



# P-values

Statistical inference

Brian Caffo, Jeffrey Leek, Roger Peng  
Johns Hopkins Bloomberg School of Public Health

# P-values

- Most common measure of "statistical significance"
- Their ubiquity, along with concern over their interpretation and use makes them controversial among statisticians
  - <http://warnercnr.colostate.edu/~anderson/thompson1.html>
  - Also see *Statistical Evidence: A Likelihood Paradigm* by Richard Royall
  - *Toward Evidence-Based Medical Statistics. 1: The P Value Fallacy* by Steve Goodman
  - The hilariously titled: *The Earth is Round ( $p < .05$ )* by Cohen.
- Some positive comments
  - [simply statistics](#)
  - [normal deviate](#)
  - [Error statistics](#)

# What is a P-value?

**Idea:** Suppose nothing is going on - how unusual is it to see the estimate we got?

**Approach:**

1. Define the hypothetical distribution of a data summary (statistic) when "nothing is going on" (*null hypothesis*)
2. Calculate the summary/statistic with the data we have (*test statistic*)
3. Compare what we calculated to our hypothetical distribution and see if the value is "extreme" (*p-value*)

# P-values

- The P-value is the probability under the null hypothesis of obtaining evidence as extreme or more extreme than would be observed by chance alone
- If the P-value is small, then either  $H_0$  is true and we have observed a rare event or  $H_0$  is false
- In our example the  $T$  statistic was 0.8.
  - What's the probability of getting a  $T$  statistic as large as 0.8?

```
pt(0.8, 15, lower.tail = FALSE)
```

```
## [1] 0.2181
```

- Therefore, the probability of seeing evidence as extreme or more extreme than that actually obtained under  $H_0$  is 0.2181

# The attained significance level

- Our test statistic was 2 for  $H_0 : \mu_0 = 30$  versus  $H_a : \mu > 30$ .
- Notice that we rejected the one sided test when  $\alpha = 0.05$ , would we reject if  $\alpha = 0.01$ , how about 0.001?
- The smallest value for alpha that you still reject the null hypothesis is called the *attained significance level*
- This is equivalent, but philosophically a little different from, the *P-value*

# Notes

- By reporting a P-value the reader can perform the hypothesis test at whatever  $\alpha$  level he or she chooses
- If the P-value is less than  $\alpha$  you reject the null hypothesis
- For two sided hypothesis test, double the smaller of the two one sided hypothesis test Pvalues

# Revisiting an earlier 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?

```
choose(8, 7) * 0.5^8 + choose(8, 8) * 0.5^8
```

```
## [1] 0.03516
```

```
pbinom(6, size = 8, prob = 0.5, lower.tail = FALSE)
```

```
## [1] 0.03516
```

# Poisson example

- Suppose that a hospital has an infection rate of 10 infections per 100 person/days at risk (rate of 0.1) during the last monitoring period.
- Assume that an infection rate of 0.05 is an important benchmark.
- Given the model, could the observed rate being larger than 0.05 be attributed to chance?
- Under  $H_0 : \lambda = 0.05$  so that  $\lambda_0 100 = 5$
- Consider  $H_a : \lambda > 0.05$ .

```
ppois(9, 5, lower.tail = FALSE)
```

```
## [1] 0.03183
```