

1. (a) This experiment was designed to test the hypothesis that female platyfish are more attracted to male platyfish who have a sword like tail. To test this, a randomized complete block design was used. Male platyfish were the experimental units and were paired into six blocks of two based on body size and coloration. Each platyfish within a block was assigned one of two treatments; the control treatment was the surgical attachment of a transparent 25mm plastic "sword" to the back of the fish, while the alternative treatment was the attachment of a yellow and black 25mm plastic "sword" to the back of the fish. In both cases, this was done three days prior to the experiment. It is not stated how assignment of treatment within a block was preformed. Each of these paired platyfish were placed simultaneously into the first and third section of an aquarium divided into equal thirds. A female platyfish was then placed in the middle section for ten minutes. This process was then repeated, except with the sides the males are on switched, resulting in the female being in the aquarium for twenty minutes. The response variable, measured for each male fish independently, was the number of seconds (the observational unit) which the female fish spent in the third of her third of the aquarium adjacent to a particular male fish's section. This entire process was replicated for each block with between nine and sixteen different female platyfish.
- (b) The standard errors in table 1 represent the standard deviation of the sampled mean around the expected time a female fish would spend in the section adjacent to a particular male fish. If we treat the time spent near one fish and another as independent, and assume each is normally distributed, we can get the standard error of the mean difference from the formula

$$SE_{\bar{X}_1 - \bar{X}_2} = \sqrt{SE_{\bar{X}_1}^2 + SE_{\bar{X}_2}^2}$$

applying this to the six pairs, we get estimates

	1	2	3	4	5	6
SE	36.1	53.8	92.7	60.1	48.0	75.7

I think the standard errors should in fact be larger than this. Since a female fish can only spend time on one side at a time, the response for the two treatments is likely inversely correlated. This would cause larger variance in the difference than one would expect if they were independent.

- (c) If we make a number of assumptions, we can take a similar approach to b to generate a confidence interval using the normal approximation. Assumptions:
  1. As before, the response for the yellow and transparent swords for a particular pair are independent. This is an obviously incorrect assumption, since by construction these variables are negatively correlated, as mentioned before.
  2. The means for yellow and transparent respectively are independent between each male and female combination. One could imagine ways in which this could be false, such as females developing preference after being used to test multiple pairs of males, but is probably reasonably safe.
  3. The mean difference for yellow and transparent for a particular pair of males is the same as for all other pairs of males. Eyeballing the data, it looks like pairwise there is sufficient data to reject this assumption in at least a few cases (pairs 1 and 4), but this is cherry picking. Whether or not we can show this to be a statistically invalid assumption though, it is unreasonably strong, and one I would not make if I had more complete data to work with.
  4. The means for yellow and transparent respectively across all males and females are normally distributed. Given assumption 2, this will be true asymptotically. There are enough trials this is reasonable in itself.

Given these assumptions, if we first calculate the mean time across all males for yellow and transparent respectively and the associated standard error, we can do a difference of means on that. This will give us a confidence interval of

$$CI = \sum_{i=1}^6 \bar{X}_{i,y} - \sum_{i=1}^6 \bar{X}_{i,t} \pm \frac{1.96}{\sum_{i=1}^6 n_i} \sqrt{\sum_{i=1}^6 (SE_{i,y}^2 + SE_{i,t}^2)} n_i^2 = 277.7 \pm 1.96 \cdot = (223.5, 331.9)$$

- (d) This additional experiment strengthens the authors argument by using the principle of self-control. When using different fish, no matter how careful the experimenter is in blocking, there is always the possibility that there is some still unaccounted for variance between experimental units. However, if both treatments can be applied to the same unit, as he does here by switching tails, any change in response should be just the result of change in treatment. Since the response changed significantly with the switch in treatment on the same unit, it is a very strong indicator that the treatment is responsible for the change.
- (e) I generally find the rest of her argument that the preexisting female bias could account for the evolution of the sword in the swordtail convincing, with the caveat that I don't think she precluded that another mechanism could have been more important. Her three points stand to reason; if the females of a species had a preference for a trait before it existed in the males, it would stand to reason this would drive its evolution. If one accepts that platyfish are a good representation of swordtails before the sword evolved, then she has established her three points well through other work and this experiment. However, it could well be the case the male trait which females are biased towards also coincidentally had other evolutionary advantages which are as powerful or more powerful than female attraction. For instance, if the sword made the fish faster, then this could also have been a more powerful force to drive its arising. This always allows the possibility that female preference was an additional motivator, rather than the main one, as the author at times suggests.
2. (a) Rejection of the null offers two possibilities: the acceleration of gravity varies over time, or that at least some of the experiments were biased. The former would contradict well established physics theory, which makes it unlikely, so I would tend to concluding the latter.
- (b) When we calculate the  $f$  statistic,

$$f = \frac{\sum_{i=1}^8 n_i (\bar{Y}_i - \bar{Y})^2 / (8 - 1)}{\sum_{i=1}^8 \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 / (81 - 8)} = 3.56$$

which corresponds to a p-value of .0024, from a F distribution with 73 and 7 degrees of freedom. On this basis we would reject the null hypothesis that each series has the same mean. This test holds true under the assumptions that each series is normally distributed with identical variance and that each observation is independent of the others.

- (c) Fitting the two models by max likelihood using the gradient optimization, I got two sets of parameters. For the null, where the mean is equal across series:

	mu	sig1	sig2	sig3	sig4	sig5	sig6	sig7	sig8
1	78.9	21.9	18.3	14.9	8.2	5.0	5.6	4.7	3.6

and where it varies by series:

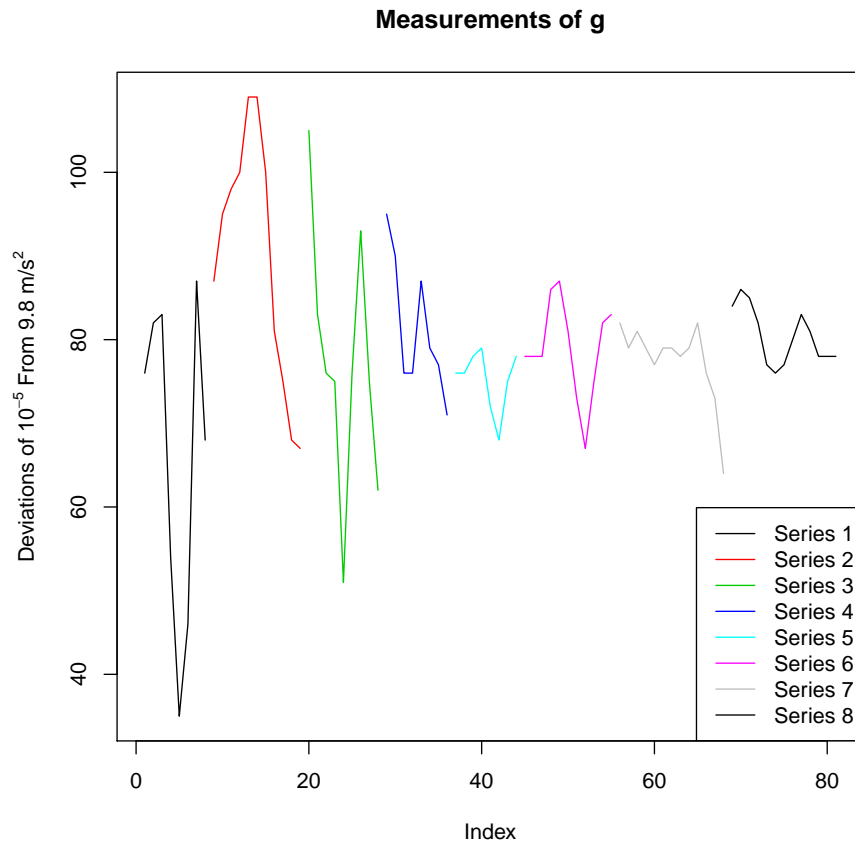
	1	2	3	4	5	6	7	8
mu	66.4	89.9	77.3	81.4	75.3	78.9	77.5	80.4
sigma	18.0	14.6	14.9	7.8	3.4	5.6	4.6	3.2

Plugging the log likelihoods to calculate the log likelihood ratio statistic, we get:

$$-2 \log(LL_0) + 2 \log(LL_1) = -2 \cdot -281.2 + 2 \cdot -271.9 = 18.7$$

Assuming this statistic is chi-square distributed with 7 degrees of freedom, we get a p-value of .0092.

- (d) When we plot the series sequentially, several of our assumptions from the prior parts look off:



from eyeballing this plot, it appears that the different series are, in fact, just sections of one series. This series appears to have decreasing variance over time, and either has a periodic component or has inverse auto-correlation (driving the systematic up and down spikes). Since all the prior tests depended on independence and identical distribution within a series, if this is correct it would invalidate their results.

- (e) To deal with the apparent dependence structure and non constant variance, I used a generalized least squares regression to predict the mean. I allowed for a weighting structure where the variance was a constant plus a power function of the number of measurements, and I adjusted for autocorrelation. The end result is an estimate of 78.75 for the mean deviation, and a standard error of 1.08

```
> summary(gls(s.ul~1,weight=varConstPower(1,form=~A),correlation=corAR1(form=~A))
Generalized least squares fit by REML
Model: s.ul ~ 1
Data: NULL
      AIC      BIC    logLik
571.7913 583.7014 -280.8956

Correlation Structure: AR(1)
Formula: ~A
Parameter estimate(s):
      Phi
0.3697967
Variance function:
Structure: Constant plus power of variance covariate
Formula: ~A
Parameter estimates:
      const      power
6.078766e-10 -6.987870e-01
```

Coefficients:

	Value	Std. Error	t-value	p-value
(Intercept)	78.74893	1.083608	72.6729	0

Standardized residuals:

	Min	Q1	Med	Q3	Max
	-3.0065206	-0.3817783	-0.1035207	0.5786388	2.2754123

Residual standard error: 93.59007

Degrees of freedom: 81 total; 80 residual