AssignmentI

Silpa Soni Nallacheruvu (19980824-5287) Hernan Aldana (20000526-4999)

2024-11-20

Summary

Exercise 1:1

Question 1:

Percentage in Favor and Against Legal Abortion by Gender :

The survey results are summarized in a 2×2 table showing responses from 500 women and 600 men regarding their opinions on legal abortion. We calculate the percentages in favor and against legal abortion separately for men and women using the formula:

Percentage (In Favor) $= \frac{\text{Count (In Favor) of the Gender}}{\text{Total Count of the Gender}} \times 100$

Percentage (Against) $= \frac{\text{Count (Against) of the Gender}}{\text{Total Count of the Gender}} \times 100$

For Women :

- `In Favor:`

Percentage (In Favor) for Women $= \frac{309}{500} \times 100 = 61.8\%$

- `Against:`

Percentage (Against) for Women $= \frac{191}{500} \times 100 = 38.2\%$

For Men:

- `In Favor:`

Percentage (In Favor) for Men $= \frac{319}{600} \times 100 = 53.2\%$

- `Against:`

Percentage (Against) for Men $= \frac{281}{600} \times 100 = 46.8\%$

Summary:

Table 1: Percentage In Favor and Against Legal Abortion by Gender

Gender	% In Favor	% Against
Women	61.8	38.2
Men	53.2	46.8

From the analysis, a higher percentage of women (61.8%) support legal abortion compared to men (53.2%). Similarly, a larger percentage of men (46.8%) are against legal abortion compared to women (38.2%).

Question 2:

Approach :

Define Hypotheses :

Null Hypothesis (H_0) : There is no difference in opinions between men and women on legal abortion.

Alternative Hypothesis (H_A) : There is a difference in opinions between men and women on legal abortion.

The formula for Pearson's Chi-Squared Statistic (X^2) is:

$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$

and

The formula for Likelihood Ratio Statistic (G^2) is:

$G^2 = 2 \sum_{i=1}^{r} \sum_{j=1}^{c} O_{ij} \log\left(\frac{O_{ij}}{E_{ij}}\right)$

where O_{ij} is the Observed Count, E_{ij} is the Expected Count of (i,j) cell with i-th index of X and j-th index of Y, r is the total number of rows and c is the total number of columns.

The observed counts are given in the 2×2 table where $O_{11} = 309$, $O_{12} = 191$, $O_{21} = 319$, $O_{22} = 281$.

The r and c values are also implied for a 2×2 table as r = 2 and c = 2.

Let us calculate the Expected Counts for each cell.

Under H_0, expected counts are calculated using:

$E_{ij} = \frac{\text{i-th Row Total}\times\text{j-th Column Total}}{\text{Grand Total}}$

Using the row and column totals:

$E_{11} = \frac{500\times628}{1100} = 285.45$

$E_{12} = \frac{500\times472}{1100} = 214.55$

$E_{21} = \frac{600\times628}{1100} = 342.55$

$E_{22} = \frac{600\times472}{1100} = 257.45$

Let us calculate the each term in Pearson's Chi-Squared Statistic (X^2) :

For cell (1,1): $\frac{(309-285.45)^2}{285.45} = 1.97$

For cell (1,2): $\frac{(191-214.55)^2}{214.55} = 2.84$

For cell (2,1): $\frac{(319-342.55)^2}{342.55} = 1.71$

For cell (2,2): $\frac{(281-257.45)^2}{257.45} = 2.14$

Summing these terms:

$X^2 = 1.97 + 2.84 + 1.71 + 2.14 = 8.66$

Degrees of freedom is defined by (r -1)(c -1) = (2-1)(2-1) = 1.

Let us calculate the P-value for X^2 using a chi-squared table, which gives us :

$P(X^2 > 8.66) = 0.0032$

Let us calculate the each term in Likelihood Ratio Statistic (G^2) :

For cell (1,1): $2 \times 309 \times \log\left(\frac{309}{285.45}\right) = 15.88$

For cell (1,2): $2 \times 191 \times \log\left(\frac{191}{214.55}\right) = 24.20$

For cell (2,1): $2 \times 319 \times \log\left(\frac{319}{342.55}\right) = 14.75$

For cell (2,2): $2 \times 281 \times \log\left(\frac{281}{257.45}\right) = 24.61$

Summing these terms:

$G^2 = 15.88 + 24.20 + 14.75 + 24.61 = 79.44$

Degrees of freedom here (r-1)(c-1) = 1 as well.

Let us calculate the P-value for G^2 using a chi-squared table as well, which gives us :

$P(G^2 > 79.44) \approx 0.003$

Conclusion :

We use a significance level $\alpha = 0.05$ to test our null hypothesis H_0. We see that both the X^2 and G^2 statistics yield small p-values, much lesser than 0.05. Hence, we reject the null hypothesis H_0, which indicates that there is a significant difference in opinions on legal abortion between men and women.

Question 3:

The odds ratio (OR) is a measure used to quantify the strength of association between two categorical variables. For the given 2×2 table regarding opinions on legal abortion among men and women, we calculate the odds ratio and its 95% confidence interval. Depending on the chosen event and conditioning, multiple odds ratios can be formulated. Below, we calculate the odds ratio of being "in favor" of legal abortion for men compared to women.

Defining the Odds :

Odds for Women:

$\text{Odds (Women)} = \frac{\text{In Favor (Women)}}{\text{Against (Women)}} = \frac{309}{191} = 1.6183$

Odds for Men:

$\text{Odds (Men)} = \frac{\text{In Favor (Men)}}{\text{Against (Men)}} = \frac{319}{281} = 1.1352$

Odds Ratio (OR) :

The odds ratio compares the odds of being "in favor" for men to that for women:

$\text{OR} = \frac{\text{Odds (Men)}}{\text{Odds (Women)}} = \frac{1.1352}{1.6183} = 0.7015$

Alternative Odds Ratios can be formulated depending on the conditioning variable:

Odds ratio of being a man vs. a woman among those "in favor":

$\text{OR} = \frac{\text{Men In Favor / Women In Favor}}{\text{Men Against / Women Against}}$

Odds ratio of being against legal abortion for men vs. women:

$\text{OR} = \frac{\text{Men Against / Women Against}}{\text{Men In Favor / Women In Favor}}$

Each odds ratio provides different insights based on the event being studied and the conditioning variable.

To compute the 95% confidence interval for the odds ratio, let us use the logarithm of odds ratio for a more accurate result and then apply the delta method to get the confidence interval of odds ratio from the confidence interval of logarithm of odds ratio.

Logarithm of Odds Ratio:

$\log(\text{OR}) = \log(0.7015) = -0.3544$

Standard Error of log(OR):

The formula for the standard error is:

$\text{SE} = \sqrt{\frac{1}{O_{11}} + \frac{1}{O_{12}} + \frac{1}{O_{21}} + \frac{1}{O_{22}}}$

where O_{ij} are the observed frequencies from the table for the (i,j) cell.

Substituting the values:

$\text{SE} = \sqrt{\frac{1}{309} + \frac{1}{191} + \frac{1}{319} + \frac{1}{281}} = 0.1191$

The 95% confidence interval for log(OR) is given by:

$\log(\text{OR}) \pm z_{0.975} \times \text{SE}$

Using $z_{0.975} = 1.96$:

$-0.3544 \pm 1.96 \times 0.1191 = (-0.5878, -0.1210)$

Exponentiating to Get the Confidence Interval for OR:

To get the confidence interval for the odds ratio, exponentiate the bounds:

$\text{Lower Bound} = e^{-0.5878} = 0.5560$

$\text{Upper Bound} = e^{-0.1210} = 0.8862$

Thus, the 95% confidence interval for the odds ratio is:

$(0.556, 0.886)$

Interpretation

The estimated odds ratio is 0.7015. This means that the odds of men being "in favor" of legal abortion are approximately 70% of the odds for women. The 95% confidence interval (0.556, 0.886) does not include 1, indicating that the difference in odds is statistically significant at the 5% significance level. The estimated odds ratio indicates that the women are significantly more likely to support legal abortion compared to men and the 95% confidence interval of the odds ratio support that indication.

Question 4:

The risk ratio (also called relative risk, RR) quantifies the likelihood of an event occurring in one group relative to another group. For the given 2×2 table regarding opinions on legal abortion among men and women, we calculate the risk ratio and its 95% confidence interval. Below, we calculate the risk ratio of being "in favor" of legal abortion for men compared to women.

Defining the Risk

The risk refers to the probability of being "in favor" of legal abortion, calculated as:

$\text{Risk} = \frac{\text{Number of Individuals "In Favor"}}{\text{Total Number of Individuals in the Gender}}$

Risk for Women:

$\text{Risk (Women)} = \frac{\text{Number of Women In Favor}}{\text{Total (Women)}} = \frac{309}{500} = 0.618$

Risk for Men:

$\text{Risk (Men)} = \frac{\text{Number of Men In Favor}}{\text{Total (Men)}} = \frac{319}{600} = 0.532$

Risk Ratio (Relative Risk) :

The risk ratio (RR) compares the risk of being "in favor" for men to that for women:

$\text{RR} = \frac{\text{Risk (Men)}}{\text{Risk (Women)}} = \frac{0.532}{0.618} = 0.861$

This value indicates that men are about 86.1% as likely as women to be “in favor” of legal abortion.

To compute the 95% confidence interval for the risk ratio, let us use the logarithm of risk ratio for a more accurate result, similar to odds ratio and then apply the delta method to get the confidence interval of risk ratio from the confidence interval of logarithm of risk ratio.

Logarithm of Risk Ratio:

$\log(\text{RR}) = \log(0.861) = -0.1493$

Standard Error of $\log(\text{RR})$:

The formula for standard error is:

$\text{SE} = \sqrt{\frac{1}{\text{In Favor (Men)}} - \frac{1}{\text{Total (Men)}} + \frac{1}{\text{In Favor (Women)}} - \frac{1}{\text{Total (Women)}}}$

Substituting the values:

$\text{SE} = \sqrt{\frac{1}{319} - \frac{1}{600} + \frac{1}{309} - \frac{1}{500}} = 0.0655$

The 95% confidence interval for $\log(\text{RR})$ is given by:

$\log(\text{RR}) \pm z_{0.975} \times \text{SE}$

Using $z_{0.975} = 1.96$:

$-0.1493 \pm 1.96 \times 0.0655 = (-0.2788, -0.0198)$

Exponentiating to Get the Confidence Interval for RR:

To get the confidence interval for the risk ratio, exponentiate the bounds:

$\text{Lower Bound} = e^{-0.2788} = 0.7565$

$\text{Upper Bound} = e^{-0.0198} = 0.9804$

Thus, the 95% confidence interval for the risk ratio is:

(0.756, 0.980)

Interpretation

The estimated risk ratio is 0.861. This means that men are about 86.1% as likely as women to be “in favor” of legal abortion. The 95% confidence interval (0.756, 0.980) does not include 1, indicating that the difference in risk is statistically significant at the 5% level. The estimated risk ratio indicates that the women are significantly more likely to support legal abortion compared to men and the 95% confidence interval of the risk ratio support that indication.

Comparison with Odds Ratio

Odds Ratio: Previously calculated as 0.7015, quantifying the odds of being “in favor” for men compared to women.

Risk Ratio: Calculated as 0.861, quantifying the relative probability of being “in favor.”

The odds ratio measures the ratio of odds, which can overstate the association, especially when the event probability is high (e.g., large proportions of people “in favor”). The risk ratio provides a more intuitive interpretation as it measures the relative likelihood.

Question 5:

Approach :

Prepare the data and calculate row percentages

```
##        opinion
## gender  favor against  Sum
##   women   309     191  500
##   men     319     281  600
##   Sum     628     472 1100
```

```
##        opinion
## gender      favor   against       Sum
##   women 0.6180000 0.3820000 1.0000000
##   men   0.5316667 0.4683333 1.0000000
```

Calculate X2, G2 and p-values

```
##
##  Pearson's Chi-squared test
##
## data:  tab1
## X-squared = 8.2979, df = 1, p-value = 0.003969
```

```
## Call:
## loglm(formula = ~gender + opinion, data = tab1)
##
## Statistics:
##                       X^2 df    P(> X^2)
## Likelihood Ratio 8.322320  1 0.003916088
## Pearson          8.297921  1 0.003969048
```

Calculate odds ratio and 95% confidence interval

```
## $data
##        opinion
## gender  favor against Total
##   women   309     191   500
##   men     319     281   600
##   Total   628     472  1100
##
## $measure
##        odds ratio with 95% C.I.
## gender  estimate    lower    upper
##   women 1.000000       NA       NA
##   men   1.425085 1.119482 1.814113
##
## $p.value
##        two-sided
## gender   midp.exact fisher.exact  chi.square
##   women          NA           NA          NA
##   men   0.003990219  0.004071121 0.003969048
##
## $correction
## [1] FALSE
##
## attr(,"method")
## [1] "Unconditional MLE & normal approximation (Wald) CI"
```

reverse the rows : odds of women to men :

```
## $data
##        opinion
```

```
## gender  favor against Total
##   men     319     281   600
##   women   309     191   500
##   Total   628     472  1100
##
## $measure
##        odds ratio with 95% C.I.
## gender   estimate     lower     upper
##   men   1.0000000        NA        NA
##   women 0.7017126 0.5512336 0.8932701
##
## $p.value
##        two-sided
## gender   midp.exact fisher.exact  chi.square
##   men            NA           NA          NA
##   women 0.003990219  0.004071121 0.003969048
##
## $correction
## [1] FALSE
##
## attr(,"method")
## [1] "Unconditional MLE & normal approximation (Wald) CI"
```

reverse the column : odds of men to women :

```
## $data
##        opinion
## gender  against favor Total
##   women     191   309   500
##   men       281   319   600
##   Total     472   628  1100
##
## $measure
##        odds ratio with 95% C.I.
## gender   estimate     lower     upper
##   women 1.0000000        NA        NA
##   men   0.7017126 0.5512336 0.8932701
##
## $p.value
##        two-sided
## gender   midp.exact fisher.exact  chi.square
##   women          NA           NA          NA
##   men   0.003990219  0.004071121 0.003969048
##
## $correction
## [1] FALSE
##
## attr(,"method")
## [1] "Unconditional MLE & normal approximation (Wald) CI"
```

reverse both : odds of women to men :

```
## $data
##        opinion
## gender  against favor Total
##   men       281   319   600
```

Table 2: Admissions Data by Gender

	admitted	not admitted
men	1198	1493
women	557	1278

```
##   women     191   309   500
##   Total     472   628  1100
##
## $measure
##         odds ratio with 95% C.I.
## gender  estimate    lower    upper
##   men   1.000000       NA       NA
##   women 1.425085 1.119482 1.814113
##
## $p.value
##         two-sided
## gender   midp.exact fisher.exact  chi.square
##   men            NA           NA          NA
##   women 0.003990219  0.004071121 0.003969048
##
## $correction
## [1] FALSE
##
## attr(,"method")
## [1] "Unconditional MLE & normal approximation (Wald) CI"
```

Exercise 1:2

Question 1:

Approach :

We are analyzing the admissions data from the University of California, Berkeley, following the 2x2 contingency table:

Our goal is to perform the following analyses:

1. Calculate the percentages of admitted and not admitted applicants separately.
2. Test for independence between gender and admission using:
    - Pearson's Chi-Squared Test
    - Likelihood Ratio Test
3. Calculate the odds ratio and its 95% confidence interval.
4. Calculate the risk ratio and its 95% confidence interval.
5. Calculating Percentages

For men:

- Total men: $ n_{Men} = 2691 $
- Admitted men: $ a_{Men} = 1198 $

- Not admitted men: $ n_{Men} - a_{Men} = 1493 $

Percentages

Percentage Admitted (Men) = $\left(\frac{1198}{2691}\right) \times 100\% \approx 44.53\%$

Percentage Not Admitted (Men) = $100\% - 44.53\% = 55.47\%$

For women:

- Total women: $ n_{Women} = 1835 $
- Admitted women: $ a_{Women} = 557 $
- Not admitted women: $ n_{Women} - a_{Women} = 1278 $

Percentages

Percentage Admitted (Women) = $\left(\frac{557}{1835}\right) \times 100\% \approx 30.35\%$

Percentage Not Admitted (Women) = $100\% - 30.35\% = 69.65\%$

Table 3: Summary Table

Gender	% Admitted	% Not Admitted
Men	44.53	55.47
Women	30.35	69.65

2. Testing for Independence

Null hypothesis (H_0): Gender and admission are independent.

Alternative hypothesis (H_1): There is an association between gender and admission Status.

Expected counts under H_0 are calculated as:

$E_{ij} = \frac{(\text{Row Total}_i)\times(\text{Column Total}_j)}{\text{Grand Total}}$

1. Men Admitted:

`-$E_{11} = \frac{R_1 \times C_1}{N} $`

2. Men not Admitted:

`-$E_{12} = \frac{R_1 \times C_2}{N}$`

3. Women Admitted:

`-$E_{21} = \frac{R_2 \times C_1}{N} $.`

4. Women not Admitted:

`-$E_{22} = \frac{R_2 \times C_2}{N} $.`

Calculating the Pearson's Chi-Squared Test Statistic:

$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$

Calculating the Likelihood Ratio Test Statistic:

$G^2 = 2 \sum_{i=1}^{2} \sum_{j=1}^{2} O_{ij} \ln\left(\frac{O_{ij}}{E_{ij}}\right)$

where:

- O_{ij}: Observed counts
- E_{ij}: Expected counts

The degrees of freedom are calculated for both statistics as $df = (r-1)(c-1)$

3. Estimating the Odds Ratio

Calculating Odds

- Odds of Admission for Men:

  $\text{Odds}_{\text{Men}} = \frac{\text{Admitted Men}}{\text{Not Admitted Men}}$

- Odds of Admission for Women:

  $\text{OR} = \frac{\text{OddsMen}}{\text{OddsWomen}}$

Calculating the Standard Error of log(OR):

$SE = \sqrt{\frac{1}{a_{\text{Men}}} + \frac{1}{b_{\text{Men}}} + \frac{1}{a_{\text{Women}}} + \frac{1}{b_{\text{Women}}}}$

The 95% confidence interval for log(OR) is calculated as:

$\ln(\text{OR}) \pm Z_{0.975} \times SE$

However we still have to get the confidence interval for the odds ratio by exponentiating the bounds.

4. Estimating the Risk Ratio

Calculating Risks (Probabilities)

- Risk of Admission for Men:

  $P_{\text{Men}} = \frac{\text{Admitted Men}}{\text{Total Men}}$

- Risk of Admission for Women:

  $P_{\text{Women}} = \frac{\text{Admitted Women}}{\text{Total Women}}$

Calculating Risk Ratio (RR)

$\text{RR} = \frac{P_{\text{Men}}}{P_{\text{Women}}}$

```
##
##  Pearson's Chi-squared test
##
## data:  tab2
## X-squared = 92.205, df = 1, p-value < 2.2e-16
```

```
## Call:
## loglm(formula = ~gender + admission, data = tab2)
##
## Statistics:
##                       X^2 df P(> X^2)
## Likelihood Ratio 93.44941  1        0
## Pearson          92.20528  1        0
```

```
## $data
##        admission
## gender  admitted not admitted Total
##   men       1198         1493  2691
##   women      557         1278  1835
##   Total     1755         2771  4526
##
## $measure
##        odds ratio with 95% C.I.
## gender  estimate    lower    upper
```

```
##   men    1.00000       NA       NA
##   women  1.84108 1.624377 2.086693
##
## $p.value
##        two-sided
## gender  midp.exact fisher.exact chi.square
##   men           NA           NA         NA
##   women          0 4.835903e-22 7.8136e-22
##
## $correction
## [1] FALSE
##
## attr(,"method")
## [1] "Unconditional MLE & normal approximation (Wald) CI"
```

```
## $data
##        admission
## gender  admitted not admitted Total
##   men       1198         1493  2691
##   women      557         1278  1835
##   Total     1755         2771  4526
##
## $measure
##        risk ratio with 95% C.I.
## gender  estimate    lower   upper
##   men   1.000000       NA      NA
##   women 1.255303 1.199631 1.31356
##
## $p.value
##        two-sided
## gender  midp.exact fisher.exact chi.square
##   men           NA           NA         NA
##   women          0 4.835903e-22 7.8136e-22
##
## $correction
## [1] FALSE
##
## attr(,"method")
## [1] "Unconditional MLE & normal approximation (Wald) CI"
```