AssignmentI

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

2024-11-23

Exercise 1:1

We are analyzing opinions on legal abortion among men and women using data from a survey, following the two-way table:

Table 1: Opinions on legal abortion

	In favor	Against
Women	309	191
Men	319	281

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 :

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 $= \frac{281}{600} \times 100 = 46.8\%$

R Output :

Table 2: 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 :

Test for independence between gender and opinions using:

- Pearson's Chi-Squared Test
- Likelihood Ratio Test

1. 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.

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 (χ^2) is:

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

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}: Observed counts of (i,j) cell
- E_{ij}: Expected counts of (i,j) cell
- r: Total number of rows
- c: Total number of columns

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

Calculations :

Expected counts under H_0 are calculated as:

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

For Women:

- In Favor: $E_{11} = \frac{R_1 \times C_1}{N} = \frac{500 \times 628}{1100} \approx 285.45$
- Against: $E_{12} = \frac{R_1 \times C_2}{N} = \frac{500 \times 472}{1100} \approx 214.55$

For Men:

- In Favor: $E_{21} = \frac{R_2 \times C_1}{N} = \frac{600 \times 628}{1100} \approx 342.55.$
- Against: $E_{22} = \frac{R_2 \times C_2}{N} = \frac{600 \times 472}{1100} \approx 257.45.$

Calculating the Pearson's Chi-Squared Test Statistic:

$\chi^2 = \frac{(309-285.45)^2}{285.45} + \frac{(191-214.55)^2}{214.55} + \frac{(319-342.55)^2}{342.55} + \frac{(281-257.45)^2}{257.45} = 1.94 + 2.58 + 1.62 + 2.15 = 8.29$

Calculating the Likelihood Ratio Test Statistic:

$G^2 = 2 \times [309 \times \log\left(\frac{309}{285.45}\right) + 191 \times \log\left(\frac{191}{214.55}\right) + 319 \times \log\left(\frac{319}{342.55}\right) + 281 \times \log\left(\frac{281}{257.45}\right)] = 48.99 - 44.41 - 45.44 + 49.19 = 8.33$

R Output :

Table 3: Hypothesis Testing Results

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

Conclusion :

- Both the X^2 and G^2 statistics yield small p-values, much lesser than 0.05 as the significance level to test our null hypothesis H_0.
- 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 95% Confidence Interval

Approach :

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

- For Women: Odds in Favor: Odds (Women) $= \frac{\text{In Favor (Women)}}{\text{Against (Women)}}$
- For Men: Odds in Favor: 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 :

- For Women: Odds (Women) $= \frac{309}{191} = 1.6178$
- For Men: Odds (Men) $= \frac{319}{281} = 1.1352$
- Odds Ratio for Men: OR $= \frac{1.1352}{1.6178} = 0.7017$

R Output :

Table 4: 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":

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:

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 :

- Standard Error of log(OR): $\mathrm{SE}_{OR} = \sqrt{\frac{1}{O_{11}} + \frac{1}{O_{12}} + \frac{1}{O_{21}} + \frac{1}{O_{22}}} = \sqrt{\frac{1}{309} + \frac{1}{191} + \frac{1}{319} + \frac{1}{281}} = 0.1225$
- 95% confidence interval for log(OR): $\log(\mathrm{OR}) \pm z_{0.975} \times \mathrm{SE}_{OR}$

R Output :

Table 5: 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 indicating that the men are significantly less likely to support legal abortion compared to women, with 70% odds to be specific.
- The 95% confidence interval (0.556, 0.886) does not include 1, indicating a statistically significant difference in the opinions.

Question 4: Risk Ratio and 95% Confidence Interval

Approach :

The risk ratio (also called relative risk, RR) quantifies the likelihood of an event occurring in one group relative to another group. Here, the risk refers to the probability of being "in favor" of legal abortion.

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

- For Women: Risk (Women) $= \frac{309}{500} = 0.618$
- For Men: Risk (Men) $= \frac{319}{600} \approx 0.5316$
- Risk Ratio for Men: RR $= \frac{0.5316}{0.618} \approx 0.8602$

R Output :

Table 6: 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.

Calculations :

- Standard error of $log(RR)$: $\text{SE}_{RR} = \sqrt{\frac{1}{\text{In Favor (Men)}} - \frac{1}{\text{Total (Men)}} + \frac{1}{\text{In Favor (Women)}} - \frac{1}{\text{Total (Women)}}}$

$= \sqrt{\frac{1}{319} - \frac{1}{600} + \frac{1}{309} - \frac{1}{500}} \approx 0.1$

- 95% Confidence Interval for $log(RR)$: $\log(\text{RR}) \pm z_{0.975} \times \text{SE}_{RR}$

R Output :

Table 7: 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 indicating 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 a statistically significant difference in the opinions.

Comparison with Odds Ratio

- Odds Ratio: highlights the relative change of odds, previously calculated as 0.7015, quantifying the odds of men being "in favor" compared to women.
- Risk Ratio: provides a more intuitive interpretation, previously calculated as 0.861, quantifying the relative probability men of being "in favor" compared to women.
- Both the ratios indicate that the women are significantly more likely to support legal abortion compared to men.

Question 5: Verification of Previous Calculations

Approach :

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.

1. Generate Frequency Table 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
```

Comparison : The row percentages are verified, and the results align with the previously calculated values.

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 previous calculations.
- (G^2) statistic and p-value are also displayed, confirming the earlier results.

3. Calculate Odds Ratio and Confidence Interval

- 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"
```

Comparison : The odds ratio of men to women and its confidence interval are verified, and the results align with the earlier values when the columns are reversed from the given table.

4. Calculate Risk Ratio and Confidence Interval

- risk of men to women :

```
## $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"
```

Comparison : The risk ratio of men to women and its confidence interval are verified, and the results align with the earlier values when the columns are reversed from the given table.

Exercise 1:2

Question 1:

Approach :

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

Table 8: Summary Table

	Admitted	Not Admitted
Men	1198	1493
Women	557	1278

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 9: 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
```

```
##        admission
## gender  Admitted Not Admitted
##   Men       1198         1493
##   Women      557         1278
```

```
## Call:
## loglm(formula = ~gender + admission, data = tab2_table)
##
## 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.

Question 2:

Approach :

The approach is similar to the previous question. We will analyze what happens if we replace all numbers in the table with one tenth of their original values and perform the same analysis, which we will compare with the previous results.

Code:

```
## $Original
## $Original$`Chi-Squared Test`
##
##  Pearson's Chi-squared test
##
## data:  data
## X-squared = 92.205, df = 1, p-value < 2.2e-16
##
##
## $Original$`Likelihood Ratio`
## Call:
## loglm(formula = ~gender + admission, data = data)
##
## Statistics:
##                       X^2 df P(> X^2)
## Likelihood Ratio 93.44941  1        0
## Pearson          92.20528  1        0
##
## $Original$`Odds Ratio`
##        odds ratio with 95% C.I.
## gender  estimate    lower    upper
##   Men    1.00000       NA       NA
```

```
##   Women  1.84108 1.624377 2.086693
##
## $Original$`Risk Ratio`
##        risk ratio with 95% C.I.
## gender  estimate    lower   upper
##   Men   1.000000       NA      NA
##   Women 1.255303 1.199631 1.31356
##
##
## $`One-Tenth`
## $`One-Tenth`$`Chi-Squared Test`
##
##  Pearson's Chi-squared test
##
## data:  data
## X-squared = 9.2409, df = 1, p-value = 0.002367
##
##
## $`One-Tenth`$`Likelihood Ratio`
## Call:
## loglm(formula = ~gender + admission, data = data)
##
## Statistics:
##                       X^2 df    P(> X^2)
## Likelihood Ratio 9.364274  1 0.002212556
## Pearson          9.240912  1 0.002366671
##
## $`One-Tenth`$`Odds Ratio`
##        odds ratio with 95% C.I.
## gender  estimate    lower    upper
##   Men   1.000000       NA       NA
##   Women 1.840844 1.239547 2.733825
##
## $`One-Tenth`$`Risk Ratio`
##        risk ratio with 95% C.I.
## gender  estimate    lower   upper
##   Men   1.000000       NA      NA
##   Women 1.255909 1.087858 1.44992
##
##
## $`One-Hundreth`
## $`One-Hundreth`$`Chi-Squared Test`
##
##  Pearson's Chi-squared test
##
## data:  data
## X-squared = 0.77499, df = 1, p-value = 0.3787
##
##
## $`One-Hundreth`$`Likelihood Ratio`
## Call:
## loglm(formula = ~gender + admission, data = data)
##
## Statistics:
```

```
##                          X^2 df  P(> X^2)
## Likelihood Ratio 0.7833632  1 0.3761145
## Pearson          0.7749930  1 0.3786768
##
## $`One-Hundreth`$`Odds Ratio`
##        odds ratio with 95% C.I.
## gender  estimate     lower    upper
##   Men   1.000000        NA       NA
##   Women 1.733333 0.5068344 5.927862
##
## $`One-Hundreth`$`Risk Ratio`
##        risk ratio with 95% C.I.
## gender  estimate     lower    upper
##   Men   1.000000        NA       NA
##   Women 1.231579 0.7812776 1.941418
```

Results and Interpretation:

1. Chi-Squared Test and Likelihood Ratio Test:

The Chi-Squared χ^2 and Likelihood Ratio G^2 statistics, as well as their associated p-values, measure the association between gender and admission. The results for the original, one-tenth, and one-hundredth tables are summarized below:

Table 10: Summary of Chi-Squared and Likelihood Ratio Tests

Dataset	X^2 Statistic	G^2 Statistic	p-value
Original	92.205	93.449	< 2.2 × 10^-16
One-Tenth	9.241	9.364	0.002
One-Hundredth	0.775	0.783	0.379

As it can be seen, both the Chi-Squared and Likelihood Ratio tests show a dramatic decrease as the table is scaled down, reflecting thus the reduced sample size. On the other hand, the p-values show a completely different behavior. While the p-value for the original table is extremely small, indicating a significant association between gender and admission, as the table is scaled down, the p-value increases, suggesting that the association cannot be reliably detected with such a small sample size.

2. Odds Ratio:

The odds ratio quantifies the strength of the association between gender and admission. The results for the original, one-tenth, and one-hundredth tables are summarized below:

Table 11: Odds Ratio for Women Across Datasets

Dataset	Odds Ratio (Women)	95% Confidence Interval
Original	1.841	[1.624, 2.087]
One-Tenth	1.841	[1.240, 2.734]
One-Hundredth	1.733	[0.507, 5.928]

The odds ratio across the three datasets shows a consistent value, varied only slightly. The confidence interval, however, widens as the sample size decreases, reflecting the increased uncertainty in the estimate. So for the original dataset, the odds ratio is 1.841, with a 95% confidence interval of [1.624, 2.087], which is very narrow, reflecting high precision. On the other hand, for the one-hundredth dataset, the odds ratio is 1.733, with a 95% confidence interval of [0.507, 5.928], which is much wider, reflecting low precision.

3. Risk Ratio:

The risk ratio compares the probability of admission for men and women. The results for the original, one-tenth, and one-hundredth tables are summarized below:

Table 12: Risk Ratio for Women Across Datasets

Dataset	Risk Ratio (Women)	95% Confidence Interval
Original	1.255	[1.200, 1.314]
One-Tenth	1.256	[1.088, 1.450]
One-Hundredth	1.232	[0.781, 1.941]

The risk ratio across the three datasets shows a consistent value, varied only slightly. The confidence interval, however, widens as the sample size decreases, reflecting the increased uncertainty in the estimate. So for the original dataset, the risk ratio is 1.255, with a 95% confidence interval of [1.200, 1.314], which is very narrow, reflecting high precision. On the other hand, for the one-hundredth dataset, the risk ratio is 1.232, with a 95% confidence interval of [0.781, 1.941], which is much wider, reflecting low precision. Moreover, for the original dataset, the confidence interval does not include 1, indicating statistical significance. However, for the one-hundredth dataset, the confidence interval includes 1, indicating no statistical significance.

Conclusion:

This question helps us highlight the importance of a sufficiently large sample size for reliabale statistical inference. While relative measures such as odds and risk ratios remain relatively stable across different sample sizes, the precision of the estimates decreases as the sample size decreases. In addition, absolute measures such as χ^2 and G^2 lose significance as the sample size decreases.

Question 3:

Approach :

For this question, we will analyze a completely new 2x2 contingency table, such that:

1. The sample odd ratio($\hat{\theta}$) lies within the interval 0.99,1.01
2. The corresponding unkown population odds ratio ($\hat{\theta}$) differs significantly from 1

So in summary, the odd ratio has to be close to 1, and its confidence interval cannot include 1, indicating statistical significance.

The general fomrula for the odds ratio as previously stated is:

Odds Ratio($\hat{\theta}$) $= \frac{(a/c)}{(b/d)} = \frac{a \cdot d}{b \cdot c}$

Where a,b,c,d are the cell counts in the 2x2 table. To achive the requirements we need $\hat{\theta} \approx 1$ meaning that $a/c \approx b/d$ or $a \cdot d \approx b \cdot c$. To ensure $\theta \neq 1$ we have to use a large sample size to narrow the confidence interval and exclude 1. Finally we have to calculate the confidence interval $\ln(\hat{\theta}) \pm Z_{0.975} \cdot \text{SE}$ and exponentiate the bounds to get the confidence interval for the odds ratio.

Code:

Table 13: Summary Table

	Admitted	Not Admitted
Men	500000	500500
Women	501515	499090

```
##        admission
## gender  Admitted Not Admitted
##   Men     500000       500500
##   Women   501515       499090
```

```
## Table:
```

```
##        admission
## gender  Admitted Not Admitted
##   Men     500000       500500
##   Women   501515       499090
```

```
##
## Likelihood Ratio Test:
```

```
## Call:
## loglm(formula = ~gender + admission, data = tab_cont)
##
## Statistics:
##                       X^2 df   P(> X^2)
## Likelihood Ratio 4.275161  1 0.03867331
## Pearson          4.275159  1 0.03867335
```

Results and Interpretation:

1. Odds Ratio($\hat{\theta}$):

Using the formula:

- $\hat{\theta} = \frac{(500{,}000 \cdot 499{,}000)}{(500{,}500 \cdot 501{,}515)} \approx 0.995$

2. Confidence Interval:

- $[0.990, 0.999]$

3. Statistical tests:

- Likelihood Ratio Test: $G^2 = 4.275, p = 0.0387$
- Chi-Squared Test: $\chi^2 = 4.275, p = 0.0386$

Conclusion:

This exercise illustrates the distinction between statistical significance and practical significance. While the odds ratio is very close to 1, the large sample size ensures a narrow confidence interval, leading to statistical significance. For an accurate report we should focus on effect sizes and their real-world implications rather than just statistical significance with p-values alone.