2004-11-24

STATS 3N03/3J04 30-1

COMMENTS ON THE SINGLE-FACTOR EXAMPLE:

EFFECT OF CODING DATA:

- SUBTRACTING A CONSTANT CHANGES NOTHING IN ANOVA

- MULTIPLYING BY k MULTIPLIES ALL SUMS OF SQUARES BY k^2 BUT LEAVES THE F-RATIOS UNCHANGED.

GETTING THE P-VALUE FROM TABLES:

$$f_{.1,3,18} = 2.42 \qquad f_{.05,3,18} = 3.16$$

$$F_0 = 2.616$$

$$\Rightarrow \quad .05 < P < .1$$

OR: $P = 0.0836$

"SOME EVIDENCE $(.05 < P < .1)$ FROM THESE DATA THAT THE MEAN DENSITY OF BRICKS IS NOT THE SAME AT EACH OF THE FIRING TEMPERATURES TESTED."

P-VALUE BY COMPUTER

1 - pf(2.616, 3, 18)

0.0826

ANOVA P-VALUES ARE ALWAYS RIGHT TAIL, BECAUSE

DENOMINATOR MSE IS ALWAYS AN ESTIMATE OF σ^2

NUMERATOR MSB IS A SUM OF SQUARES THAT IS AN ESTIMATE OF σ^2 IF H_0 IS TRUE, BUT IS INFLATED IF H_0 IS FALSE.

$$E(MSB \mid H_0) = \sigma^2$$

$$E(MSB \mid H_1) > \sigma^2$$

COMPARE TO 5% ACCEPT-REJECT TEST:

"WE ACCEPT THE HYPOTHESIS THAT MEAN DENSITY IS THE SAME AT EACH OF THE FIRING TEMPERATURES TESTED, AT THE 5% LEVEL OF SIGNIFICANCE"

IF THERE ARE ONLY 2 GROUPS $(a=2)$ COULD USE 2-SAMPLE t-TEST OR SINGLE-FACTOR ANOVA

SHOW: $t_0^2 = F_0$

$t \sim t(n-1)$ $\quad$ $F(1, n-1)$

SEE: "DEFINITIONS" WEB PAGE. RELATIONS BETWEEN NORMAL, χ^2, t AND F. SELF STUDY.

UNBIASED ESTIMATE:

"A STATISTIC IS AN UNBIASED ESTIMATE OF A PARAMETER IF ITS EXPECTED VALUE EQUALS THE TRUE VALUE OF THE PARAMETER"

EX

$E(\bar{X}) = \mu \Rightarrow$ UNBIASED

$E(s^2) = \sigma^2 \Rightarrow$ UNBIASED

$E(s) \approx \sigma\left(1 - \frac{1}{4n}\right) \Rightarrow$ BIASED

$E\left(\frac{1}{n}\sum(x_i - \bar{x})^2\right) = \frac{n-1}{n}\sigma^2 \Rightarrow$ BIASED

EMPIRICAL VERIFICATION

$n = 5 \qquad E(s) \approx \sigma\left(1 - \frac{1}{20}\right) = 0.95\sigma$

```
> mean(sqrt(apply(matrix(rnorm(1000*5),ncol=5),1,var)))
[1] 0.9490131
> mean(sqrt(apply(matrix(rnorm(1000*5),ncol=5),1,var)))
[1] 0.938579
> mean(sqrt(apply(matrix(rnorm(1000*5),ncol=5),1,var)))
[1] 0.935386
```

$n = 10 \qquad E(s) \approx \sigma\left(1 - \frac{1}{40}\right) = 0.975\sigma$

```
> mean(sqrt(apply(matrix(rnorm(1000*10),ncol=10),1,var)))
[1] 0.9726705
> mean(sqrt(apply(matrix(rnorm(1000*10),ncol=10),1,var)))
[1] 0.977451
> mean(sqrt(apply(matrix(rnorm(1000*10),ncol=10),1,var)))
[1] 0.9809756
```

2-FACTOR ANOVA WITH INTERACTION

F2

F1		H	L
	H	1, 2	7, 6
	L	9, 7	2, 5

$a = 2$ LEVELS OF FACTOR 1 (A)

$b = 2$ " " " 2 (B)

$\Rightarrow ab = 4$ DESIGN POINTS

$n = 2$ REPLICATIONS AT EACH DESIGN POINT.

* BALANCED
* REPLICATION AT EACH DESIGN POINT.

ANALYSIS "BY HAND"

1-FACTOR ANOVA OF ROWS $\Rightarrow$ SSA

1-FACTOR ANOVA OF COLS $\Rightarrow$ SSB

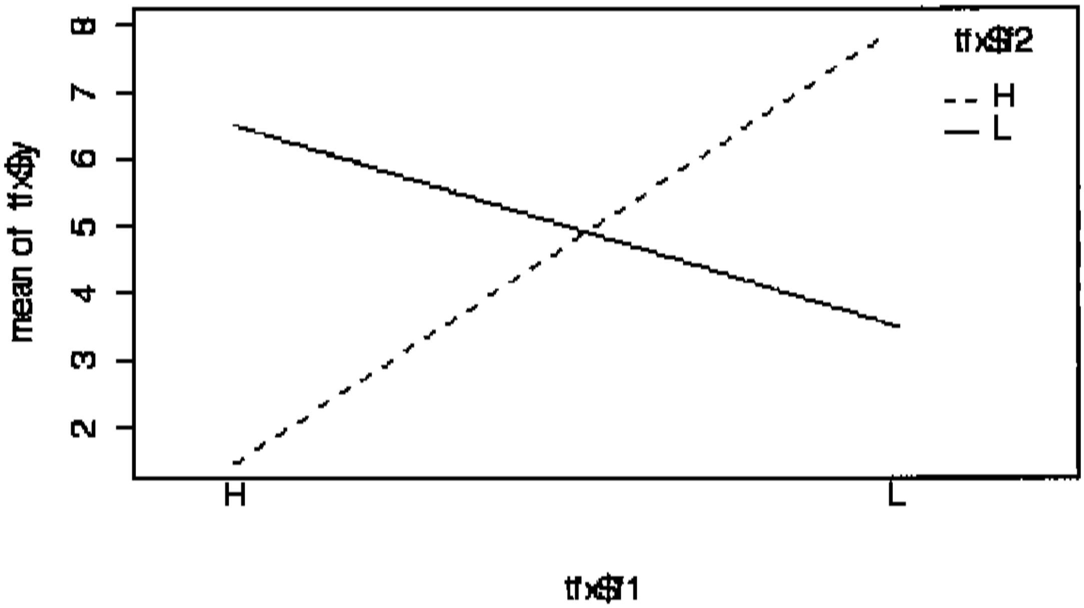
POOLED VARIANCE EST. OVER DESIGN PTS $\Rightarrow$ SSE

DIFFERENC FROM TOTAL $\Rightarrow$ SSI

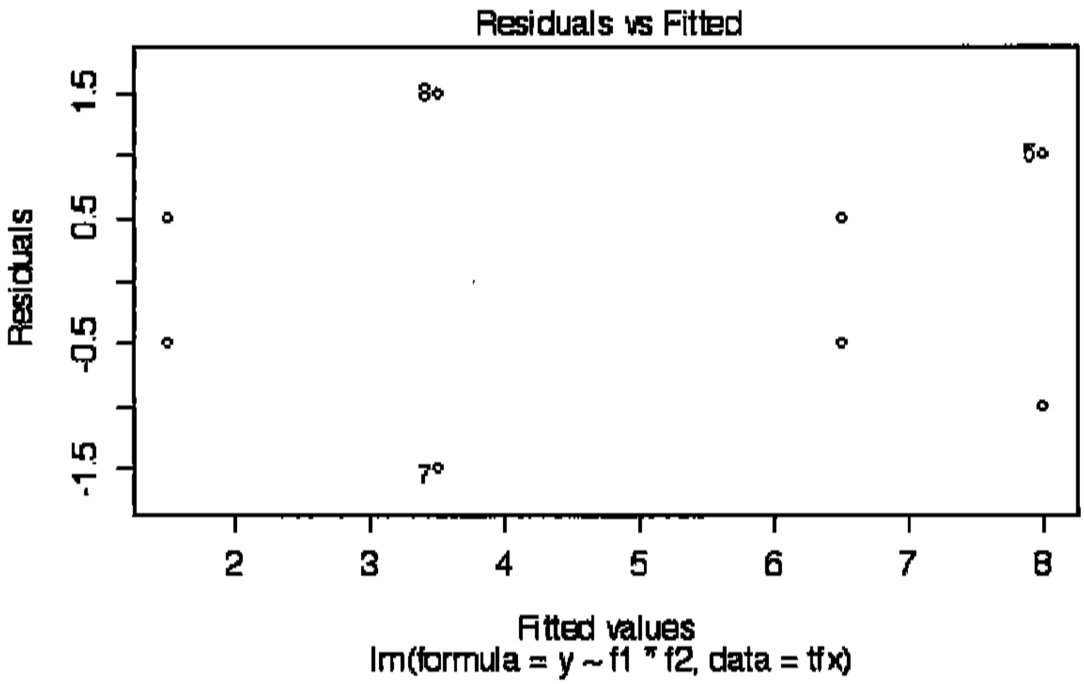
Two-factor Anova with Interaction

```
> tfx <- data.frame(y=c(1,2,7,6,9,7,2,5),
f1=factor(rep(c("H","L"),c(4,4))),
f2=factor(rep(rep(c("H","L"),c(2,2)),2)))

> tfx
  y f1 f2
1 1  H  H
2 2  H  H
3 7  H  L
4 6  H  L
5 9  L  H
6 7  L  H
7 2  L  L
8 5  L  L

> anova(lm(y~f1+f2, data=tfx))
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value Pr(>F)
f1         1  6.125   6.125  0.5819 0.4800
f2         1  0.125   0.125  0.0119 0.9175
Residuals  5 52.625  10.525

> anova(lm(y~f1*f2, data=tfx))
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value  Pr(>F)
f1         1  6.125   6.125  3.2667 0.14499
f2         1  0.125   0.125  0.0667 0.80899
f1:f2      1 45.125  45.125 24.0667 0.00801 **
Residuals  4  7.500   1.875
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1
```

```
> interaction.plot(tfx$f1,tfx$f2,tfx$y)
```

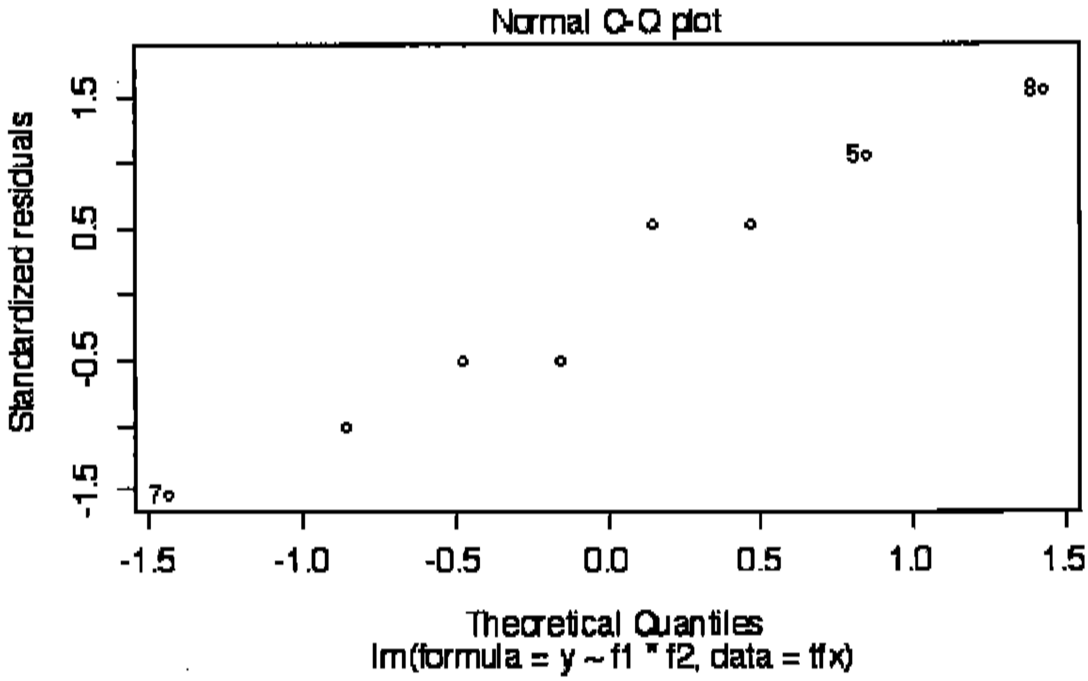



```
> plot(lm(y~f1*f2, data=tfx))
Hit &lt;Return&gt; to see next plot:
```

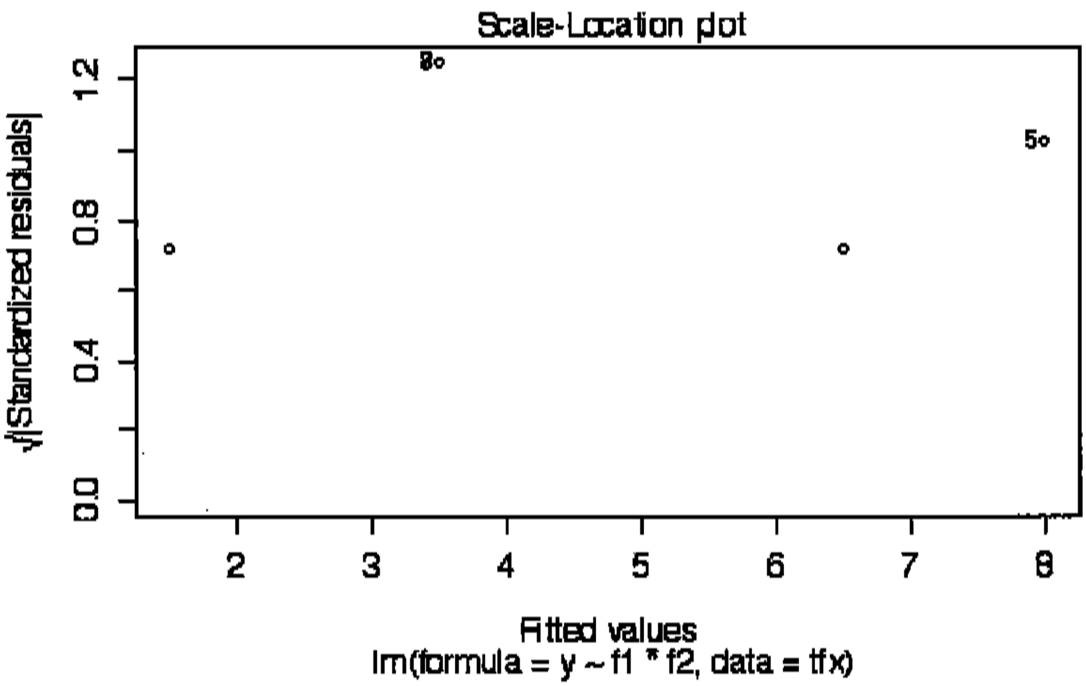


30-8

Hit <Return> to see next plot:



Hit <Return> to see next plot:



30-9

Hit <Return> to see next plot: