



# **A Handbook of Statistical Analyses Using R**

---

Brian S. Everitt and Torsten Hothorn



---

## CHAPTER 4

# Analysis of Variance: Weight Gain, Foster Feeding in Rats, Water Hardness and Male Egyptian Skulls

---

### 4.1 Introduction

### 4.2 Analysis of Variance

### 4.3 Analysis Using R

#### 4.3.1 Weight Gain in Rats

Before applying analysis of variance to the data in Table ?? we should try to summarise the main features of the data by calculating means and standard deviations and by producing some hopefully informative graphs. The data is available in the *data.frame* `weightgain`. The following R code produces the required summary statistics

```
R> data("weightgain", package = "HSAUR")
R> tapply(weightgain$weightgain, list(weightgain$source, weightgain$type),
+ mean)
```

```
      High  Low
Beef    100.0 79.2
Cereal   85.9 83.9
```

```
R> tapply(weightgain$weightgain, list(weightgain$source, weightgain$type),
+ sd)
```

```
      High      Low
Beef  15.13642 13.88684
Cereal 15.02184 15.70881
```

To apply analysis of variance to the data we can use the `aov` function in R and then the `summary` method to give us the usual analysis of variance table. The model *formula* specifies a two-way layout with interaction terms, where the first factor is `source`, and the second factor is `type`.

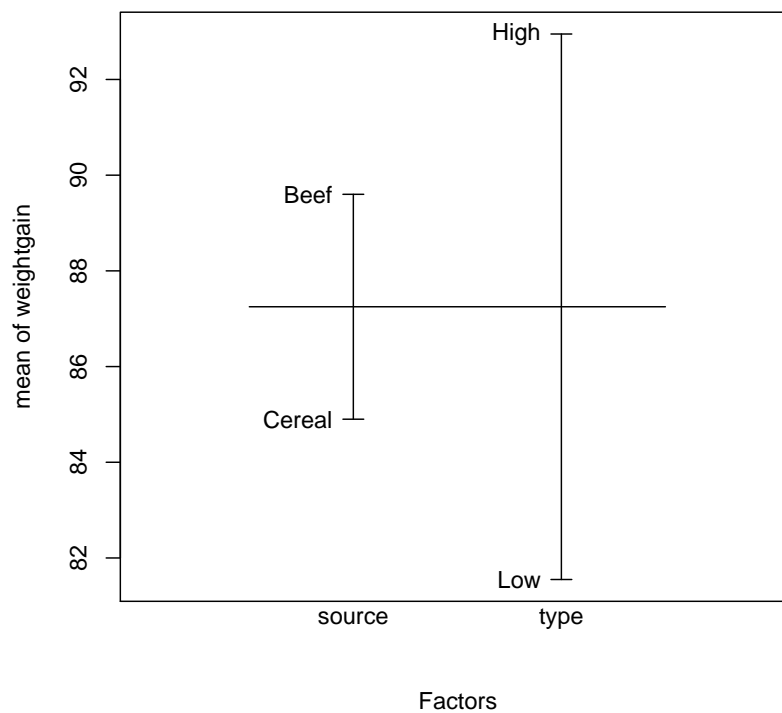
```
R> wg_aov <- aov(weightgain ~ source * type, data = weightgain)
```

The estimates of the intercept and the main and interaction effects can be extracted from the model fit by

```
R> coef(wg_aov)
```

(Intercept)	sourceCereal	typeLow
100.0	-14.1	-20.8

```
R> plot.design(weightgain)
```



**Figure 4.1** Plot of mean weight gain for each level of the two factors.

```
sourceCereal:typeLow
18.8
```

Note that the model was fitted with the restrictions  $\gamma_1 = 0$  (corresponding to **Beef**) and  $\beta_1 = 0$  (corresponding to **High**) because treatment contrasts were used as default as can be seen from

```
R> options("contrasts")
```

```
$contrasts
      unordered      ordered
"contr.treatment"  "contr.poly"
```

Thus, the coefficient for **source** of  $-14.1$  can be interpreted as an estimate of the difference  $\gamma_2 - \gamma_1$ . Alternatively, we can use the restriction  $\sum_i \gamma_i = 0$  by

```
R> coef(aov(weightgain ~ source + type + source:type, data = weightgain,
+          contrasts = list(source = contr.sum)))
```

```
R> summary(wg_aov)

              Df Sum Sq Mean Sq F value    Pr(>F)
source         1  220.9    220.9    0.9879 0.32688
type           1 1299.6   1299.6    5.8123 0.02114 *
source:type     1   883.6    883.6    3.9518 0.05447 .
Residuals      36 8049.4    223.6
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

**Figure 4.2** R output of the ANOVA fit for the `weightgain` data.

```
(Intercept)          source1          typeLow
          92.95              7.05          -11.40
source1:typeLow
          -9.40
```

#### 4.3.2 Foster Feeding of Rats of Different Genotype

As in the previous subsection we will begin the analysis of the foster feeding data in Table ?? with a plot of the mean litter weight for the different genotypes of mother and litter (see Figure 4.4). The data are in the *data.frame* `foster`

```
R> data("foster", package = "HSAUR")
```

We can derive the two analyses of variance tables for the foster feeding example by applying the R code

```
R> summary(aov(weight ~ litgen * motgen, data = foster))
```

to give

```
              Df Sum Sq Mean Sq F value    Pr(>F)
litgen         3   60.16    20.05    0.3697 0.775221
motgen         3  775.08   258.36    4.7632 0.005736 **
litgen:motgen   9  824.07    91.56    1.6881 0.120053
Residuals      45 2440.82    54.24
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

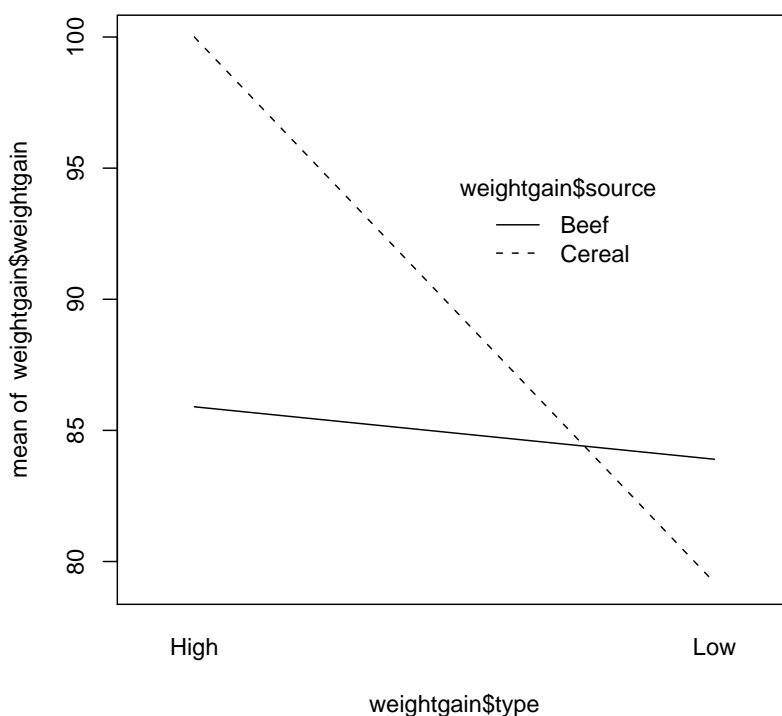
and then the code

```
R> summary(aov(weight ~ motgen * litgen, data = foster))
```

to give

```
              Df Sum Sq Mean Sq F value    Pr(>F)
motgen         3  771.61   257.20    4.7419 0.005869 **
litgen         3   63.63    21.21    0.3911 0.760004
motgen:litgen   9  824.07    91.56    1.6881 0.120053
Residuals      45 2440.82    54.24
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
R> interaction.plot(weightgain$type, weightgain$source, weightgain$weightgain)
```



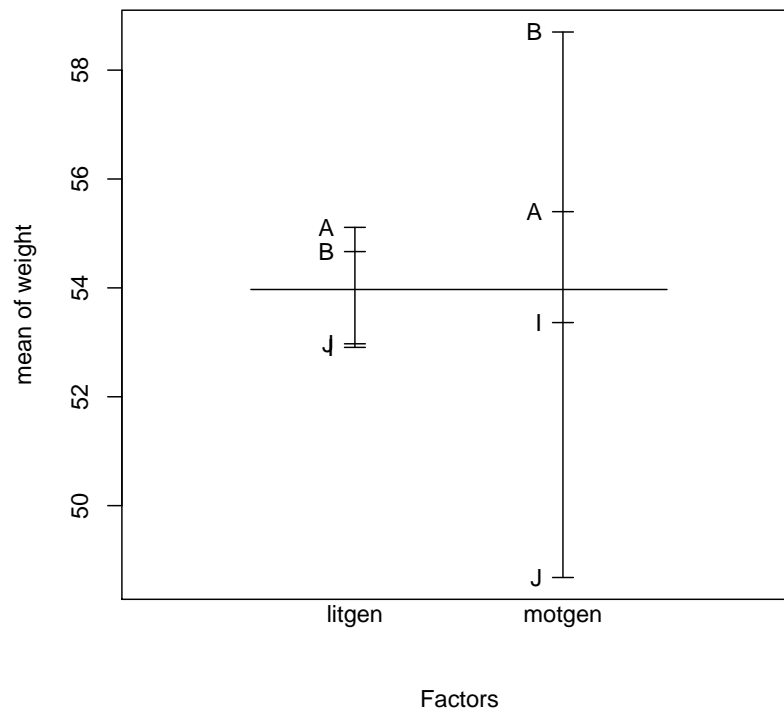
**Figure 4.3** Interaction plot of type  $\times$  source.

There are (small) differences in the sum of squares for the two main effects and, consequently, in the associated  $F$ -tests and  $p$ -values. This would not be true if in the previous example in Subsection 4.3.1 we had used the code

```
R> summary(aov(weightgain ~ type * source, data = weightgain))
```

instead of the code which produced Figure 4.2 (readers should confirm that this is the case). We can investigate the effect of genotype B on litter weight in more detail by the use of *multiple comparison procedures* (see [Everitt, 1996](#)). Such procedures allow a comparison of all pairs of levels of a factor whilst maintaining the nominal significance level at its selected value and producing adjusted confidence intervals for mean differences. One such procedure is called *Tukey honest significant differences* suggested by [Tukey \(1953\)](#), see [Hochberg and Tamhane \(1987\)](#) also. Here, we are interested in simultaneous confidence

```
R> plot.design(foster)
```



**Figure 4.4** Plot of mean litter weight for each level of the two factors for the **foster** data.

intervals for the weight differences between all four genotypes of the mother. First, an ANOVA model is fitted

```
R> foster_aov <- aov(weight ~ litgen * motgen, data = foster)
```

which serves as the basis of the multiple comparisons, here with allpair differences by

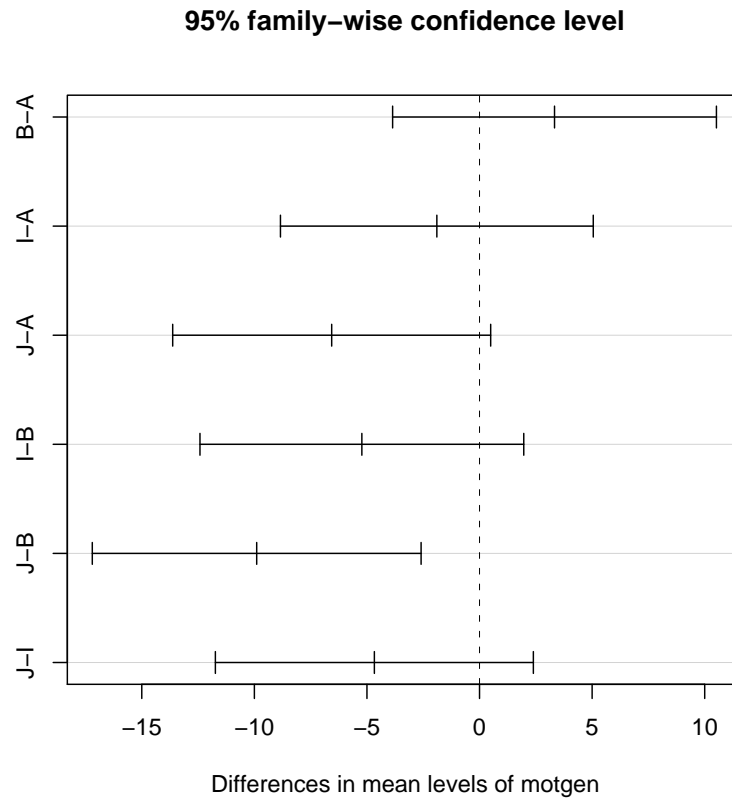
```
R> foster_hsd <- TukeyHSD(foster_aov, "motgen")
```

```
R> foster_hsd
```

```
Tukey multiple comparisons of means
 95% family-wise confidence level
```

```
Fit: aov(formula = weight ~ litgen * motgen, data = foster)
```

```
R> plot(foster_hsd)
```



**Figure 4.5** Graphical presentation of multiple comparison results for the `foster` feeding data.

```
$motgen
      diff      lwr      upr    p adj
B-A  3.330369 -3.859729 10.5204672 0.6078581
I-A -1.895574 -8.841869  5.0507207 0.8853702
J-A -6.566168 -13.627285  0.4949498 0.0767540
I-B -5.225943 -12.416041  1.9641552 0.2266493
J-B -9.896537 -17.197624 -2.5954489 0.0040509
J-I -4.670593 -11.731711  2.3905240 0.3035490
```

A convenient `plot` method exists for this object and we can get a graphical representation of the multiple confidence intervals as shown in Figure 4.5. It appears that there is only evidence for a difference in the B and J genotypes.



### 4.3.3 Water Hardness and Mortality

The water hardness and mortality data for 61 large towns in England and Wales (see Table 2.3) was analysed in Chapter 2 and here we will extend the analysis by an assessment of the differences of both hardness and mortality in the North or South. The hypothesis that the two-dimensional mean-vector of water hardness and mortality is the same for cities in the North and the South can be tested by *Hotelling-Lawley* test in a multivariate analysis of variance framework. The R function `manova` can be used to fit such a model and the corresponding `summary` method performs the test specified by the `test` argument

```
R> data("water", package = "HSAUR")
R> summary(manova(cbind(hardness, mortality) ~ location, data = water), test = "Hotelling-Lawley")
```

	Df	Hotelling-Lawley	approx F	num Df	den Df	Pr(>F)
location	1	0.9002	26.1062	2	58	8.217e-09
Residuals	59					

```
location ***
Residuals
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The `cbind` statement in the left hand side of the formula indicates that a *multivariate* response variable is to be modelled. The *p*-value associated with the *Hotelling-Lawley* statistic is very small and there is strong evidence that the mean vectors of the two variables are not the same in the two regions. Looking at the sample means

```
R> tapply(water$hardness, water$location, mean)
```

	North	South
mean	30.40000	69.76923

```
R> tapply(water$mortality, water$location, mean)
```

	North	South
mean	1633.600	1376.808

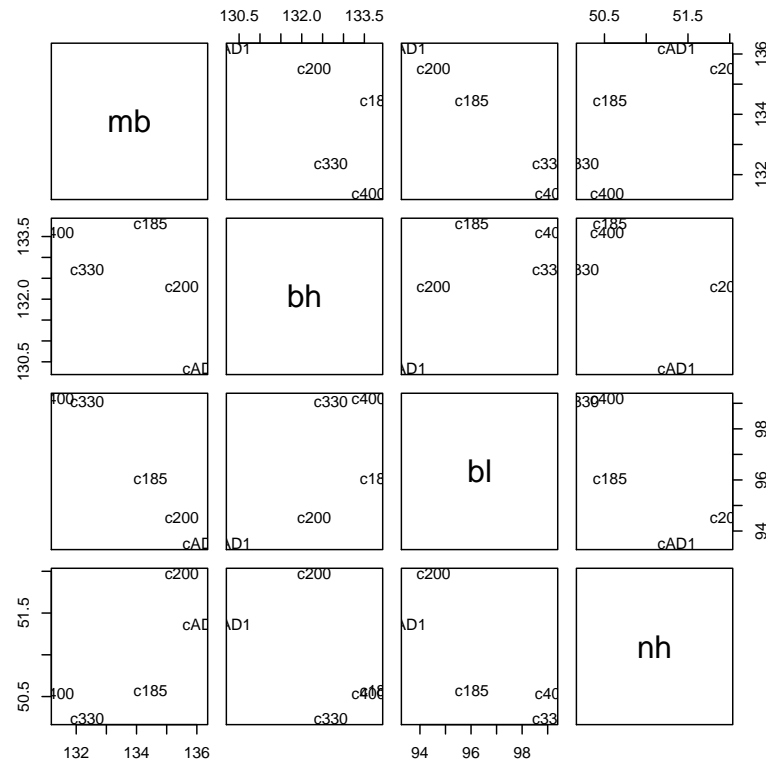
we see large differences in the two regions both in water hardness and mortality, where low mortality is associated with hard water in the South and high mortality with soft water in the North (see Figure ?? also).

### 4.3.4 Male Egyptian Skulls

We can begin by looking at a table of mean values for the four measurements within each of the five epochs. The measurements are available in the *data.frame* `skulls` and we can compute the means over all epochs by

```
R> data("skulls", package = "HSAUR")
R> means <- aggregate(skulls[,c("mb", "bh", "bl", "nh")], list(epoch = skulls$epoch), mean)
R> means
```

```
R> pairs(means[, -1],
+       panel = function(x, y) {
+         text(x, y, abbreviate(levels(skulls$epoch)))
+       })
```



**Figure 4.6** Scatterplot matrix of epoch means for Egyptian skulls data.

epoch	mb	bh	bl	nh
1 c4000BC	131.3667	133.6000	99.16667	50.53333
2 c3300BC	132.3667	132.7000	99.06667	50.23333
3 c1850BC	134.4667	133.8000	96.03333	50.56667
4 c200BC	135.5000	132.3000	94.53333	51.96667
5 cAD150	136.1667	130.3333	93.50000	51.36667

It may also be useful to look at these means graphically and this could be done in a variety of ways. Here we construct a scatterplot matrix of the means using the code attached to Figure 4.6. There appear to be quite large differences between the epoch means, at least on some of the four measurements. We can

now test for a difference more formally by using MANOVA with the following R code to apply each of the four possible test criteria mentioned earlier;

```
R> skulls_manova <- manova(cbind(mb, bh, bl, nh) ~ epoch, data = skulls)
R> summary(skulls_manova, test = "Pillai")
```

```
              Df Pillai approx F num Df den Df    Pr(>F)
epoch          4 0.3533   3.5120    16   580 4.675e-06 ***
Residuals 145
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
R> summary(skulls_manova, test = "Wilks")
```

```
              Df Wilks approx F num Df den Df    Pr(>F)
epoch          4 0.6636   3.9009   16.00 434.45 7.01e-07 ***
Residuals 145.00
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
R> summary(skulls_manova, test = "Hotelling-Lawley")
```

```
              Df Hotelling-Lawley approx F num Df den Df
epoch          4          0.4818   4.2310    16   562
Residuals 145
              Pr(>F)
epoch      8.278e-08 ***
Residuals
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
R> summary(skulls_manova, test = "Roy")
```

```
              Df      Roy approx F num Df den Df    Pr(>F)
epoch          4 0.4251  15.4097     4   145 1.588e-10 ***
Residuals 145
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The  $p$ -value associated with each four test criteria is very small and there is strong evidence that the skull measurements differ between the five epochs. We might now move on to investigate which epochs differ and on which variables. We can look at the univariate  $F$ -tests for each of the four variables by using the code

```
R> summary.aov(manova(cbind(mb, bh, bl, nh) ~ epoch, data = skulls))
```

```
Response mb :
              Df Sum Sq Mean Sq F value    Pr(>F)
epoch          4  502.83   125.71   5.9546 0.0001826 ***
Residuals 145 3061.07    21.11
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Response bh :
```

```

              Df Sum Sq Mean Sq F value    Pr(>F)
epoch           4   229.9     57.5   2.4474 0.04897 *
Residuals     145  3405.3     23.5
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Response bl :
              Df Sum Sq Mean Sq F value    Pr(>F)
epoch           4   803.3    200.8   8.3057 4.636e-06 ***
Residuals     145  3506.0     24.2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Response nh :
              Df Sum Sq Mean Sq F value    Pr(>F)
epoch           4    61.20    15.30   1.507 0.2032
Residuals     145 1472.13    10.15

```

We see that the results for the maximum breadths (**mb**) and basialveolar length (**bl**) are highly significant, with those for the other two variables, in particular for nasal heights (**nh**), suggesting little evidence of a difference. To look at the pairwise multivariate tests (any of the four test criteria are equivalent in the case of a one-way layout with two levels only) we can use the `summary` method and `manova` function as follows:

```

R> summary(manova(cbind(mb, bh, bl, nh) ~ epoch, data = skulls,
+                 subset = epoch %in% c("c4000BC", "c3300BC")))

```

```

              Df Pillai approx F num Df den Df Pr(>F)
epoch           1 0.02767  0.39135      4    55 0.814
Residuals     58

```

```

R> summary(manova(cbind(mb, bh, bl, nh) ~ epoch, data = skulls,
+                 subset = epoch %in% c("c4000BC", "c1850BC")))

```

```

              Df Pillai approx F num Df den Df Pr(>F)
epoch           1 0.1876  3.1744      4    55 0.02035 *
Residuals     58

```

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

R> summary(manova(cbind(mb, bh, bl, nh) ~ epoch, data = skulls,
+                 subset = epoch %in% c("c4000BC", "c200BC")))

```

```

              Df Pillai approx F num Df den Df Pr(>F)
epoch           1 0.3030  5.9766      4    55 0.0004564 ***
Residuals     58

```

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

R> summary(manova(cbind(mb, bh, bl, nh) ~ epoch, data = skulls,
+                 subset = epoch %in% c("c4000BC", "cAD150")))

```

```

              Df Pillai approx F num Df den Df      Pr(>F)
epoch          1 0.3618    7.7956      4    55 4.736e-05 ***
Residuals 58
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

To keep the overall significance level for the set of all pairwise multivariate tests under some control (and still maintain a reasonable power), [Stevens \(2001\)](#) recommends setting the nominal level  $\alpha = 0.15$  and carrying out each test at the  $\alpha/m$  level where  $m$  is the number of tests performed. The results of the four pairwise tests suggest that as the epochs become further separated in time the four skull measurements become increasingly distinct.



---

## Bibliography

---

- Everitt, B. S. (1996), *Making Sense of Statistics in Psychology: A Second-Level Course*, Oxford, UK: Oxford University Press.
- Hochberg, Y. and Tamhane, A. C. (1987), *Multiple Comparison Procedures*, New York, USA: John Wiley & Sons.
- Stevens, J. (2001), *Applied Multivariate Statistics for the Social Sciences*, Mahwah, New Jersey, USA: Lawrence Erlbaum, 4th edition.
- Tukey, J. W. (1953), "The problem of multiple comparisons (unpublished manuscript)," in *The Collected Works of John W. Tukey VIII. Multiple Comparisons: 1948-1983*, New York, USA: Chapman & Hall.