Chapter 11. Hypothesis Testing (II)

11.1 Likelihood Ratio Tests — one of the most popular ways of constructing tests when both null and alternative hypotheses are composite (i.e. not a single point).

Let $\mathbf{X} \sim f(\cdot, \boldsymbol{\theta})$. Consider hypotheses

$$H_0: \ \boldsymbol{\theta} \in \Theta_0 \quad \text{vs} \quad H_1: \ \boldsymbol{\theta} \in \Theta - \Theta_0.$$

The likelihood ratio test will reject H_0 for the large values of the statistic

$$LR = LR(\mathbf{X}) \equiv \frac{\sup_{\boldsymbol{\theta} \in \Theta} f(\mathbf{X}, \boldsymbol{\theta})}{\sup_{\boldsymbol{\theta} \in \Theta_0} f(\mathbf{X}, \boldsymbol{\theta})} = f(\mathbf{X}, \widehat{\boldsymbol{\theta}}) \big/ f(\mathbf{X}, \widetilde{\boldsymbol{\theta}}),$$

where $\widehat{\boldsymbol{\theta}}$ the (unconstrained) MLE, and $\widetilde{\boldsymbol{\theta}}$ is the constrained MLE under hypothesis H_0.

Remark. (i) It is easy to see that $LR \geq 1$.

(ii) The exact sampling distributions of LR are usually unknown, except in a few special cases.

Example 1. **(One-sample t-test)**
Let $\mathbf{X} = (X_1, \cdots, X_n)^\tau$ be a random sample from $N(\mu, \sigma^2)$. We are interested in testing hypotheses

$$H_0: \mu = \mu_0 \quad \text{against} \quad H_1: \mu \neq \mu_0,$$

where μ_0 is given, and σ^2 is unknown and is a nuisance parameter. Now both H_0 and H_1 are composite. The likelihood function is

$$L(\mu, \sigma^2) = C\sigma^{-n} \exp\Big\{ -\frac{1}{2\sigma^2} \sum_{j=1}^{n} (X_i - \mu)^2 \Big\}.$$

The unconstrained MLEs are

$$\widehat{\mu} = \bar{X}, \quad \widehat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}(X_i - \bar{X})^2,$$

and the constrained MLE is

$$\widetilde{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}(X_i - \mu_0)^2.$$

The LR-ratio statistic is then

$$LR = \frac{L(\widehat{\mu}, \widehat{\sigma}^2)}{L(\mu_0, \widetilde{\sigma}^2)} = (\widetilde{\sigma}^2 / \widehat{\sigma}^2)^{n/2}.$$

Since

$$n\widetilde{\sigma}^2 = n\widehat{\sigma}^2 + n(\bar{X} - \mu_0)^2,$$

it holds that $\widetilde{\sigma}^2/\widehat{\sigma}^2 = 1 + T^2/(n-1)$, where

$$T = \sqrt{n}(\bar{X} - \mu_0) \Big/ \Big\{\frac{1}{n-1}\sum_{j=1}^{n}(X_i - \bar{X})^2\Big\}^{1/2}.$$

Note that $T \sim t_{n-1}$ under H_0. The LRT will reject H_0 iff $|T| > t_{n-1,\alpha/2}$, where $t_{k,\alpha}$ is the upper α-point of the t-distribution with k degrees of freedom.

Asymptotic Distribution of Likelihood ratio test statistic

Let $\mathbf{X} = (X_1, \cdots, X_n)^\tau$, and assume certain regularity conditions. Then as $n \to \infty$, the distribution of $2\log(LR)$ under H_0 converges to the χ^2-distribution with $d - d_0$ degrees of freedom, where d is the 'dimension' of Θ and d_0 is the 'dimension' of Θ_0.

To make the computation of 'dimension' easy, **reparametrisation** is often adopted. Suppose that the parameter θ may be written in two parts

$$\boldsymbol{\theta} = (\boldsymbol{\psi}, \boldsymbol{\lambda})$$

where $\boldsymbol{\psi}$ is $k \times 1$ parameter of interest, and $\boldsymbol{\lambda}$ is of little interest and is called *nuisance parameters*. The hypotheses to be tested may be expressed as

$$H_0: \boldsymbol{\psi} = \boldsymbol{\psi}_0 \quad \text{vs} \quad H_1: \boldsymbol{\psi} \neq \boldsymbol{\psi}_0.$$

Now the LR-statistic is of the form

$$LR = \frac{L(\widehat{\boldsymbol{\psi}}, \widehat{\boldsymbol{\lambda}}; \mathbf{X})}{L(\boldsymbol{\psi}_0, \widetilde{\boldsymbol{\lambda}}; \mathbf{X})},$$

where $(\widehat{\boldsymbol{\psi}}, \widehat{\boldsymbol{\lambda}})$ is unconstrained MLE while $\widetilde{\boldsymbol{\lambda}}$ is the constrained MLE of $\boldsymbol{\lambda}$ subject to $\boldsymbol{\psi} = \boldsymbol{\psi}_0$. Then as $n \to \infty$,

$$2\log(LR) \xrightarrow{D} \chi_k^2 \quad \text{under } H_0.$$

Example 2. Let $X_1, \cdots, X_n$ be independent, and $X_j \sim N(\mu_j, 1)$. Consider the null hypothesis

$$H_0: \ \mu_1 = \cdots = \mu_n.$$

The likelihood function is

$$L(\mu_1, \cdots, \mu_n) = C \ \exp\Big\{ -\frac{1}{2}\sum_{j=1}^{n}(X_j - \mu_j)^2\Big\},$$

where $C > 0$ is a constant independent of μ_j. Then the unconstrained MLE are $\widehat{\mu}_j = X_j$ and the constrained MLE is $\widetilde{\mu} = \bar{X}$. Hence

$$LR = \frac{L(\widehat{\mu}_1, \cdots, \widehat{\mu}_n)}{L(\widetilde{\mu}, \cdots, \widetilde{\mu})} = \exp\Big\{\frac{1}{2}\sum_{j=1}^{n}(X_j - \bar{X})^2\Big\}.$$

Hence

$$2\log(LR) = \sum_{j=1}^{n}(X_j - \bar{X})^2 \sim \chi^2_{n-1} \quad \text{under } H_0,$$

which is true for any finite n as well.

How to *calculate the degree of freedom*?

Since $d = n, \; d_0 = 1$, the d.f. is $d - d_0 = n - 1$.

Alternatively we may adopt the following reparametrisation:

$$\mu_j = \mu_1 + \psi_j \quad \text{for } 2 \le j \le n.$$

Then the null hypothesis can be expressed as

$$H_0: \; \psi_2 = \cdots = \psi_n = 0.$$

Therefore $\boldsymbol{\psi} = (\psi_2, \cdots, \psi_n)^\tau$ has $n - 1$ component, i.e. $k = n - 1$.

11.2 The permutation test — a nonparametric method for testing if two distributions are the same. It is particularly appealing when sample sizes are small, as it does not rely on any asymptotic theory.

Let $X_1, \cdots, X_m$ be sample from distribution F_x and $Y_1, \cdots, Y_n$ be a sample from distribution F_y. We are interested in testing

$$H_0 : F_x = F_y \qquad \text{versus} \qquad H_1 : F_x \neq F_y.$$

Key idea: under H_0, $\{X_1, \cdots, X_m, Y_1, \cdots, Y_n\}$ form a sample of size $m + n$ from a single distribution.

Choose a test statistic

$$T = T(X_1, \cdots, X_m, Y_1, \cdots, Y_n)$$

which is capable to tell the difference between the two distribution, e.g. $T = |\bar{X} - \bar{Y}|$, or $T = |\bar{X} - \bar{Y}|^2 + |S_x^2 - S_y^2|$.

Consider all $(m + n)!$ permutations of $(X_1, \cdots, X_m, Y_1, \cdots, Y_n)$, compute the test statistic T for each permutation, yielding the values $T_1, \cdots, T_{(m+n)!}$.

The p-value of the test is defined as

$$p = \frac{1}{(m+n)!} \sum_{j=1}^{(m+n)!} I(T_j > t_{obs}),$$

where $t_{obs} = T(X_1, \cdots, X_m, Y_1, \cdots, Y_n)$. We reject H_0 at the significance level α if $p \le \alpha$.

Note. When H_0 holds, all those $(m+n)!$ T_j's are on the equal footing, and $t_{obs} = T(X_1, \cdots, X_m, Y_1, \cdots, Y_n)$ is one of them. Therefore t_{obs} is unlikely to be an extreme value among T_j's.

Algorithm for Permutation Tests:

1. Compute $t_{obs} = T(X_1, \cdots, X_m, Y_1, \cdots, Y_n)$.

2. Randomly permute the data. Compute T again using the permuted date.

3. Repeat Step 2 B times, and let $T_1, \cdots, T_B$ denote the resulting values.

4. The approximate p-value is $B^{-1} \sum_{1 \le j \le B} I(T_j > t_{obs})$.

Remark. Let $Z = (X_1, \cdots, X_m, Y_1, \cdots, Y_n)$ (`Z <- c(X,Y)`). A permutation of Z may be obtained in R as

```
Zp <- sample(Z, n+m)
```

You may also use the R-function sample.int:

```
k <- sample.int(n+m, n+m)
```

Now k is a permutation of $\{1, 2, \cdots, n+m\}$.

Example 3. Class A was taught using detailed PowerPoint slides. The marks in the final exam are

$$45,\ 55,\ 39,\ 60,\ 64,\ 85,\ 80,\ 64,\ 48,\ 62,\ 75,\ 77,\ 50.$$

Students in Class B were required to read books and answer questions in class discussions. The marks in the final exam are

$$45,\ 59,\ 48,\ 74,\ 73,\ 78,\ 66,\ 69,\ 79,\ 81,\ 60,\ 52.$$

Can we infer that the marks from the two classes are significantly different?

We conduct the permutation test using the test statistic $T = |\bar{X} - \bar{Y}|$ in R:

```
> x <- c(45, 55, 39, 60, 64, 85, 80, 64, 48, 62, 75, 77, 50)
> y <- c(45, 59, 48, 74, 73, 78, 66, 69, 79, 81, 60, 52)
> length(x); length(y)
```

```
[1] 13
[1] 12
> summary(x)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  39.00   50.00   62.00   61.85   75.00   85.00
> summary(y)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  45.00   57.25   67.50   65.33   75.00   81.00
> Tobs <- abs(mean(x)-mean(y))
> z <- c(x,y)
> k <- 0
> for(i in 1:5000) {
+ zp <- sample(z, 25)       # zp is a permutation of z
+ T <- abs(mean(zp[1:13])-mean(zp[14:25]))
+ if(T>Tobs) k <- k+1
```

```
+ }
cat("p-value:", k/5000, "\n")
p-value: 0.5194
```

Since *p*-value is 0.5194, we cannot reject the null-hypothesis that the mark distributions of the two classes are the same.

We also apply the *t*-sample, obtaining the similar results:

```
> t.test(x, y, var.equal=T)   #  mu=0 is the default
Two Sample t-test
data:  x and y
t = -0.6472, df = 23, p-value = 0.5239
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-14.632967   7.658608
```

11.3 χ^2-tests

11.3.1 Goodness-of-fit tests: to test if a given distribution fits the data.

Let $\{X_1, \cdots, X_n\}$ be a random sample from a discrete distribution of k categories denoted by $1, \cdots, k$. Denote the probability function

$$p_j = P(X_i = j), \qquad j = 1, \cdots, k.$$

Then $p_j \geq 0$ and $\sum_{j=1}^{k} p_j = 1$.

Typically $n >> k$. Therefore the data are often compressed into a table:

Category	1	2	$\cdots$	k
Frequency	Z_1	Z_2	$\cdots$	Z_k

where

$$Z_j = \text{ No. of } X_i\text{'s equal to } j, \qquad j = 1, \cdots, k.$$

Obviously $\sum_{j=1}^{k} Z_j = n$.

To test the null hypothesis

$$H_0: \; p_i = p_i(\theta), \qquad i = 1, \cdots, k,$$

where the function forms of $p_i(\theta)$ are known but the parameter θ is unknown. For example, $p_i(\theta) = \theta^{i-1} e^{-\theta}/(i-1)!$ (i.e. Poisson distribution).

We first estimate θ by, for example, its MLE $\widehat{\theta}$. The expected frequencies under H_0 are

$$E_i = np_i(\widehat{\theta}), \quad i = 1, \cdots, k.$$

Listing them together with observed frequencies, we have

Category	1	2	$\cdots$	k
Frequency	Z_1	Z_2	$\cdots$	Z_k
Expected frequency	E_1	E_2	$\cdots$	E_k

If H_0 is true, we expect $Z_j \approx E_j = np_j(\widehat{\theta})$ when n is large, as, by the LLN, it holds

$$\frac{Z_j}{n} = \frac{1}{n}\sum_{i=1}^{n} I(X_i = j) \longrightarrow E\{I(X_i = j)\} = P(X_i = j) = p_j(\theta).$$

Test statistic: $T = \sum_{j=1}^{k}(Z_j - E_j)^2 / E_j$.

Theorem. Under H_0, $T \xrightarrow{D} \chi^2_{k-1-d}$ as $n \to \infty$, where d is the number of components in θ.

Remark. (i) It is important that $E_i \geq 5$ at least. This may be achieved by combining together the categories with smaller expected frequencies.

(ii) When p_j are completely specified (i.e. known) under H_0, $d = 0$.

Example 4. A supermarket recorded the numbers of arrivals over 100 one-minute intervals. The data were summarized as follows

No. of arrivals	0	1	2	3	4	5	7
Frequency	13	29	32	20	4	1	1

Do the data match a Poisson distribution?

The null hypothesis is $H_0: \; p_i = \lambda^i e^{-\lambda}/i!$ for $i = 0, 1, \cdots$. We find the MLE for λ first.

The likelihood function: $L(\lambda) = \prod_{i=1}^{100} \frac{\lambda^{X_i}}{X_i!} e^{-\lambda} \propto \lambda^{\sum_{i=1}^{100} X_i} e^{-100\lambda}$.

The log-likelihood function: $l(\lambda) = \log(\lambda) \sum_{i=1}^{100} X_i - 100\lambda$.

Let $\frac{d}{d\lambda} l(\lambda) = 0$, leading to $\widehat{\lambda} = \frac{1}{100} \sum_{i=1}^{100} X_i = \bar{X}$.

Since we are only given the counts Z_j instead of X_i, we need to compute $\bar{X}$ from Z_j. Recall Z_j = no. of X_i equal to j. Hence

$$
\begin{aligned}
\bar{X} &= \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{n} \sum_{j=1}^{k} j \cdot Z_j \\
&= \frac{1}{100} (0 \times 13 + 1 \times 29 + 2 \times 32 + 3 \times 20 + 4 \times 4 \\
&\quad +5 \times 1 + 7 \times 1) = 1.81.
\end{aligned}
$$

With $\widehat{\lambda} = 1.81$, the expected frequencies are

$$E_i = n \cdot p_i(\widehat{\lambda}) = 100 \times \frac{(1.81)^i}{i!} e^{-1.81}, \quad i = 0, 1, \cdots .$$

We combine the last three categories to make sure $E_i \geq 5$.

No. of arrivals	0	1	2	3	≥ 4	Total
Frequency Z_i	13	29	32	20	6	100
$p_i(\widehat{\lambda}) = \widehat{\lambda}^i e^{-\widehat{\lambda}}/i!$	0.164	0.296	0.268	0.162	0.110	1
Expected frequency E_i	16.4	29.6	26.8	16.2	11.0	100
Difference $Z_i - E_i$	-3.4	-0.6	5.2	3.8	-5	0
$(Z_i - E_i)^2/E_i$	0.705	0.012	1.01	0.891	2.273	4.891

Note under H_0, $T = \sum_{i=0}^{4}(Z_i - E_i)^2/E_i \sim \chi^2_{5-1-1} = \chi^2_3$. Since $T = 4.891 < \chi^2_{0.10,3} = 6.25$, we cannot reject the assumption that the data follow a Poisson distribution.

Remark. (i) The goodness-of-fit test has been widely used in practice. However we should bear in mind that when H_0 cannot be rejected, *we are not in the position to conclude that the assumed distribution is true,* as

“not reject” ≠ “accept"

(ii) The above test may be used to test the goodness-of-fit of a continuous distribution via discretization. However there exist more appropriate methods such as *Kolmogorov-Smirnov test* and *Cramér-von Mises test*, which deal with the goodness-of-fit for continuous distributions directly.

11.3.2 Tests for contingency tables

Tests for independence of two discrete random variables

Let (X, Y) be two discrete random variables, and X have r categories and Y have c categories. Let

$$p_{ij} = P(X = i,\ Y = j), \quad i = 1, \cdots, r,\ j = 1, \cdots, c.$$

Then $p_{ij} \geq 0$ and $\sum_{i,j} p_{ij} \equiv \sum_{i=1}^{r} \sum_{j=1}^{c} p_{ij} = 1$.

Let $p_{i\cdot} = P(X = i)$ and $p_{\cdot j} = P(Y = j)$. It is easy to see that

$$p_{i\cdot} = \sum_{j=1}^{c} P(X = i,\ Y = j) = \sum_{j=1}^{c} p_{ij} = \sum_{j} p_{ij}$$

Similarly, $p_{\cdot j} = \sum_i p_{ij}$

X and Y are independent iff

$$p_{ij} = p_{i\cdot}p_{\cdot j} \text{ for } i = 1, \cdots, r \text{ and } j = 1, \cdots, c.$$

Suppose we have n pairs of observations from (X, Y). The data are presented in a contingency table below

$$\begin{array}{cc|cccc} & & & Y & & \\ & & 1 & 2 & \cdots & c \\ \hline & 1 & Z_{11} & Z_{12} & \cdots & Z_{1c} \\ X & 2 & Z_{21} & Z_{22} & \cdots & Z_{2c} \\ & \vdots & \vdots & \vdots & \vdots & \vdots \\ & r & Z_{r1} & Z_{r2} & \cdots & Z_{rc} \end{array}$$

where Z_{ij} = no. of the pairs equal to (i, j).

It is often useful to add the marginals into the table:

$$Z_{i\cdot} = \sum_{j=1}^{c} Z_{ij}, \quad Z_{\cdot j} = \sum_{i=1}^{r} Z_{ij}, \quad Z_{\cdot\cdot} = \sum_{i=1}^{r} Z_{i\cdot} = \sum_{j=1}^{c} Z_{\cdot j} = n$$

		Y				
		1	2	$\cdots$	c	
	1	Z_{11}	Z_{12}	$\cdots$	Z_{1c}	$Z_{1\cdot}$
X	2	Z_{21}	Z_{22}	$\cdots$	Z_{2c}	$Z_{2\cdot}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	r	Z_{r1}	Z_{r2}	$\cdots$	Z_{rc}	$Z_{r\cdot}$
		$Z_{\cdot 1}$	$Z_{\cdot 2}$	$\cdots$	$Z_{\cdot c}$	$Z_{\cdot\cdot} = n$

We are interested in testing the independence

$$H_0: \; p_{ij} = p_{i\cdot}p_{\cdot j}, \; i = 1, \cdots, r, \; j = 1, \cdots, c.$$

Under H_0, a natural estimator for p_{ij} is

$$\widetilde{p}_{ij} = \widehat{p}_{i\cdot}\widehat{p}_{\cdot j} = \frac{Z_{i\cdot}}{n}\frac{Z_{\cdot j}}{n}$$

Hence the expected frequency at the (i, j)-th cell is

$$E_{ij} = n\widetilde{p}_{ij} = Z_{i\cdot}Z_{\cdot j}/n = Z_{i\cdot}Z_{\cdot j}/Z_{\cdot\cdot}, \quad i = 1, \cdots, r, \;\; j = 1, \cdots, c.$$

If H_0 is true, we expect $Z_{ij} \approx E_{ij}$. The goodness-of-fit test statistic is defined as

$$T = \sum_{i=1}^{r}\sum_{j=1}^{c}(Z_{ij} - E_{ij})^2 / E_{ij}.$$

We reject H_0 for large values of T.

Under H_0, $T \sim \chi^2_{p-d}$, where

- p = no. of cells -1 = $rc - 1$
- d = no. of estimated 'free' parameters = $r + c - 2$.

Note. 1. The sum of $r \times c$ counts Z_{ij} is n fixed. So knowing $rc - 1$ of them, the other one is also known. This is why $p = rc - 1$.

2. The estimated parameters are $p_{i\cdot}$ and $p_{\cdot j}$. But $\sum_{i=1}^r p_{i\cdot} = 1$ and $\sum_{j=1}^c p_{\cdot j} = 1$. Hence $d = (r-1) + (c-1) = r + c - 2$.

3. For testing independence, it always holds that

$$Z_{i\cdot} - E_{i\cdot} = 0 \qquad \text{and} \qquad Z_{\cdot j} - E_{\cdot j} = 0.$$

Those are useful facts in checking for computational errors. The proofs are simple, as, for example,

$$Z_{i\cdot} - E_{i\cdot} = Z_{i\cdot} - \sum_j E_{ij} = Z_{i\cdot} - \sum_j \frac{Z_{i\cdot} Z_{\cdot j}}{Z_{\cdot\cdot}} = Z_{i\cdot} - \frac{Z_{i\cdot} Z_{\cdot\cdot}}{Z_{\cdot\cdot}} = 0.$$

Theorem. Under H_0, the limiting distribution of T is χ^2 with $(r-1)(c-1)$ degrees of freedom, as $n \to \infty$.

Example. The table below lists the counts on the beer preference and gender of beer drinker from randomly selected 150 individuals. Test at the 5% significance level the hypothesis that the preference is independent of the gender.

		Beer preference			
		Light ale	Lager	Bitter	Total
Gender	Male	20	40	20	80
	Female	30	30	10	70
	Total	50	70	30	150

The expected frequencies are:

$$E_{11} = \frac{80 \cdot 50}{150} = 26.67, \quad E_{12} = \frac{80 \cdot 70}{150} = 37.33, \quad E_{13} = \frac{80 \cdot 30}{150} = 16,$$

$$E_{21} = \frac{70 \cdot 50}{150} = 23.33, \quad E_{22} = \frac{70 \cdot 70}{150} = 32.67, \quad E_{33} = \frac{70 \cdot 30}{150} = 14.$$

E_{ij}				
	26.67	37.33	16	80
	23.33	32.67	14	70
	50	70	30	150

$Z_{ij} - E_{ij}$				
	-6.67	2.67	4	0
	6.67	-2.67	-4	0
	0	0	0	0

$(Z_{ij} - E_{ij})^2/E_{ij}$				
	1.668	0.191	1.000	2.859
	1.907	0.218	1.142	3.267
				6.126

Under the null hypothesis of independence, $T = \sum_{i,j}(Z_{ij} - E_{ij})^2/E_{ij} \sim \chi^2_2$. Note the degree freedom is $(2-1)(3-1) = 2$.

Since $T = 6.126 > \chi^2_{0.05,2} = 5.991$, we reject the null hypothesis, i.e. there is significant evidence from the data indicating that the beer preference and the gender of beer drinker are not independent.

Tests for several binomial distributions

Consider a real example: Three independent samples of sizes 80, 120 and 100 are taken respectively from single, married, and widowed or divorced persons. Each individual was asked to if "friends and social life" or "job and primary activity" contributes most to their general well-being. The counts from the three samples are summarized in the table below.

	Single	Married	Widowed or divorced
Friends and social life	47	59	56
Job or primary activity	33	61	44
Total	80	120	100

Conditional Inference: Sometimes we conduct inference under the assumption that all the row (or column) margins are fixed.

Different from the tables for independent tests, now

$$Z_{1j} \sim Bin(Z_{\cdot j},\ p_{1j}), \qquad j = 1, 2, 3,$$

where $Z_{\cdot j}$ are fixed constants — sample sizes. Furthermore, $p_{2j} = 1 - p_{1j}$.

We are interested in testing hypothesis

$$H_0:\ p_{11} = p_{12} = p_{13}.$$

Under H_0, the three independent samples may be seen from the same population. Furthermore,

$$Z_{11} + Z_{12} + Z_{13} \sim Bin(Z_{\cdot 1} + Z_{\cdot 2} + Z_{\cdot 3},\ p),$$

where p denotes the common value of p_{11}, p_{12} and p_{13}.

Therefore the MLE is

$$\widehat{p} = \frac{Z_{11} + Z_{12} + Z_{13}}{Z_{\cdot 1} + Z_{\cdot 2} + Z_{\cdot 3}} = \frac{47 + 59 + 56}{80 + 120 + 100} = 0.54.$$

The expected frequencies are

$$E_{1j} = \widehat{p} Z_{\cdot j} \quad \text{and} \quad E_{2j} = Z_{\cdot j} - E_{1j}, \qquad j = 1, 2, 3.$$

E_{ij}			
	43.2	64.8	54.0
	36.8	55.2	46.0
Total	80	120	100

$Z_{ij} - E_{ij}$			
	3.8	-5.8	2.0
	-3.8	5.8	-2.0
Total	0	0	0

$(Z_{ij} - E_{ij})^2 / E_{ij}$				
	0.334	0.519	0.074	
	0.392	0.609	0.087	
Total	0.726	1.128	0.161	2.015

Under H_0, $T = \sum_{i,j}(Z_{ij} - E_{ij})^2/E_{ij} \sim \chi^2_{p-d} = \chi^2_2$, where

- p = no. of free counts $Z_{ij} = 3$
- d = no. of estimated free parameters = 1.

Since $T = 2.015 < \chi^2_{0.10,2} = 4.605$, we cannot reject H_0, i.e. there is no significant difference among the three populations in terms of choosing between F&SL and J&PA as the more important factor towards their general well-being.

Remark. Similar to the independence tests, it holds that $Z_{i\cdot} - E_{i\cdot} = 0$ and $Z_{\cdot j} - E_{\cdot j} = 0$.

Tests for $r \times c$ tables – a general description

In general, we may test for different types of the structure in a $r \times c$ table, for example, the symmetry ($p_{ij} = p_{ji}$).

The key is to compute expected frequencies E_{ij} under null hypothesis H_0.

Under H_0, the test statistic

$$T = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(Z_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{p-d},$$

- p = no. of 'free' counts among Z_{ij} ,
- d = no. of the estimated 'free' parameters.

We reject H_0 if $T > \chi^2_{\alpha,\, p-d}$.

Remark. The R-function `chisq.test` performs both the goodness-of-fit test and the contingency table test.