

# i0002 Rlang

## 1 prod: Data

### 1.1 Spec

```
readr::spec(df0)
```

```
cols(  
  Idx0 = col_factor(levels = NULL, ordered = FALSE, include_na = FALSE),  
  ActiveVoice = col_character(),  
  PassiveVoice = col_character(),  
  ba0Act = col_character(),  
  ba0Pas = col_character(),  
  ba0ActTot = col_double(),  
  ba0PasTot = col_double()  
)
```

### 1.2 Sample

```
cogsys::set.w(8888)  
df0 %>% dplyr::slice_head(n=4)
```

```
# A tibble: 4 × 7
```

	Idx0	ActiveVoice	PassiveVoice	b
	<fct>	<chr>	<chr>	<
1	i1	['The chef cooked the meal.']	['The meal was cooked by the chef.']	[
2	i2	['The teacher praised the student.']	['The student was praised by the teacher.']	[
3	i3	['The storm destroyed the village.']	['The village was destroyed by the storm.']	[
4	i4	['She wrote a letter.']	['A letter was written by her.']	[

### 1.3 Spec

```
cogsys::set.w(111)  
skimr::skim(df0)
```

```

Data Summary
Name      Values
Number of rows      df0
Number of columns    75
Number of columns    7

-----
Column type frequency:
  character      4
  factor         1
  numeric        2
-----
Group variables      None

Variable type: character
skim_variable n_missing complete_rate min max empty n_unique whitespace
1 ActiveVoice      0             1 20 42      0       75           0
2 PassiveVoice     0             1 27 49      0       75           0
3 ba0Act           0             1 19 25      0       75           0
4 ba0Pas           0             1 20 23      0       75           0

Variable type: factor
skim_variable n_missing complete_rate ordered n_unique top_counts
1 Idx0         0             1 FALSE           75 i1: 1, i2: 1, i3: 1, i4: 1

Variable type: numeric
skim_variable n_missing complete_rate mean sd p0 p25 p50 p75 p100 hist
1 ba0ActTot     0             1 0.228 0.139 -0.144 0.124 0.233 0.319 0.581
2 ba0PasTot     0             1 0.147 0.121 -0.160 0.0588 0.143 0.196 0.483

```

## 1.4 DF2

```
df2
```

```

# A tibble: 150 × 7
  Idx0 ActiveVoice PassiveVoice ba0Act b
  <fct> <chr> <chr> <chr> <
1 i1 ['The chef cooked the meal.'] ['The meal was cooked by the chef... [0.24... [
2 i1 ['The chef cooked the meal.'] ['The meal was cooked by the chef... [0.24... [
3 i2 ['The teacher praised the student.'] ['The student was praised by the t... [0.28... [
4 i2 ['The teacher praised the student.'] ['The student was praised by the t... [0.28... [
5 i3 ['The storm destroyed the village.'] ['The village was destroyed by the... [-0.1... [
6 i3 ['The storm destroyed the village.'] ['The village was destroyed by the... [-0.1... [
7 i4 ['She wrote a letter.'] ['A letter was written by her.'] [0.06... [
8 i4 ['She wrote a letter.'] ['A letter was written by her.'] [0.06... [
9 i5 ['They built the bridge.'] ['The bridge was built by them.'] [0.18... [
10 i5 ['They built the bridge.'] ['The bridge was built by them.'] [0.18... [
# i 140 more rows
# i Use `print(n = ...)` to see more rows

```

## 1.5 Test

```
t_test_result <- t.test(value ~ feature, data = df2, paired = TRUE)
t_test_result
```

Paired t-test

```
data: value by feature
t = 9.6011, df = 74, p-value = 1.229e-14
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 0.06414016 0.09773468
sample estimates:
mean difference
 0.08093742
```

## 1.6 Test 2

```
suppressWarnings(rm(list=ls(pattern="^model")))
model_lm <- lm(value ~ feature, data = df2)
summary(model_lm)
```

Call:

```
lm(formula = value ~ feature, data = df2)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.37139	-0.09036	0.00009	0.07834	0.35321

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.22758	0.01506	15.11	< 2e-16 ***
featureba0PasTot	-0.08094	0.02130	-3.80	0.000211 ***

---

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

Residual standard error: 0.1304 on 148 degrees of freedom

Multiple R-squared: 0.08888, Adjusted R-squared: 0.08272

F-statistic: 14.44 on 1 and 148 DF, p-value: 0.0002111

## 1.7 Test 2

```
library(lmerTest)
suppressWarnings(rm(list=ls(pattern="^model_lmer")))
model_lmer <- lmer(value ~ feature + (1 | Idx0), data = df2)
summary(model_lmer)
```

Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']

Formula: value ~ feature + (1 | Idx0)

Data: df2

REML criterion at convergence: -266.2

Scaled residuals:

Min	1Q	Median	3Q	Max
-1.77396	-0.43996	-0.06451	0.47196	1.91774

Random effects:

Groups	Name	Variance	Std.Dev.
Idx0	(Intercept)	0.014350	0.11979
	Residual	0.002665	0.05162

Number of obs: 150, groups: Idx0, 75

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t )
(Intercept)	0.22758	0.01506	86.48461	15.110	< 2e-16 ***
featureba0PasTot	-0.08094	0.00843	74.00000	-9.601	1.23e-14 ***

---

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

Correlation of Fixed Effects:

	(Intr)
featrb0PsTt	-0.280