

# Analysis of Income and Gender Data

Here's a detailed document that includes the R code, along with explanations and formulas used in each step of the analysis.

## 1. Marginal Distributions of Gender and Income, Expected Yearly Income, and the Square Root of its Variance

### Formulae:

- **Marginal Distribution of Gender:**

$$P(\text{Men}) = \sum P(\text{Income Level}, \text{Men})$$

$$P(\text{Women}) = \sum P(\text{Income Level}, \text{Women})$$

- **Marginal Distribution of Income:**

$$P(\text{Income}) = P(\text{Men}) \cdot P(\text{Income} \mid \text{Men}) + P(\text{Women}) \cdot P(\text{Income} \mid \text{Women})$$

- **Expected Yearly Income:**

$$E(\text{Income}) = \sum P(\text{Income}) \cdot \text{Income Level}$$

- **Variance of Income:**

$$\text{Var}(\text{Income}) = \sum P(\text{Income}) \cdot (\text{Income Level})^2 - (E(\text{Income}))^2$$

- **Standard Deviation:**

$$SD(\text{Income}) = \sqrt{\text{Var}(\text{Income})}$$

### R Code:

```
# Data
income_levels <- c(0, 13, 36, 65, 91, 104)
joint_prob_men <- c(0.01, 0.05, 0.06, 0.16, 0.14, 0.15)
joint_prob_women <- c(0.00, 0.09, 0.11, 0.12, 0.07, 0.04)

# Marginal probabilities of men and women
P_men <- sum(joint_prob_men)
P_women <- sum(joint_prob_women)

# Marginal distribution of income
marginal_prob_income <- P_men * joint_prob_men + P_women * joint_prob_women

# Expected yearly income
expected_income <- sum(income_levels * marginal_prob_income)
```

```

# Variance and Standard Deviation
variance_income <- sum((income_levels^2) * marginal_prob_income) -
expected_income^2
std_income <- sqrt(variance_income)

# Output results
cat("Marginal Probability for Men: ", P_men, "\n")
cat("Marginal Probability for Women: ", P_women, "\n")
cat("Marginal Distribution of Income: ", marginal_prob_income, "\n")
cat("Expected Yearly Income: ", expected_income, "\n")
cat("Variance of Income: ", variance_income, "\n")
cat("Standard Deviation of Income: ", std_income, "\n")

```

## 2. Conditional Yearly Income Distribution for Men and Women, and Visualization of Cumulative Distribution Function (CDF)

### Formulae:

- **Conditional Distribution for Men:**

$$P(\text{Income} \mid \text{Men}) = \frac{P(\text{Income}, \text{Men})}{P(\text{Men})}$$

- **Conditional Distribution for Women:**

$$P(\text{Income} \mid \text{Women}) = \frac{P(\text{Income}, \text{Women})}{P(\text{Women})}$$

- **CDF:**

The cumulative distribution function is calculated as:

$$\text{CDF} = \sum P(\text{Income Level})$$

### R Code:

```

# Conditional distributions for men and women
conditional_income_men <- joint_prob_men / P_men
conditional_income_women <- joint_prob_women / P_women

# CDFs
cdf_men <- cumsum(conditional_income_men)
cdf_women <- cumsum(conditional_income_women)

# Plotting CDF
plot(income_levels, cdf_men, type = "o", col = "blue", ylim = c(0, 1),
      xlab = "Income", ylab = "Cumulative Probability", main = "CDF of Income

```

```
by Gender")
lines(income_levels, cdf_women, type = "o", col = "red")
legend("bottomright", legend = c("Men", "Women"), col = c("blue", "red"),
      lty = 1)
```

### 3. Conditional Expectations and Square Roots of the Variances

#### Formulae:

- **Conditional Expected Income for Men:**

$$E(\text{Income} \mid \text{Men}) = \sum P(\text{Income} \mid \text{Men}) \cdot \text{Income Level}$$

- **Conditional Expected Income for Women:**

$$E(\text{Income} \mid \text{Women}) = \sum P(\text{Income} \mid \text{Women}) \cdot \text{Income Level}$$

- **Variance of Conditional Income for Men:**

$$\text{Var}(\text{Income} \mid \text{Men}) = \sum P(\text{Income} \mid \text{Men}) \cdot (\text{Income Level})^2 - (E(\text{Income} \mid \text{Men}))^2$$

- **Variance of Conditional Income for Women:**

$$\text{Var}(\text{Income} \mid \text{Women}) = \sum P(\text{Income} \mid \text{Women}) \cdot (\text{Income Level})^2 - (E(\text{Income} \mid \text{Women}))^2$$

#### R Code:

```
# Conditional expectations and square roots of variances
expected_income_men <- sum(income_levels * conditional_income_men)
variance_income_men <- sum((income_levels^2) * conditional_income_men) -
  expected_income_men^2
std_income_men <- sqrt(variance_income_men)

expected_income_women <- sum(income_levels * conditional_income_women)
variance_income_women <- sum((income_levels^2) * conditional_income_women) -
  expected_income_women^2
std_income_women <- sqrt(variance_income_women)

# Output results
cat("Expected Yearly Income for Men: ", expected_income_men, "\n")
cat("Variance of Income for Men: ", variance_income_men, "\n")
cat("Standard Deviation of Income for Men: ", std_income_men, "\n")
cat("Expected Yearly Income for Women: ", expected_income_women, "\n")
cat("Variance of Income for Women: ", variance_income_women, "\n")
cat("Standard Deviation of Income for Women: ", std_income_women, "\n")
```

## 4. Covariance and Correlation Between Income and Gender

### Formulae:

- **Covariance:**

$$\text{Cov}(\text{Income}, \text{Gender}) = E(\text{Income} \cdot \text{Gender}) - E(\text{Income}) \cdot E(\text{Gender})$$

- **Correlation:**

$$\text{Corr}(\text{Income}, \text{Gender}) = \frac{\text{Cov}(\text{Income}, \text{Gender})}{\sqrt{\text{Var}(\text{Income})} \cdot \sqrt{\text{Var}(\text{Gender})}}$$

### R Code:

```
# Treat male as 0 and female as 1 for gender dummy variable
gender_dummy <- c(rep(0, length(joint_prob_men)), rep(1,
length(joint_prob_women)))
income_all <- rep(income_levels, 2) # Combine income levels for both men and
women
joint_prob_all <- c(joint_prob_men, joint_prob_women) # Combined joint
probabilities

# Covariance calculation
expected_income_total <- sum(income_all * joint_prob_all)
expected_gender <- sum(gender_dummy * joint_prob_all)
cov_income_gender <- sum(income_all * gender_dummy * joint_prob_all) -
(expected_income_total * expected_gender)

# Variance of gender and income
var_gender <- sum((gender_dummy^2) * joint_prob_all) - expected_gender^2
var_income <- sum((income_all^2) * joint_prob_all) - expected_income_total^2

# Correlation
cor_income_gender <- cov_income_gender / (sqrt(var_gender) *
sqrt(var_income))

# Output results
cat("Covariance between Income and Gender: ", cov_income_gender, "\n")
cat("Correlation between Income and Gender: ", cor_income_gender, "\n")
```

---

## 5. Independence Test

### Test for Independence:

To check if income and gender are independent, compare the joint distribution to the product of their marginals. If the joint distribution equals the product of the marginal probabilities, then the variables are independent.

## R Code:

```
# Two variables are independent if their joint distribution equals the
product of their marginals
marginal_gender_men <- rep(P_men, length(income_levels))
marginal_gender_women <- rep(P_women, length(income_levels))

product_marginals_men <- marginal_gender_men * marginal_prob_income
product_marginals_women <- marginal_gender_women * marginal_prob_income

# Check independence by comparing joint probabilities with product of
marginals
independence_test_men <- all(joint_prob_men == product_marginals_men)
independence_test_women <- all(joint_prob_women == product_marginals_women)

cat("Are income and gender independent for men? ", independence_test_men,
"\n")
cat("Are
income and gender independent for women? ", independence_test_women, "\n")
```

---

## Summary of Results:

1. **Marginal Distributions:** These represent the overall probabilities for men and women as well as for the different income levels.
2. **Conditional Distributions:** The income distribution conditional on gender, visualized using CDFs.
3. **Covariance and Correlation:** Quantifying the relationship between income and gender.
4. **Independence Test:** Evaluating if income and gender are independent variables.