University of Colorado
Boulder

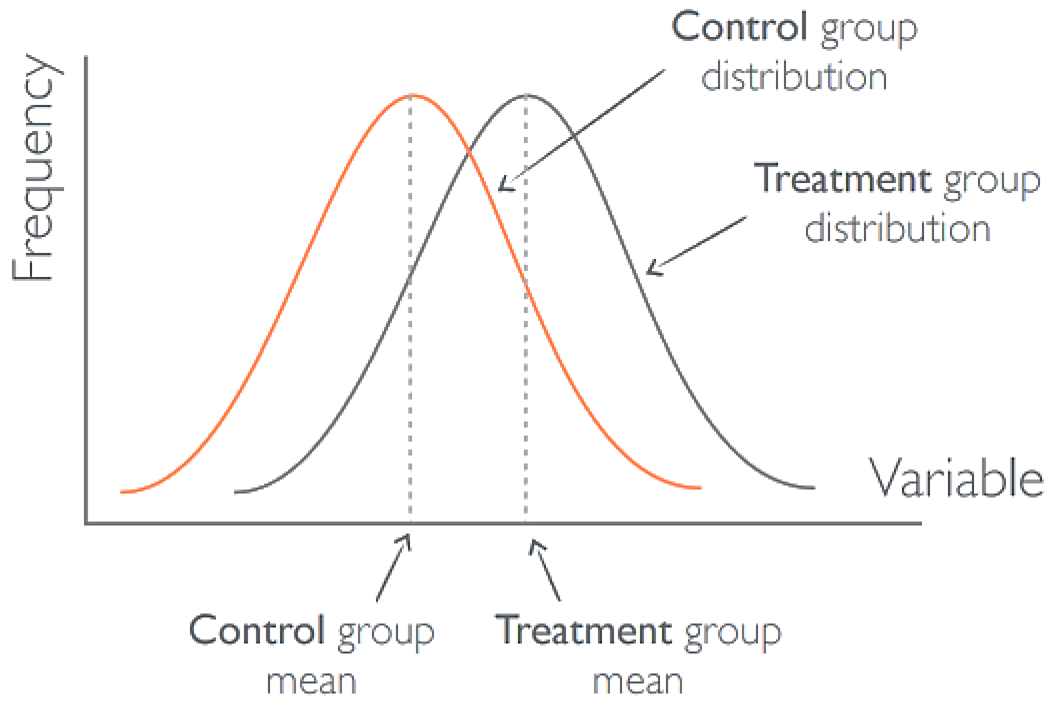
Human-Robot Interaction

T-Tests

Professor **Dan Szafir**

Computer Science & ATLAS Institute
University of Colorado Boulder

Comparing Distributions



IV: Categorical | DV: **Categorical or Ordinal**

- Example: did people rate a robot with mutual gaze more highly than one without?
- Example 2: of three teleoperation interfaces used, which one did participants rank most highly in terms of preference?
- Use nonparametric tests
- Compare distributions
  - 1 Factor with 2 Levels (i.e., 2 conditions)
    - Independent measures (between-subjects design): Mann-Whitney U
    - Repeated Measures (within-subjects design): Wilcoxon Signed-Rank
  - 1 Factor with >2 Levels (i.e., >2 conditions)
    - Independent measures: Kruskal-Wallace
    - Repeated measures: Friedman Test

IV: Categorical | DV: **Continuous**

Example: Did a human-robot team where the robot's behavior was governed by algorithm 1 have faster task completion times than a human-robot team where the robot's behavior was governed by algorithm 2?

Use parametric tests

Compare distributions

- 1 Factor with 2 Levels (i.e., 2 conditions)
  - Independent measures (between-subjects design): Independent t-test
  - Repeated Measures (within-subjects design): Paired t-test
- 1 Factor with >2 Levels (i.e., >2 conditions)
  - Independent measures: One factor independent measures ANOVA
  - Repeated measures: One factor repeated measures ANOVA
- 2 Factors
  - Independent measures: Two-way independent measures ANOVA
  - Repeated measures: Two-way repeated measures ANOVA
  - Mixed design: Two-way mixed model ANOVA
- > 2 Factors: multi-way ANOVA
- 1 Factor, multiple DVs: MANOVA

Student’s t-test

Output Variable

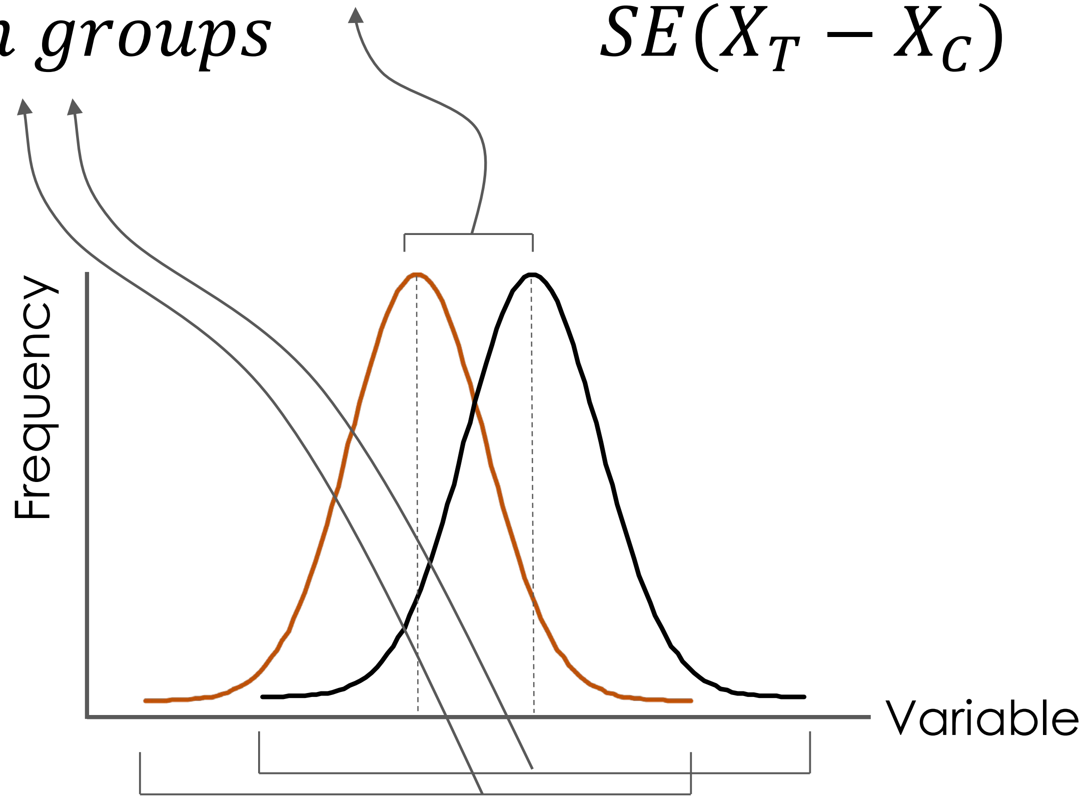
Input Variable

	Nominal	Categorical (>2 categories)	Ordinal	Quantitative Discrete	Quantitative Non-Normal	Quantitative Normal
Nominal	X^2 or Fisher's	X^2	X^2 trend or Mann-Whitney	Mann-Whitney	Mann-Whitney or log-rank (a)	**Student's t-test**
Categorical (>2 categories)	X^2	X^2	Kruskal-Wallis (b)	Kruskal-Wallis (b)	Kruskal-Wallis (b)	Analysis of variance (c)
Ordinal	X^2 trend or Mann-Whitney	(e)	Spearman rank	Spearman rank	Spearman rank	Spearman rank or linear regression (d)
Quantitative Discrete	Logistic regression	(e)	(e)	Spearman rank	Spearman rank	Spearman rank or linear regression (d)
Quantitative Non-Normal	Logistic regression	(e)	(e)	(e)	Plot data and Pearson or Spearman rank	Plot data and Pearson or Spearman rank and linear regression
Quantitative Normal	Logistic regression	(e)	(e)	(e)	Linear regression (d)	Pearson and linear regression

T-statistic Calculation

$$\frac{signal}{noise} = \frac{difference\ between\ group\ means}{variability\ in\ groups} = \frac{X_T - X_C}{SE(X_T - X_C)}$$

$$t = \frac{X_T - X_C}{\sqrt{\frac{var_T}{n_T} + \frac{var_C}{n_C}}}$$



Example

Do men and women have different body temperatures?

$\boldsymbol{\mu}_M$ = mean body temperature of adult males

$\boldsymbol{\mu}_F$ = mean body temperature of adult females

Test difference between:

- Null hypothesis – H_0: $\mu_M = \mu_F$
- Alternative hypothesis – H_1: $\mu_M \neq \mu_F$

Is this one-tailed or two-tailed?

Calculate *T*-statistic

```
> temp = read.delim("Temps.txt")
> aggregate(temp$bodytemp~temp$gender, temp, mean)
  temp$gender temp$bodytemp
1      female      98.39385
2        male      98.10462
```

$\bar{F}$

$\bar{M}$

```
> aggregate(temp$bodytemp~temp$gender, temp, var)
  temp$gender temp$bodytemp
1      female     0.5527740
2        male     0.4882596
```

s_F^2

s_M^2

```
> summary(temp$gender)
female   male
    65     65
```

n_F n_M

$$t = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{s_X^2/n_X + s_Y^2/n_Y}}$$

$$t = \frac{(98.10462 - 98.39385) - (0 - 0)}{\sqrt{\frac{.4882596}{65} + \frac{.5527740}{65}}}$$

t = -2.285

Degrees of Freedom

$$\nu = \frac{\left(\dfrac{s_X^2}{n_X} + \dfrac{s_Y^2}{n_Y}\right)^2}{\dfrac{(s_X^2/n_X)^2}{n_X - 1} + \dfrac{(s_Y^2/n_Y)^2}{n_Y - 1}}$$

rounded down to the nearest integer.

$$\nu = \frac{\left(\dfrac{.4882596}{65} + \dfrac{.5527740}{65}\right)^2}{\dfrac{\left({.4882596}/{65}\right)^2}{65 - 1} + \dfrac{\left({.5527740}/{65}\right)^2}{65 - 1}} = 128$$

Calculate P-Value

Using a t-table:

α =.05

v = 128

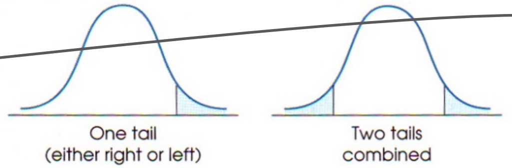
t = -2.285

Thus:

> 0.05 > P($\mu_M \neq \mu_F$) > 0.02
>
> We reject the null hypothesis at the **α** =.05 level

THE *t* DISTRIBUTION

Table entries are values of t corresponding to proportions in one tail or in two tails combined.

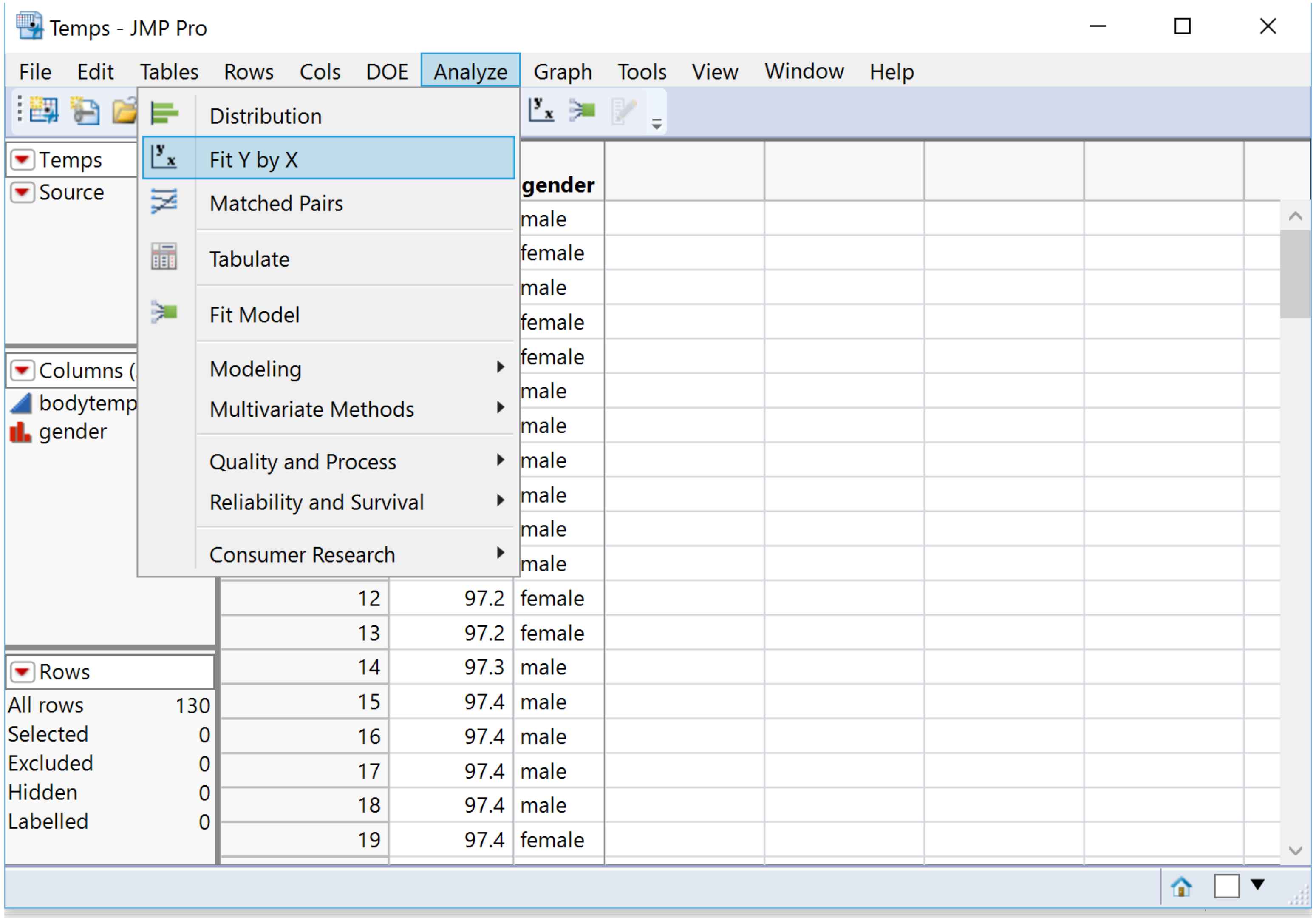


	PROPORTION IN ONE TAIL					
	0.25	0.10	0.05	0.025	0.01	0.005
	PROPORTION IN TWO TAILS COMBINED					
df	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.319	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
40	0.681	1.303	1.684	2.021	2.423	2.704
60	0.679	1.296	1.671	2.000	2.390	2.660
120	0.677	1.289	1.658	1.980	2.358	2.617
∞	0.674	1.282	1.645	1.960	2.326	2.576

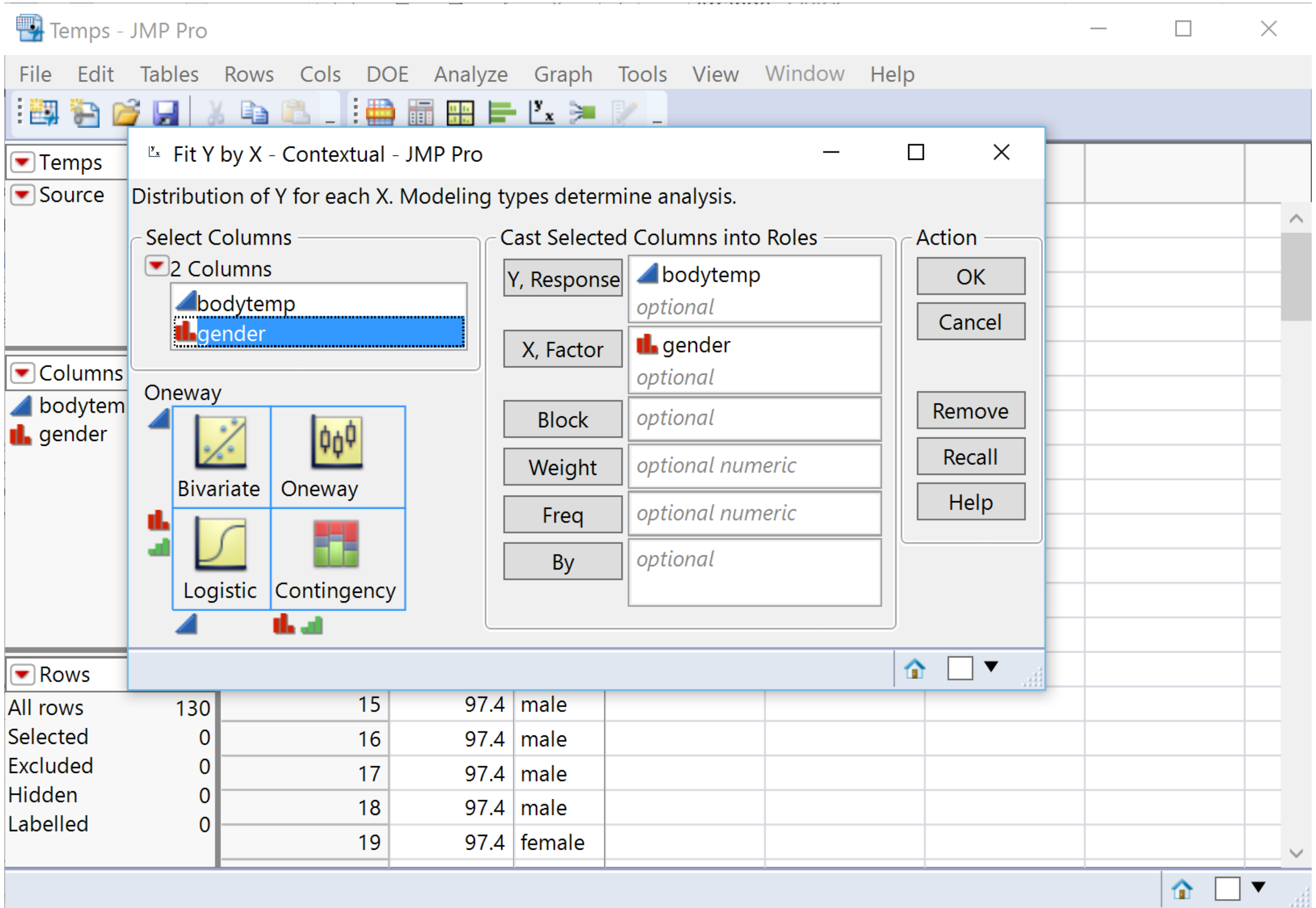
Unpaired T-test in R

```
> temps <- read.table("Temps.txt", header=TRUE)
> attach(temps)
> t.test(bodytemp~gender, alternative="two.sided", var.equal=TRUE,
conf.level=.95)
      Two Sample t-test
data:  bodytemp by gender
t = 2.2854, df = 128, p-value = 0.02393
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.03882216 0.53963938
sample estimates:
mean in group female   mean in group male
            98.39385             98.10462
```

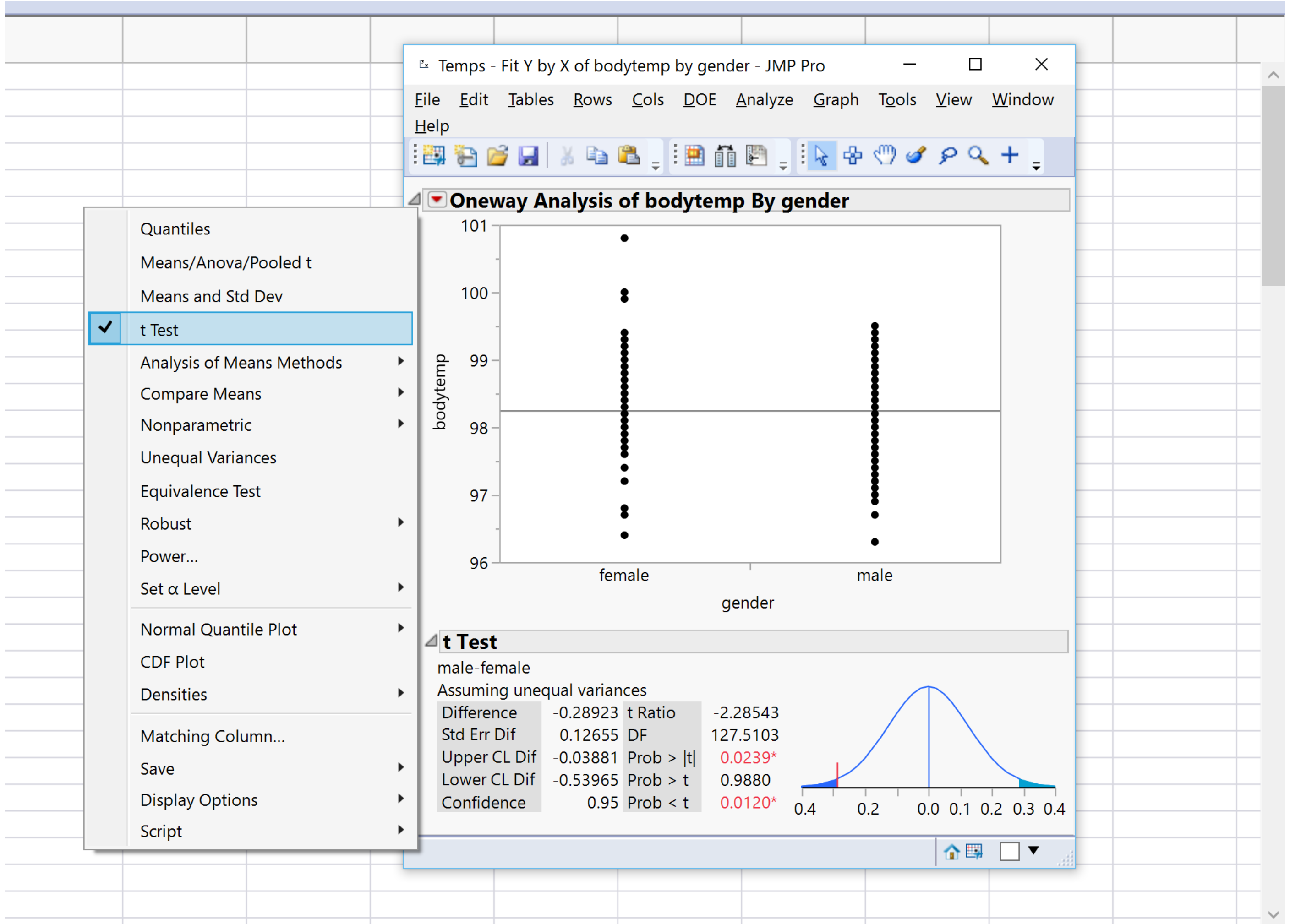
Unpaired T-test in JMP



Unpaired T-test in JMP



Unpaired T-test in JMP



Questions?

Another Example

Hypothesis

- A robot will be perceived as less warm when it produces a gaze cue
  - Null Hypothesis – H_0: $\mu_{Cue} = \mu_{NoCue}$
  - Alternative Hypothesis – H_1: $\mu_{Cue} < \mu_{NoCue}$
- A robot will be perceived as less attentive when it produces a gaze cue
  - Null Hypothesis – H_0: H_0: $\mu_{Cue} = \mu_{NoCue}$
  - Alternative Hypothesis – H_1: $\mu_{Cue} < \mu_{NoCue}$

What kind of t-test?

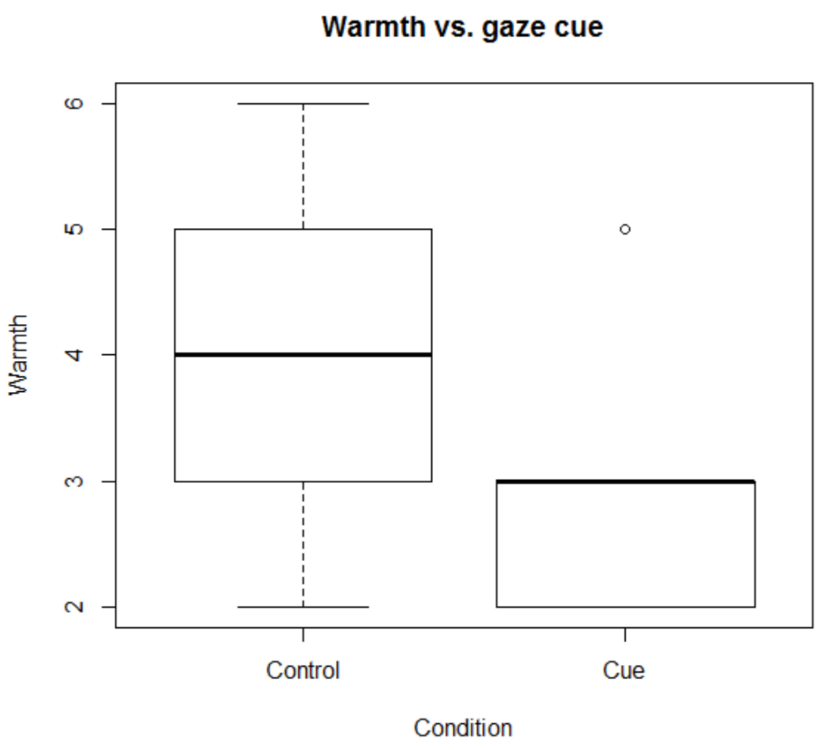
- Unpaired or paired?
- One-tailed or two-tailed?

Go do it!

- Use the FactorAnalysis_Data.csv data

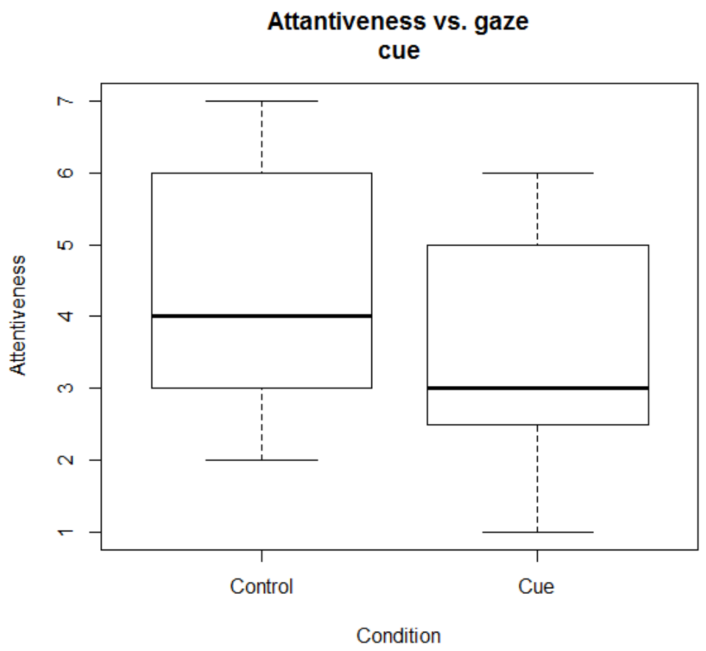
1. Visualize the data

```
> boxplot(qdata$Warm ~ qdata$Condition,
main="Warmth vs. gaze cue", xlab="Condition",
ylab="Warmth")
```



1. Visualize the data

```
> boxplot(qdata$Attentive ~ qdata$Condition,
main="Attantiveness vs. gaze cue",
xlab="Condition", ylab="Attentiveness")
```



2. T-Tests

```
> t.test(qdata$Warm~qdata$Condition, paired=FALSE,
alternative="less",conf.level=0.95)
     Welch Two Sample t-test
data:  qdata$Warm by qdata$Condition
t = 2.2734, df = 26.827, p-value = 0.9844
alternative hypothesis: true difference in means is less
than 0
95 percent confidence interval:
    -Inf 1.63277
sample estimates:
mean in group Control     mean in group Cue
             4.000000              3.066667
```

Huh?

2. T-Tests

If you expect control to be greater than treatment, use “greater”!

i.e. H_1: $\mu_{Treatment} < \mu_{Control}$

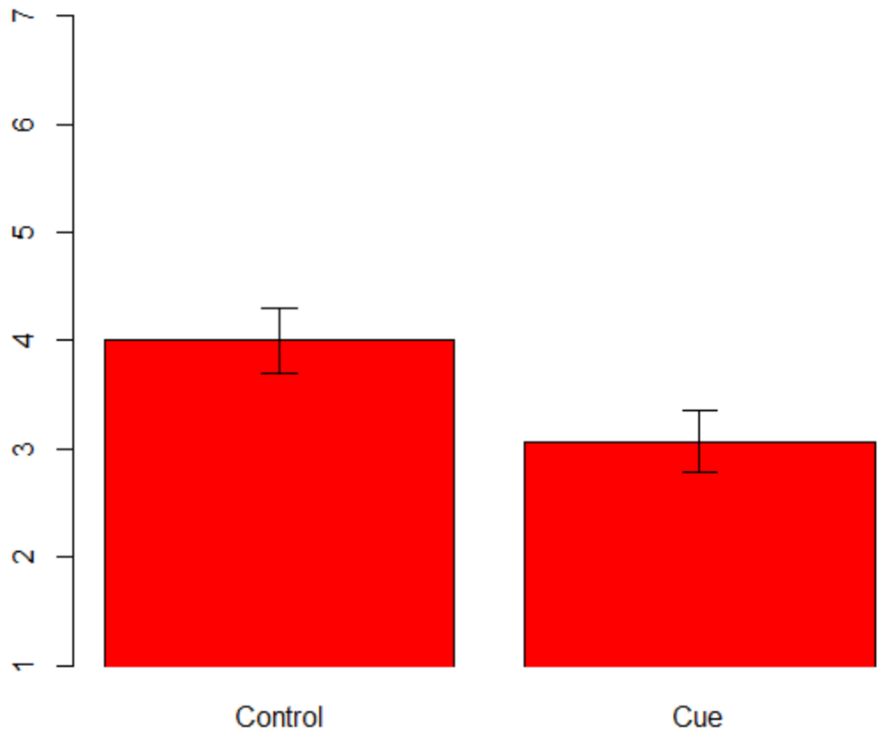
```
> t.test(qdata$Warm~qdata$Condition, paired=FALSE, alternative="greater",conf.level=0.95)

            Welch Two Sample t-test

data:  qdata$Warm by qdata$Condition

t = 2.2734, df = 26.827, p-value = 0.01561

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

 0.2338971       Inf

sample estimates:

mean in group Control     mean in group Cue

             4.000000              3.066667
```

2. T-Tests

```
> t.test(qdata$Attentive~qdata$Condition, paired=FALSE,
alternative="greater", conf.level=0.95)
     Welch Two Sample t-test
data:  qdata$Attentive by qdata$Condition
t = 1.7504, df = 26.066, p-value = 0.04591
alternative hypothesis: true difference in means is
greater than 0
95 percent confidence interval:
 0.02678412        Inf
sample estimates:
mean in group Control     mean in group Cue
             4.642857              3.600000
```

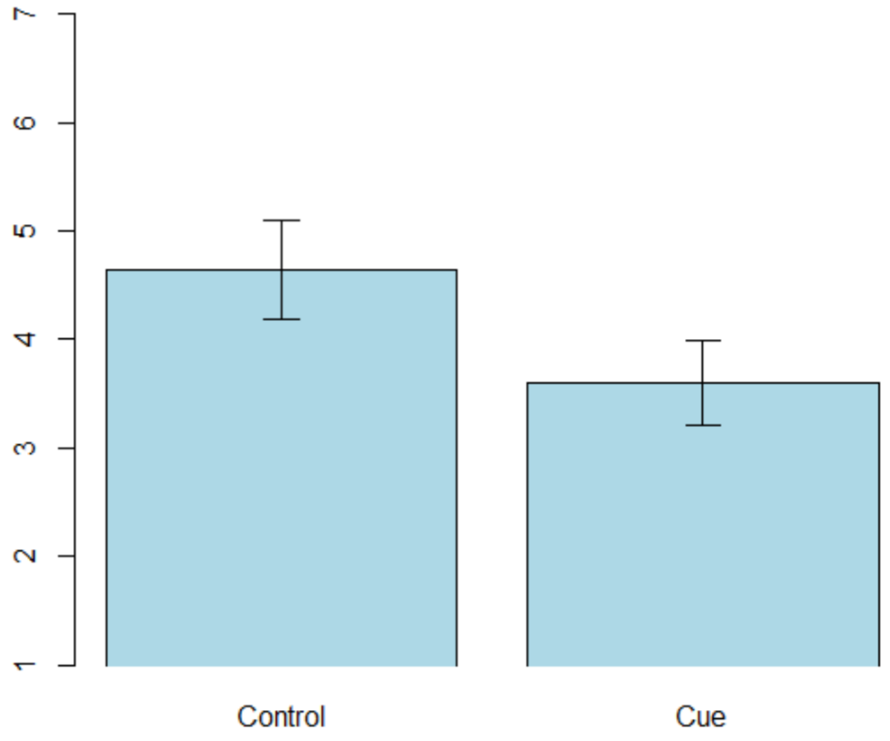
3. Graph Results

```
> library(sciplot)
> bargraph.CI(qdata$Condition, qdata$Warm,
ylim=c(1,7), col="Red")
```

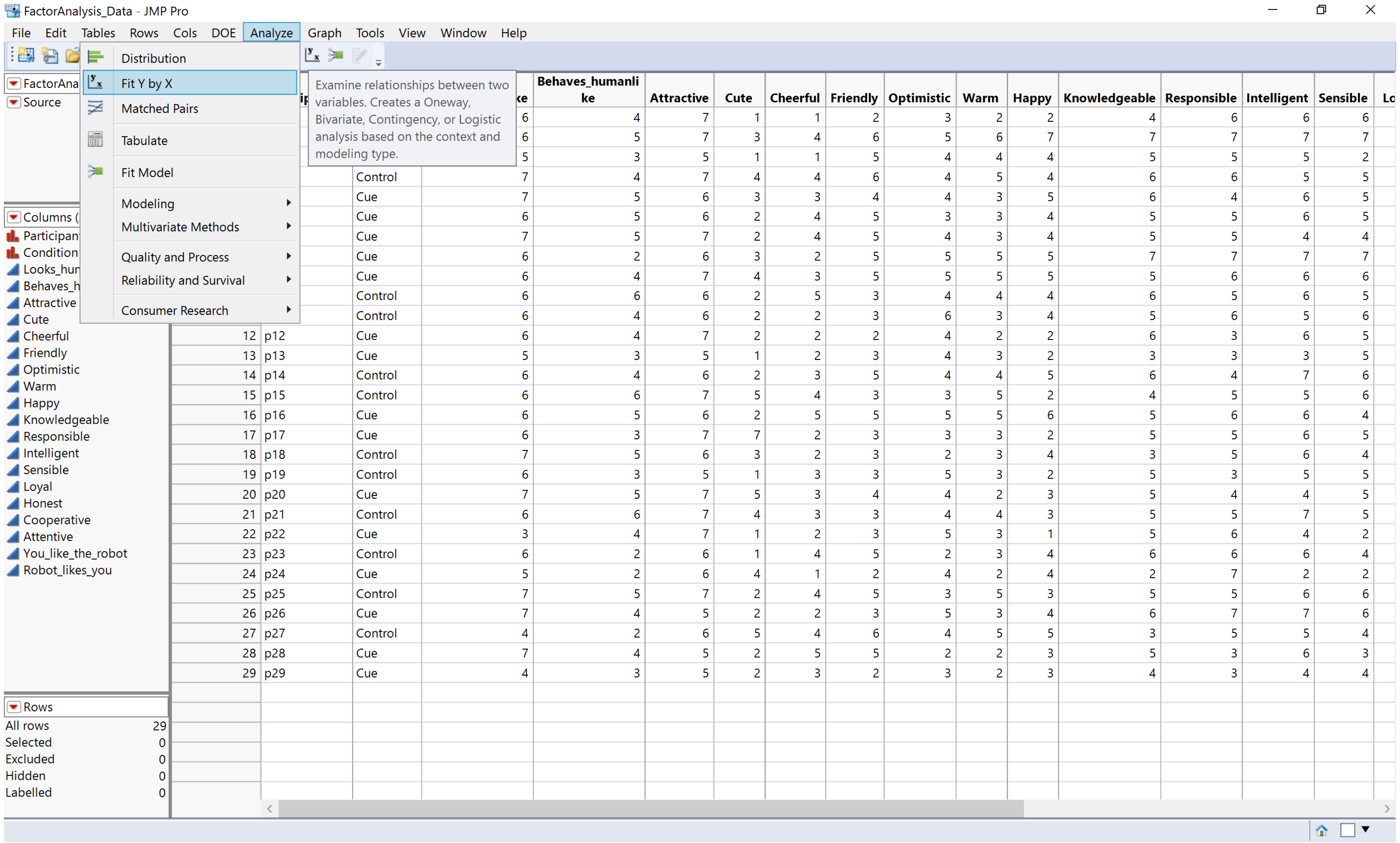


3. Graph Results

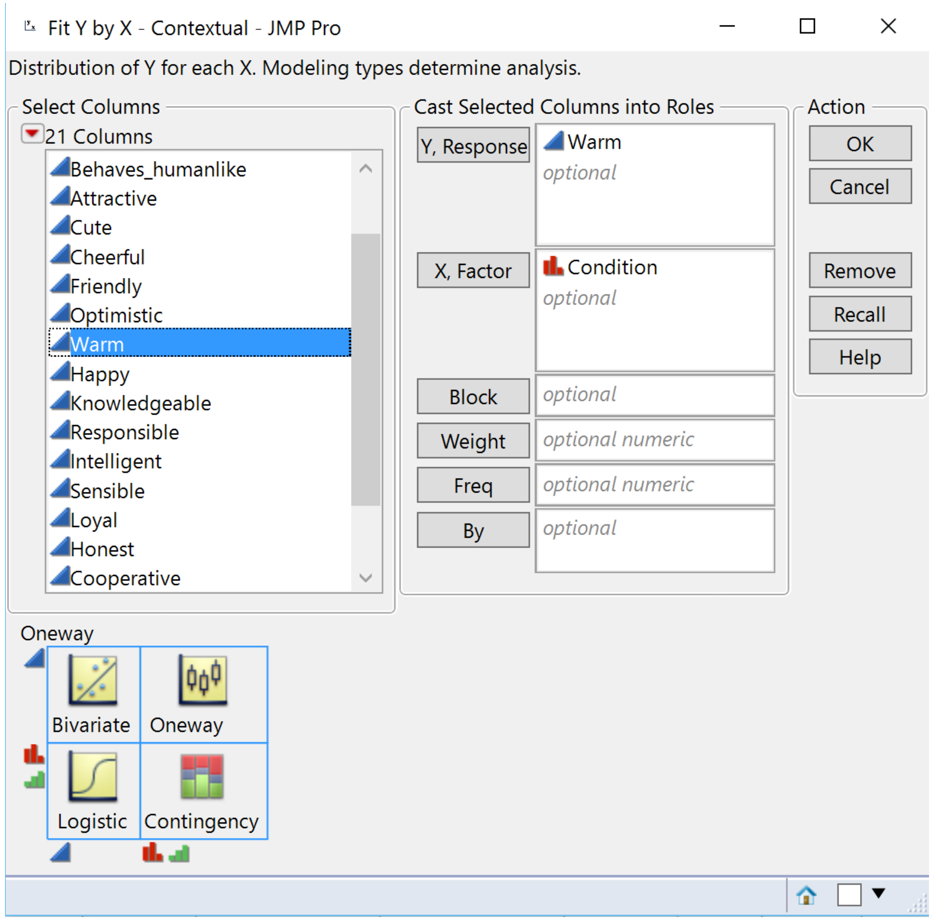
```
> bargraph.CI(qdata$Condition, qdata$Attentive,
ylim=c(1,7), col="lightblue")
```



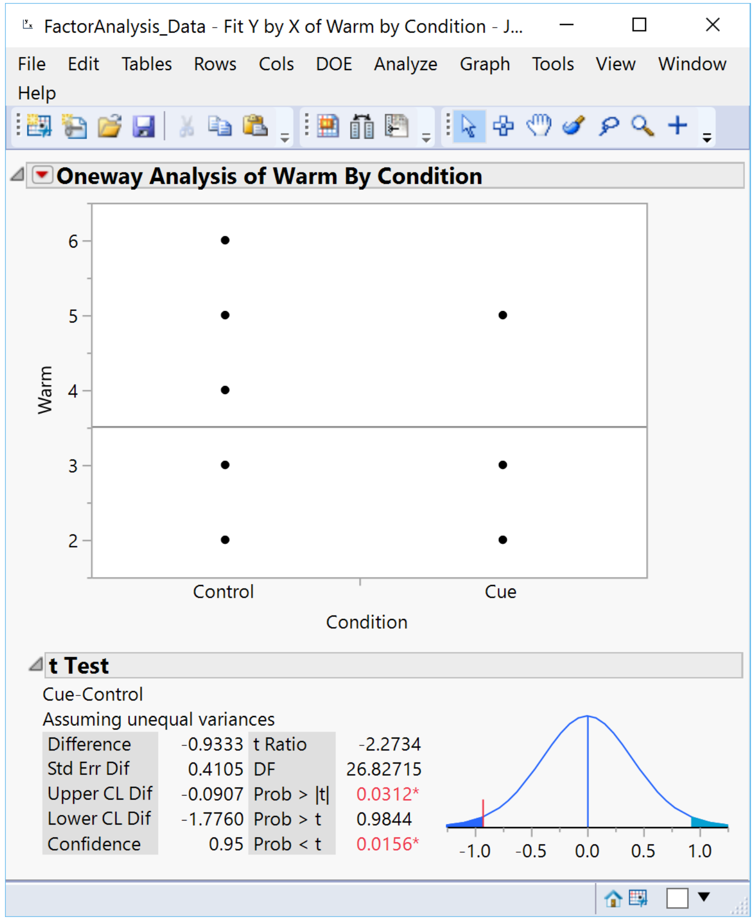
Unpaired T-Test in JMP



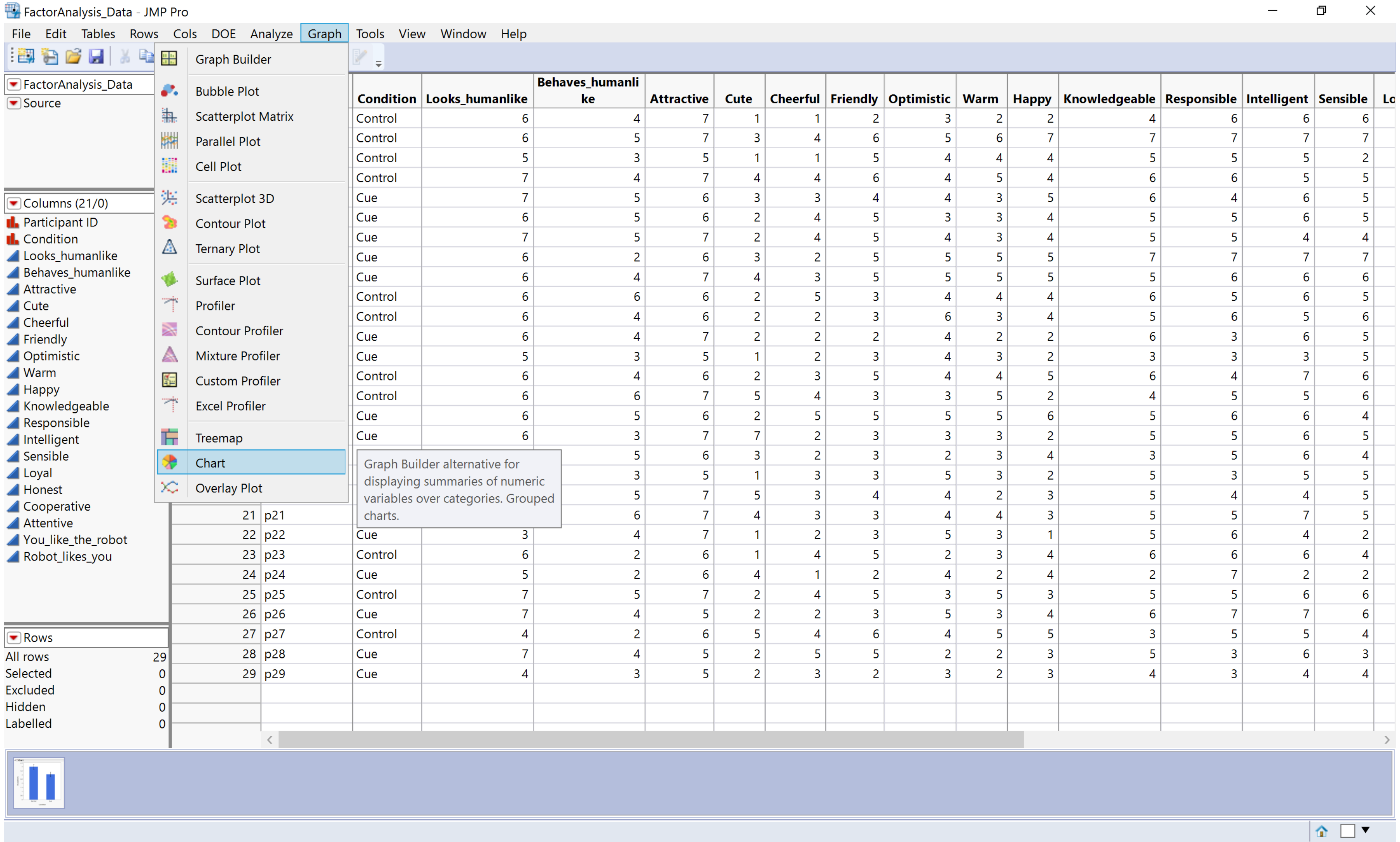
Unpaired T-Test in JMP



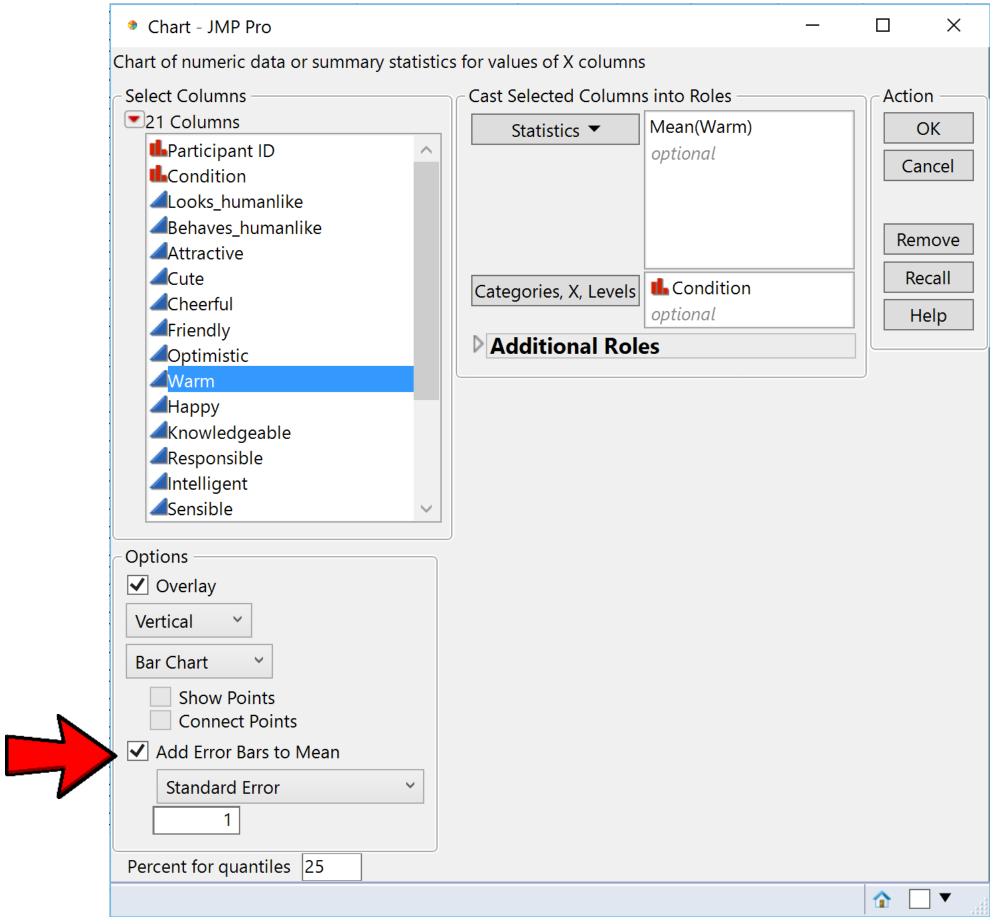
Unpaired T-Test in JMP



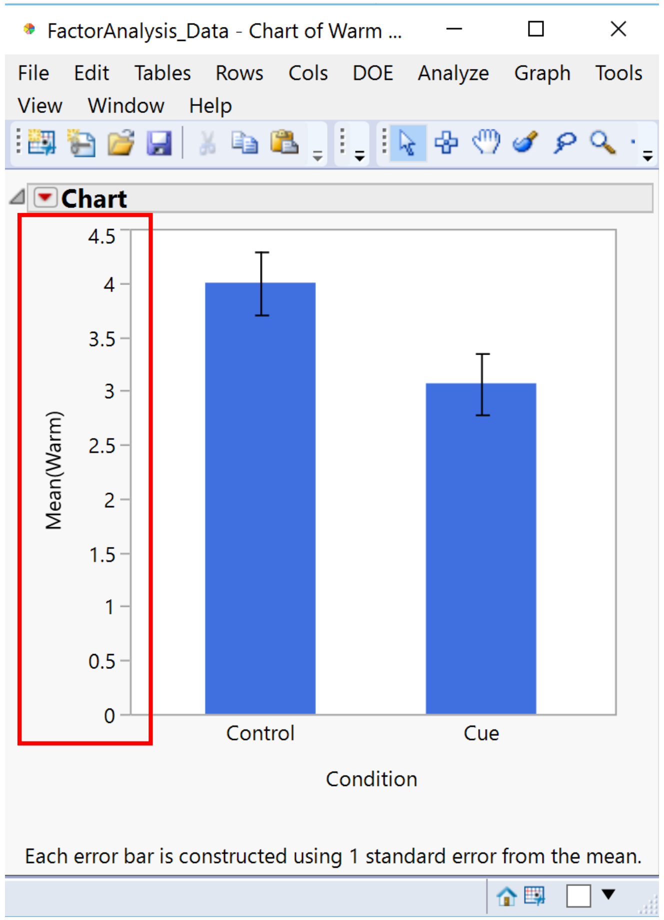
Graphing the Result



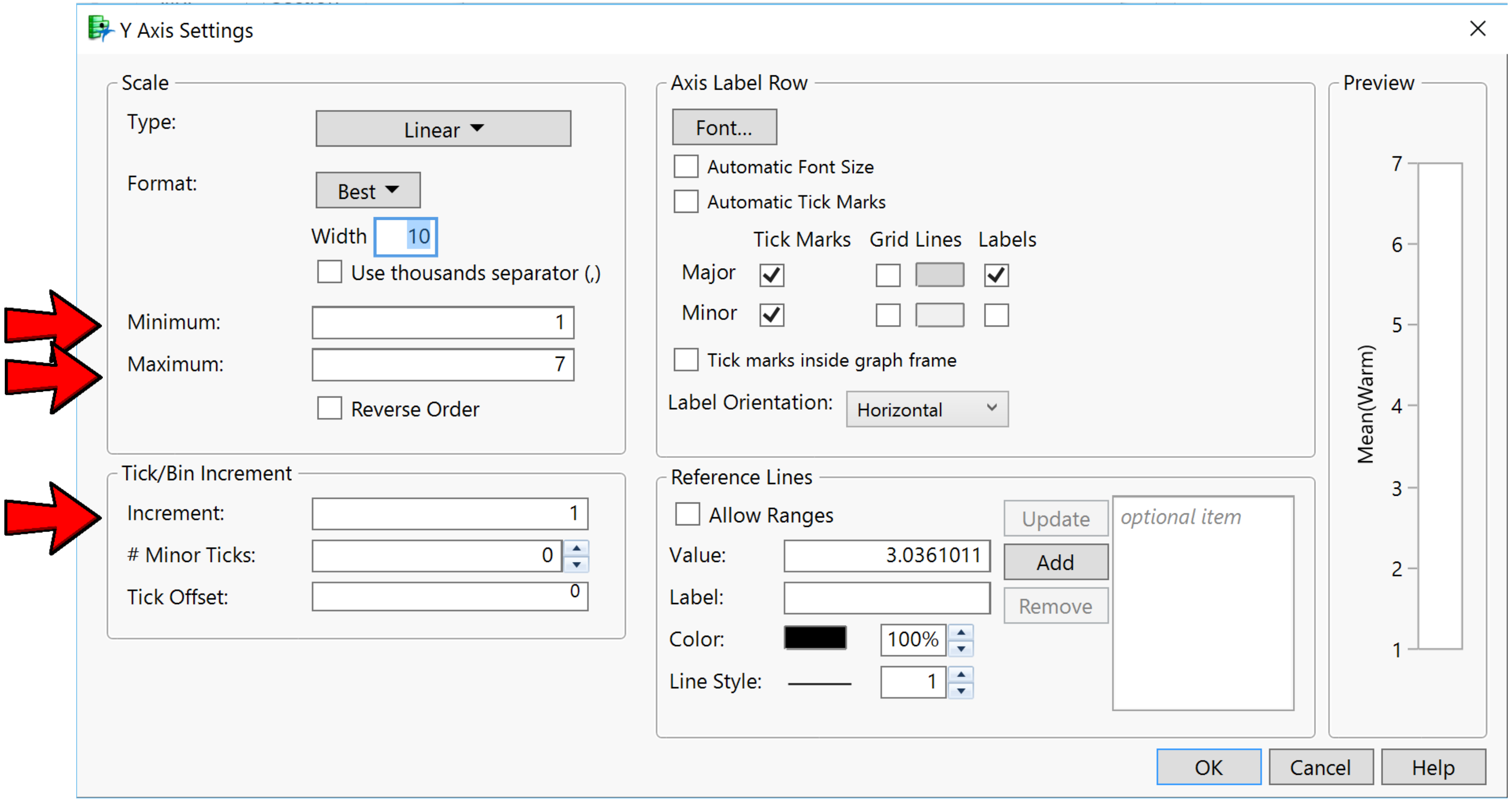
Graphing the Result



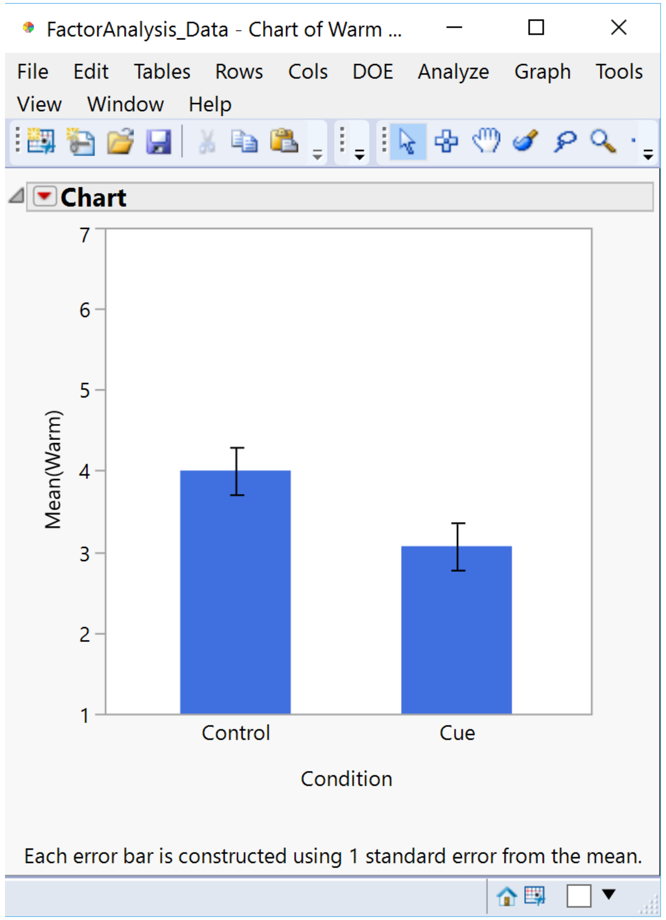
Graphing the Result



Graphing the Result



Graphing the Result



Questions?

Paired T-Test

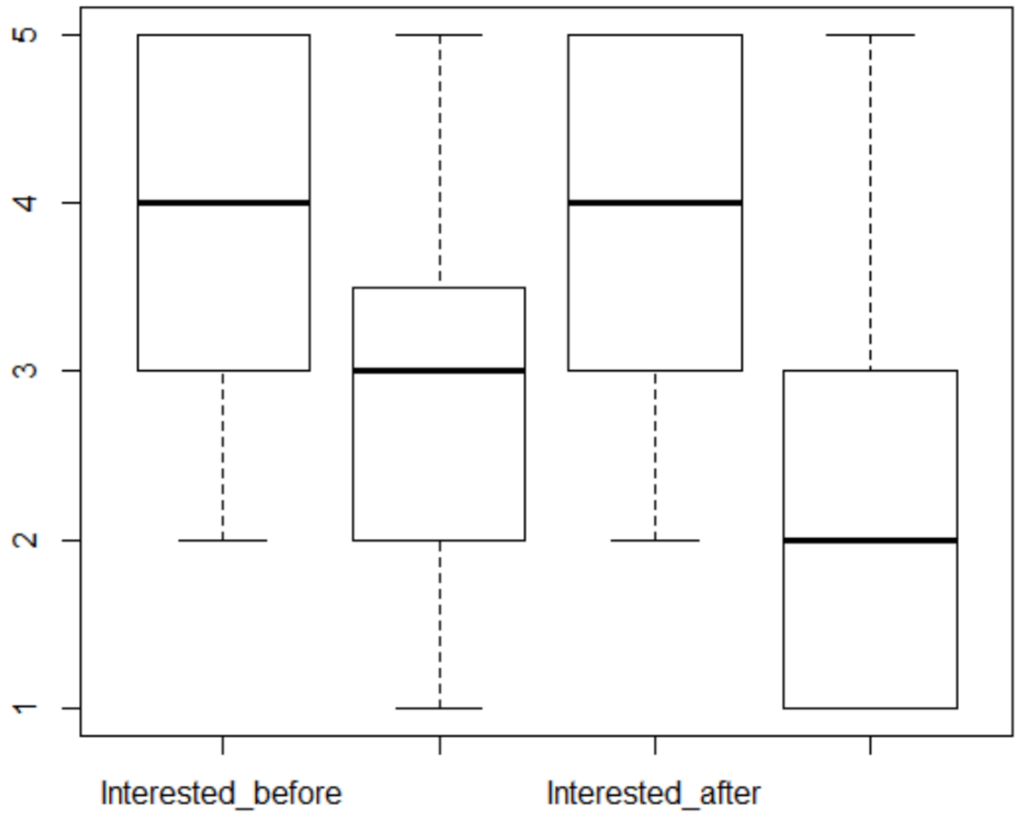
Paired T-Test

Within-participants

- Quasi-experimental designs
- Repeated measures
  - Example: emotions before and after people interact with a robot

Paired T-Test R Example

```
> emotions <- read.table("RepeatedMeasures.csv", sep=",", header=TRUE)
> attach(emotions)
> boxplot(emotions)
```



Paired T-Test R Example

```
> t.test(emotions$Interested_before, emotions$Interested_after, paired=TRUE, conf.level=0.95)

      Paired t-test

data:  emotions$Interested_before and emotions$Interested_after
t = -1.4314, df = 42, p-value = 0.1597
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.22417757  0.03813105
sample estimates:
mean of the differences
            -0.09302326
```

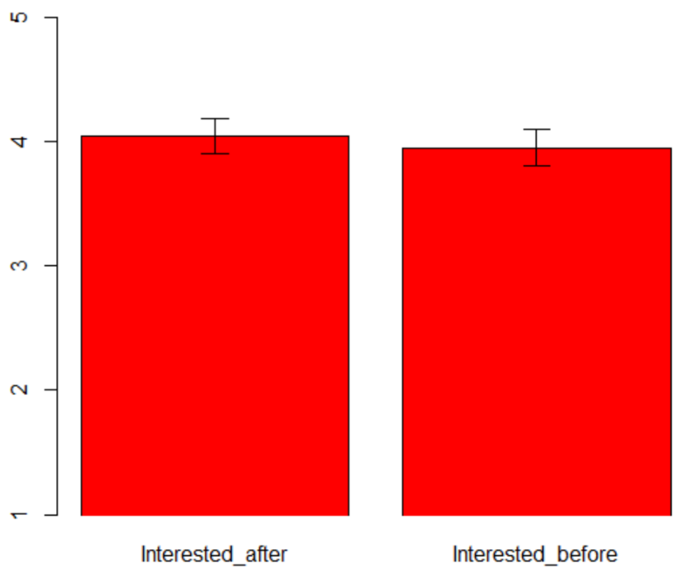
Paired T-Test R Example

```
> t.test(emotions$Worried_before, emotions$Worried_after,
paired=TRUE, conf.level=0.95)
```

```
     Paired t-test

data:  emotions$Worried_before and emotions$Worried_after
t = 3.4931, df = 42, p-value = 0.001138
alternative hypothesis: true difference in means is not equal
to 0
95 percent confidence interval:
 0.2160424 0.8072134
sample estimates:
mean of the differences
              0.5116279
```

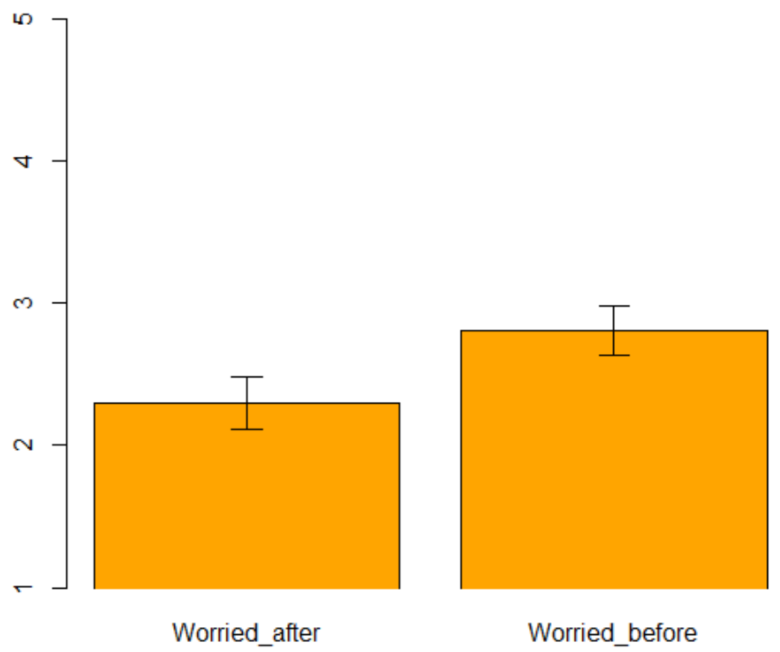
Paired T-Test R Example

```
>bargraph.CI(c(rep("Interested_before",times=43)
,rep("Interested_after",times=43)),
c(emotions$Interested_before,emotions$Interested
_after), ylim=c(1,5), col="Red")
```

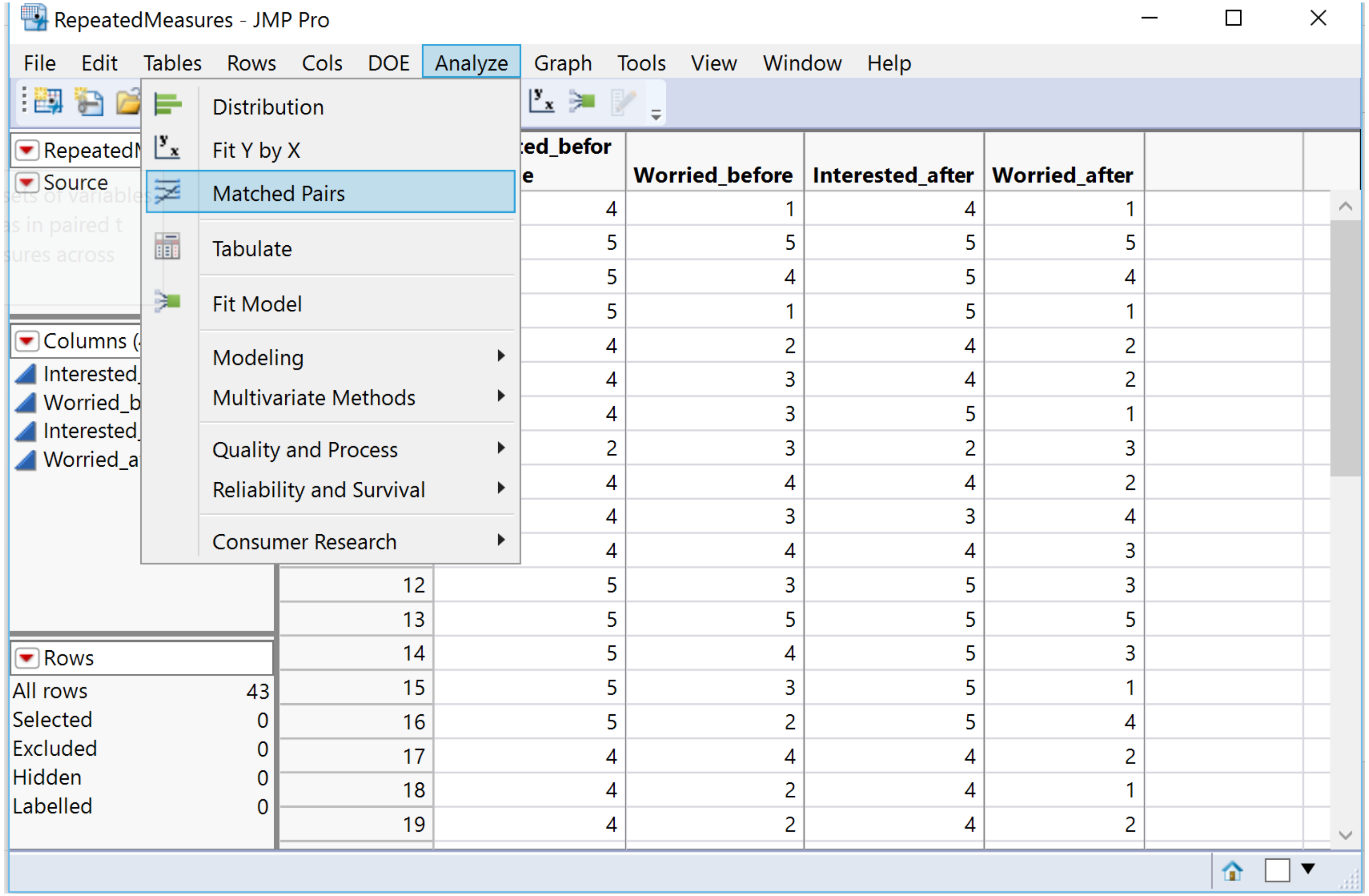


Paired T-Test R Example

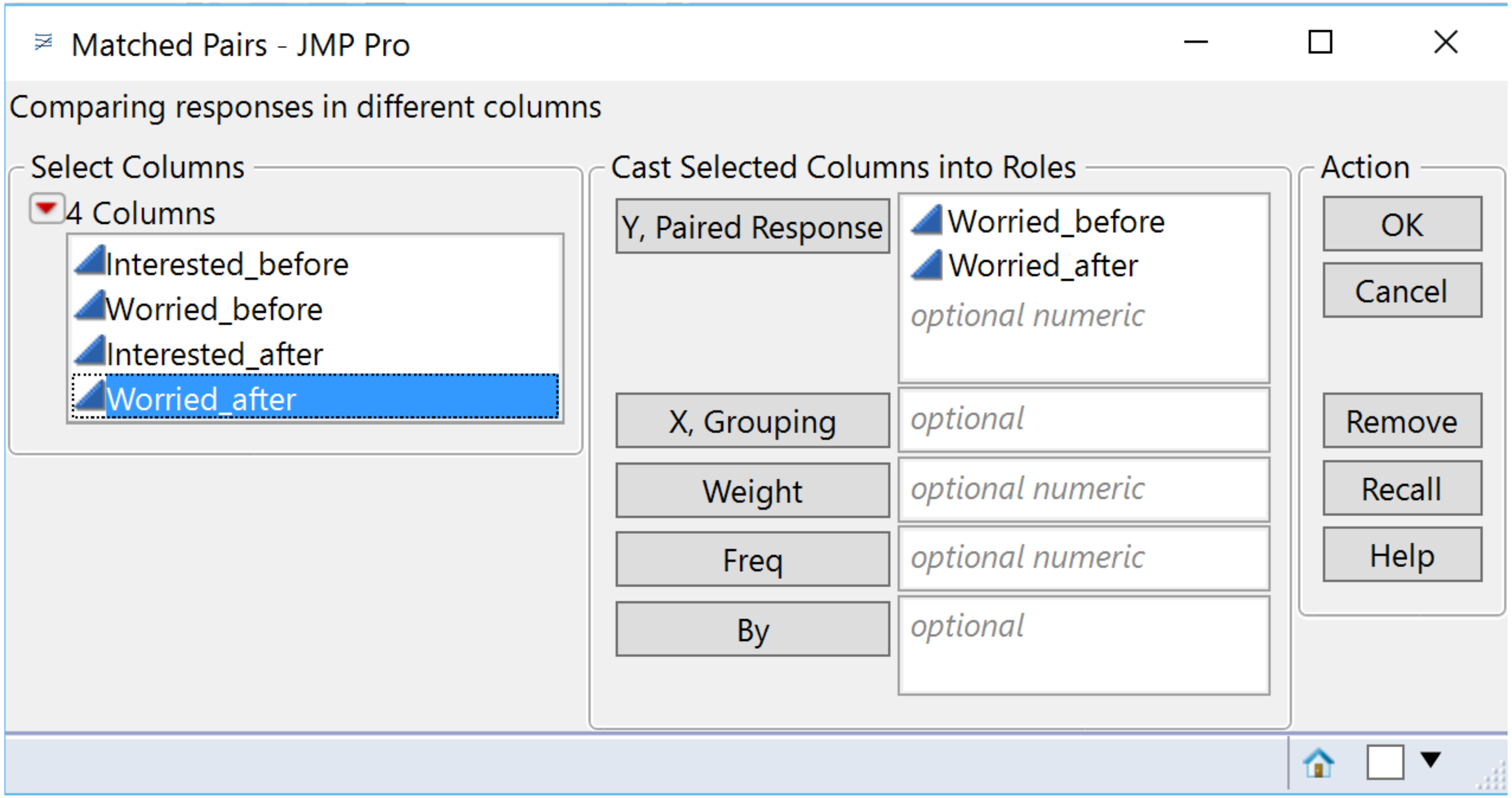
```
>bargraph.CI(c(rep("Worried_before",times=43),rep("Worried_after",times=43)),
c(emotions$Worried_before,emotions$Worried_after), ylim=c(1,5),col="Orange")
```



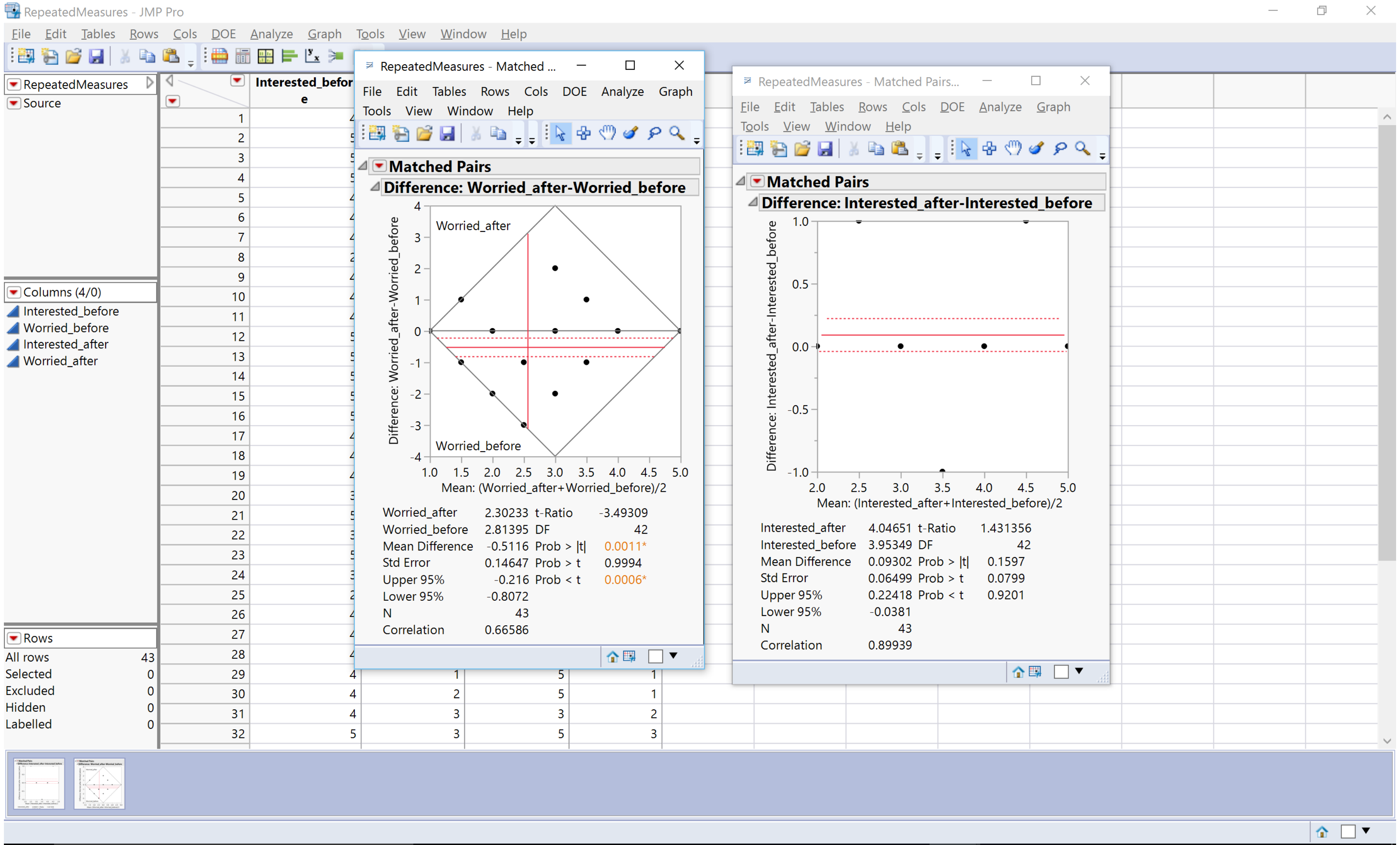
Paired T-Test in JMP



Paired T-Test in JMP



Paired T-Test in JMP



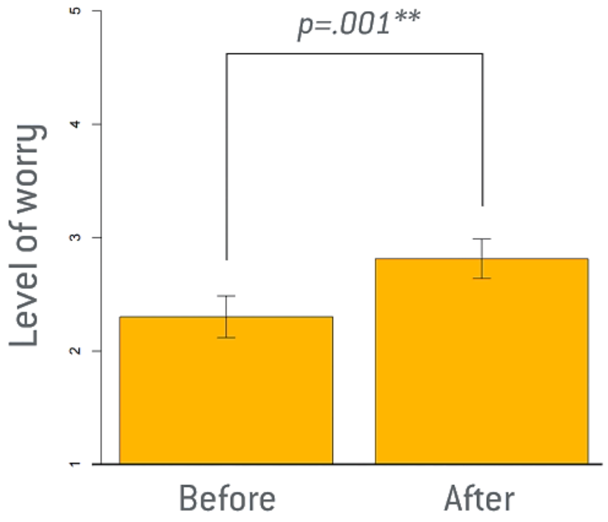
Questions?

Reporting Tests

Statistic	Example
Mean and standard deviation	$M = 3.45$, $SD = 1.21$
Mann-Whitney	$U = 67.5$, $p = .034$, $r = .38$
Wilcoxon signed-rank	$Z = 4.21$, $p < .001$
Sign test	$Z = 3.47$, $p = .001$
T-test	$t(19) = 2.45$, $p = .031$, $d = 0.54$
ANOVA	$F(2, 1279) = 6.15$, $p = .002$, $\eta_p^2 = 0.010$
Pearson's correlation	$r(1282) = .13$, $p < .001$

Reporting T-tests

"We conducted a **paired two-tailed t-test** to test whether participating in the experiment affected participant's level of worry and found that participants were significantly more worried after the experiment that they were before the experiment , **$t(42) = 3.49$**, **$p = .001$**."



Summary

Choosing the right test

What question do you want to answer?

What are the **type** and **number** of variables you are comparing?

Is your experimental design **within** (likely paired observations) or **between** (likely independent observations)

Questions?

University of Colorado
Boulder

THANKS!

Professor **Dan Szafir**

Computer Science & ATLAS Institute
University of Colorado Boulder