STATS 3N03/3J04

2004-09-29
7-1

INTERACTION PLOTS

- EASY TO MAKE BY HAND
- DO BY HAND ON FINAL EXAM
- SIMPLE WAY IN R:
  - COMPUTE MEAN FOR EACH POSITION-TEMP. COMBINATION ON CALCULATOR
  - MAKE NEW DATA FRAME
  - PLOT DENSITY VS. TEMP, DIFFERENT PLOT SYMBOL FOR EACH POSITION
  - ADD LINES BY HAND
- CHANCE TO LEARN SOME ADVANCED FUNCTIONS WITH MY interactplot() FUNCTION.

TEXT COVERED TO DATE:

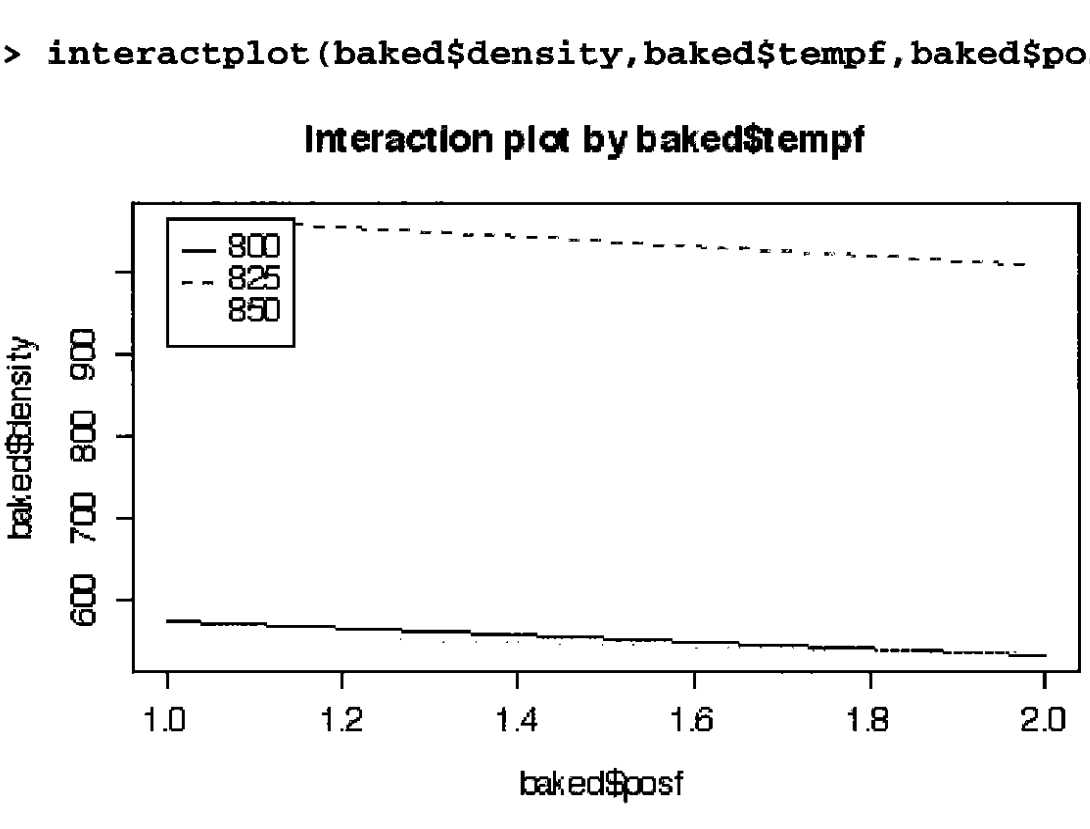
- ALL OF CHAPT. 1
- ALL OF CHAPT. 6 (EXCEPT PROBABILITY PLOTS)
- START OF CHAPT. 4 (4.1, 4.2, 4.6)

```
> baked
   density position temp
1      570        1  800
2      565        1  800
3      583        1  800
4     1063        1  825
5     1080        1  825
6     1043        1  825
7      565        1  850
8      510        1  850
9      590        1  850
10     528        2  800
11     547        2  800
12     521        2  800
13     988        2  825
14    1026        2  825
15    1004        2  825
16     526        2  850
17     538        2  850
18     532        2  850
> baked$posf <- as.factor(baked$position)
> baked$tempf <- as.factor(baked$temp)
> baked
   density position temp posf tempf
1      570        1  800    1   800
2      565        1  800    1   800
3      583        1  800    1   800
4     1063        1  825    1   825
5     1080        1  825    1   825
6     1043        1  825    1   825
7      565        1  850    1   850
8      510        1  850    1   850
9      590        1  850    1   850
10     528        2  800    2   800
```

```
11     547         2  800     2   800
12     521         2  800     2   800
13     988         2  825     2   825
14    1026         2  825     2   825
15    1004         2  825     2   825
16     526         2  850     2   850
17     538         2  850     2   850
18     532         2  850     2   850
> save.image()
> baked$posf
 [1] 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2
Levels: 1 2
> baked$tempf
 [1] 800 800 800 825 825 825 850 850 850 800 800 800 825 825 825 850 850 850
Levels: 800 825 850
> baked$posf:baked$tempf
 [1] 1:800 1:800 1:800 1:825 1:825 1:825 1:850 1:850 1:850 2:800 2:800 2:800
[13] 2:825 2:825 2:825 2:850 2:850 2:850
Levels: 1:800 1:825 1:850 2:800 2:825 2:850
> split(baked$density, baked$posf:baked$tempf)
$"1:800"
[1] 570 565 583

$"1:825"
[1] 1063 1080 1043

$"1:850"
[1] 565 510 590

$"2:800"
[1] 528 547 521

$"2:825"
[1]  988 1026 1004
```

```
$"2:850"
[1] 526 538 532

> lapply(split(baked$density, baked$posf:baked$tempf), mean)
$"1:800"
[1] 572.6667

$"1:825"
[1] 1062

$"1:850"
[1] 555

$"2:800"
[1] 532

$"2:825"
[1] 1006

$"2:850"
[1] 532

> sapply(split(baked$density, baked$posf:baked$tempf), mean)
    1:800     1:825     1:850     2:800     2:825     2:850
 572.6667 1062.0000  555.0000  532.0000 1006.0000  532.0000
> matrix(sapply(split(baked$density, baked$posf:baked$tempf), mean),
ncol=nlevels(baked$posf))
          [,1] [,2]
[1,]  572.6667  532
[2,] 1062.0000 1006
[3,]  555.0000  532
> matplot(matrix(sapply(split(baked$density, baked$posf:baked$tempf), mean),
ncol=nlevels(baked$posf)), type="l")
```

```
> interactplot
function (y, facta, factb, xlab = deparse(substitute(factb)),
    ylab = deparse(substitute(y)), main = paste("Interaction plot by",
        deparse(substitute(facta))), ...)
{
    values <- sapply(split(y, facta:factb), mean)
    matplot(matrix(values, ncol = nlevels(facta)), type = "l",
        xlab = xlab, ylab = ylab, ...)
    title(main = main)
    legend(1, max(values), levels(facta), lty = 1:nlevels(facta),
        col = 1:nlevels(facta))
    invisible()
}

> interactplot(baked$density,baked$tempf,baked$posf)
```



SAMPLE MEAN, VARIANCE, STANDARD DEVIATION, CORRELATION COEFFICIENT, SIMPLE LINEAR REGRESSION

- BY HAND
- BY CALCULATOR **
- BY R

DATA:

$$x_1, \ldots, x_n$$

SAMPLE MEAN:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

SAMPLE VARIANCE:

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= \frac{1}{n-1}\left\{ \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} \right\}$$

SAMPLE STANDARD DEVIATION:

$$s_x = \sqrt{s_x^2}$$

EX 20, 26, 32, 32, 45, 51 ← FIND THESE!!

$n = 6 \quad \sum x_i = 206 \quad \sum x_i^2 = 7750$

$$\bar{x} = \frac{206}{6} = 34.3\dot{3}$$

$$s_x^2 = \frac{1}{6-1}\left\{ 7750 - \frac{(206)^2}{6} \right\}$$

$$= 135.4\dot{6}$$

σ_{n-1} $s_x = \sqrt{135.4\dot{6}} = 11.64$

NUMBER OF SIGNIFICANT DIGITS:

- MEAN 1-3 MORE THAN ORIGINAL UNITS
- STD. DEVIATION SAME AS MEAN
- DON'T ROUND UNTIL THE END!!

ON CALCULATOR:

- ENTER DATA WITH $\Sigma+$
- REMOVE WITH $\Sigma-$
- KEYS FOR $\bar{x}$, s
- FIND WHERE n, Σx_i, Σx_i^2 ARE STORED !!!

BIVARIATE DATA

$(x_1, y_1), \ldots, (x_n, y_n)$

ALSO GET $\bar{y}$, s_y^2, s_y

PRODUCT MOMENT

$$s_{xy} = \frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y})$$

$$= \frac{1}{n-1} \left\{ \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n} \right\}$$

CORRELATION COEFFICIENT

$$r = \frac{s_{xy}}{s_x s_y}$$

REGRESSION COEFFICIENTS

SLOPE: $b = \dfrac{S_{xy}}{S_x^2}$

INTERCEPT: $a = \bar{y} - b\bar{x}$

EX

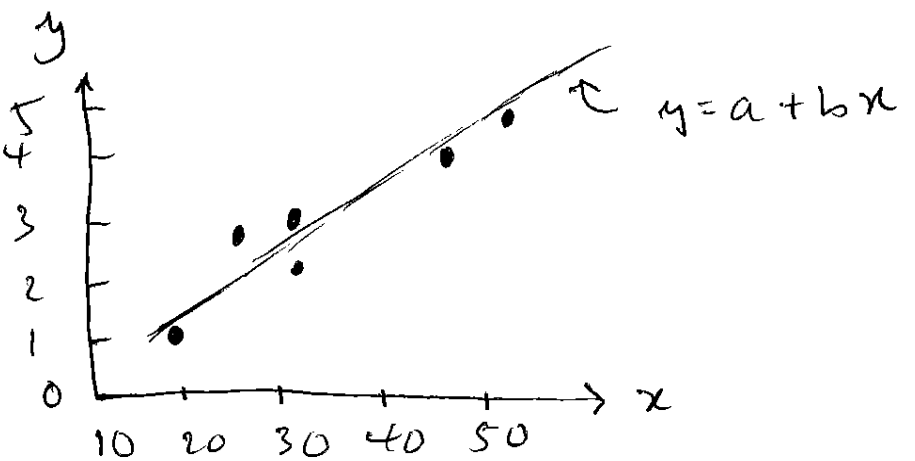
x	y
20	1.0
26	2.8
32	2.2
32	3.0
45	3.8
51	4.1

$n = 6$

$\Sigma x = 206 \quad \Sigma y = 16.9$

$\Sigma x^2 = 7750 \quad \Sigma y^2 = 53.93$

$\Sigma xy = 639.3$



COMPUTE $\bar{x}$, $\bar{y}$, S_x^2, S_x, S_y^2, S_y, S_{xy}, r, a, b BY HAND AND ON YOUR CALCULATOR!

7-10

```
> slrdata<-data.frame(x=c(20,26,32,32,45,51),
y=c(1,2.8,2.2,3,3.8,4.1))
> slrdata
   x   y
1 20 1.0
2 26 2.8
3 32 2.2
4 32 3.0
5 45 3.8
6 51 4.1
> mean(slrdata)
        x         y
34.333333  2.816667
> var(slrdata)
          x         y
x 135.46667 11.813333
y  11.81333  1.265667
> diag(var(slrdata))
         x          y
135.466667   1.265667
> sqrt(diag(var(slrdata)))
        x         y
11.639015  1.125019
> cor(slrdata)
         x        y
x 1.000000 0.902187
y 0.902187 1.000000
> coef(lm(y~x, data=slrdata))
(Intercept)           x
-0.17736220  0.08720472
> plot(y~x, data=slrdata)
> abline(lm(y~x, data=slrdata))
```

$$\begin{pmatrix} s_x^2 & s_{xy} \\ s_{xy} & s_y^2 \end{pmatrix}$$

s_x^2 s_y^2

s_x s_y