Quiz 4

1 – With a continuous X-variable, the meaning of the beta is the slope, or the amount of change in Y for each 1 unit change in X. What is the meaning of the beta when the X-variable is categorical?

* **Difference between that group and the reference group**
* Slope
* Difference between all the groups
* Average Y for that group

2 – What does the β0​ coefficient represent in a linear model with a categorical X-variable?

* **The average value of Y for the reference group**
* The difference between two groups
* The residual error in the model
* The significance of the X-variable

3 – For an analysis with a categorical X-variable, what is ‘dummy-coding’?

* A technique used to code categorical variables into a single continuous variable.
* **A method to convert a categorical variable with multiple levels into a series of binary (0 or 1) variables, each representing one of the levels.**
* A coding method where all categories are assigned a value of 1.
* A technique that combines multiple categorical variables into one new variable.

4 – If you have a categorical X-variable with three groups (A, B, and C) and you want to test if groups B and C are different using ‘lm()’ in R, what should you do?

* Run ‘lm(Y ~ X)’ and examine the betas
* Add an interaction term and examine the betas
* Run ‘TukeyHSD(lm(Y ~ X))’ and examine the betas
* **Change the reference group to group B, run ‘lm(Y ~ X)’, and examine the betas**

5 – Why are there so many different post-hoc tests? Note: we haven’t talked about Type II error much in this class, but Type II error is failing to reject the null hypothesis when it is false and is related to power.

* None of them has a perfect Type II error rate
* Statisticians need something to do to justify their job
* **They all attempt to balance Type I and Type II error, and none of them is perfect**
* None of them has a perfect Type I error rate

6 – What is the ultimate purpose of conducting ‘post-hoc tests’?

* To determine which groups are significantly different from each other
* To generate estimated differences between all the groups
* **To adjust pairwise p-values to maintain the experiment-wise error rate**
* To statistically test all pairwise comparisons

7 – If a relationship is known to be linear, how does Cottingham suggest you should distribute your treatments in an experiment?

* **Many treatment levels with few replicates**
* All replicates in two treatment levels at the extremes in natural variation
* Many treatment levels with no replicates
* Three treatment levels with replicates equally divided among them

This way, you can evaluate linearity and test for linearity if you want to.

8 – What do Steury and Murray recommend instead? If you know it’s linear, then just stack your replicates at either end of the linear variation and test for the single effect with more replicates in those two groups.