STAT 581 - Exam 1: Due Dec 2, 2021

Alex Towell (atowell@siue.edu)

Contents

Problem 1

An experiment is conducted to study the effect of fitness level on ego strength. Random samples of college faculty members are selected from each fitness level, and an ego score is observed for each member in the sample. Higher values indicate greater ego. The data is provided as an attachment.

Preliminary analysis

We are interested in whether

$$\text{input} = \text{fitness level}$$

has an effect on

$$\text{response} = \text{ego strength}.$$

We have two populations, one in which the mean fitness level is `high` and another in which the mean fitness level is `low`, denoted respectively by group 1 and group 2.

We take a sample of size n_1 from group 1 (`high`)

$$Y_{1j} = \mu_1 + \epsilon_{1j}$$

and a sample of size n_2 from group 2 (`low`),

$$Y_{2j} = \mu_2 + \epsilon_{2j},$$

where

$$\begin{aligned} &Y_{ij} \text{ is the } j\text{-th ego strength response for } i\text{-th group (fitness level or treatment)},\\ &\mu_i \text{ is the mean response for the } i\text{-th group},\\ &\epsilon_{ij} \text{ is iid normal with zero mean.} \end{aligned}$$

The result is two samples:

group 1	group 2
y_{11}	y_{21}
y_{12}	y_{22}
$\vdots$	$\vdots$
y_{1n_1}	y_{2n_2}

Part (a)

> State the hypotheses of interest. Provide an interpretation, stated in the context of the problem.

Reproduce

The hypothesis of interest is whether fitness level is effected by ego strength. We may formulate this as a hypothesis test of the form

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 && \text{(fitness level has no effect on ego strength)},\\ H_A &: \mu_1 \neq \mu_2 && \text{(fitness level does have an effect on ego strength)}. \end{aligned}$$

where μ_1 and μ_2 are respectively the expected ego strengths for the `high` and `low` fitness level groups.

Part (b)

> Compute the t_0 statistic and the p-value. Provide an interpretation, statedin the context of the problem.

Analysis

Let $W = \bar{y}_{1\cdot} - \bar{y}_{2\cdot}$. Then,

$$E(W) = \mu_1 - \mu_2$$

and

$$\sigma_W^2 = \sigma^2(\bar{y}_{1\cdot}) + \sigma^2(\bar{y}_{2\cdot}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Assuming $\sigma_1 = \sigma_2 = \sigma$,

$$\sigma_W^2 = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right).$$

Since linear combinations of normal random variates are normal,

$$\frac{\bar{y}_{1\cdot} - \bar{y}_{2\cdot}}{\sigma_W} \sim \mathcal{N}(\mu_1 - \mu_2, 1).$$

We do not know σ^2, so we estimate it with the pooled estimator

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

where

$$s_j^2 = \mathrm{SS}_1 / (n_j - 1).$$

We are interested in $H_0 : \mu_1 - \mu_2 = 0$, and thus the appropriate test statistic is given by

$$t_0 = \frac{\bar{y}_{1\cdot} - \bar{y}_{2\cdot}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

which under H_0 has the reference distribution $t(n_1 + n_2 - 2)$.

```r
library("readxl")
data = read_excel("exam1data.xlsx")
fitness = as.factor(na.omit(data$fitness.level))
ego = na.omit(data$ego.score)
t.star = t.test(ego ~ fitness, var.equal=T)
t.star
```

```
##
##  Two Sample t-test
##
## data:  ego by fitness
## t = 3.7092, df = 28, p-value = 0.0009111
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.6268509 2.1731491
## sample estimates:
## mean in group high  mean in group low
##                5.2                3.8
```

Reproduce: statistics

We see that $t_0 = 3.71$ with a p-value $= .001$.

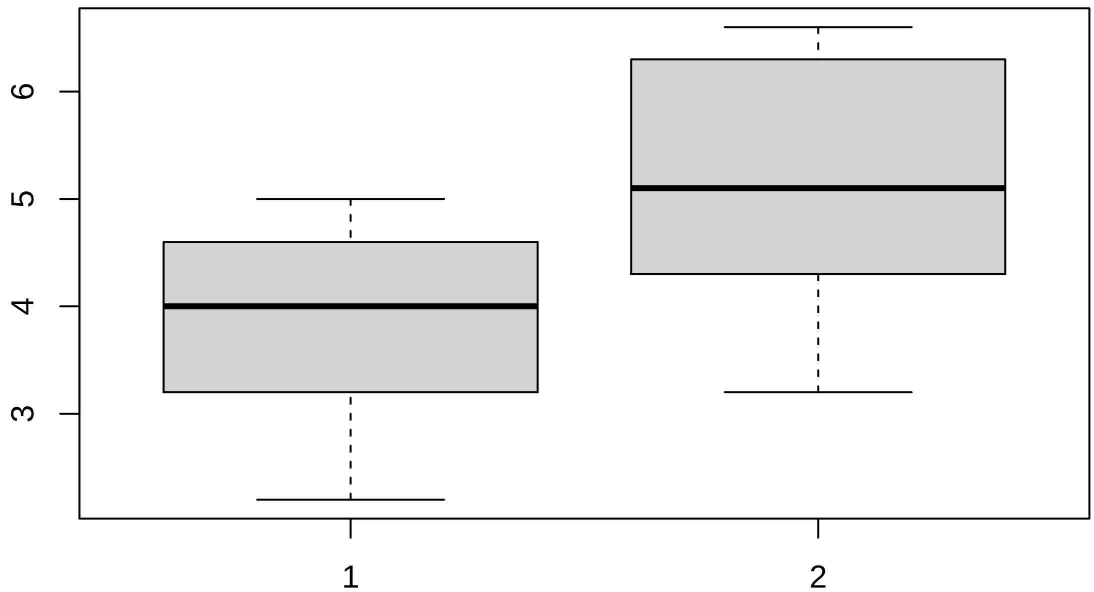
Reproduce: interpretation

The experiment finds that the fitness level affects ego strength. A high fitness level leads to an increased ego strength.

Part (c)

> Create a Boxplot as a graphical display of the data. Is it true that all high fitness faculty members have greater egos than low fitness faculty members?

```
boxplot(ego[fitness=='low'],ego[fitness=='high'])
```



Reproduce

The experimental finding is based on a comparison of means. It is not true that all ego strengths in the `high` fitness level group exceeds all ego strengths in the `low` fitness level group.

Part (d)

> Compute a 95% confidence interval for $\delta = \mu_1 - \mu_2$. Provide an interpretation, stated in the context of the problem.

We computed the results earlier to be:

```
t.star
```

```
##
##  Two Sample t-test
##
## data:  ego by fitness
## t = 3.7092, df = 28, p-value = 0.0009111
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.6268509 2.1731491
## sample estimates:
## mean in group high  mean in group low
##                5.2                3.8
```

Reproduce: CI (see output)

We see that $\text{CI}(\delta) = [.63, 2.17]$.

Reproduce: interpretation

Based on the observed data, we estimate that the difference in mean ego score, (`high fitness` − `low fitness`), is between .63 and 2.17 units.

Part (e)

Explain how a confidence interval provides a complementary result to a hypothesis test.

Reproduce

A hypothesis test looks to determine if an effect exists (dichotomous). A confidence interval looks to determine the *size* of the effect. (Also, a CI provides a measure of evidence strength.)

Part (f)

Explain how a confidence interval can be used in testing $H_0 : \mu_1 = \mu_2$.

Reproduce

If 0 is contained in CI(δ), decide in favor of $H_0 : \mu_1 = \mu_2$. Otherwise, decide in favor of $H_A : \mu_1 \neq \mu_2$.

Problem 2

A completely randomized design is used to investigate the effect of drug dosage on the activity level of lab rats. Each dose level is applied to $n = 4$ rats, and an activity score is observed for each rat in the sample. Higher values indicate greater activity. The data is provided as an attachment.

Preliminary analysis

The input factor is `drug dosage` with $a = 4$ levels (1 = `control`, 2 = `high`, 3 = `low`, 4 = `medium`).

The response is `activity level`.

We have $n = 4$ replicates for a total of $N = an = 16$ observations.

Part (a)

State the statistical hypotheses of interest. Briefly explain how the form of the alternative hypothesis requires a need for further investigation.

The hypothesis of interest is whether `drug dosage` level effects `activity level`. We may formulate this as a hypothesis test of the form

$$\begin{aligned} &H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 && \text{(‘drug dosage‘ has no effect on ‘activity level‘)}, \\ &H_A : \mu_i \neq \mu_j \text{ for some pair } (i, j) && \text{(‘drug dosage‘ does have an effect on ‘activity level‘)}. \end{aligned}$$

where μ_j is the expected activity level given a `drug dosage` level j.

If we decide H_A, i.e., there are differences in the dosage level means, then further investigation is required to determine where the differences occur.

Part (b)

> Compute the F_0 statistic and the p-value. Provide an interpretation, stated in the context of the problem. Create a Boxplot as a graphical display of the data.

This is a one-factor experiment with $a = 4$ levels of the factor and $n = 4$ replicates for a total of $N = na = 16$ observations. The appropriate test statistic is

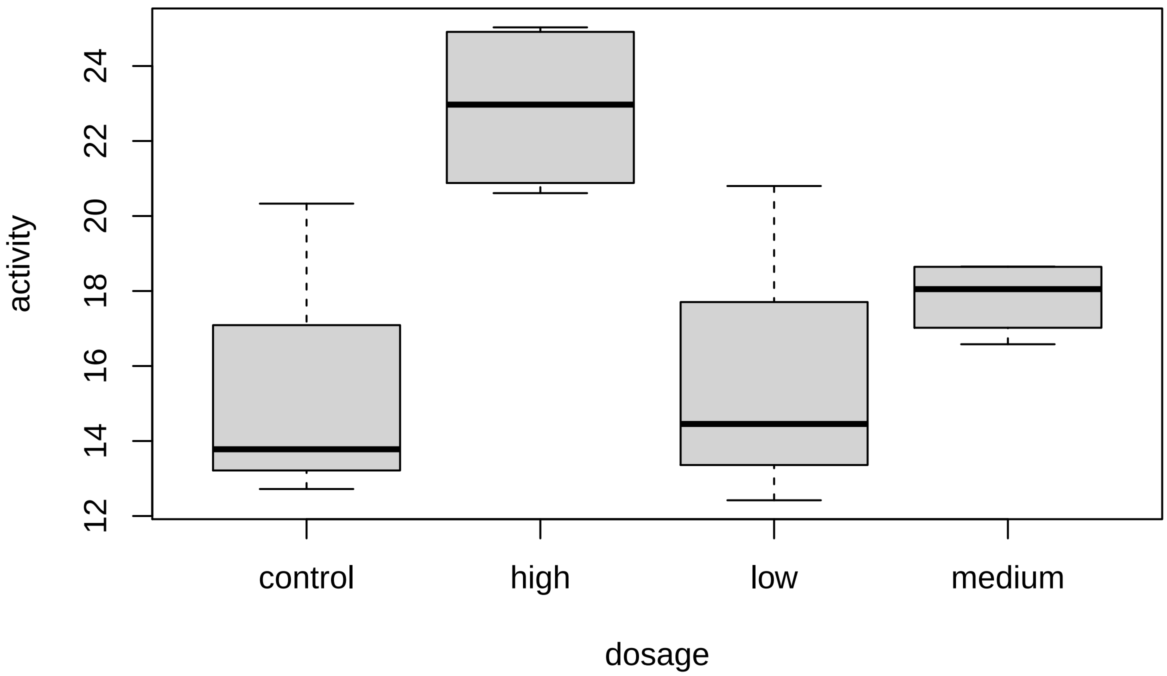
$$F_0 = \mathrm{MS}_\mathrm{A} \,/\, \mathrm{MS}_\mathrm{E}$$

which under $H_0 : \mu_1 = \cdots = \mu_4$ has the reference distribution

$$F_0 \sim F(a - 1 = 3, N - a = 12).$$

We compute the observed test statistic F_0 with:

```
dosage = as.factor(na.omit(data$dose.level))  # factor A
activity = na.omit(data$activity.score)       # response
boxplot(activity ~ dosage)
```




```
model = aov(activity ~ dosage)
summary(model)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## dosage       3 152.40   50.80   6.367 0.00791 **
## Residuals   12  95.75    7.98
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

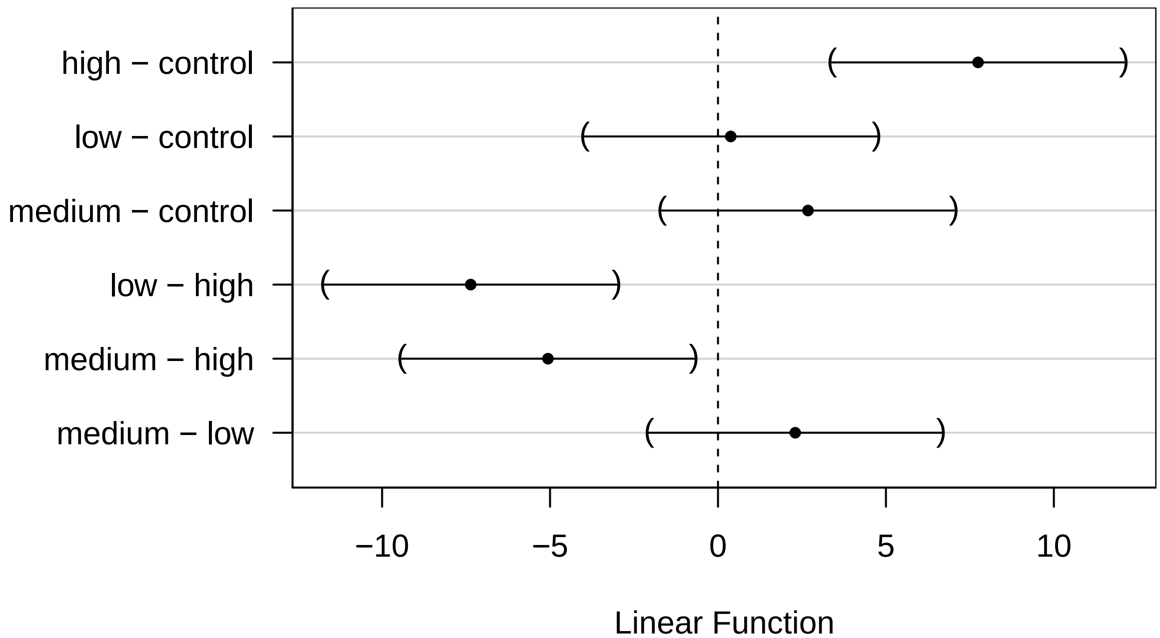
We see that $F_0 = 6.367$ with a p-value $= .008$ under H_0.

Interpretation (rep)

The experiment finds that the dosage level has an effect on the activity score.

Part (c)

Compute and display 95% confidence intervals for all pairwise comparisons. Explain how computing multiple intervals impacts the probability of committing an error.




```
##
##   Simultaneous Confidence Intervals
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: aov(formula = activity ~ dosage)
##
## Quantile = 2.1788
## 95% confidence level
##
##
## Linear Hypotheses:
##                       Estimate lwr      upr
## high - control == 0     7.7425   3.3906  12.0944
## low - control == 0      0.3800  -3.9719   4.7319
## medium - control == 0   2.6800  -1.6719   7.0319
## low - high == 0        -7.3625 -11.7144  -3.0106
## medium - high == 0     -5.0625  -9.4144  -0.7106
## medium - low == 0       2.3000  -2.0519   6.6519
```

Reproduce: Multiple intervals impact on committing an error

When multiple intervals are computed, then

$$\Pr\{\text{at least 1 type I error}\} > \alpha$$

where

$$\Pr\{\text{type I error } (i,j)\} = \alpha.$$

Part (d)

Perform pairwise comparisons using the Fisher LSD method, and the Tukey method. Provide grouping information for each method. Comment on the seemingly contradictory nature of a pairwise comparisons analysis.

Output: Fisher LSD method

```
summary(comps,test=univariate())
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: aov(formula = activity ~ dosage)
##
## Linear Hypotheses:
##                         Estimate Std. Error t value Pr(>|t|)
## high - control == 0        7.742      1.997   3.876  0.00220 **
## low - control == 0         0.380      1.997   0.190  0.85229
## medium - control == 0      2.680      1.997   1.342  0.20451
## low - high == 0           -7.362      1.997  -3.686  0.00311 **
## medium - high == 0        -5.062      1.997  -2.535  0.02620 *
## medium - low == 0          2.300      1.997   1.152  0.27195
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Univariate p values reported)
```

Output: Tukey HSD method

```
summary(comps)
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: aov(formula = activity ~ dosage)
##
## Linear Hypotheses:
##                         Estimate Std. Error t value Pr(>|t|)
## high - control == 0        7.742      1.997   3.876   0.0102 *
## low - control == 0         0.380      1.997   0.190   0.9974
## medium - control == 0      2.680      1.997   1.342   0.5559
## low - high == 0           -7.362      1.997  -3.686   0.0142 *
## medium - high == 0        -5.062      1.997  -2.535   0.1044
## medium - low == 0          2.300      1.997   1.152   0.6665
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Reproduce

Grouping information:

	Fisher	Tukey
high	b	b
medium	a	a b
low	a	a
control	a	a

See output of pairwise comparisons.

Contradictory nature: when multiple decisions are made in the presence of uncertainty, a measure of belief/evidence is necessary to avoid contradiction.

Part (e)

Describe the defining characteristics for each of the above pairwise comparison methods.

Reproduce

Fisher controls the probability of a type I error for each pairwise comparison. **Tukey** controls the overall probability of a type I error across all pairwise comparisons.

(f)

Compute the margin of error and comparison-wise error rate for the Tukey method in this problem.

Output

```
a = 4 # levels
n = 4 # replicates
N = a*n
df = N-a
mse = 7.98
alpha = .05

# margin of error for tukey method (hsd)
qtukey(alpha,a,df,lower.tail=F)*sqrt(mse/n)
```

```
## [1] 5.930375
```

```
# comparison-wise error rate for tukey
2*pt(qtukey(alpha,a,df,lower.tail=F)/sqrt(2),df,lower.tail=F)
```

```
## [1] 0.0117251
```

(g)

Compute the margin of error and family-wise error rate for the Fisher LSD method in this problem.

Output

```
# margin of error for fisher (lsd)
qt(alpha/2,df,lower.tail=F)*sqrt(2*mse/n)
```

```
## [1] 4.352175
```

```
# family-wise error rate
ptukey(qt(alpha/2,df,lower.tail=F)*sqrt(2),a,df,lower.tail=F)
```

```
## [1] 0.1843343
```

Problem 3

> A factorial experiment is used to investigate the effect of pressure, temperature, and time on the yield from a chemical reaction. Two levels (`low`, `high`) of each factor are set and $n = 2$ runs of a 2^3 design are completed. The data is provided as an attachment.

Preliminary analysis

This is a 2^3 factorial design ($n = 2$) in which we consider whether factors A (pressure), B (temp), and C (time) have an effect on response (yield).

Part (a)

> Perform tests for all main effects and for all interaction effects. State the F-statistic and p-value for each test of an effect deemed to be important. Fit a reduced model with the main effects and the statistically significant interaction effect.

Output

```
pressure = as.factor(na.omit(data$pressure))
temp = as.factor(na.omit(data$temperature))
time = as.factor(na.omit(data$time))
yield = na.omit(data$yield)

mod.F = aov(yield ~ pressure*temp*time)
summary(mod.F)
```

```
##                    Df Sum Sq Mean Sq F value  Pr(>F)
## pressure            1  51.25   51.25   8.862 0.01768 *
## temp                1 106.88  106.88  18.479 0.00262 **
## time                1  47.47   47.47   8.208 0.02099 *
## pressure:temp       1   1.85    1.85   0.320 0.58709
## pressure:time       1   2.95    2.95   0.510 0.49555
## temp:time           1  57.01   57.01   9.857 0.01381 *
## pressure:temp:time  1   0.92    0.92   0.159 0.70055
## Residuals           8  46.27    5.78
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Reproduce

F statistics and p-values of the important effects:

effect	F-statistic	p-value
pressure (A)	8.862	.018
temp (B)	18.479	.003
time (C)	8.208	.021
$B \times C$	9.857	.014

See output.

Part (b)

Provide a general definition of an interaction effect. Explain how an interaction plot is used in studying an interaction effect.

Reproduce

An interaction effect occurs when the effect of one factor depends on the level of the other factors.

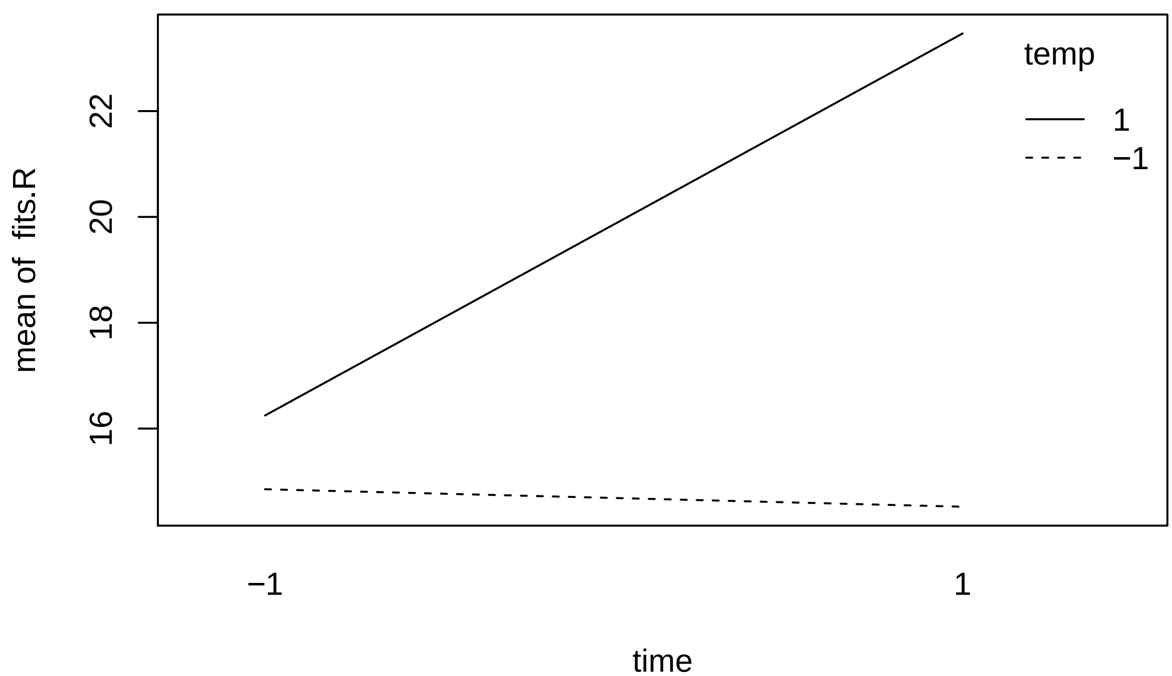
If an interaction plot is parallel, or nearly so, then there is no need to include the interaction terms.

Part (c)

Create a plot for the interaction effect deemed important. Provide an interpretation, stated in the context of the problem.

Output

```
mod.R = aov(yield ~ pressure+temp+time+temp:time)
fits.R = predict(mod.R)
interaction.plot(time,temp,fits.R)
```



Reproduce

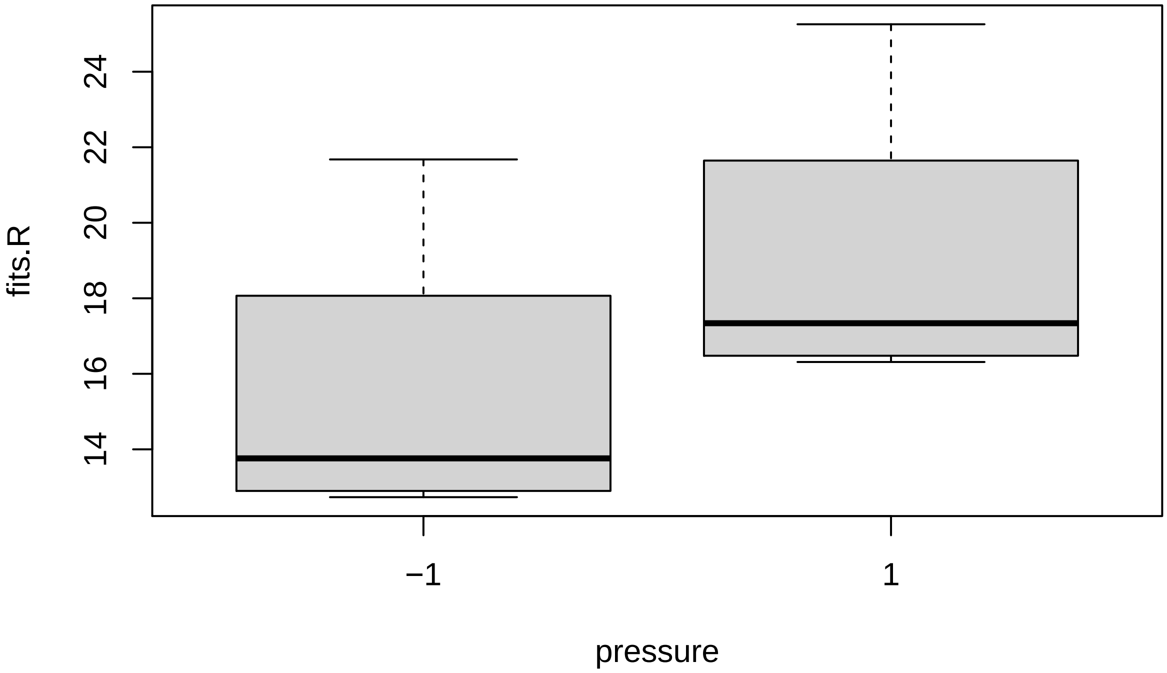
Time has a positive effect when temperature is at a `high` level. See output.

Part (d)

Create a Boxplot showing the main effect for the remaining factor. Provide an interpretation, stated in the context of the problem.

Output

```
plot(fits.R ~ pressure)
```



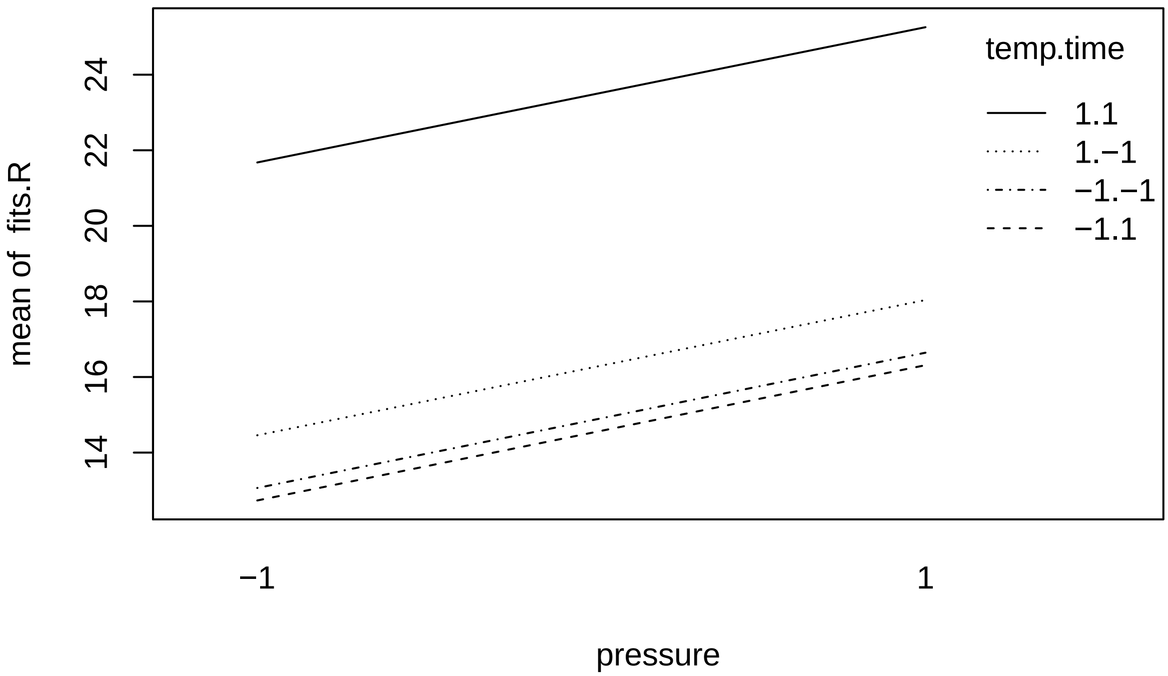
The experiment finds that `pressure` (factor A) has a positive effect on yield (response). See output.

Part (e)

Create a plot of the fitted values for the reduced model. Which setting of the factors should be used if the goal is to maximize yield?

Output

```
temp.time = interaction(temp,time)
interaction.plot(pressure,temp.time,fits.R)
```



Reproduce

The optimal setting is *high* `temp`, *high* `time`, and *high* `pressure`. See output.

Part (f)

Explain how the analysis is providing a simplification to the observed data.

Reproduce

The model smooths over the randomness in the data, simplifying the analysis.

Problem 4

An experiment to compare a new drug to a standard is in the planning stages. The response variable of interest is the clotting time (in minutes) of blood drawn from the subject. The experimenters want to perform a two sample t test at level $\alpha = .05$ with power $\pi = .90$ at $\delta_A = 0.5$, for standard deviation $\sigma = .7$.

(a)

Determine the sample size for each drug in order to achieve the stated test specifications.

$$n = 2(z_{\alpha/2} + z_\beta)^2 \sigma^2 / \delta_A^2.$$

where $\beta = 1 - \pi = .9$.

Output

```
sd = .7
alpha=.05
pwr=.9
```

```r
# beta = 1-pwr
delta.A=.5

# n = 2*(qnorm(1-alpha/2)+qnorm(1-beta))^2*sd^2/delta.A^2
power.t.test(n=NULL,delta=delta.A,sd=sd,sig.level=alpha,power=pwr,type="two.sample")
```

```
##
##      Two-sample t test power calculation
##
##               n = 42.17301
##           delta = 0.5
##              sd = 0.7
##       sig.level = 0.05
##           power = 0.9
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

Reproduce

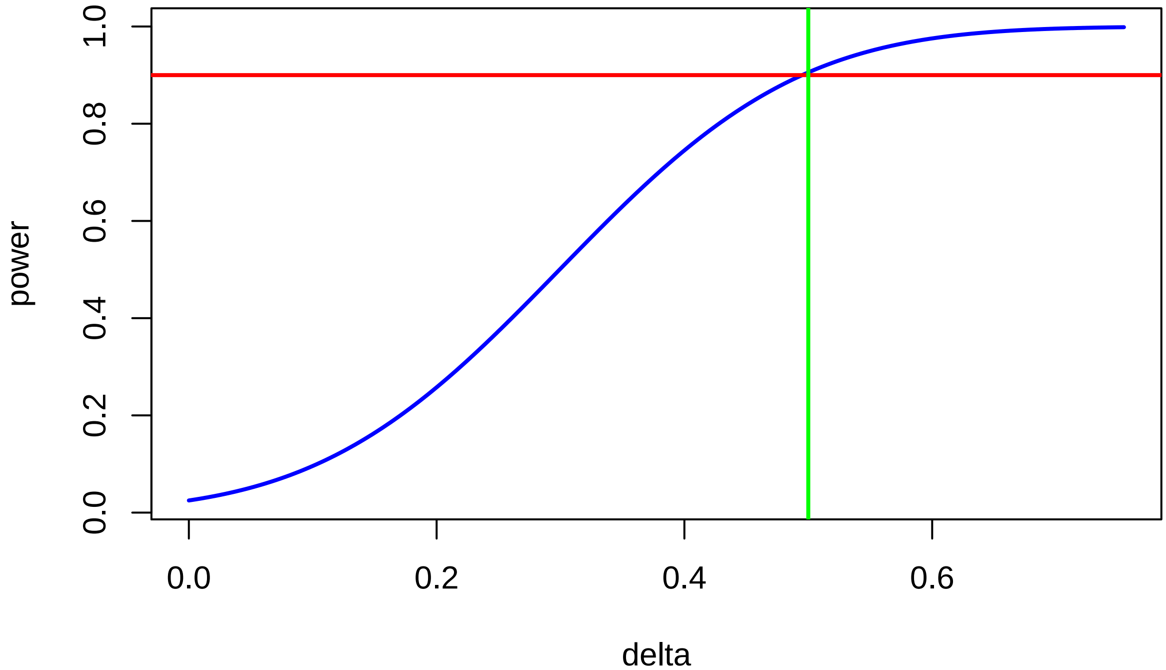
We see that $n = 43$ (when we round up to ensure the specifications are at least met).

(b)

> Graph the power curve for the chosen sample size. Explain how the power curve displays the desired properties of the test.

Output

```r
# power.curve
#
# create the power curve for the chosen sample size.
#
# arguments:
#   n: sample size,
#   sd: standard deviation
#   alpha: significance level
#   h: power (shows as a horizontal line)
#   v: specific alternative (shows as a vertical line)
# output:
#   graph of power curve
power.curve = function(n, sd, alpha, h, v)
{
  df = 2*(n-1)
  delta = seq(from=0,to=5*sd/sqrt(n/2),length.out = 1000)
  power = 1 - pt(qt(1-alpha/2,df),df,ncp = sqrt(n/2)*(delta/sd))
  plot(delta,power,type = "l",lwd=2,col="blue")
  abline(h=h,col="red",lwd=2)
  abline(v=v,col="green",lwd=2)
}
power.curve(n=43,sd=sd,alpha=alpha,h=pwr,v=delta.A)
```



Reproduce

We computed $n = 43$.

(c)

Provide a general explanation of how A can be determined.

(d)

Briefly discuss some other issues that may provide additional insight to an experimental result beyond a finding of statistical significance.

(e)

Briefly comment on the additional information provided by the p-value, beyond a determination of statistical significance alone.