Hypothesis Testing (Part II)

A. Introduction to P-values

P-values are the most common measure of statistical significance. Their ubiquity, along with concern over their interpretation and use makes them controversial among statisticians. The following manuscripts are interesting reads about P-values.

B. What is a P-value?

The central idea of a P-value is to assume that the null hypothesis is true and calculate how unusual it would be to see data (in the form of a test statistic) as extreme as was seen in favor of the alternative hypothesis. The formal definition is:

A P-value is the probability of observing a test statistic as or more extreme in favor of the alternative than was actually obtained, where the probability is calculated assuming that the null hypothesis is true.

A P-value then requires a few steps. 1. Decide on a statistic that evaluates support of the null or alternative hypothesis. 2. Decide on a distribution of that statistic under the null hypothesis (null distribution). 3. Calculate the probability of obtaining a statistic as or more extreme as was observed using the distribution in 2.

The way to interpret P-values is as follows. If the P-value is small, then either H_0 is true and we have observed a rare event or H_0 is false (or possibly the null model is incorrect).

Let's do a quick example. Suppose that you get a t statistic of 2.5 for 15 degrees of freedom testing $H_0: \mu = \mu_0$ versus $H_a: \mu > \mu_0$. What's the probability of getting a t statistic as large as 2.5?

```
pt(2.5, 15, lower.tail = FALSE)
## [1] 0.0122529
```

Therefore, the probability of seeing evidence as extreme or more extreme than that actually obtained under H_0 is 0.0123. So, (assuming our model is correct) either we observed data that was pretty unlikely under the null, or the null hypothesis if false.

C. Binomial P-value example

Suppose a friend has 8 children, 7 of which are girls and none are twins. If each gender has an independent 50 probability for each birth, what's the probability of getting 7 or more girls out of 8 births?

This calculation is a P-value where the statistic is the number of girls and the null distribution is a fair coin flip for each gender. We want to test $H_0: p = 0.5$ versus $H_a: p > 0.5$, where p is the probability of having a girl for each birth.

Recall here's the calculation:

```
pbinom(6, size = 8, prob = 0.5, lower.tail = FALSE)
## [1] 0.03515625
```

Since our P-value is less than 0.05 we would reject at a 5 error rate. Note, however, if we were doing a two sided test, we would have to double the P-value and thus would then fail to reject.