

# The Statistical Sleuth in R:

## Chapter 1

Ruobing Zhang

Kate Aloisio

Nicholas J. Horton\*

August 24, 2024

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Motivation and Creativity</b>	<b>2</b>
2.1	Statistical summary and graphical display . . . . .	2
2.2	Inferential procedures (two-sample t-test) . . . . .	4
2.3	Permutation test . . . . .	5
<b>3</b>	<b>Gender Discrimination</b>	<b>6</b>
3.1	Statistical summary and graphical display . . . . .	6
3.2	Inferential procedures (two-sample t-test) . . . . .	8
3.3	Permutation test . . . . .	9

## 1 Introduction

This document is intended to help describe how to undertake analyses introduced as examples in the Second Edition of the *Statistical Sleuth* (2002) by Fred Ramsey and Dan Schafer. More information about the book can be found at <http://www.proaxis.com/~panorama/home.htm>. This file as well as the associated `knitr` reproducible analysis source file can be found at <http://www.amherst.edu/~nhorton/sleuth>.

This work leverages initiatives undertaken by Project MOSAIC (<http://www.mosaic-web.org>), an NSF-funded effort to improve the teaching of statistics, calculus, science and computing in the undergraduate curriculum. In particular, we utilize the `mosaic` package, which was written to simplify the use of R for introductory statistics courses. A short summary of the R needed to teach introductory statistics can be found in the `mosaic` package vignette (<http://cran.r-project.org/web/packages/mosaic/vignettes/MinimalR.pdf>).

To use a package within R, it must be installed (one time), and loaded (each session). The package can be installed using the following command:

---

\*Department of Mathematics, Amherst College, [nhorton@amherst.edu](mailto:nhorton@amherst.edu)

```
> install.packages('mosaic') # note the quotation marks
```

Once this is installed, it can be loaded by running the command:

```
> require(mosaic)
```

This needs to be done once per session.

In addition the data files for the *Sleuth* case studies can be accessed by installing the **Sleuth2** package.

```
> install.packages('Sleuth2') # note the quotation marks
```

```
> require(Sleuth2)
```

We also set some options to improve legibility of graphs and output.

```
> trellis.par.set(theme=col.mosaic()) # get a better color scheme for lattice  
> options(digits=3)
```

The specific goal of this document is to demonstrate how to calculate the quantities described in Chapter 1: Drawing Statistical Conclusions using R.

## 2 Motivation and Creativity

For Case Study 1: Motivation and Creativity, the following questions are posed: Do grading systems promote creativity in students? Do ranking systems and incentive awards increase productivity among employees? Do rewards and praise stimulate children to learn? Do rewards and praise stimulate children to learn?

The data for Case Study 1 was collected by psychologist Teresa Amabile in an experiment concerning the effects of intrinsic and extrinsic motivation on creativity (page 2 of the *Sleuth*).

### 2.1 Statistical summary and graphical display

We begin by reading the data and summarizing the variables.

```
> summary(case0101)
```

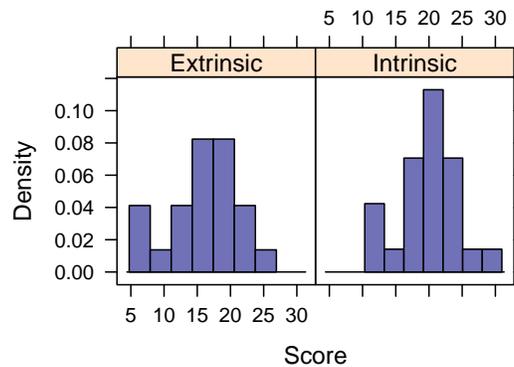
Score	Treatment
Min. : 5.0	Extrinsic:23
1st Qu.:14.9	Intrinsic:24
Median :18.7	
Mean :17.9	
3rd Qu.:21.2	
Max. :29.7	

A total of 47 subjects with considerable experience in creative writing were randomly assigned to one of two treatment groups: 23 were placed into the “extrinsic” treatment group and 24 were placed into the “intrinsic” treatment group. As summarized in Display 1.1 (*Sleuth*, page 2)

```
> favstats(Score ~ Treatment, data=case0101)

Treatment min   Q1 median   Q3  max mean   sd  n missing
1 Extrinsic   5 12.2  17.2 19.0 24.0 15.7 5.25 23      0
2 Intrinsic  12 17.4  20.4 22.3 29.7 19.9 4.44 24      0

> histogram(~ Score | Treatment, data=case0101)
```



```
> with(subset(case0101, Treatment=="Extrinsic"), stem(Score, scale=5))
```

The decimal point is at the |

```
5 | 04
6 | 1
7 |
8 |
9 |
10 | 9
11 | 8
12 | 03
13 |
14 | 8
15 | 0
16 | 8
17 | 2245
18 | 577
19 | 25
20 | 7
```

```
21 | 2
22 | 1
23 |
24 | 0

> with(subset(case0101, Treatment=="Intrinsic"), stem(Score, scale=5))

The decimal point is at the |

12 | 009
13 | 6
14 |
15 |
16 | 6
17 | 25
18 | 2
19 | 138
20 | 356
21 | 36
22 | 126
23 | 1
24 | 03
25 |
26 | 7
27 |
28 |
29 | 7
```

Similar output can be generated using the following code:

```
> maggregate(Score ~ Treatment, FUN=stem, data=case0101)
```

The extrinsic group (n=23) has an average creativity score that is 4.1 points less than the intrinsic group (n=24). The extrinsic group has relatively larger spread than the intrinsic group (sd=5.25 for extrinsic group and sd=4.44 for intrinsic group). Both distributions are approximately normally distributed.

## 2.2 Inferential procedures (two-sample t-test)

```
> t.test(Score ~ Treatment, alternative="two.sided", data=case0101)
```

```
Welch Two Sample t-test
```

```

data: Score by Treatment
t = -3, df = 43, p-value = 0.006
alternative hypothesis: true difference in means between group Extrinsic and group Intrinsic is
95 percent confidence interval:
 -7.01 -1.28
sample estimates:
mean in group Extrinsic mean in group Intrinsic
          15.7              19.9

```

The two-sample  $t$ -test shows strong evidence that a subject would receive a lower creativity score for a poem written after the extrinsic motivation questionnaire than for one written after the intrinsic motivation questionnaire. The two-sided  $p$ -value is 0.006, which is small enough to reject the null hypothesis.

Thus, we can conclude that there is a difference between the population mean in extrinsic group and the population mean for the intrinsic group; the estimated difference between these two scores is 4.1 points on the 0-40 point scale. A 95% confidence interval for the decrease in score due to having extrinsic motivation rather than intrinsic motivation is between -1.28 and -7.01 points (*Sleuth*, page 3).

```

> summary(lm(Score ~ Treatment, data=case0101))

Call:
lm(formula = Score ~ Treatment, data = case0101)

Residuals:
    Min     1Q  Median     3Q     Max
-10.74  -2.98   1.06   2.96   9.82

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      15.74      1.01    15.55 <2e-16 ***
TreatmentIntrinsic  4.14      1.42     2.93  0.0054 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.85 on 45 degrees of freedom
Multiple R-squared:  0.16, Adjusted R-squared:  0.141
F-statistic: 8.56 on 1 and 45 DF, p-value: 0.00537

```

In the creativity study, the question that whether there is a treatment effect becomes a question about whether the parameter has a nonzero value. The value of the test statistic for the creativity scores is 4.14.

### 2.3 Permutation test

```

> diffmeans = diff(mean(Score ~ Treatment, data=case0101))
> diffmeans      # observed difference

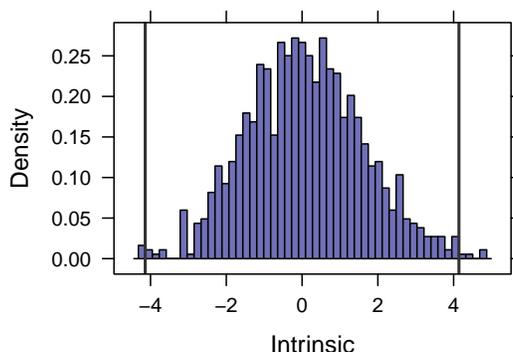
Intrinsic
  4.14

> numsim = 1000      # set to a sufficient number
> nulldist = do(numsim)*diff(mean(Score~shuffle(Treatment), data=case0101))
> confint(nulldist)

      name lower upper level      method estimate
1 Intrinsic -2.73  3.28  0.95 percentile      4.14

> # Display 1.8 Sleuth
> histogram(~ Intrinsic, nint=50, data=nulldist, v=c(-4.14,4.14))

```



As described in the Sleuth on page 12, if the group assignment changes, we will get different results. First, the test statistics will be just as likely to be negative as positive. Second, the majority of values fall in the range from -3.0 to +3.0. Third, only few of the 1,000 randomization produced test statistics as large as 4.14. This last point indicates that 4.14 is a value corresponding to an unusually uneven randomization outcome, if the null hypothesis is correct.

### 3 Gender Discrimination

For Case Study 2: Gender Discrimination the following questions are posed: Did a bank discriminatorily pay higher starting salaries to men than to women? Display 1.3 (page 4 of the *Sleuth*) displays the beginning salaries for male and female skilled entry level clerical employees hired between 1969 and 1977.

#### 3.1 Statistical summary and graphical display

We begin by reading the data and summarizing the variables.

```
> summary(case0102) # Display 1.3 Sleuth p4
```

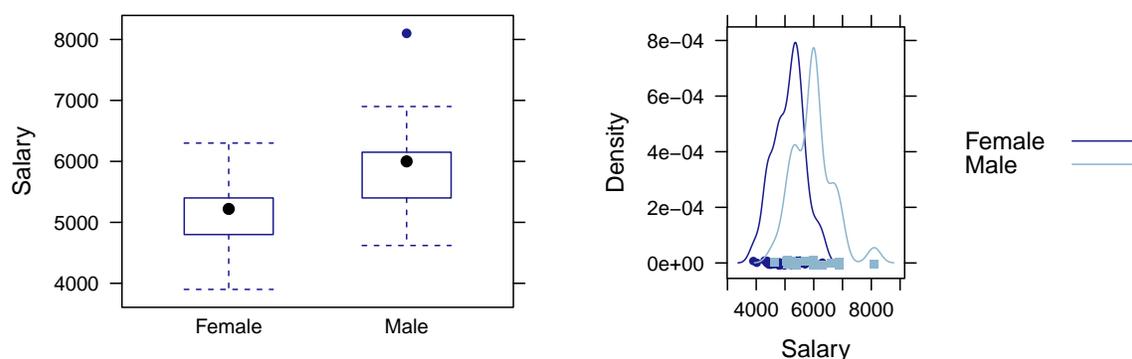
```
      Salary      Sex
Min.   :3900  Female:61
1st Qu.:4980  Male  :32
Median :5400
Mean   :5420
3rd Qu.:6000
Max.   :8100
```

```
> favstats(Salary ~ Sex, data=case0102)
```

```
      Sex min  Q1 median  Q3 max mean  sd  n missing
1 Female 3900 4800  5220 5400 6300 5139 540 61      0
2  Male 4620 5400  6000 6075 8100 5957 691 32      0
```

```
> bwplot(Salary ~ Sex, data=case0102)
```

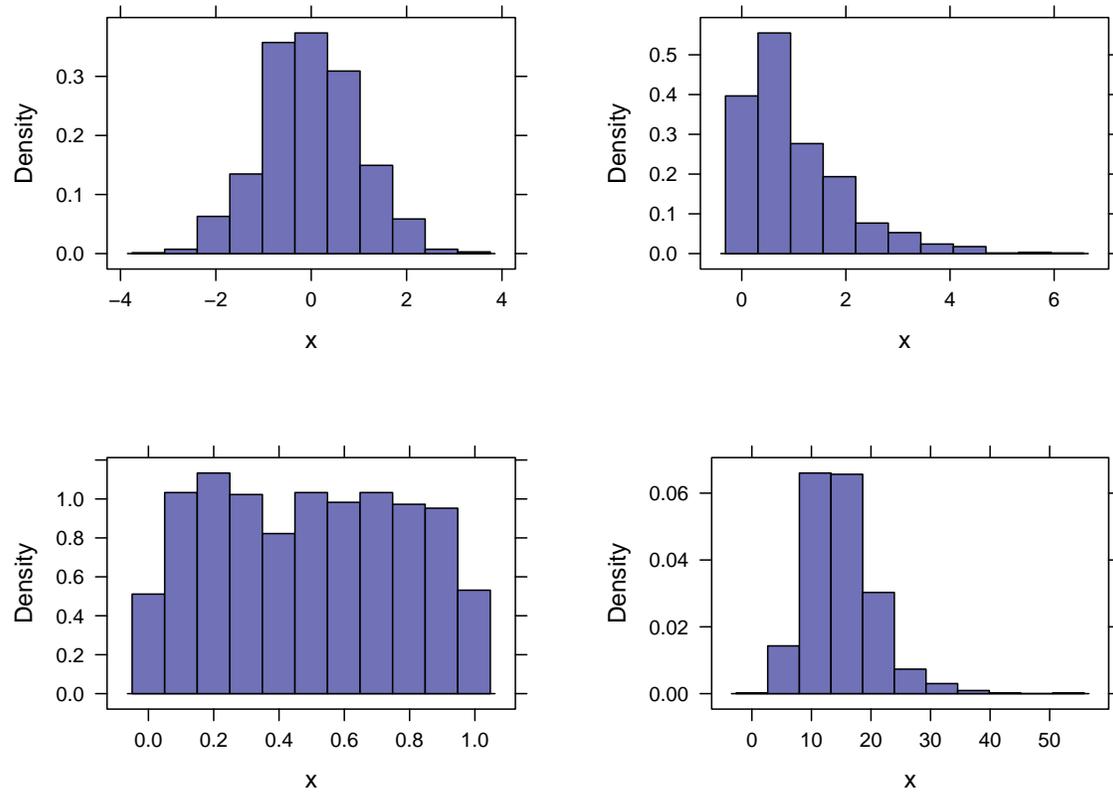
```
> densityplot(~ Salary, groups=Sex, auto.key=TRUE, data=case0102)
```



The 0 men have an average starting salary that is \$818 more than the 61 women (\$5957 vs \$5139). Both distributions have similar spread ( $sd = \$539.87$  for women and  $sd = \$690.73$  for men) and distributions that are approximately normally distributed (see density plot). The key difference between the groups is the shift (as indicated by the parallel boxplots).

To show Display 1.13

```
> x = rnorm(1000)
> histogram(~ x) # Normal
> x = rexp(1000)
> histogram(~ x) # Long-tailed
> x = runif(1000)
> histogram(~ x) # Short-tailed
> x = rchisq(1000, df=15)
> histogram(~ x) # Skewed
```



### 3.2 Inferential procedures (two-sample t-test)

The *t*-test on page 4 of Sleuth can be replicated using the following commands (note that the equal-variance *t*-test is specified by `var.equal=TRUE` which is not the default).

```
> t.test(Salary ~ Sex, var.equal=TRUE, data=case0102)
```

```
Two Sample t-test
```

```
data: Salary by Sex
```

```
t = -6, df = 91, p-value = 1e-08
```

```
alternative hypothesis: true difference in means between group Female and group Male is not equal to 0
```

```
95 percent confidence interval:
```

```
-1076 -560
```

```
sample estimates:
```

```
mean in group Female    mean in group Male
                5139                5957
```

### 3.3 Permutation test

We undertake a permutation test to assess whether the differences in the center of these samples that we are observing are due to chance, if the distributions are actually equivalent back in the populations of male and female possible clerical hires. We start by calculating our test statistic (the difference in means) for the observed data, then simulate from the null distribution (where the labels can be interchanged) and calculate our  $p$ -value.

```
> obsdiff = diff(mean(Salary ~ Sex, data=case0102)); obsdiff

Male
818

> numsim = 1000
> res = do(numsim) * diff(mean(Salary~shuffle(Sex), data=case0102))
> densityplot(~ Male, data=res)
> confint(res)

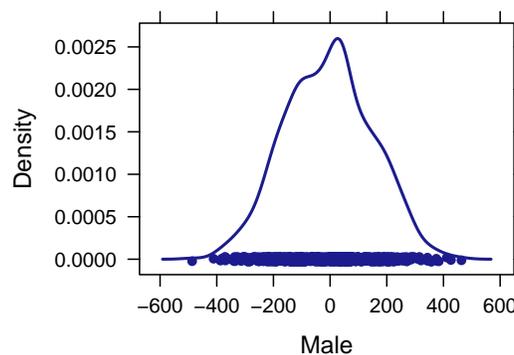
  name lower upper level    method estimate
1 Male  -303   281 0.95 percentile      818

> favstats(~ Male, data=res)

  min  Q1 median  Q3 max  mean  sd    n missing
-487 -115   -2.4 103 464 -5.92 154 1000      0

> p = sum(abs(res$Male) >= abs(obsdiff))/numsim; p

[1] 0
```



Through the permutation test, we observe that the mean starting salary for males is estimated to be \$515.4 to \$1098.7 larger than the mean starting salary for females (95% confidence interval). We never see a permuted difference in means close to our observed value. Therefore, we reject the null hypothesis ( $p < 0.001$ ) and conclude that the salaries of the men are significantly higher than that of the women.

```
> t.test(Salary ~ Sex, alternative="less", data=case0102)

Welch Two Sample t-test

data: Salary by Sex
t = -6, df = 51, p-value = 2e-07
alternative hypothesis: true difference in means between group Female and group Male is less than 0
95 percent confidence interval:
 -Inf -583
sample estimates:
mean in group Female    mean in group Male
                5139                5957
```

The  $p$ -value ( $< 0.001$ ) from the two-sample  $t$ -test shows that the large difference between estimated salaries for males and females is unlikely to be due to chance.