

# Comparing Several Means



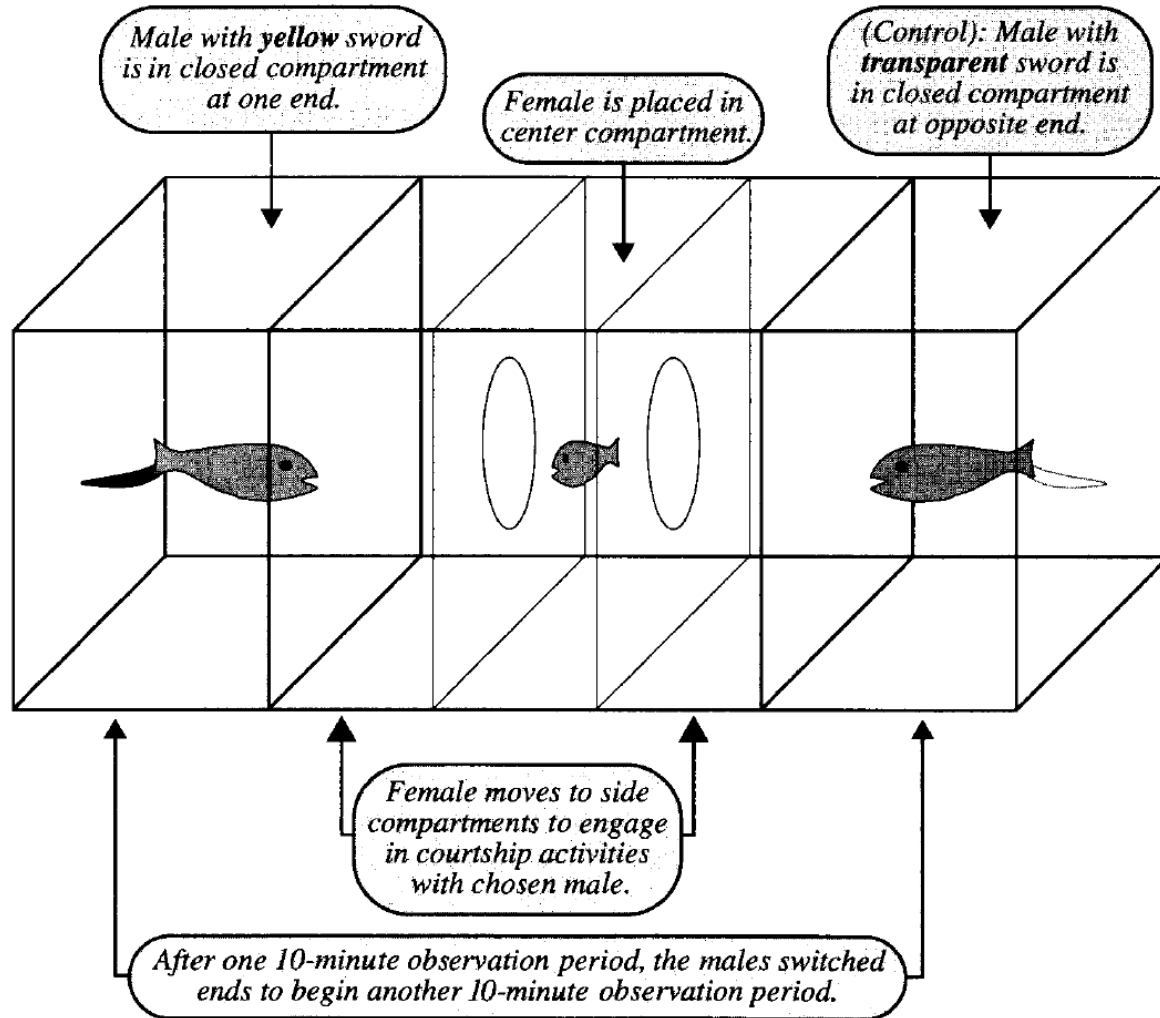
Some slides from R. Pruim

SML201: Introduction to Data Science, Spring 2019

Michael Guerzhoy

# The Dating World of Swordtail Fish

- In some species of swordtail fish, males develop brightly coloured swordtails
- Southern Platyfish do not
- Want to know: will female Southern Platyfish prefer males with artificial brightly-coloured swordtails to males with transparent swordtails?
  - If they do, that's evidence that males in other species evolved as a result of female preference
- Experiment: multiple pairs of males, one with a transparent artificial tail, one with a bright yellow artificial swordtail. Measure the percentage of time the female spends courting with the male with the yellow tail. There are 84 females in total.



	<b>Pair 1</b> (35 mm)	<b>Pair 2</b> (31 mm)	<b>Pair 3</b> (33 mm)	<b>Pair 4</b> (34 mm)	<b>Pair 5</b> (28 mm)	<b>Pair 6</b> (34 mm)
	43.7	52.5	91.0	72.2	78.3	33.4
	54.0	65.6	62.0	58.5	66.0	42.2
	49.8	68.5	10.0	51.0	47.7	35.6
	65.5	45.9	83.8	56.8	77.5	79.9
	53.1	80.2	91.3	92.4	58.3	59.0
	53.0	67.0	56.3	55.3	61.1	58.1
	62.3	73.0	83.6	59.3	65.1	64.2
	49.4	71.7	53.3	42.0	62.9	82.8
	45.7	55.0	36.5	68.5	61.0	75.7
	56.6	70.0	65.4	78.4		66.3
	59.0	63.2	48.1	69.6		56.3
	67.8	39.6	50.6	89.2		84.5
	73.3	41.0	40.4	67.3		61.1
	43.8	59.2	90.6	77.5		87.6
	67.4		74.9			
	58.1		56.0			
			67.5			
<b>Average:</b>	56.41	60.89	62.43	67.00	64.21	63.34
<b>SD:</b>	9.02	12.48	22.29	14.33	9.41	17.68
<b>n:</b>	16	14	17	14	9	14

# Model Assumptions

- Have multiple measurements for each condition
- The measurements in each condition are normally distributed
- Null hypothesis
  - E.g., the average of the group averages is 0.5
  - Can construct a confidence interval

# More

- Lots of situations, lots of models
- Pick the appropriate model
- Be careful about multiple hypotheses and multiple comparisons
- Be careful about making conclusions about causation without controlled experiments
- Recommended references
  - Fred Ramsey and Daniel Schafer, *The Statistical Sleuth*
  - Andrew Gelman and Jennifer Hill, *Data Analysis using Regression and Multilevel/Hierarchical Models*