

Problem Set #5:
Power, GLMs, and Frequencies

(1) (5pts) You plan to test the effect of a protein-supplemented diet on the growth of green iguanas. You will have two treatment groups: **one that receives a normal diet** of various plants and **another that receives a diet supplemented** with processed soy protein. You will use a **two-sample t-test** to evaluate the **effect of the supplemented diet**. Based on a small preliminary experiment, you think that average growth over the course of the experiment will be about 10 grams, with a standard deviation of 1.5g, and you'd like to be able to detect whether the difference between treatments is at least 1g.

- a. How many replicates of each treatment will you need to have a statistical power of 0.8 at $\alpha = 0.05$? (Assume equal replication in the two treatments.)

With power analysis test for a t-test, I found that we need at least **37 (n=36.27042) replicates per group** (no supplement vs. supplement diet).

- b. Maintaining the same power (0.8) and α (0.05), how many replicates would you need to detect a difference of 3g between the two treatments?

We need at least **6 (n=5.089995) replicates per group**.

- c. What if everything was the same as in (b), but you decided to set α at 0.01? How many replicates would you need?

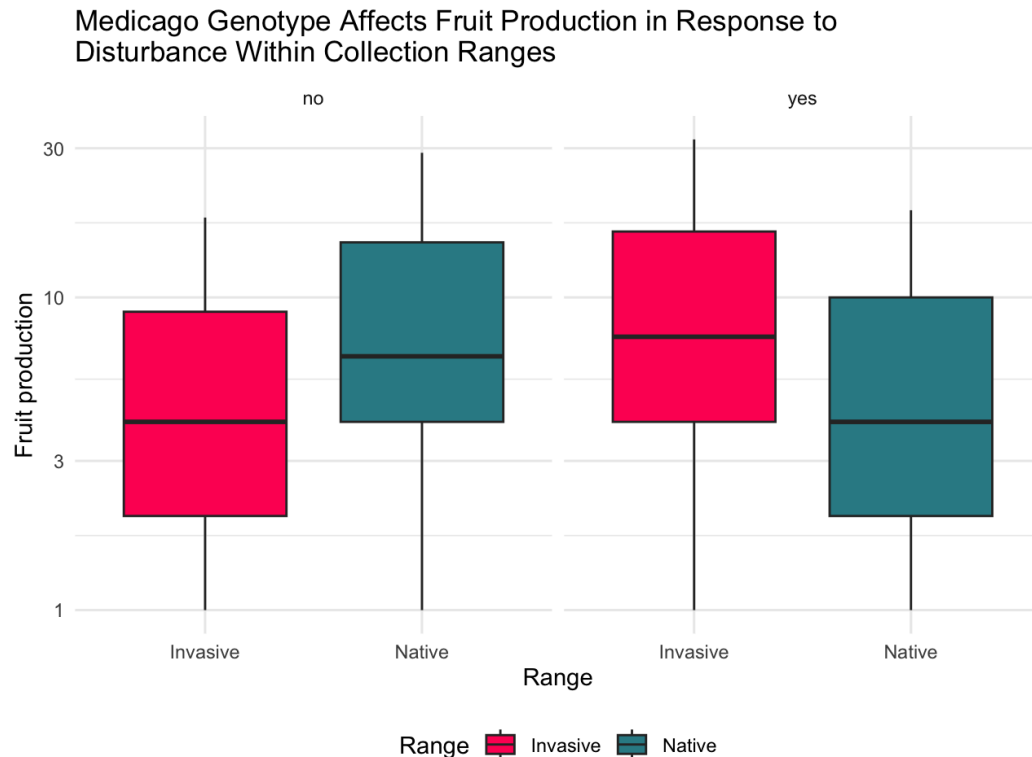
We need at least **8 (n=7.66057) replicates per group**.

- d. What if instead you had only 20 iguanas total (10 for each treatment) to work with? What would the minimum detectable difference between treatments be with power 0.8, $\alpha = 0.05$, and SD = 1.5?

After doing a power analysis test for a t-test, and working backward with Cohen's d ($d=1.324947$) and the standard deviation ($sd=1.5$), I found that the **minimum detectable difference is 1.987**.

(2) (10 pts) Shannon was a former student in Casey terHorst's lab who studied an invasive plant, *Medicago polymorpha*. She wants to know whether some genotypes of this plant are better able to invade disturbed areas. She conducts a field experiment in which she plants seedlings of eight different genotypes, planting five replicate seedlings of each genotype into each of two treatments: plots that have been disturbed, with vegetation removed and the soil raked to a depth of 2cm, and plots that were undisturbed, with vegetation and soil only altered as much as necessary to plant the seedlings. Some of the genotypes she uses were collected from the invasive range of this plant (North America, Asia, South America, Africa) and the others were collected from the native range of the plant, around the Mediterranean basin. After six months, Shannon measures the number of fruits produced by each plant. Use Shannon's data, the file *Medicago.csv*, to fit a generalized linear mixed model and test whether the effects of disturbance and collection range (invasive/native) affect fruit production. Also test whether genotypes (a random effect, nested within collection range) produce different numbers of fruits.

Type III Anova on a glmm shows statistically significant effects of **Disturbance** ($p<0.01$), **Range** ($p<0.05$), and their interaction ($p<0.001$) on fruit production. After fitting models with and without the random effect, ANOVA shows that genotype has a statistically significant ($p<0.001$) on fruit production numbers.



(3) (10 pts) Bridgette is a former student in Larry Allen's lab who was interested in courtship behavior of black surfperch. She notices that large males seemed to receive more attention from females. To evaluate the effect of male size on probability of mating, she watches courtship behavior of 50 male surfperch with 50 females. She records the size of each male and whether it was successful (mated) or not after its courtship with the female. Analyze the data in the file *surfperch_mating.csv* to determine whether the likelihood of mating was related to male size. Write a brief statement of results and include a graph illustrating your regression.

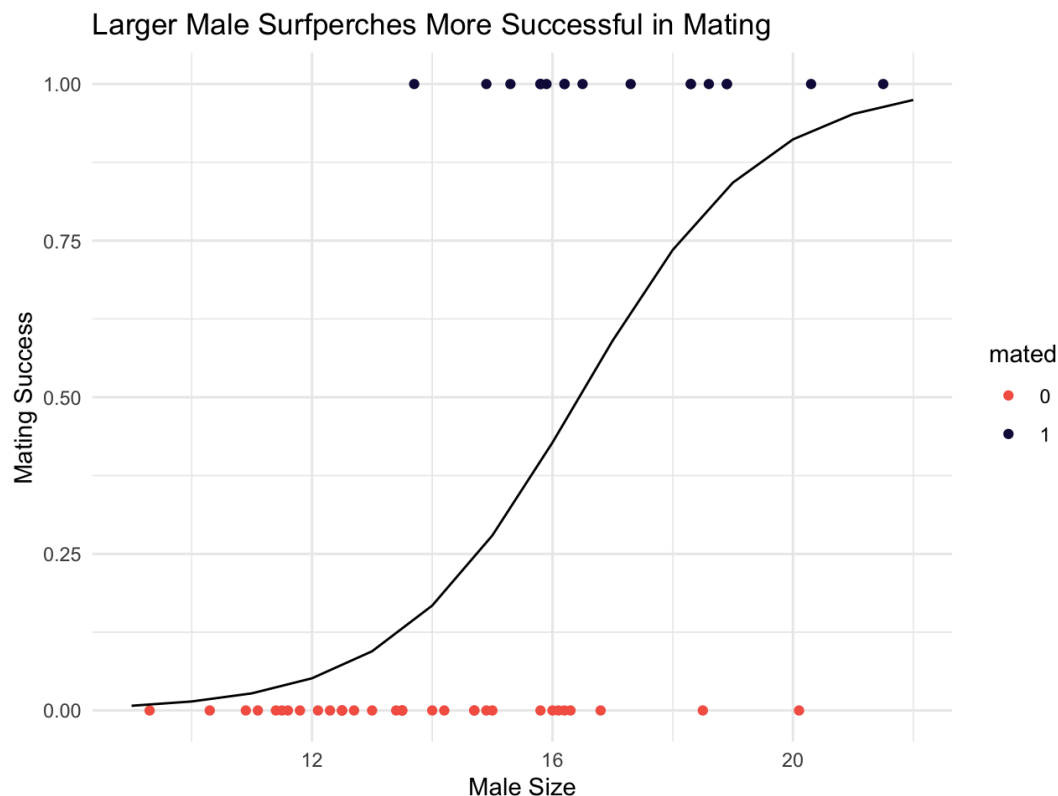
Logistic regression shows us that **male size** ($p<0.001$) has a significant effect on **mating success**. Further R2squared analysis confirms by showing that the model explains a significant amount of predicted power (McFadden=0.3224830). The model is a better fit than the null by 32%.

```

Analysis of Deviance Table (Type III tests)

Response: mated
      LR Chisq Df Pr(>Chisq)
`male size`  20.672 1  5.45e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> pR2(surfmod1)
fitting null model for pseudo-r2
      llh      llhNull      G2      McFadden      r2ML      r2CU
-21.7156206 -32.0517739  20.6723067  0.3224830  0.3386328  0.4686714

```



(4) (5 pts) The sex ratio of non-sex-changing (gonochoristic) fishes is typically near 1:1. The local kelp forest fish, *Oxyjulis californica*, is a gonochore. Mark Steele's lab collected 621 *O. californica* off the coast of Southern California in 2015. Of these, 371 were female and 250 were male. Use a Chi-square test to determine if this species adheres to the expectations of the general model for the sex ratio of gonochoristic fishes?

The species observed deviates from the expected ratio because $p < 0.05$.

```
> X2test <- sum((observed-expected)^2/expected)
> X2test
[1] 23.57649
> 1-pchisq(X2test, df=1)
[1] 1.200453e-06
> |
```

(5) (5 pts) You conduct an experiment to evaluate the mode of inheritance of the vestigial wing mutation in fruit flies. You obtain 161 wild type and 33 mutant offspring from a cross of the F_1 generation. Do the data agree with a 3:1 Mendelian ratio? Does the answer depend on whether you do a Chi-square or G test?

The data with G value and Chi-square **both show deviation from the Medelian ratio** ($p < 0.05$).

```
> #G test
> Gvalue <- 2*sum(observed*log(observed/expected))
> 1-pchisq(Gvalue, df=1)
[1] 0.007364703
> #Chisquared
> X2test <- sum((observed-expected)^2/expected)
> 1-pchisq(X2test, df=1)
[1] 0.01017036
```

(6) (5 pts) Marta is curious about how to cure a hangover. On the day after St. Patrick's Day, she polls a large number of students across campus. First, she asks them if they've been drinking the night before, and only uses the responses of students who say yes. Then she asks two questions: (1) Did you drink water or Gatorade this morning? (2) Do you have a headache right now? The data are as follows:

	Water	Gatorade
No Headache	58	33
Headache	61	21
Total	119	54

Analyze these data to determine if a student's headache depends on what they drank.

Both the Chi-square test alone and within ANOVA show that there is **no significant relationship between headache and the effect of a morning beverage** after a night of drinking ($p > 0.05$).

Analysis of Deviance Table

Model: binomial, link: logit

Response: Headache

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			172	239.36	
Drink	1	2.2963	171	237.06	0.1297

```

      Water Gatorade
No Headache  58      33
Headache     61      21
> chi_square_test <- chisq.test(hanging)
> chi_square_test

Pearson's Chi-squared test with Yates' continuity correction

data: hanging
X-squared = 1.811, df = 1, p-value = 0.1784

```

(7) (10 pts) In the 2020 election, exit pollsters asked 22,457 voters of different race/ethnicities leaving the polls which candidate they voted for. The data for White, Black, and Latino participants are below. Use a log linear analysis to determine whether voting preference is dependent on sex or gender, or whether the effect of gender depends on race. Write a brief statement of your results.

After doing log-linear analysis and ANOVA with Chi-Squared, I found that **voting preference is statistically and significantly affected by sex ($p < 0.05$)**.

The way that sex affects voting is also significantly explained by race ($p < 0.05$).

So, if we were to rank, voting depends on race first and then sex.

Analysis of Deviance Table

Model: poisson, link: log

Response: votes

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			11	20627	
candidate:sex	3	453.8	8	20173	< 2.2e-16 ***
candidate:sex:race	8	20173.2	0	0	< 2.2e-16 ***

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

	Voted for Biden	Voted for Trump
White Men	2588	5177
White Women	3907	5239
Black Men	1007	160
Black Women	1616	69
Latino Men	774	393
Latina Women	1017	510