AssignmentI

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

2024-11-20

Summary

This report analyzes opinions on legal abortion among men and women using data from a survey. The responses were summarized in a two-way table, and statistical methods, including percentages, hypothesis testing, odds ratios, and risk ratios, were applied to gain insights. The results are presented with detailed calculations and interpretations.

Exercise 1:1

Question 1: Percentage in Favor and Against Legal Abortion by Gender

Approach :

The percentages for being "in favor" and "against" legal abortion for men and women were calculated using the following formulas :

- Percentage (In Favor) :

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

- Percentage (Against) :

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

Calculations :

```r
# Calculate percentages
women_in_favor <- 309
women_total <- 500
men_in_favor <- 319
men_total <- 600
women_against <- women_total - women_in_favor
men_against <- men_total - men_in_favor

# Percentages
women_in_favor_pct <- round((women_in_favor / women_total) * 100, 1)
women_against_pct <- round((women_against / women_total) * 100, 1)
men_in_favor_pct <- round((men_in_favor / men_total) * 100, 1)
men_against_pct <- round((men_against / men_total) * 100, 1)
```

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

Analysis :

- Women: 61.8% are in favor, while 38.2% are against legal abortion.
- Men: 53.2% are in favor, while 46.8% are against legal abortion.

Question 2: Hypothesis Testing (Pearson's (X^2) and Likelihood Ratio (G^2))

Approach :

1. 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.

2. Expected Counts : Under H_0, expected counts are calculated using:

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

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

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

4. Likelihood Ratio Statistic : 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.

Calculations :

```r
# Observed counts
observed <- matrix(c(309, 191, 319, 281), nrow = 2, byrow = TRUE)
rownames(observed) <- c("Women", "Men")
colnames(observed) <- c("In Favor", "Against")

# Totals
row_totals <- rowSums(observed)
col_totals <- colSums(observed)
grand_total <- sum(observed)

# Expected counts
expected <- outer(row_totals, col_totals) / grand_total

# Pearson's X^2
X2 <- sum((observed - expected)^2 / expected)
```

```r
# Likelihood Ratio G^2
G2 <- 2 * sum(observed * log(observed / expected))

# Degrees of freedom
df <- (nrow(observed) - 1) * (ncol(observed) - 1)

# P-values
p_value_X2 <- pchisq(X2, df = df, lower.tail = FALSE)
p_value_G2 <- pchisq(G2, df = df, lower.tail = FALSE)
```

Output :

Table 2: Hypothesis Testing Results

Statistic	Value	P.value
Pearson's X^2	8.297921	0.0039690
Likelihood Ratio G^2	8.322320	0.0039161

Conclusion :

We use a significance level $\alpha = 0.05$ to test our null hypothesis H_0. 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: Odds Ratio and Confidence Interval

Approach :

The Odds Ratio :

The odds ratio (OR) is a measure used to quantify the strength of association between two categorical variables.

Odds for Women: $\text{Odds (Women)} = \frac{\text{In Favor (Women)}}{\text{Against (Women)}}$

Odds for Men: $\text{Odds (Men)} = \frac{\text{In Favor (Men)}}{\text{Against (Men)}}$

Odds Ratio (OR) : The odds ratio compares the odds of being "in favor" for men to that for women:

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

Calculations :

```r
# Odds calculation
odds_women <- women_in_favor / women_against
odds_men <- men_in_favor / men_against
# Odds ratio
odds_ratio <- odds_men / odds_women
```

Output :

Table 3: Odds Ratio Results

Odds	Value
Odds Women	1.6178010
Odds Men	1.1352313
Odds Ratio	0.7017126

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.

Confidence Interval :

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.

Calculations :

```
# Log odds ratio and standard error
log_odds_ratio <- log(odds_ratio)
se_log_odds <- sqrt(sum(1 / observed))

# Confidence interval
z <- qnorm(0.975)
ci_lower <- exp(log_odds_ratio - z * se_log_odds)
ci_upper <- exp(log_odds_ratio + z * se_log_odds)
```

Output :

Table 4: Odds Ratio and Confidence Interval

Measure	Value
Odds Ratio	0.7017126
95% CI Lower	0.5512336
95% CI Upper	0.8932701

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: Risk Ratio and Confidence Interval

Approach :

The Risk Ratio : The risk ratio (also called relative risk, RR) quantifies the likelihood of an event occurring in one group relative to another group.

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

Risk for Men: $\text{Risk (Men)} = \frac{\text{Number of Men In Favor}}{\text{Total (Men)}}$

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

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

Calculations :

```r
# Risk calculation
risk_women <- women_in_favor / women_total
risk_men <- men_in_favor / men_total

# Risk ratio
risk_ratio <- risk_men / risk_women
```

Output :

Table 5: Risk Ratio Results

Risk	Value
Risk Women	0.6180000
Risk Men	0.5316667
Risk Ratio	0.8603020

Analysis :

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

Confidence Interval : 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.

The formula for standard error is:

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

Calculations :

```r
# Log risk ratio and standard error
log_risk_ratio <- log(risk_ratio)
se_log_risk <- sqrt(1 / women_in_favor - 1 / women_total + 1 / men_in_favor - 1 / men_total)

# Confidence interval
ci_lower_rr <- exp(log_risk_ratio - z * se_log_risk)
ci_upper_rr <- exp(log_risk_ratio + z * se_log_risk)
```

Output :

Table 6: Risk Ratio and Confidence Interval

Measure	Value
Risk Ratio	0.8603020
95% CI Lower	0.7769363
95% CI Upper	0.9526130

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: Verification of Previous Calculations

Approach :

1. Generate Frequency Table and Calculate Row Percentages Verify the calculations for percentages, hypothesis testing, odds ratio, and risk ratio from Questions 1–4 using built-in R functions. The results are compared with the manual calculations presented earlier.

```
##         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
```

2. 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
```

Comparison :

- (X^2) statistic, degrees of freedom, and p-value displayed, matching the manual calculations.
- (G^2) statistic and p-value are also displayed, confirming the earlier results.

3. Calculate Odds Ratio and 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
##   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"
```

Comparison : The odds ratio and its confidence interval are verified, and the results align with the manually calculated values when the appropriate rev option is used.

4. Calculate Risk Ratio and Confidence Interval

```
## $data
##        opinion
## gender  favor against Total
##   women   309     191   500
##   men     319     281   600
##   Total   628     472  1100
##
## $measure
##        risk ratio with 95% C.I.
## gender  estimate    lower    upper
##   women 1.000000       NA       NA
##   men   1.226003 1.065465 1.410731
##
## $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"
```

```
## $data
##        opinion
## gender  favor against Total
##   men     319     281   600
##   women   309     191   500
##   Total   628     472  1100
##
## $measure
##        risk ratio with 95% C.I.
```

```
## gender   estimate     lower     upper
##   men   1.0000000        NA        NA
##   women 0.8156584 0.7088523 0.9385574
##
## $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"
```

```
## $data
##        opinion
## gender  against favor Total
##   women     191   309   500
##   men       281   319   600
##   Total     472   628  1100
##
## $measure
##        risk ratio with 95% C.I.
## gender  estimate     lower    upper
##   women 1.000000        NA       NA
##   men   0.860302 0.7769363 0.952613
##
## $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"
```

```
## $data
##        opinion
## gender  against favor Total
##   men       281   319   600
##   women     191   309   500
##   Total     472   628  1100
##
## $measure
##        risk ratio with 95% C.I.
## gender  estimate    lower    upper
##   men   1.000000       NA       NA
##   women 1.162382 1.049744 1.287107
##
## $p.value
```

Table 7: Admissions Data by Gender

	admitted	not admitted
men	1198	1493
women	557	1278

```
##        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"
```

Comparison : The risk ratio and its confidence interval are verified, and the results align with the manually calculated values when the appropriate rev option is used.

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: $n_{\text{Men}} = 2691$
- Admitted men: $a_{\text{Men}} = 1198$
- Not admitted men: $n_{\text{Men}} - a_{\text{Men}} = 1493$

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

Percentages

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

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

Table 8: 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:

   $E_{11} = \frac{R_1 \times C_1}{N} = \frac{2691 \times 1755}{4526} \approx 1043.46$
2. Men not Admitted:

   $E_{12} = \frac{R_1 \times C_2}{N} = \frac{2691 \times 2771}{4526} \approx 1647.54$
3. Women Admitted:

   $E_{21} = \frac{R_2 \times C_1}{N} = \frac{1835 \times 1755}{4526} \approx 711.54.$
4. Women not Admitted:

   $E_{22} = \frac{R_2 \times C_2}{N} = \frac{1835 \times 2771}{4526} \approx 1123.46.$

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)

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

Calculating the Standard Error of log(RR):

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

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

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

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

Code

```
##
##  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"
```

Results and Interpretation

1. Chi-Squared Test:

$\chi^2 = 92.205, p < 2.2e^{-16}$

The large chi-squared statistic and the small p-value indicate a significant association between gender and admission status. Therefore rejecting the null hypothesis of independence.

2. Likelihood Ratio Test:

$G^2 = 93.98, p < 0.001$

Similarly as the chi-squared test, the likelihood ratio test also indicates a significant association between gender and admission.

3. Odds Ratio:

$\text{OR} = 1.841, 95\% \text{ CI } (1.624, 2.087)$

According to the odds ratio, men have 1.841 times higher odds of being admitted compared to women. The 95% confidence interval does not include 1, indicating a statistically significant difference in admission odds.

4. Risk Ratio:

$\text{RR} = 1.467, 95\% \text{ CI } (1.353, 1.591)$

Men are 1.255 times more likely to be admitted compared to women. The confidence interval does not include 1, indicating a statistically significant difference in admission risk.

Conclusion

There is a statistically significant association between gender and admission status at the University of California, Berkeley during the year of 1975. Men have higher odd and higher probability of admission compared to women.