No.</td>
<td>Prevalence, % (95% CI)</td>
<td>No.</td>
</tr>
<tr>
<td>Asthma or wheezy bronchitis&lt;sup&gt;b&lt;/sup&gt;</td>
<td>191</td>
<td>4.7 (2.2, 8.8)</td>
<td>1173</td>
</tr>
<tr>
<td>Wheezing (more than once)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>173</td>
<td>8.7 (4.9, 13.9)</td>
<td>1117</td>
</tr>
<tr>
<td>Eczema&lt;sup&gt;b&lt;/sup&gt;</td>
<td>190</td>
<td>4.2 (1.8, 8.1)</td>
<td>1180</td>
</tr>
<tr>
<td>Eczematous rash&lt;sup&gt;b&lt;/sup&gt;</td>
<td>189</td>
<td>5.5 (4.9, 13.4)</td>
<td>1170</td>
</tr>
<tr>
<td>Allergic rhinitis&lt;sup&gt;b&lt;/sup&gt;</td>
<td>190</td>
<td>1.1 (0.1, 3.8)</td>
<td>1180</td>
</tr>
<tr>
<td>Allergic rhinitis&lt;sup&gt;b&lt;/sup&gt;</td>
<td>190</td>
<td>10.5 (6.5, 15.8)</td>
<td>1170</td>
</tr>
<tr>
<td>Atopic dermatitis (dermatologic examination)</td>
<td>163</td>
<td>0.0 (0.0, 2.2)</td>
<td>1053</td>
</tr>
<tr>
<td>Allergic sensitization (skin-prick test)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At least 1 of 12 tested allergens</td>
<td>165</td>
<td>11.5 (7.1, 17.4)</td>
<td>1057</td>
</tr>
<tr>
<td>Mite allergens only (der p1 or der 11)</td>
<td>165</td>
<td>3.0 (1.0, 6.9)</td>
<td>1057</td>
</tr>
<tr>
<td>Fungi allergens only (aspergillus, alternaria)</td>
<td>165</td>
<td>5.5 (2.5, 10.1)</td>
<td>1057</td>
</tr>
<tr>
<td>Pollen (local grasses, hazel, birch, plantain, mugwort)</td>
<td>165</td>
<td>7.3 (3.8, 12.4)</td>
<td>1057</td>
</tr>
<tr>
<td>Can epithelium</td>
<td>165</td>
<td>0.0 (0.0, 2.2)</td>
<td>1057</td>
</tr>
<tr>
<td>At least 2 of 12 tested allergens</td>
<td>165</td>
<td>6.5 (4.7, 13.8)</td>
<td>1057</td>
</tr>
<tr>
<td>Total serum IgE ≥ 140 IU/mL&lt;sup&gt;c&lt;/sup&gt;</td>
<td>151</td>
<td>27.8 (20.8, 35.7)</td>
<td>904</td>
</tr>
</tbody>
</table>

Note. CI = confidence interval.

<sup>a</sup>The social level of a family was defined by the highest school grade completed by either the mother or the father of the study subject.

<sup>b</sup>Families with parental education of less than 10 grades were defined as belonging to social class I; 10 grades, class II; more than 10 grades, class III.

<sup>c</sup>Physician-diagnosed diseases as reported by parents.

<sup>d</sup>Perceived by parents.

<sup>e</sup>This value represented the 80th percentile of the entire population.
For Social Class I, Asthma or Wheezy bronchitis

\[ n = 191 \quad p = 0.047 \text{ or } 4.7\% \]

1. The hypothesis: \[ H_0: P = 0.10 \text{ vs } H_1: P \neq 0.10 \]

2. The assumptions: Independence

3. The \( \alpha \)-level: \[ \alpha = 0.05 \]

4. The test statistic:
\[
z = \frac{p - 0.10}{\sqrt{\frac{p(1-p)}{n}}}\]

5. The rejection region: Reject \( H_0: P = 0.10 \), if \( z \) is not between \( \pm 1.96 \)
6. The result:

\[ z = \frac{0.047 - 0.100}{\sqrt{0.047(1 - 0.047)} \over 191} \]

\[ z = \frac{0.053}{\sqrt{0.047(0.953)} \over 191} \]

\[ z = \frac{0.053}{\sqrt{0.0448}} = \frac{0.053}{\sqrt{0.0002}} = \frac{0.053}{0.0153} \]

\[ z = 3.46 \]

7. The conclusion: Reject \( H_0: P = 0.10 \) since \( z \) is not between ±1.96
Confidence Interval, with $\alpha = 0.05$

\[ n = 191, \ p = 0.047, \ p(1-p) = 0.047(0.953) = 0.0448 \]

\[ C \left[ p - 1.96 \sqrt{\frac{p(1-p)}{n}} \ < \ p < \ p + 1.96 \sqrt{\frac{p(1-p)}{n}} \right] = 0.95 \]

\[ C \left[ 0.047 - 1.96 \sqrt{\frac{0.0448}{191}} < p < 0.047 + 1.96 \sqrt{\frac{0.0448}{191}} \right] = 0.95 \]

\[ C \left[ 0.047 - 1.96(0.0153) < p < 0.047 + 1.96(0.0153) \right] = 0.95 \]

\[ C \left[ 0.047 - 0.03 < p < 0.047 + 0.03 \right] = 0.95 \]

\[ C \left[ 0.017 < p < 0.077 \right] = 0.95 \quad \text{or} \]

\[ C \left[ 1.7\% < p < 7.7\% \right] = 0.95 \quad \text{vs} \quad C \left[ 2.2\% < p < 8.8\% \right] = 0.95 \]