

# Cogsci 109

Data Analysis and Computational Modeling

Virginia de Sa  
desa at cogsci

# Descriptive Statistics

Useful summary statistics for your data

**Sample Mean**

**Median**

**Mode**

**Range:** max - min

**Mean or Average Deviation:**

$$\sum abs(X - \bar{X})$$

# Hypothesis Testing

(after <http://www.uwsp.edu/psych/stat/9/hyptestd.htm>)

Consider a binary random variable (can only take 2 values) e.g. flipping a coin

Question: You flip a coin and get 10 Heads and 7 tails. Is the coin normal (50/50 Heads and tails)?

How do we answer this question?

Coin flips follow a binomial distribution

$$P(n \text{ successes in } N) = \binom{N}{n} p^n (1 - p)^{N-n}$$

# Hypothesis Testing

Pick a **significance level**  $\alpha$  (How conservative do you want to be? – How often can you afford to be wrong 1/20 is common (.05))

Hypotheses:

**Null Hypothesis** ( $H_0$ ): the coin is fair **Alternate Hypothesis** ( $H_1$ ): the coin is funny

Assume the null hypothesis

$$P(n \text{ successes in } N) = \binom{N}{n} p^n (1 - p)^{N-n}$$

binomial applet

Now we see where our particular value falls on the distribution for the Null Hypothesis.

## Another Alternate Hypothesis

Now let's consider Alternate Hypothesis ( $H_2$ ): the coin is biased towards heads

This is the difference between doing a one-tailed or a two-tailed test

NOTE: We don't prove the null hypothesis we either

- reject the null and assert the alternate
- or fail to reject the null

Type I Error – Reject the Null hypothesis when it is actually true

Type II Error – Don't reject the Null when it is not true

What is the P(type 1 error)?

What is the P(type 2 error)?

## Another Alternate Hypothesis

Now let's consider Alternate Hypothesis ( $H_2$ ): the coin is biased towards heads

This is the difference between doing a one-tailed or a two-tailed test

NOTE: We don't prove the null hypothesis we either

- reject the null and assert the alternate
- or fail to reject the null

Type I Error – Reject the Null hypothesis when it is actually true

Type II Error – Don't reject the Null when it is not true

What is the P(type 1 error)?  $\alpha$

What is the P(type 2 error)?

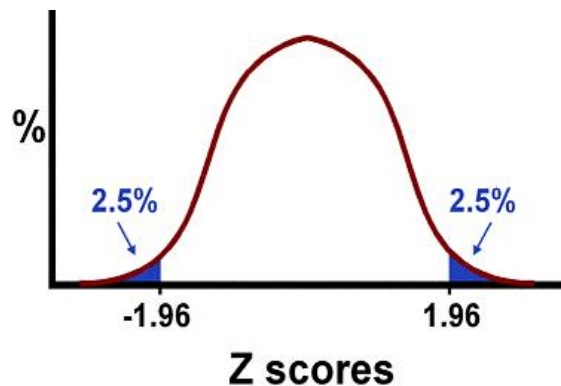
**Power** of a test is (1- the Probability of committing a Type II error)

## Example 2: Z-test

Z-test assumes that the data is from a Gaussian distribution with known variance.

Question: You know that humans have IQs that are normally distributed with mean 100 and standard deviation 15. You have an individual of unknown species and you want to predict whether they are human or not based on their IQ. The null hypothesis is that their IQ is drawn from the human distribution.

Pick a significance level



## T-test: when you don't know the variance/standard deviation

In this case your estimate of variance ( $\sigma^2$ ) is off Using the sample variance ( $s^2$ )

$$\frac{Y - \mu}{s}$$

is distributed as a t-distribution

T-distributions have fatter tails but get more normal as degrees of freedom approach infinity



# Confidence interval

Confidence interval for a mean can be given as

$$\bar{X} \pm t \sqrt{\frac{s^2}{n}}$$

where  $t$  is the  $(1 - \alpha/2)$  quantile of Student's T distribution

# Mann-Whitney U test AKA Wilcoxon rank-sum test

The T-test uses as its null hypothesis that the two sets of samples are drawn from the same Gaussian distribution with unknown variance.

If the Gaussian assumption is not true, but you want to compare a difference in central tendency, you can use the Mann-Whitney U test

In matlab use `ranksum` (and `signrank` to test for zero median)

# Bootstrap

Basic idea: In order to get confidence intervals or standard errors of statistics, we would love to be able to resample many times from the distribution from which the data were sampled (the real distribution) and compute the statistic for each sample.

In bootstrap methods we replace the real distribution with the **empirical distribution**—that obtained by placing  $1/n$  probability at each sample point.

**bootstrap sample of size n:** random sample of  $n$  observations *with replacement* from the empirical distribution

# Matlab's bootstrp function

`bootstrp(N,bootfun,data)` generates bootstrap samples of size N from data and applies bootfun to them.

```
>> bootstrp(10,'mean', [2 4 4 5 6 ])
```

```
ans =
```

```
5.0000  
4.2000  
4.6000  
5.2000  
3.2000  
3.4000  
4.2000  
4.6000  
4.6000  
3.8000
```

**mean of a bootstrap sample** is written  $\bar{X}^*$

$$\bar{X}^* = \frac{1}{n} \sum X_i^*$$

where  $X_i^*$  is the  $i$ th member of a bootstrap sample of size  $n$

# Percentile Bootstrap Method

Generates  $B$  bootstrap samples of size  $n$ , compute the statistic (e.g. sample mean) on them, then estimate a  $1 - \alpha$  conf interval as the range that includes  $(1 - \alpha)\%$  of the values computed on the bootstrap samples

(e.g. for mean)

$$(\bar{X}^*_{(l+1)}, \bar{X}^*_{(u)})$$

$l = \alpha B/2$  rounded to the nearest integer  $u = B - l$

Note : The percentile bootstrap method does not work well for the sample mean

There are many other fancier methods that are beyond the scope of this course.

# Multiple Comparison Issues

If you test 100 things at the .05 level, how many “significant” findings do you expect?

How many voxels are recorded in fMRI?

Bonferonni correction <http://mathworld.wolfram.com/BonferroniCorrection.html> is very conservative – If you want an  $\alpha$  value of .05 and you are doing 100 tests, you must perform each test with an  $\alpha$  of .05/100.

This is too conservative for fMRI as it assumes that all the voxels are independent.  
<http://imaging.mrc-cbu.cam.ac.uk/imaging/PrinciplesMultipleComparisons>