# Solutions to Problems

1. A study is run to estimate the association between exposure to lead paint in childhood and attention-deficit hyperactivity disorder (ADHD). Data on n=400 children are collected and data on exposure and ADHD diagnosis are shown below.

|  |  |  |
| --- | --- | --- |
| Exposure to Lead Paint | ADHD | No ADHD |
| Yes | 34 | 71 |
| No | 29 | 266 |

1. Estimate the crude or unadjusted relative risk.

=3.31.

1. Estimate the crude or unadjusted odds ratio.

= 4.39.

1. In the study described in Problem 1, data are also available reflecting whether or not the child’s father has ADHD. The relationship between the father’s diagnosis, the child’s exposure to lead paint and the child’s diagnosis are shown below.

|  |  |  |
| --- | --- | --- |
| Father with ADHD |  |  |
| Exposure to Lead Paint | ADHD | No ADHD |
| Yes | 27 | 39 |
| No | 7 | 37 |
|  |  |  |
| Father without ADHD |  |  |
| Exposure to Lead Paint | ADHD | No ADHD |
| Yes | 7 | 32 |
| No | 22 | 229 |

* 1. Estimate the relative risks and odds ratios for a child’s diagnosis of ADHD relative to exposure to lead paint, stratified by the father’s diagnosis of ADHD (i.e., for fathers with and without the diagnosis).

In fathers with ADHD:

=2.57.

= 3.66.

In fathers without ADHD:

=2.03.

= 2.27.

* 1. How do the results in part (a) compare to the results in problem #1?

The crude estimates overstate the magnitude of the associations.

1. Use the data in Problems 1 and Problem 2 to conduct tests of hypothesis to determine whether the father’s diagnosis of ADHD is a confounder. (Hint: Test if there is a relationship between the father’s diagnosis and the child’s exposure and between the father’s diagnosis and the child’s diagnosis.) Is father’s diagnosis of ADHD a confounder? Justify your conclusion.

|  |  |  |
| --- | --- | --- |
| Father’s diagnosis and child’s exposure | Exposed to Lead Paint | Not Exposed to Lead Paint |
| Father with ADHD | 66 | 44 |
| Father without ADHD | 39 | 251 |

Step 1. Set up hypotheses and determine level of significance.

H0: Father’s diagnosis and child’s exposure are independent

H1: H0 is false. =0.05

Step 2. Select the appropriate test statistic.

The formula for the test statistic is in Table 7.8 and is given below.



The condition for appropriate use of the above test statistic is that each expected frequency is 5 or more. In Step 4 we will compute the expected frequencies and ensure that the condition is met.

Step 3. Set up decision rule.

df=(2-1)(2-1)=1. Reject H0 if 2 > 3.84.

Step 4. Compute the test statistic.

|  |  |  |
| --- | --- | --- |
| Father’s diagnosis and child’s exposure | Exposed to Lead Paint | Not Exposed to Lead Paint |
| Father with ADHD | 66  (28.9) | 44  (81.1) |
| Father without ADHD | 39  (76.1) | 251  (213.9) |

The test statistic is computed as follows:



2 = 47.63 + 16.97 + 18.09 + 6.43 = 89.12.

Step 5. Conclusion.

We reject H0 because 89.12 > 3.84. We have statistically significant evidence at =0.05 to show that H0 is false; father’s diagnosis and child’s exposure are not independent (i.e., they are related). Using Table 3, the p-value is p < 0.005.

|  |  |  |
| --- | --- | --- |
| Father’s diagnosis and child’s diagnosis | Child with ADHD | Child without ADHD |
| Father with ADHD | 34 | 76 |
| Father without ADHD | 29 | 261 |

Step 1. Set up hypotheses and determine level of significance.

H0: Father’s diagnosis and child’s diagnosis are independent

H1: H0 is false. =0.05

Step 2. Select the appropriate test statistic.

The formula for the test statistic is in Table 7.8 and is given below.



The condition for appropriate use of the above test statistic is that each expected frequency is 5 or more. In Step 4 we will compute the expected frequencies and ensure that the condition is met.

Step 3. Set up decision rule.

df=(2-1)(2-1)=1. Reject H0 if 2 > 3.84.

Step 4. Compute the test statistic.

|  |  |  |
| --- | --- | --- |
| Father’s diagnosis and child’s diagnosis | Child with ADHD | Child without ADHD |
| Father with ADHD | 34  (17.3) | 76  (92.7) |
| Father without ADHD | 29  (45.7) | 261  (244.3) |

The test statistic is computed as follows:



2 = 16.12 + 3.01 + 6.10 + 1.14 = 26.37

Step 5. Conclusion.

We reject H0 because 26.37 > 3.84. We have statistically significant evidence at =0.05 to show that H0 is false; father’s diagnosis and child’s diagnosis are not independent (i.e., they are related). Using Table 3, the p-value is p < 0.005.

Father’s diagnosis is a confounder as it is related to both the child’s exposure and to the child’s diagnosis.

1. Use the data in Problem 2 to estimate the relative risk and odds ratio for a child’s diagnosis of ADHD relative to exposure to lead paint, adjusted for the father’s diagnosis using the Cochran-Mantel-Haenszel method.





1. The data presented in Problem 1 and Problem 2 are analyzed using multiple logistic regression analysis and the models are shown below. In the models below, the data are coded as follows: p = the proportion of children with a diagnosis of ADHD, Child Exposed and Father’s Diagnosis are coded as 1=yes and 0=no.

(1) = -2.216 + 1.480 Child Exposed

(2) Father with Diagnosis: = -1.665 + 1.297 Child Exposed

(3) Father Without Diagnosis: = -2.343 + 0.823 Child Exposed

(4) = -2.398 + 1.056 Child Exposed + 0.906 Father’s Diagnosis

1. Which model produces the unadjusted odds ratio? Compute the unadjusted odds ratio using the regression coefficient from appropriate model.

= -2.216 + 1.480 Child Exposed

=exp(1.480) = 4.39.

1. Which model produces the odds ratio in father’s without a diagnosis of ADHD. Compute the unadjusted odds ratio using the regression coefficient from appropriate model.

= -2.343 + 0.823 Child Exposed

=exp(0.823) = 2.28.

NOTE: The difference between the odds ratio estimated from the logistic regression model and that computed directly (in problem #2) is due to rounding.

1. What is the odds ratio adjusted for father’s diagnosis?

=exp(1.056) = 2.88.

1. A study is conducted in patients with HIV. The primary outcome is CD4 cell count which is a measure of the stage of the disease. Lower CD4 counts are associated with more advanced disease. The investigators are interested in the association between vitamin and mineral supplements and CD4 count. A multiple regression analysis is performed relating CD4 count to use of supplements (coded as 1=yes, 0=no) and to duration of HIV, in years (i.e., the number of years between the diagnosis of HIV and the study date). For the analysis, Y=CD4 count.

= 501.41 + 12.67 Supplements – 30.23 Duration of HIV

1. What is the expected CD4 count for a patient taking supplements who has had HIV for 2.5 years?

= 501.41 + 12.67 (1) – 30.23 (2.5) = 438.51.

1. What is the expected CD4 count for a patient not taking supplements who was diagnosed with HIV at study enrollment?

= 501.41 + 12.67 (0) – 30.23 (0) = 501.41.

1. What is the expected CD4 count for a patient not taking supplements who has had HIV for 2.5 years.

= 501.41 + 12.67 (0) – 30.23 (2.5) = 425.84.

1. If we compare two patients and one has had HIV for 5 years longer than the other, what is the expected difference in their CD4 counts ?

The patient who has had HIV for 5 years longer will have a lower CD4 count by approximately 5(30.23) = 151.15 units.

1. A clinical trial is run to evaluate the efficacy of a new medication to relieve pain in patients undergoing total knee replacement surgery. In the trial, patients are randomly assigned to receive either the new medication or the standard medication. After receiving the assigned medication, patients are asked to report their pain on a scale of 0-100 with higher scores indicative of more pain. Data on the primary outcome are shown below.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Sample Size | Mean Pain Score | Standard Deviation of Pain Score |
| New Medication | 60 | 30.31 | 7.52 |
| Standard Medication | 60 | 53.85 | 7.44 |

Because procedures can be more complicated in older patients, the investigators are concerned about confounding by age. For analysis, patients are classified into two age groups, less than 65 and 65 years of age and older. The data are shown below.

|  |  |  |  |
| --- | --- | --- | --- |
| Age < 65 Years | Sample Size | Mean Pain Score | Standard Deviation of Pain Score |
| New Medication | 40 | 25.30 | 2.46 |
| Standard Medication | 25 | 45.51 | 1.83 |
| Total: Age < 65 Years | 65 | 33.07 | 10.16 |
|  |  |  |  |
| Age 65+ Years | Sample Size | Mean Pain Score | Standard Deviation of Pain Score |
| New Medication | 20 | 40.33 | 2.16 |
| Standard Medication | 35 | 59.80 | 2.49 |
| Total: Age 65+ | 55 | 52.72 | 9.74 |

1. Is there a statistically significant difference in mean pain scores between patients assigned to the new medication as compared to the standard medication? Run the appropriate test at =0.05. (Ignore age in this analysis.)

Step 1. Set up hypotheses and determine level of significance

H0: = 2

H1: ≠ 2 =0.05

Step 2. Select the appropriate test statistic.

The appropriate test statistic is.

Check whether the assumption of equality of population variances is reasonable: s12/s22 = 7.522/7.442 = 1.02. The ratio of the sample variances is between 0.5 and 2 suggesting that the assumption of equality of population variances is reasonable.

Step 3. Set up decision rule.

This is a two-tailed test, using a Z statistic and a 5% level of significance. The appropriate critical values can be found in Table 1C and the decision rule is as follows:

Reject H0 if Z < -1.960 or is Z > 1.960.

Step 4. Compute the test statistic.

We first compute Sp, the pooled estimate of the common standard deviation.



=7.48.

Now the test statistic,



Step 5. Conclusion.

We reject H0 because -17.18 < -1.960. We have statistically significant evidence at =0.05 to show that there is a difference in mean pain scores between treatments. The p-value can be found in Table 1C and is equal to p < 0.0001.

1. Use the data in Problem 7 to determine whether age a confounding variable. Run the tests of hypothesis to determine whether the age is related to treatment assignment and whether there is a difference in mean pain scores by age group. Is age a confounder? Justify your conclusion.

|  |  |  |
| --- | --- | --- |
| Age Group and Treatment | New  Medication | Standard Medication |
| Age < 65 | 40 | 25 |
| Age 65+ | 20 | 35 |

Step 1. Set up hypotheses and determine level of significance.

H0: Age and treatment are independent

H1: H0 is false. =0.05

Step 2. Select the appropriate test statistic.

The formula for the test statistic is in Table 7.8 and is given below.



The condition for appropriate use of the above test statistic is that each expected frequency is 5 or more. In Step 4 we will compute the expected frequencies and ensure that the condition is met.

Step 3. Set up decision rule.

df=(2-1)(2-1)=1. Reject H0 if 2 > 3.84.

Step 4. Compute the test statistic.

|  |  |  |
| --- | --- | --- |
| Age Group and Treatment | New  Medication | Standard Medication |
| Age < 65 | 40  (32.5) | 25  (32.5) |
| Age 65+ | 20  (27.5) | 35  (27.5) |

The test statistic is computed as follows:



2 = 1.73 + 1.73 + 2.05 + 2.05 = 7.56

Step 5. Conclusion.

We reject H0 because 7.56 > 3.84. We have statistically significant evidence at =0.05 to show that H0 is false; age and treatment are not independent (i.e., they are related). Using Table 3, the p-value is p < 0.005.

Age Group and Mean Pain Scores

Step 1. Set up hypotheses and determine level of significance

H0: = 2

H1: ≠ 2 =0.05

Step 2. Select the appropriate test statistic.

The appropriate test statistic is.

Check whether the assumption of equality of population variances is reasonable: s12/s22 = 10.162/9.742 = 1.16. The ratio of the sample variances is between 0.5 and 2 suggesting that the assumption of equality of population variances is reasonable.

Step 3. Set up decision rule.

This is a two-tailed test, using a Z statistic and a 5% level of significance. The appropriate critical values can be found in Table 1C and the decision rule is as follows:

Reject H0 if Z < -1.960 or is Z > 1.960.

Step 4. Compute the test statistic.

We first compute Sp, the pooled estimate of the common standard deviation.



= 9.97.

Now the test statistic,



Step 5. Conclusion.

We reject H0 because -10.74 < -1.960. We have statistically significant evidence at =0.05 to show that there is a difference in mean pain scores between age groups. The p-value can be found in Table 1C and is equal to p < 0.0001.

Age is confounder as it is related to both to treatment and to pain scores.

1. The data presented in Problem 7 are analyzed using multiple linear regression analysis and the models are shown below. In the models below, the data are coded as follows: treatment is coded as 1=new medication and 0=standard medication and age 65+ is coded as 1=yes and 0=no.

= 53.85 – 23.54 Medication

= 45.31 – 19.88 Medication + 14.64 Age 65+

= 45.51 – 20.21 Medication + 14.29 Age > 65 + 0.75 Medication \*Age 65+

Patients < 65: = 45.51 – 20.21 Medication

Patients > 65: = 59.80 – 19.47 Medication

Does it appear that there is effect modification by age? Justify your response using the models above.

It does not appear that there is effect modification by age because the regression coefficient associated with the treatment by age interaction is b3 = 0.75 which is close to zero (and likely non-significant in a statistical test). In addition, the magnitude of the treatment effect is similar in patients < 65 years of age as compared to patients 65 years and older. In both groups, mean pain scores are lower in patients assigned to the new medication. The differences in means are 20.21 and 19.47 units, respectively.

1. Based on your answers to Problem 8 and Problem 9, how should the effect of the treatment be summarized? Should results be reported separately by age group or combined? Should the effect of treatment be adjusted for age? Justify your response using the models presented in Problem 9.

Because there is no effect modification by age, the data can be pooled. However, age is a confounder so the estimate of treatment effect should be adjusted fro age. The difference in mean pain scores adjusted for age is 19.88, with patients assigned to the new medication group scoring, on average, 19.88 units lower than patients assigned to the standard medication group.

1. Using the data in Problem 7, generate a plot to display the mean pain scores by treatment and age group. (Hint: Use Figure 9.4 as an example.)
2. An open label study (where participants are aware of the treatment they are taking) is run to assess the time to pain relief following treatment in patients with arthritis. The following linear regression equations are estimated relating time to pain relief measured in minutes (dependent variable) to participant’s age (in years), gender (coded 1 for males and 0 for females) and severity of disease (a score ranging from 0 to 100 with higher scores indicative of more severe arthritis):

Time to Pain Relief = -24.2 + 0.9 Age

Time to Pain Relief = 11.8 + 19.3 Male Gender

Time to Pain Relief = 3.2 + 0.4 Severity

Time to Pain Relief = -19.8 + 0.50 Age + 10.9 Male Gender + 0.2 Severity

1. What is the expected time to pain relief for a male following treatment?

Time to Pain Relief = 11.8 + 19.3 (1) = 31.1 minutes.

1. What is the expected time to pain relief for a participant aged 50 following treatment?

Time to Pain Relief = -24.2 + 0.9 (50) = 20.8 minutes.

1. In assessing the association between gender and time to pain relief, is there evidence of confounding by age or severity? Briefly justify your answer.

The unadjusted (or crude) association between gender and time to pain relief is 19.3 minutes. After adjustment for age and severity, the association is 10.9, a change of 43% in the parameter estimate. Because the change in the regression coefficient exceeds 10%, this is evidence of confounding by age and severity.

1. A study is run to evaluate risk factors for incident hypertension. All participants are free of hypertension at the start of the study and are followed for 4 years at which time they are re-assessed for hypertension. Risk factors are measured in all participants at the start of the study. A total of n=3182 participants enroll and 1123 develop hypertension over 4 years. A multiple logistic regression model is run and the results are as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| Risk Factor | Regression Coefficient | Chi-Square | P-value |
| Intercept | -18.416 | 746.103 | 0.0001 |
| Age, years | 0.0533 | 95.004 | 0.0001 |
| Male Gender\* | -0.2524 | 6.189 | 0.0129 |
| Systolic Blood Pressure | 0.0629 | 141.417 | 0.0001 |
| Diastolic Blood Pressure | 0.0752 | 80.237 | 0.0001 |
| BMI | 0.0637 | 29.209 | 0.0001 |
| Current Smoker\* | 0.3270 | 10.116 | 0.0015 |

Male gender is coded 1=male, 0=female. Current smoker is coded 1=yes, 0=no.

* 1. What is the relative importance (statistical significance) of the risk factors? Justify your answer.

Using the chi-square statistics to judge relative importance, systolic blood pressure is the most important, followed by age, diastolic blood pressure, BMI, current smoking status and gender.

* 1. Estimate adjusted odds ratios to quantify the effect of gender and current smoking status on incident hypertension.

OR for gender = exp(-0.2524) = 0.78

OR for current smoking status = exp(0.3270) = 1.39

* 1. Who is more likely to develop hypertension, a man or a woman? Justify your answer.

There is an inverse association between gender and incident hypertension. Because gender is coded as 1=male and 0=female, men are less likely to develop hypertension. In fact, men are 22% less likely to develop hypertension as compared to women (OR=0.78), or alternatively the odds of developing hypertension in men are 0.78 times the odds of developing hypertension in women.

1. The following table displays the numbers of participants who develop hypertension by gender and age group in the study described in Problem 13.

|  |  |  |  |
| --- | --- | --- | --- |
| Women | Develop  Hypertension | Do Not Develop Hypertension | Total |
| Age 50+ Years | 255 | 237 | 492 |
| Age < 50 Years | 241 | 916 | 1157 |
| Total | 496 | 1153 | 1649 |
|  |  |  |  |
| Men | Develop  Hypertension | Do Not Develop Hypertension | Total |
| Age 50+ Years | 283 | 188 | 471 |
| Age < 50 Years | 344 | 718 | 1062 |
| Total | 627 | 906 | 1533 |

1. What is the relative risk for hypertension in women 50+ years versus women < 50 years of age?

= 2.49.

1. What is the relative risk for hypertension in men 50+ years versus men < 50 years of age?

= 1.85.

1. Use the data in Problem 14 to estimate the relative risk for hypertension in participants < 50 years of age versus 50 years of age and older, adjusted for gender using the Cochran-Mantel-Haenszel method.

