Completely randomized designs

Course website

I'll put slides and labs here:

`http://vicpena.github.io/doe`

Lab groups

We have made two lab groups

- Group A: Meets from 3pm to 5pm **at PC2**
- Group B: Meets from 5pm to 7pm **at PC1** (different classroom)

You can find your group on my website

Last time

We reviewed the model

$$Y_j = \mu + \varepsilon_j, \qquad \varepsilon_j \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

- μ is a deterministic component, the expectation of Y
- ε_j is a random "noise" term, centered at zero
- We know how to estimate μ and σ^2, find confidence intervals for them, etc.

Today

We work with the model

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \qquad \varepsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2),$$

- i indexes group membership, j indexes observations within a group
- There are k groups, each with its own expected value $\mu_1, \mu_2, \ldots, \mu_k$; that is, $\mathbb{E}(Y_{ij}) = \mu_i$
- Within each group, there are n_i observations, if $n_1 = \cdots = n_k$, we say that the design is balanced
- The total sample size is $N = n_1 + n_2 + \ldots + n_k$
- The variance of the data within groups is the same; that is, $\mathrm{Var}(Y_{ij}) = \sigma^2$ for all i and j

Example

- A pharmaceutical wants to compare the effectiveness of 3 treatments (T1, T2, and T3).
- The treatments are **randomly** assigned to 24 patients. Large values of the outcome are associated with high effectiveness.

T1	T2	T3
4	7	9
2	6	12
6	5	6
6	7	11
5	6	10
6	4	11
2	7	9
6	5	10

Model assumptions

The model comes with assumptions:

- The data are **normally distributed**
- The data are **independent** (there is no time dependence, for example)
- The variance is the **same** for all groups: $\mathrm{Var}(y_{ij}) = \sigma^2$

Designed experiments vs observational studies

- The model

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \qquad \varepsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2),$$

can be used for both designed experiments and observational studies

- In this course, we will assume that the group assignment is **controlled by us and made at random.** In observational studies, we cannot assign observations to groups ourselves
- As we saw last time, randomizing can protect us from the effect of confounding variables
- In our context, this model is called a **completely randomized design** or **one-way ANOVA** (both names refer to the same thing)

Inferential goals

We want to

- estimate the parameters $\mu_1, \mu_2, \ldots, \mu_k$ and σ^2
- find confidence intervals for $\mu_1, \mu_2, \ldots, \mu_k$
- know whether the μ_i are all equal or not, and to see which ones are different, if any
- check that the model assumptions are satisfied

Point estimation

- The data come from groups with different means but equal variance σ^2
- It won't surprise you that we'll estimate μ_i with the sample mean of the observations coming from group i

$$\widehat{\mu}_i = \overline{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$$

- How to estimate σ^2? Let the sample variance for group i be

$$S_i^2 = \frac{\sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_i)^2}{n_i - 1}$$

Then, we estimate σ^2 with a weighted average of the S_i^2:

$$\widehat{\sigma}^2 = \frac{\sum_{i=1}^{k} (n_i - 1) S_i^2}{\sum_{i=1}^{k} (n_i - 1)}$$

Confidence intervals for μ_i

- A $100 \cdot (1-\alpha)\%$ confidence interval for μ_i is

$$\mathrm{CI}_{1-\alpha}(\mu_i) = \overline{Y}_i \pm \mathtt{qt}(N-k, 1-\alpha/2)\frac{\widehat{\sigma}}{\sqrt{n_i}}$$

- In practice, we can use the `confint` function in `R`

Back to example: Reading in data

Reading in and formatting the data

```
library(tidyverse)
wide = matrix(c(4, 7, 9,
                2, 6, 12,
                6, 5, 6,
                6, 7, 11,
                5, 6, 10,
                6, 4, 11,
                2, 7, 9,
                6, 5, 10), byrow = T, ncol = 3)
wide = as.data.frame(wide)
colnames(wide) = c("T1", "T2", "T3")
df = wide %>% pivot_longer(cols = c(T1, T2, T3),
                           names_to = "treat",
                           values_to = "outcome")
df$treat = factor(df$treat)
```

Example: coefficients and confidence intervals

By default, `confint` does 95% confidence intervals

```
mod = aov(outcome ~ treat - 1, data = df)
dummy.coef(mod)
```

```
## Full coefficients are
##
## treat:       T1    T2    T3
##           4.625 5.875 9.750
```

```
confint(mod)
```

```
##              2.5 %    97.5 %
## treatT1 3.443244  5.806756
## treatT2 4.693244  7.056756
## treatT3 8.568244 10.931756
```

Alternative parametrization: sum-to-zero

The model

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \qquad \varepsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2),$$

can be reparametrized as

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \qquad \varepsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2), \qquad \sum_{i=1}^{k} \tau_i = 0$$

In this parametrization,

- $\mathbb{E}(Y_{ij}) = \mu + \tau_i$
- μ is a "grand mean" common to all groups and τ_i is a group-specific effect
- The models are equivalent: given $\mu_1, \mu_2, \ldots \mu_k$, we can find μ and $\tau_1, \tau_2, \ldots, \tau_k$ (and viceversa)

$$\mu = \frac{\sum_{i=1}^{k} \mu_i}{k}, \qquad \tau_i = \mu_i - \frac{\sum_{i=1}^{k} \mu_i}{k}$$

Sum-to-zero: estimation

Point estimates

$$\widehat{\mu} = \frac{1}{k}\sum_{j=1}^{k} \overline{Y}_j, \qquad \tau_i = \overline{Y}_i - \widehat{\mu}$$

As before, we can find point estimates and intervals with `R`

Sum-to-zero parametrization in R

Only CIs for μ, τ_1, and τ_2 because $\tau_3 = -\tau_1 - \tau_2$

```
options(contrasts = c("contr.sum", "contr.poly"))
mod = aov(outcome ~ treat, data = df)
dummy.coef(mod)
```

```
## Full coefficients are
##
## (Intercept):      6.75
## treat:              T1      T2      T3
##                 -2.125  -0.875   3.000
```

```
confint(mod)
```

```
##                   2.5 %      97.5 %
## (Intercept)  6.067713  7.43228732
## treat1      -3.089900 -1.16010001
## treat2      -1.839900  0.08989999
```

Hypothesis tests

We'll see two types of hypothesis tests

- **Global test:**

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k \qquad H_1 : \mu_i \neq \mu_j \text{ for some } i, j,$$

which, in the sum-to-zero parametrization, is equivalent to

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_k = 0 \qquad H_1 : \tau_i \neq 0 \text{ for some } i$$

- **Pairwise tests:** for all pairs $\{i, j\} \subset \{1, \ldots, k\}$, test

$$H_{0,ij} : \mu_i = \mu_j \qquad H_{1,ij} : \mu_i \neq \mu_j$$

Global test

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k = \mu \qquad H_1 : \mu_i \neq \mu_j \text{ for some } i, j,$$

How we do the test:

1. Check that model assumptions are satisfied; if not, do not perform test
2. Set a significance level α
3. Given the data, compute a test statistic with a known distribution under H_0
4. Find p-value: probability of finding a t-statistic as extreme or more extreme than the one we observed assuming that H_0 is true; if the p-value is less than α, reject H_0; otherwise, do not reject H_0

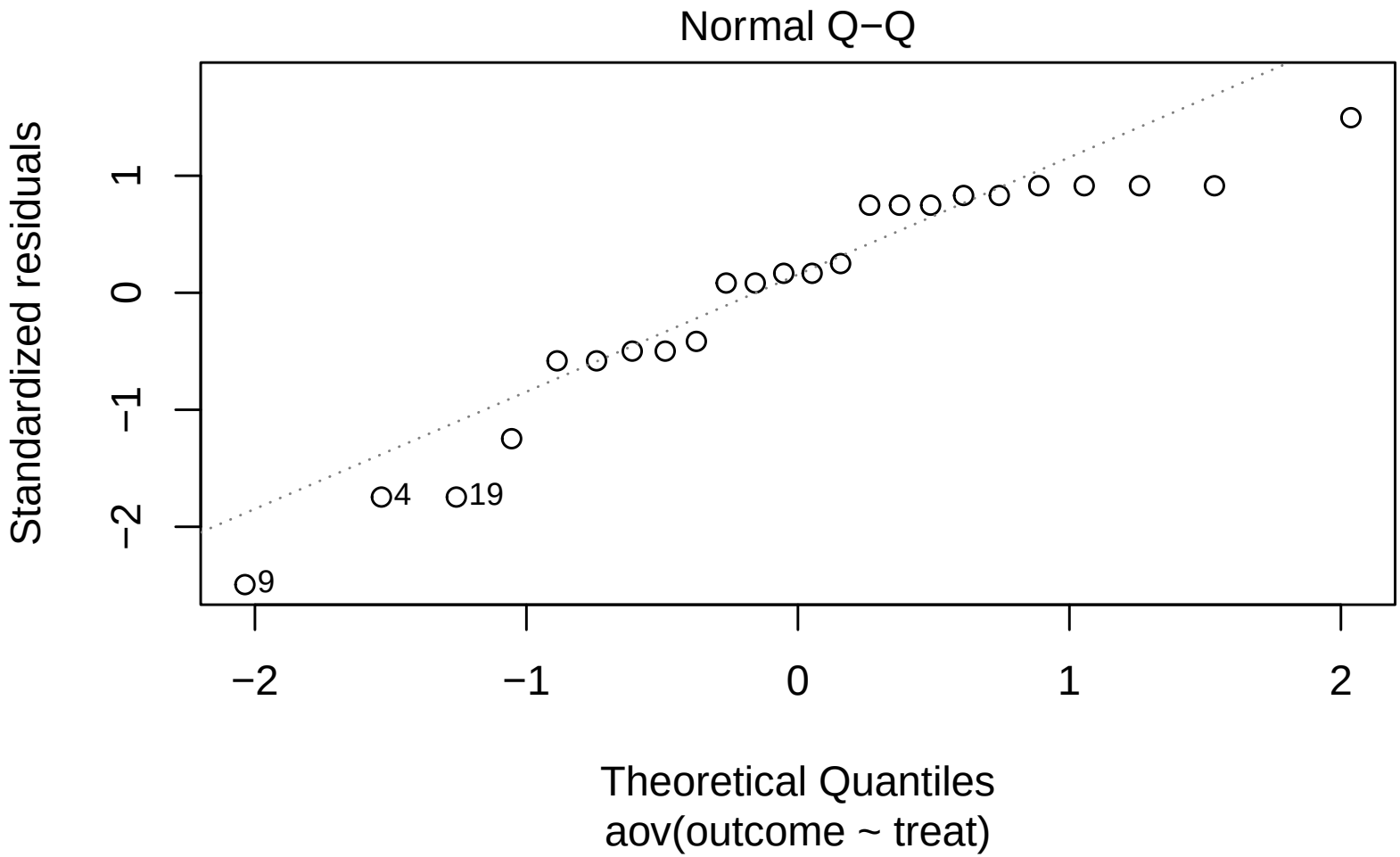
Checking model assumptions

Model assumptions and how to check them:

1. **Normality:** qq-plot of model residuals
2. **Equality of variance:** visually, inspecting that the data from the groups have similar variance; there are formal tests as well, like `bartlett.test` in `R`
3. **Independence:** Harder to check... If we have recorded the order in which the experiments have been run, we can check for time dependence

Example: Checking normality

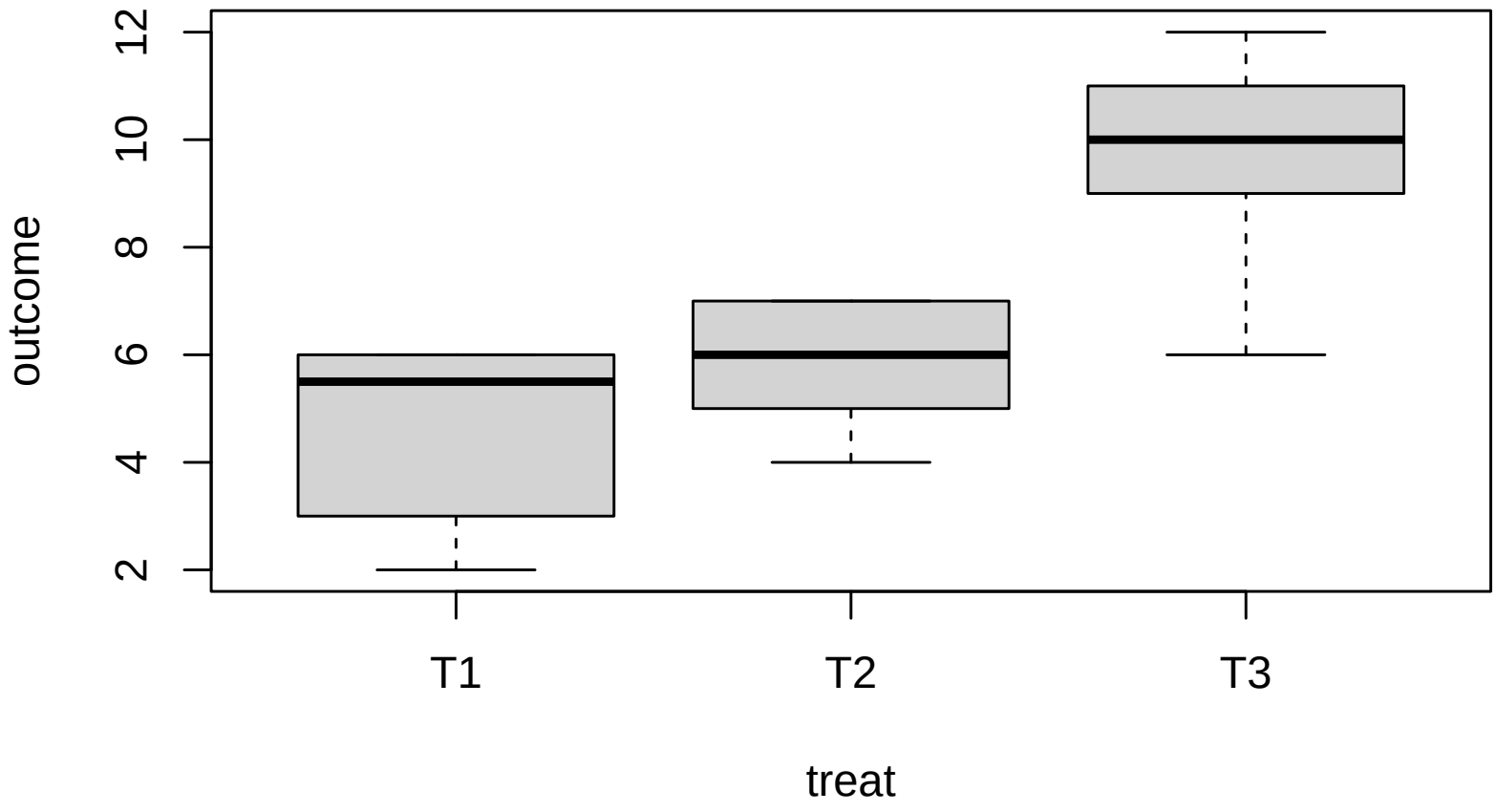
```
plot(mod, which = 2)
```



If the assumption is met, most points should be near the dashed line

Example: Checking equality of variance

```
boxplot(outcome ~ treat,  data = df)
```



If the assumption is met, the groups should have similar variability

Finding a test statistic

Let

$$\overline{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}, \qquad \overline{Y} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}.$$

Then, the total sum of squares (SS total) as

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y})^2 = \sum_{i=1}^{k} n_i(\overline{Y}_i - \overline{Y})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_i)^2$$

that is,

$$\text{SS total} = \text{SS between groups} + \text{SS within groups}$$

Finding a test statistic

We are testing:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k \qquad H_1 : \mu_i \neq \mu_j \text{ for some } i, j,$$

Intuitively...

- If H_0 is true $\overline{Y}_1 \approx \cdots \approx \overline{Y}_k \approx \overline{Y}$, so we expect to have more variability within groups than between groups
- If H_0 is not true, we expect to see more variability between groups

The test statistic will take that into account

The test statistic

Define the between groups (SSB), and within groups (SSW) sum of squares as below:

$$\mathrm{SSB} = \sum_{i=1}^{k} n_i(\overline{Y}_i - \overline{Y})^2, \qquad \mathrm{SSW} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_i)^2$$

And then define the "mean squares"

$$\mathrm{MSB} = \frac{\mathrm{SSB}}{k-1}, \qquad \mathrm{MSW} = \frac{\mathrm{SSW}}{N-k}$$

If H_0 is true,

$$F = \frac{\mathrm{MSB}}{\mathrm{MSW}} \overset{H_0}{\sim} F_{k-1,N-k},$$

where $F_{k-1,N-k}$ is the F-distribution with $k-1$ and $N-k$ degrees of freedom.

Finding the p-value

The p-value is one-sided: given data y_{ij}, we can compute the observed F-statistic f_{obs} and compute the p-value as

$$P(F_{k-1,N-k} > f_{\text{obs}})$$

Why is it one-sided? Recall that the F-statistic is

$$F = \frac{\text{MSB}}{\text{MSW}},$$

- If $F \approx 0$, we shouldn't be rejecting H_0 because those are cases where the variability between groups is small, which is consistent H_0.
- We should only reject H_0 for large values of F

ANOVA (Analysis of Variance) table

	df	SS	MS	F	p-value
Between	$k-1$	SSB	$\mathrm{MSB} = \frac{\mathrm{SSB}}{k-1}$	$f_{obs} = \frac{\mathrm{MSB}}{\mathrm{MSW}}$	p
Within	$N-k$	SSW	$\mathrm{MSW} = \frac{\mathrm{SSW}}{N-k}$		

where the p-value is

$$p = P(F_{k-1,N-k} > f_{\mathrm{obs}})$$

A note on terminology

- Some authors refer to SSB as SST (sum of squares treatment) and SSW as SSE (sum of squares error)
- `R` will use the term `Residuals` to refer to SSW

Example: ANOVA table

Suppose our significance level is $\alpha = 0.05$

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## treat         2 114.25   57.12   22.11 6.79e-06 ***
## Residuals    21  54.25    2.58
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

p-value is $6.79 \cdot 10^{-6} < \alpha$, so we reject H_0

Pairwise comparisons

- The global test can detect if there are at least two group means that are different, but it doesn't tell us which ones are different
- If we're interested in seeing where the differences between groups lie, we can perform all $\binom{k}{2}$ pairwise comparisons

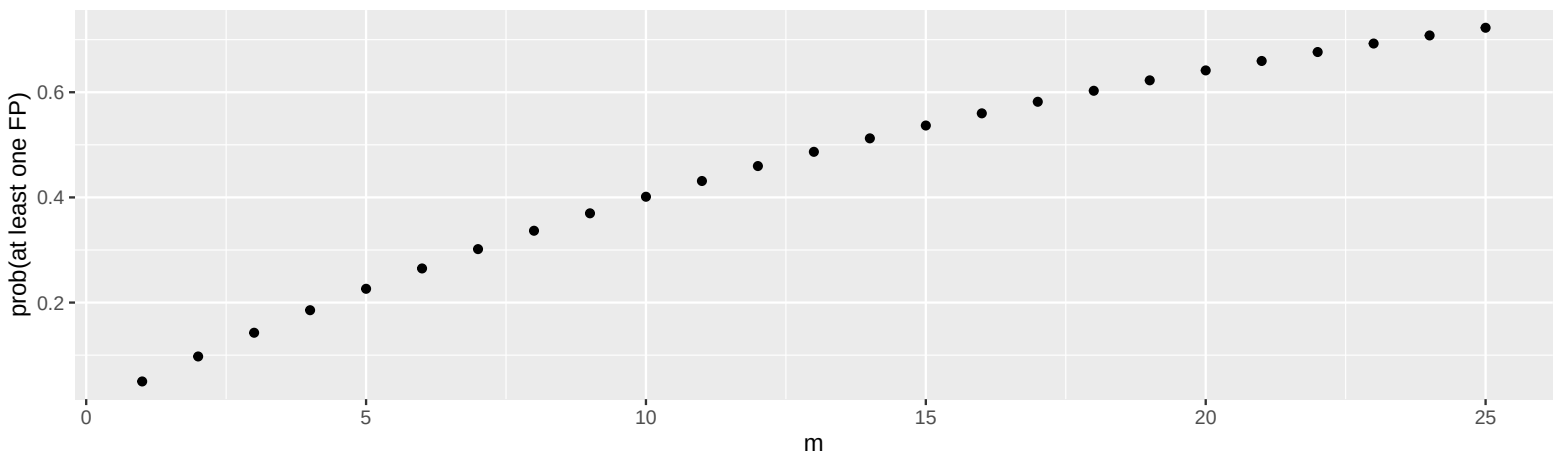
$$H_{0,ij} : \mu_i = \mu_j, \qquad H_{1,ij} : \mu_i \neq \mu_j$$

Probability of a false positive

- We have to be careful when we perform such tests
- If we perform m independent tests at significance level γ and the null hypothesis is true for all of them, the probability of getting at least one false positive is

$$P(\text{at least one FP}) = 1 - P(\text{no FP}) = 1 - (1-\gamma)^m$$

If $\gamma = 0.05$, here's a graph of the probability



Controlling false positive rate

- There are many methods to perform the pairwise tests we're interested in while controlling that

$$P(\text{at least one FP}) = \alpha$$

- One such method is the Tukey Honestly Significant Difference test: `TukeyHSD` in `R`
- `TukeyHSD` reports point estimates for all pairwise differences, along with adjusted confidence intervals and p-values. If we reject an adjusted p-value if it is less than α, the probability of at least one false positive is α.
- `TukeyHSD` is based on the distribution of the Studentized range $q = \sqrt{N}\,(\overline{Y}_{\text{max}} - \overline{Y}_{\text{min}})/\widehat{\sigma}$ under the same assumptions we've been making

Example: TukeyHSD

The default confidence level $(1 - \alpha)$ is 0.95

```
TukeyHSD(mod, conf.level = 0.95)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = outcome ~ treat, data = df)
##
## $treat
##         diff         lwr      upr     p adj
## T2-T1 1.250 -0.7756249 3.275625 0.2865063
## T3-T1 5.125  3.0993751 7.150625 0.0000073
## T3-T2 3.875  1.8493751 5.900625 0.0002592
```

Significant differences between T3 and T1 and T3 and T2 at $\alpha = 0.05$