EXERCISES: FINAL REVIEW

STATISTICS 1403: PROBABILITY AND STATISTICS FOR THE BIOSCIENCES

Abstract. Examples of examination problems.

1. Unpaired Two-Sample Tests

1.1. Two-Way Tables. Students were asked whether they drove after drinking in 1983, another group was asked in 1987.

<table>
<thead>
<tr>
<th>drove</th>
<th>year 1983</th>
<th>year 1987</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>1250</td>
<td>991</td>
<td>2241</td>
</tr>
<tr>
<td>no</td>
<td>1387</td>
<td>1666</td>
<td>3053</td>
</tr>
<tr>
<td>total</td>
<td>2637</td>
<td>2657</td>
<td>5294</td>
</tr>
</tbody>
</table>

\[ \hat{\omega} = \frac{n_{1,1} n_{2,2}}{n_{1,2} n_{2,1}} \]

\[ SE(\ln \hat{\omega}) = \sqrt{\frac{1}{n_{1,1}} + \frac{1}{n_{1,2}} + \frac{1}{n_{2,1}} + \frac{1}{n_{2,2}}} \]

\[ \ln(\omega) = \ln(\hat{\omega}) \pm 1.96 SE(\ln \hat{\omega}) \]

(a) Do the proportions differ? What test(s) would be appropriate to test this hypothesis?

(b) If a chi-square test were applied, what would be the critical value for a test at the 95% level?

(c) For the conjectured chi-square test, what is the expected value for students who drove in 1997?
1.2. **Fisher’s Exact Test.** Chipmunks were released either 10 or 100 meters from their home burrow, then chased them (to simulate predator pursuit). Out of 24 female chipmunks released 10 m from their burrow, 16 trilled and 8 did not trill. When released 100 m from their burrow, only 3 female chipmunks trilled, while 18 did not trill.

<table>
<thead>
<tr>
<th>distance</th>
<th>trilled</th>
<th>did not</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 m</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>100 m</td>
<td>3</td>
<td>18</td>
</tr>
</tbody>
</table>

(a) What distribution is used to calculate the p-value for Fisher’s Exact Test?

(b) What is the probability of observing the exact values in the table above, given the row and column totals?
2. Probability Models

2.1. Binomial Distribution. You’re totally unprepared for the quiz. It has 10 multiple-choice questions, each with 5 alternatives. In desperation, you guess all the answers.

(a) What is your expected score out of 10?

(b) What is the probability you get 4 correct?

\[ \Pr(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad \mu = np \quad \sigma^2 = np(1 - p) \]
2.2. **Hypergeometric Distribution.** An experimenter is designing a balanced experiment using completely randomized design. He has 30 mice to allocate to two treatment groups. Unknown to him, 9 of the mice have a birth defect that is sensitive to the treatments in his experiment. What is the probability that one of the groups has 7 defective mice, and the other has 2?

\[ Pr(X = x) = \frac{\binom{A}{x} \binom{B}{n-x}}{\binom{A+B}{n}} \]
2.3. Poisson Distribution. Mesquite seedlings are distributed completely at random in a grassland at a rate of 0.002 per square meter.

\[ \Pr(X = x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!} \quad x = 0, 1, \ldots \quad \mu = \sigma^2 = \lambda t \]

(a) What is the probability of a randomly-chosen hectare containing 10 seedlings? (1 Ha = 10^4 m^2)

(b) What is the probability of a randomly-chosen hectare containing NO seedlings?
Consider this data on the mating patterns by color for a variety of tropical bird:

<table>
<thead>
<tr>
<th></th>
<th>females</th>
<th>red</th>
<th>blue</th>
<th>green</th>
</tr>
</thead>
<tbody>
<tr>
<td>red</td>
<td>25</td>
<td>6</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>blue</td>
<td>10</td>
<td>14</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>green</td>
<td>12</td>
<td>5</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

(a) Assuming independence, what are the expected values for the counts in the table?

(b) What is the chi-square statistic for the observed vs. expected counts?

**Useful Formulas.**

\[ < n_{i,j} > = \frac{n_i n_j}{n_+} \]

\[ D^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \]

\[ df = (r - 1)(c - 1) \]

1Magic numbers: 4, 8.42, 15.83
4. Probability

4.1. Law of Total Probability. Exclusive cell phone use (no land line to their residence) varies by age group:

<table>
<thead>
<tr>
<th>age</th>
<th>% phones</th>
<th>% users</th>
</tr>
</thead>
<tbody>
<tr>
<td>15-24</td>
<td>92</td>
<td>25</td>
</tr>
<tr>
<td>25-44</td>
<td>47</td>
<td>30</td>
</tr>
<tr>
<td>45+</td>
<td>32</td>
<td>45</td>
</tr>
</tbody>
</table>

What is the overall proportion of cell phone users who have only a cell phone?

\[
Pr(A) = \sum_i Pr(A|B_i) Pr(B_i) \quad \sum_i Pr(B_i) = 1
\]

4.2. Bayes’ Theorem. Using the table above, what is the probability that a randomly-selected cell phone customer is over 45 years old, given he does not have a land line to his residence?

\[
Pr(B_i|A) = \frac{Pr(A|B_i) Pr(B_i)}{\sum_k Pr(A|B_k) Pr(B_k)} = \frac{Pr(A|B_i) Pr(B_i)}{Pr(A)}
\]
4.3. **Diagnostic Tests.** For a diagnostic test with a sensitivity of 0.995 and a specificity of 0.99, what is the positive predictive value in a population with a prevalence of 5%?

\[
PPV = \frac{sensitivity \times prevalence}{sensitivity \times prevalence + (1 - specificity) \times (1 - prevalence)}
\]