Stats Revision:

Test:

* Will need to do calculations (need calculator with ln and log functions)
* Timed (1 hour) but anytime from 9-5
* All questions need quite a short answer. Questions are differently weighted.

Questions and Topics:

* Why are all models wrong? 1.3/8
* Stats vs biological significance: 1.3/13, 1.3/16
  + Probability of finding a difference this large or larger is very small, but difference in the mean can be very small. With big sample size, can find these smaller differences to be statistically significance.
  + Need to say how big the difference is.
* Difference between correlation and causation when interpreting LM, GLM, GLMM: 3.3/15
  + Could have strong significance, but might be something else driving the relationship.
  + Need to do an experiment to show causality

Null hypothesis testins:

* Can we prove null hypothesis is true: 1.3/11-13
  + We are trying to reject it (ie no difference or relationship)
  + Difference, slope, or variance = 0 if null hypothesis is true
  + We collect data to show how much evidence there is against the null hypothesis.
  + Can never prove that the null hypothesis is true
* Type I and II error: 1.3/15
  + I = incorrect rejection of true null hypotheses =. False alarm (equal to the p value -> if we find p value of 0.01 we say that in 1 in 100 cases we would find the difference)
  + II = failure to reject a false null hypothesis (1-power). There is a difference but the p value does not detect it as statistically significant. Note enough evidence to reject. Directly related to the statistical power (probability of rejecting false null hypothesis.
    - Greater probability with a small sample size
    - Ignoring alarm

P values

* What is a p value : 1.3/13
  + Probability that this or something more extreme would happen = Prob of finding a result this large or large if null hypothesis were true
  + Can compare it to our significance threshold (0.05 or 0.01)
  + Measure of strength of evidence, not of the effect size
  + Is a result of the effect size and the sample size.
* Give two explanations for p> 0.05 : 1.3/15
  + No relationship
  + Type II error -> Using 95% CI is handy if no stat sig, could get a CI of slope or effect size.

Distributions

* What is the difference between standard normal (0r z) distribution and t distribution: 2.2/12, 2.2/14, 3.2/9
  + Probability density function
  + T distribution = thicker tails if sample size is smaller
    - T value = estimate/ st erro of the estimate
    - Expect a normal distribution. Under null hypothesis, mean value of t = 0.
    - What is the prob of finding a value of 3 if the null hypothesis were true. Given by the area under the curve of +3 and -3. In example would be a small probability. This is our p value
  + Z distribution is t distribution with infinite DF
* X axis of PDF is standard deviation
  + Area under the curve =. How many individuals we expect to have a height/value within a certain number of st devs
* Degrees of freedom = sample size – number of things we estimate from the data.

Describing Data

Standard Devations and Standard Errors:

* Variance = provides measure of how much variaition around the mean. Calculate by devaiation of each data point from the mean. Square it, add all data points for sum of squares. Then divide by sample size – 1.
  + Larger = more variation around the mean. Hard to interpret bc deviations are squared.
  + Units are original^2 ie cm^2
* ST Dev = square root of variance
  + How much variation in a sample?
  + Variation among individuals
* 95% CI = derived from St dev
  + Can be estimated
* Coefficient of variation
  + Divide st dev by mean (2.1/21)
* St error: how precise is the estimate of mean or slope:
  + Estimates in models come with a st error
  + If we redid the model many times, we would get a distribution of slopes. The st error = st dev of the slopes. /

Basic Stats

* Correlation coefficient = no units,
* Regression coefficient = can be any value, expected increase or decrease on y variable if x variable goes up 1 unit. Has units (ie days/year)
  + Implies causal effect of x on y, but does not prove causality

T tests:

* 1 sample = is a mean different from 0? Rare test
* 2 sample = test mean between two groups.
* Paired = pairing in data (measured individuals twice at different time point). First calculates the difference between the measurements for each measurement. Gives a number of differences, then test the mean of all differences against 0. Ends up being a 1 sample t test.

What if you have > 2 groups?

* ANOVA = allows us to test if there is more variation among groups than expected by chance. Get a single p value and says if the groups differ, but not which groups differ or by how much
* After variation is established can do pairwise tests of each of the groups against each other
* In r = aov() or anova(). Latter has more flexibility

P threshold = 0.05 expect 1 in 20 to be significant under null -> leads to problem of multiple testing. We can change our threshold: if we do 20 tests, need to divide the threshold by 20 -> 0.05/20

Choosing the right model:

1. What is the response variable?
2. What is the process that generated the data?
   1. Count data = poisson glm. No need to look diagnostic plots or distribution.
   2. Number of successes from limited trial = binomial glm
   3. Any other continuous = linear model
3. If we fit a lm -> assumptions check
   1. Distribution of residuals need to be normal
   2. Homogeneity of variances
   3. Any points with high leverage (outliers)?
   4. Can use tests like Shapiro.
4. If not met ->
   1. non parametric tests -> But can’t fit multiple predictors
   2. transform the data, if you need it to be lm
5. Poisson or binomial -> variance no longer independent of the mean
   1. Expect mean and variance to increase together. They are dependent on each other.
   2. Often you find more variation than expected based on mean = overdispersion. Gives us too small p values.
   3. Can use a quasipoisson or quasibinomial, estimates variance straight from the data.

Linear Models

Estimates: always need to ask what it tells us (using mouse example)

* Intercept = expected mass of mouse 0 days old and a female. Often outside of the range of the measured samples.
  + Can mean center our continuous predictors to make this more informatice
* Slope
* if we fit quadratic term, always also fit linear term.
* Can fit correlated predictors, (a+b) measure the individual effects of a and b (while you keep the other constant)
* Lm does not assume 0 collinearity

Interpreting model output:

* If there is an effect on x on y, that is not the same as an interaction
* Interaction = testing if two slopes are different from each other.
  + Difference between the two slopes is the interaction estimate.
  + Prob of finding a t value or larger if null hypothesis was true?
  + If slopes look different, need to test if the interaction is significant

Reporting Results:

Check the practicals

Rules:

* First report the estimate if possible. If you have tested for a relationship between two groups, but if more than two groups, no single estimates, would refer to a figure.
  + Tells us if the difference was positive or negative…
  + Provide st error of the estimate.
* Is the estimate different from 0?
  + Need the test statistic (t, f, chi sq). All have a PDF
  + Also need the df.
    - Numerator df = number of parameters estimated by predictor
    - Residual df = sample size – all the things we estimated from the data
  + In t test, numerator is always 1, only resid df vary, so only need to report resid df
  + In f test, we can test multiple levels of predictors. So need to provide numerator and residual degrees of freedom
  + Chi square -> when comparing glm or mixed models. Assumes that our sample size is very large. Doesn’t need residual degrees of freedom (Assumed to be very large). Always difference between two models.