

Set R08 - Experimental design

STAT 401 (Engineering) - Iowa State University

April 7, 2017

Strength of wood glue

You are interested in testing two different wood glues:

- Gorilla Wood Glue
- Titebond 1413 Wood glue

On a scarf joint:



So you collect up some wood, glue the pieces together, and determine the force required to break the joint. (There are lots of details missing here.)

Inspiration: https://woodgears.ca/joint_strength/glue.html

Completely Randomized Design (CRD)

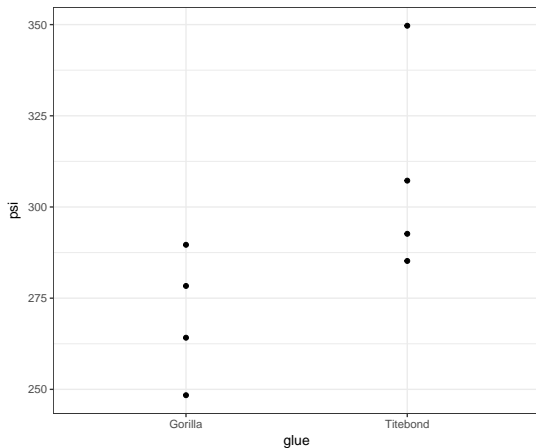
Suppose I have 8 pieces of wood laying around. I cut each piece and **randomly** use either Gorilla or Titebond glue to recombine the pieces. I do the randomization in such a way that I have exactly 4 Gorilla and 4 Titebond results, e.g.

```
# A tibble: 8  2
  woodID      glue
<fctr>    <chr>
1 wood1 Gorilla
2 wood2 Titebond
3 wood3 Gorilla
4 wood4 Titebond
5 wood5 Titebond
6 wood6 Titebond
7 wood7 Gorilla
8 wood8 Gorilla
```

This is called a **completely randomized design (CRD)**.

Visualize the data

```
ggplot(d, aes(glue, psi)) + geom_point() + theme_bw()
```



Model

Let

- F_w be the force needed to break wood w ,
- T_w be an indicator that the Titebond glue was used on wood w , i.e.

$$T_w = I(\text{glue}_w = \text{Titebond}).$$

Then a regression model for these data is

$$F_w \stackrel{\text{ind}}{\sim} N(\beta_0 + \beta_1 T_w, \sigma)$$

where

- β_1 is the expected difference in force when using Titebond glue compared to using Gorilla glue.

Perform analysis

```
m <- lm(psi ~ glue, data = d)
summary(m)
```

```
Call:
lm(formula = psi ~ glue, data = d)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-23.486 -17.468  -3.722  11.038  40.997
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    270.14      11.99   22.537 4.99e-07 ***
glueTitebond     38.56       16.95    2.275  0.0633 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 23.97 on 6 degrees of freedom
Multiple R-squared:  0.463, Adjusted R-squared:  0.3735
F-statistic: 5.174 on 1 and 6 DF,  p-value: 0.06327
```

```
confint(m)
```

```
            2.5 %    97.5 %
(Intercept) 240.806326 299.46474
glueTitebond -2.921249  80.03428
```

Randomized complete block design (RCBD)

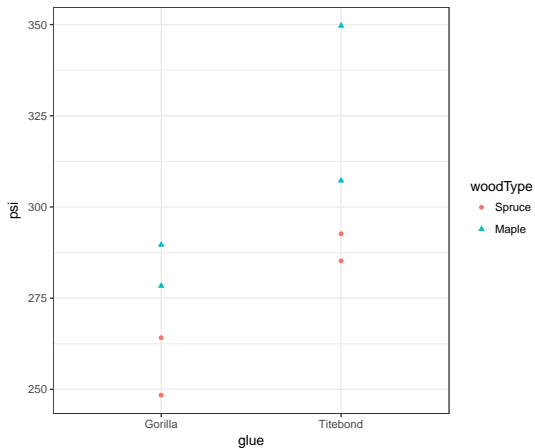
Suppose the wood actually came from two different types: Maple and Spruce. And perhaps you have reason to believe the glue will work differently depending on the type of wood. In this case, you would want to **block** by wood type and perform the randomization within each block, i.e.

```
# A tibble: 8  3
  woodID woodType   glue
<fctr>  <fctr>    <chr>
1  wood1   Spruce Gorilla
2  wood2   Spruce Titebond
3  wood3   Spruce Gorilla
4  wood4   Spruce Titebond
5  wood5    Maple Titebond
6  wood6    Maple Titebond
7  wood7    Maple Gorilla
8  wood8    Maple Gorilla
```

This is called a **randomized complete block design (RCBD)**.

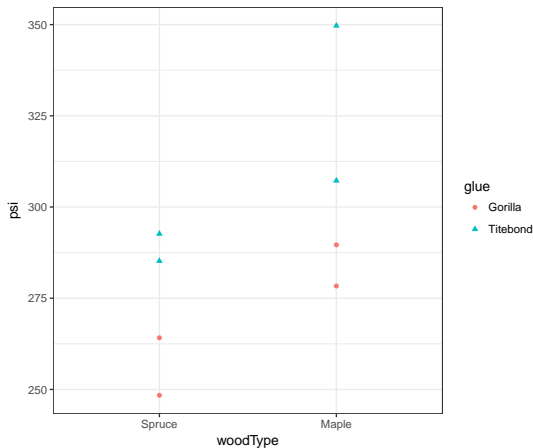
Visualize the data

```
ggplot(d, aes(glue, psi, color=woodType, shape=woodType)) + geom_point() + theme_bw()
```



Visualize the data - a more direct comparison

```
ggplot(d, aes(woodType, psi, color=glue, shape=glue)) + geom_point() + theme_bw()
```



Model

Let

- F_w be the force needed to break wood w
- T_w be an indicator that Titebond glue was used on wood w , and
- M_w be an indicator that wood w was Maple.

Then a regression model for these data is

$$F_w \stackrel{ind}{\sim} N(\beta_0 + \beta_1 T_w + \beta_2 M_w, \sigma)$$

where

- β_1 is the expected difference in force when using Titebond glue compared to using Gorilla glue when adjusted for type of wood, i.e. the type of wood is held constant, and
- β_2 is the expected difference in force when using Spruce compared to Maple when adjusted for type of glue, i.e. the glue is held constant.

Perform analysis

```
m <- lm(psi ~ glue + woodType, data = d)
summary(m)
```

```
Call:
lm(formula = psi ~ glue + woodType, data = d)
```

Residuals:

1	2	3	4	5	6	7	8
-4.929	0.768	10.835	-6.674	24.186	-18.279	-8.594	2.688

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	253.324	9.435	26.848	1.34e-06 ***
glueTitebond	38.557	10.895	3.539	0.0166 *
woodTypeMaple	33.623	10.895	3.086	0.0273 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.41 on 5 degrees of freedom

Multiple R-squared: 0.8151, Adjusted R-squared: 0.7412

F-statistic: 11.02 on 2 and 5 DF, p-value: 0.01469

```
confint(m)
```

	2.5 %	97.5 %
(Intercept)	229.069570	277.57817
glueTitebond	10.550061	66.56297
woodTypeMaple	5.616873	61.62978

Replication

Since there are more than one observation for each woodType-glue combination, the design is **replicated**:

```
d %>% group_by(woodType, glue) %>% summarize(n = n())
```

Source: local data frame [4 x 3]

Groups: woodType [?]

	woodType	glue	n
	<fctr>	<chr>	<int>
1	Spruce	Gorilla	2
2	Spruce	Titebond	2
3	Maple	Gorilla	2
4	Maple	Titebond	2

When the design is replicated, we can consider assessing an interaction. In this example, an interaction between glue and woodType would indicate that the effect of glue depends on the woodType, i.e. the difference in expected force between the two glues depends on woodType. At an extreme, it could be that Gorilla works better on Spruce and Titebond works better on Maple.

Assessing an interaction using a t-test

```
m <- lm(psi ~ glue*woodType, data = d)
summary(m)
```

```
Call:
lm(formula = psi ~ glue * woodType, data = d)
```

Residuals:

1	2	3	4	5	6	7	8
-7.882	3.721	7.882	-3.721	21.233	-21.233	-5.641	5.641

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	256.28	11.82	21.686	2.67e-05 ***
glueTitebond	32.65	16.71	1.954	0.122
woodTypeMaple	27.72	16.71	1.658	0.173
glueTitebond:woodTypeMaple	11.81	23.64	0.500	0.643

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.71 on 4 degrees of freedom

Multiple R-squared: 0.826, Adjusted R-squared: 0.6955

F-statistic: 6.33 on 3 and 4 DF, p-value: 0.05335

Assessing an interaction using an F-test

```
anova(m)
```

Analysis of Variance Table

Response: psi

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
glue	1	2973.21	2973.21	10.6449	0.03100 *
woodType	1	2261.06	2261.06	8.0952	0.04662 *
glue:woodType	1	69.77	69.77	0.2498	0.64346
Residuals	4	1117.24	279.31		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
drop1(m, test='F')
```

Single term deletions

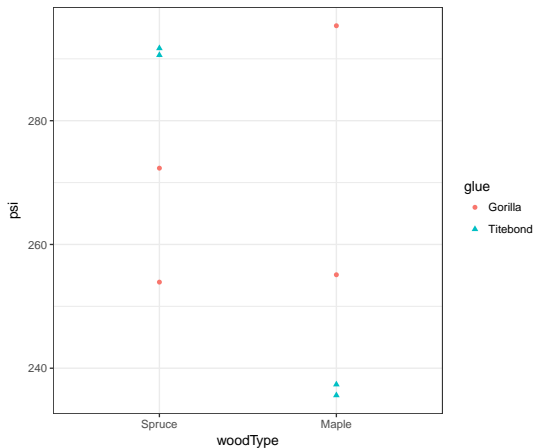
Model:

```
psi ~ glue * woodType
```

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
<none>			1117.2	47.513		
glue:woodType	1	69.769	1187.0	45.998	0.2498	0.6435

What if this had been your data?

```
ggplot(d, aes(woodType, psi, color=glue, shape=glue)) + geom_point() + theme_bw()
```



Assessing an interaction using a t-test

```
m <- lm(psi ~ glue*woodType, data = d)
summary(m)
```

```
Call:
lm(formula = psi ~ glue * woodType, data = d)
```

Residuals:

1	2	3	4	5	6	7	8
-9.2083	-0.5529	0.5529	9.2083	-0.8764	20.1215	-20.1215	0.8764

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	263.12	11.08	23.755	1.86e-05 ***
glueTitebond	28.03	15.66	1.790	0.1480
woodTypeMaple	12.10	15.66	0.773	0.4829
glueTitebond:woodTypeMaple	-66.76	22.15	-3.014	0.0394 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.66 on 4 degrees of freedom

Multiple R-squared: 0.7648, Adjusted R-squared: 0.5883

F-statistic: 4.335 on 3 and 4 DF, p-value: 0.09522

Unreplicated study

Suppose you now have

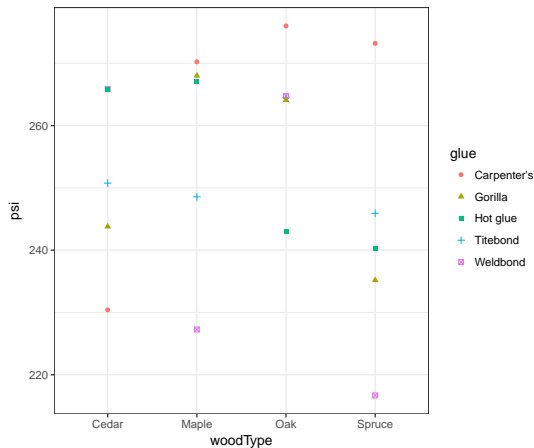
- 5 glue choices
- 4 different types of wood with
- 5 samples of each type of wood.

Thus you can only run each glue choice once on each type of wood.

Then you can run an unreplicated RCBD.

Visualize

```
ggplot(d, aes(woodType, psi, color=glue, shape=glue)) +  
  geom_point() + theme_bw()
```



Fit the main effects (or additive) model

```
m <- lm(psi ~ glue + woodType, data = d)
anova(m)
```

Analysis of Variance Table

Response: psi

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
glue	4	714.8	178.71	0.5636	0.6937
woodType	3	1091.4	363.80	1.1474	0.3697
Residuals	12	3804.9	317.07		

Fit the main effects (or additive) model

```
summary(m)
```

Call:

```
lm(formula = psi ~ glue + woodType, data = d)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-30.302	-7.093	2.316	10.326	23.992

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	260.717	11.262	23.150	2.51e-11 ***
glueGorilla	-9.696	12.591	-0.770	0.456
glueHot glue	-8.460	12.591	-0.672	0.514
glueTitebond	-10.018	12.591	-0.796	0.442
glueWeldbond	-18.834	12.591	-1.496	0.161
woodTypeMaple	4.907	11.262	0.436	0.671
woodTypeOak	11.157	11.262	0.991	0.341
woodTypeSpruce	-9.056	11.262	-0.804	0.437

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17.81 on 12 degrees of freedom

Multiple R-squared: 0.3219, Adjusted R-squared: -0.07366

F-statistic: 0.8138 on 7 and 12 DF, p-value: 0.5931

Fit the full (with interaction) model

```
m <- lm(psi ~ glue * woodType, data = d)
anova(m)
```

Warning in anova.lm(m): ANOVA F-tests on an essentially perfect fit are unreliable

Analysis of Variance Table

Response: psi

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
glue	4	714.8	178.71		
woodType	3	1091.4	363.80		
glue:woodType	12	3804.9	317.07		
Residuals	0	0.0			

Fit the full (with interaction) model

```
summary(m)
```

Call:

```
lm(formula = psi ~ glue * woodType, data = d)
```

Residuals:

ALL 20 residuals are 0: no residual degrees of freedom!

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	230.41	NA	NA	NA
glueGorilla	13.38	NA	NA	NA
glueHot glue	35.32	NA	NA	NA
glueTitebond	20.35	NA	NA	NA
glueWeldbond	35.46	NA	NA	NA
woodTypeMaple	39.83	NA	NA	NA
woodTypeOak	45.59	NA	NA	NA
woodTypeSpruce	42.80	NA	NA	NA
glueGorilla:woodTypeMaple	-15.61	NA	NA	NA
glueHot glue:woodTypeMaple	-38.52	NA	NA	NA
glueTitebond:woodTypeMaple	-42.03	NA	NA	NA
glueWeldbond:woodTypeMaple	-78.44	NA	NA	NA
glueGorilla:woodTypeOak	-25.27	NA	NA	NA
glueHot glue:woodTypeOak	-68.37	NA	NA	NA
glueTitebond:woodTypeOak	-31.80	NA	NA	NA
glueWeldbond:woodTypeOak	-46.74	NA	NA	NA
glueGorilla:woodTypeSpruce	-51.41	NA	NA	NA
glueHot glue:woodTypeSpruce	-68.22	NA	NA	NA
glueTitebond:woodTypeSpruce	-47.64	NA	NA	NA
glueWeldbond:woodTypeSpruce	-92.00	NA	NA	NA