Hypothesis testing

Is this a fair game? What if the coin isn't fair?

Consider the outcome of several games. If confidence interval doesn't overlap with 0, you might be suspicious..

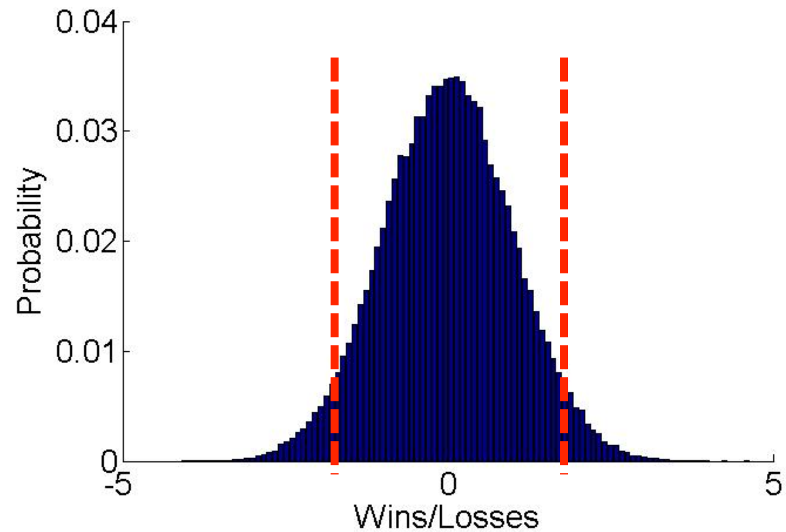
But how do you reject the hypothesis that the coin (and betting game) is far?

Null hypothesis: Game is fair

Alternative hypothesis: Game is not fair

Select cut-off such that the probability of incorrectly rejecting the null hypothesis is < 0.05

Null distribution

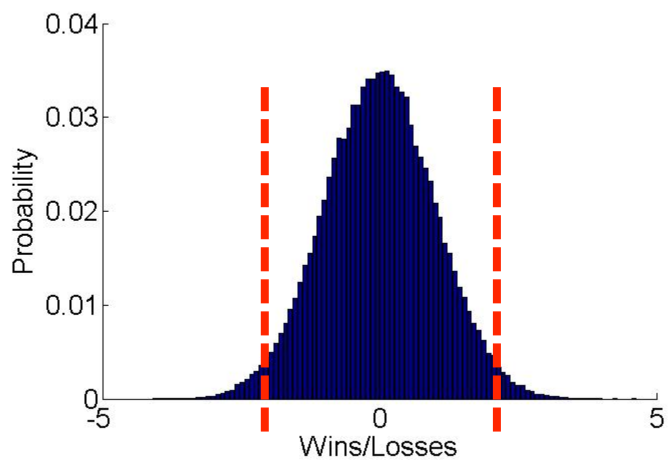


Hypothesis testing

STEPS:

1. **Formulate the null hypothesis**
2. **Select a statistical test**
3. Identify the test statistic under the null hypothesis
4. Determine rejection region of the test statistic
5. Calculate test statistic
6. Determine whether or not the test statistic falls within the rejection region

One MATLAB command!



Hypothesis testing

STEPS:

1. Formulate the null hypothesis:

2. **Select a statistical test:**

3. Identify the test statistic under the null hypothesis

4. Determine rejection region of the test statistic

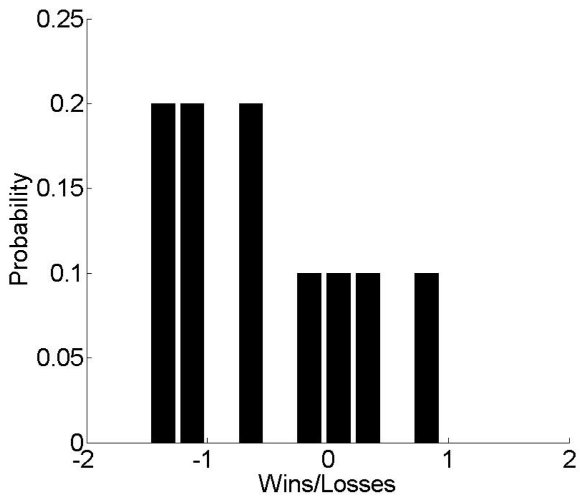
5. Calculate test statistic

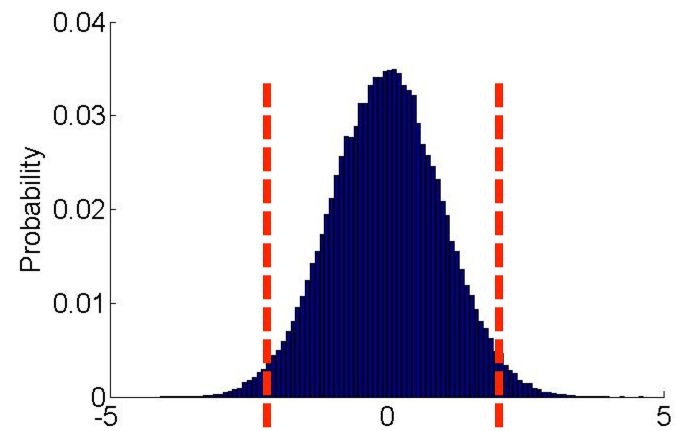
6. Determine whether or not the test statistic falls within the rejection region

H0: mean = 0; H1: mean ≠ 0

T-test

$$t = \frac{\hat{\mu} - \mu}{\hat{\sigma}/\sqrt{n}}$$





MATLAB: ttest(data)

Hypothesis testing

Handout on website:

What you're testing	Assumptions	Test Name	Hypotheses	Notes
Something about μ	Normal distribution Variance known	z-test	H_0: $\mu_{test} = \mu_{null}$ H_1: $\mu_{test} \neq \mu_{null}$ or $\mu_{test} > \mu_{null}$ or $\mu_{test} > \mu_{null}$	Use a two-tailed t-test if H_1 is an inequality. Use a one-tailed test if H_1 is directional (i.e. you have previous evidence which lets you rule out one direction).
Something about μ	Normal distribution Variance unknown	t-test	H_0: $\mu_{test} = \mu_{null}$ H_1: $\mu_{test} \neq \mu_{null}$ or $\mu_{test} > \mu_{null}$ or $\mu_{test} > \mu_{null}$	Use a two-tailed t-test if H_1 is an inequality. Use a one-tailed test if H_1 is directional (i.e. you have previous evidence which lets you rule out one direction). # degrees of freedom = n-1
How well does an observed frequency distribution of certain events fit the distribution predicted by the null hypothesis?	*n* is larger than 30. Expected frequencies are all larger than 5. Error is normally distributed.	Pearson's chi-square test	H_0: observed distribution comes from predicted distribution H_1: observed distribution doesn't come from predicted distribution	# degrees of freedom = # categories – 1
Are all of the means from several distributions the same?	Everything is normal and has equal variance	ANOVA	H_0: all μ's are equal H_1: at least one pair of μ's are not equal (but DOESN'T tell you which one)	Based on idea that observed value = average + variation within group + variation across groups. Null hypothesis is that variation across groups is 0 Degrees of freedom within groups = sample size within a group -1 Degrees of freedom between groups = # groups -1

What if the data aren't normally distributed?

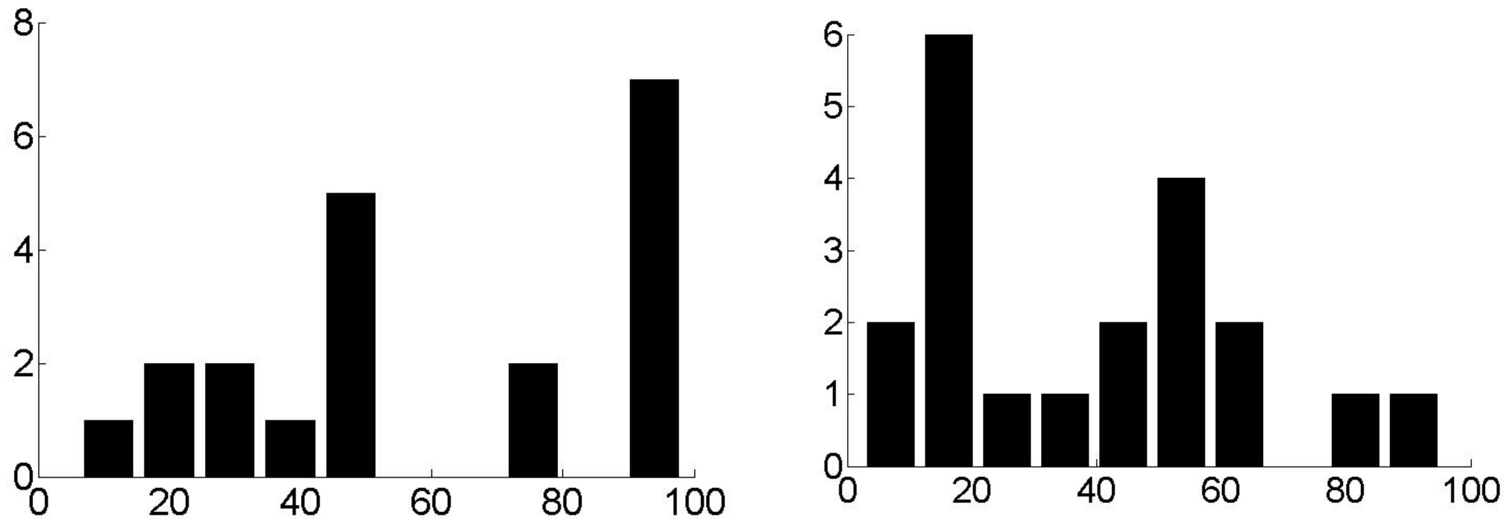
Parametric tests: rely on strong assumptions about shape of underlying distribution (like normality)

Nonparametric tests: rely on fewer assumptions

Analysis	Parametric procedure	Nonparametric procedure
Compare means between groups	Two-sample t-test	Wilcoxon rank-sum test
Compare paired means between groups	Paired t-test	Wilcoxon signed-rank test
Compare means between 3+ groups	ANOVA	Kruskal-Wallis test
Estimate degree of association between variables	Pearson coefficient of correlation	Spearman's rank correlation

http://www.mayo.edu/mayo-edu-docs/center-for-translational-science-activities-documents/berd-5-6.pdf

What if the data aren't normally distributed?



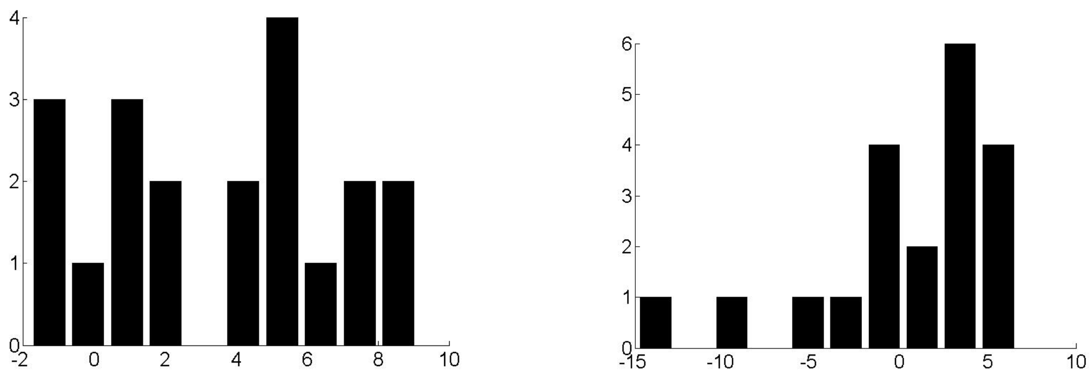
P-value from t-test: **0.02**
P-value from rank-sum test: **0.06**

Actually drawn from same, uniform distribution: can get false positives!

Why don’t we always use nonparametric tests?

Two reasons:

1. Nonparametric tests are less powerful



Samples were drawn from different distributions

P-value from t-test: **0.03**

P-value from rank-sum test: **0.07**

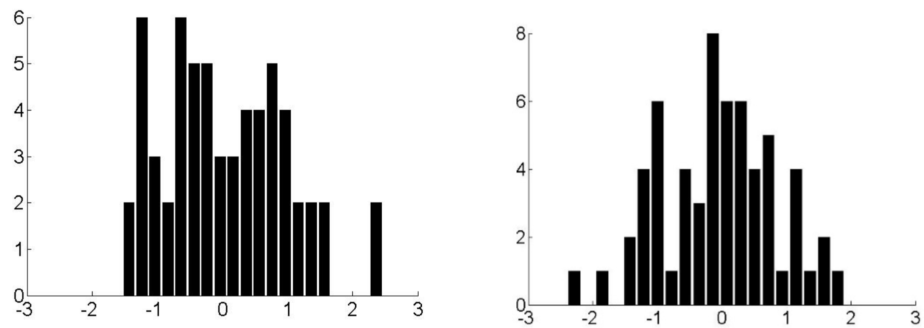
2. Based on things that are harder to interpret: medians versus means

http://www.mayo.edu/mayo-edu-docs/center-for-translational-science-activities-documents/berd-5-6.pdf

More nonparametric tests…

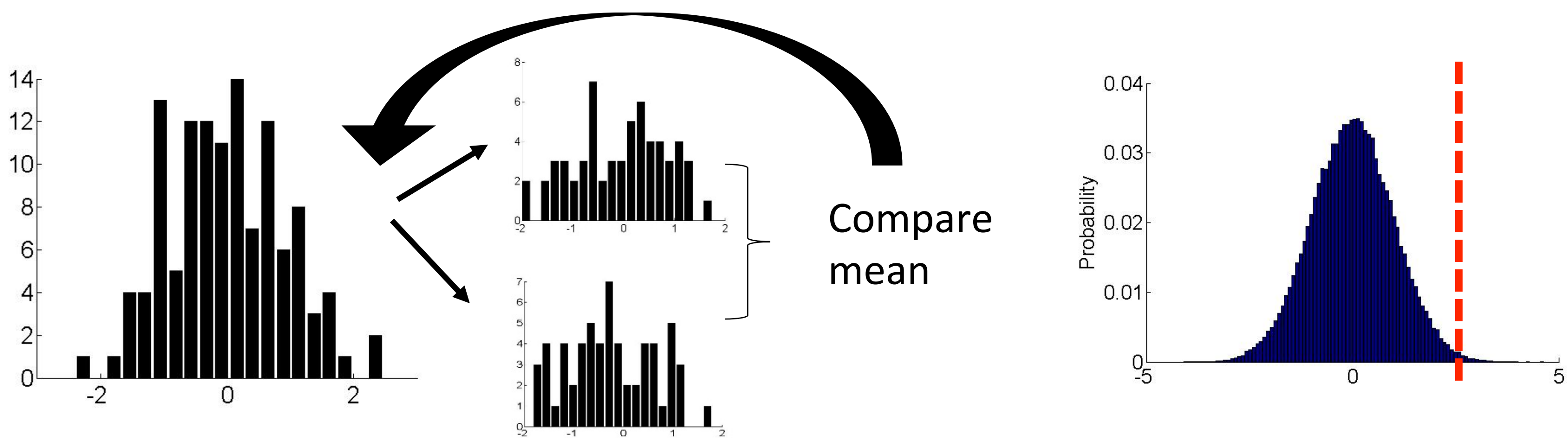
Randomization (or permutation) tests:

Consider two sets of data:



Null hypothesis: two sets of data from same distribution (or have the same mean)

If we aggregate data, randomly split it, and then compare the differences in the mean, we will get a null distribution of mean differences that we can compare to

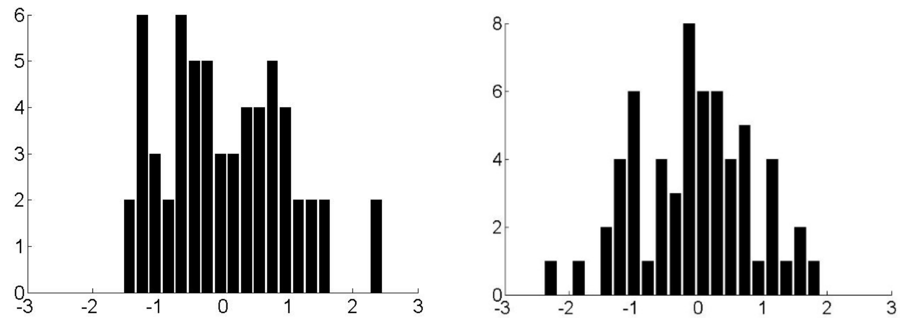


http://white.stanford.edu/~knk/Psych216A/Psych216ALecture2.pdf

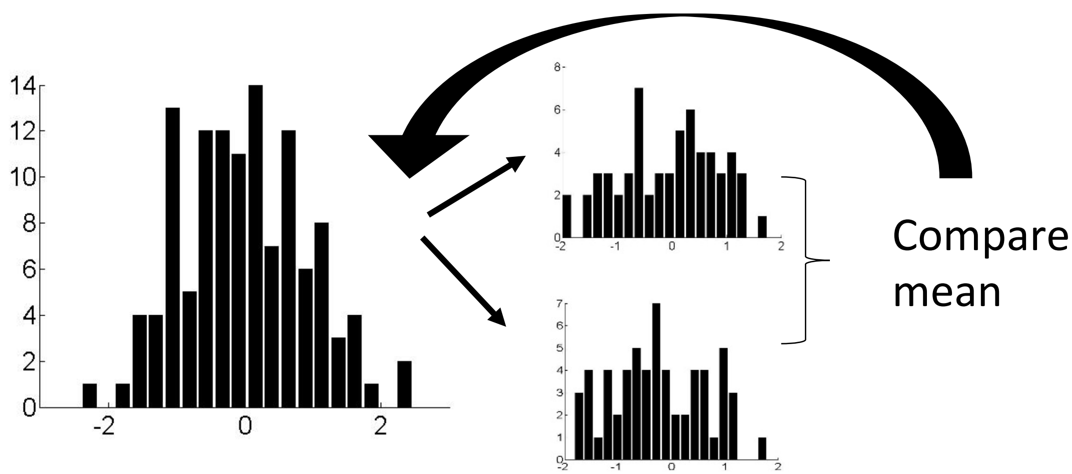
More nonparametric tests…

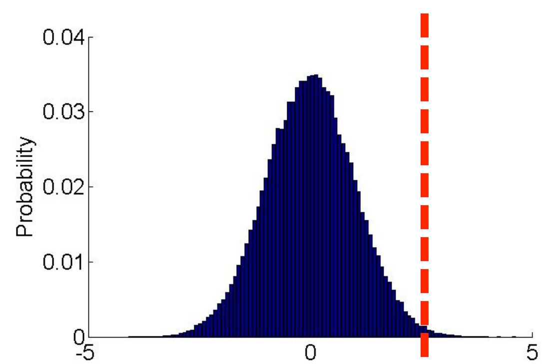
Bootstrapping tests:

Same data sets:



Same procedure, but sample **with replacement**!



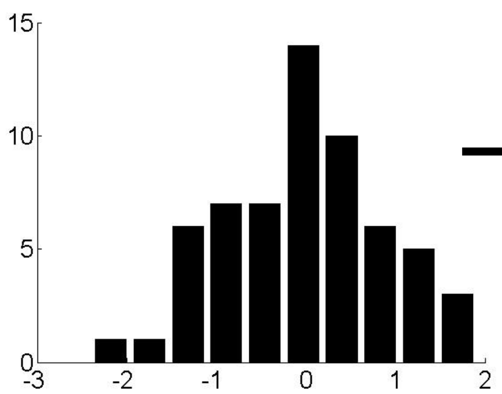


More nonparametric tests…

More with bootstrapping:

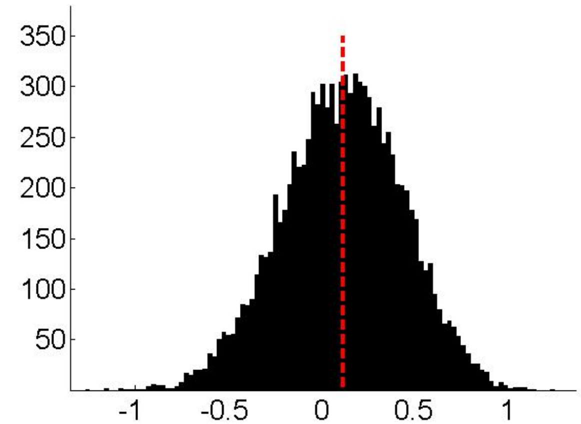
Estimate distribution of the mean:

Sample population:



Resample and compute mean 10,000 times:

Confidence interval on mean:



Can use this method to see if the mean is sufficiently close to 0

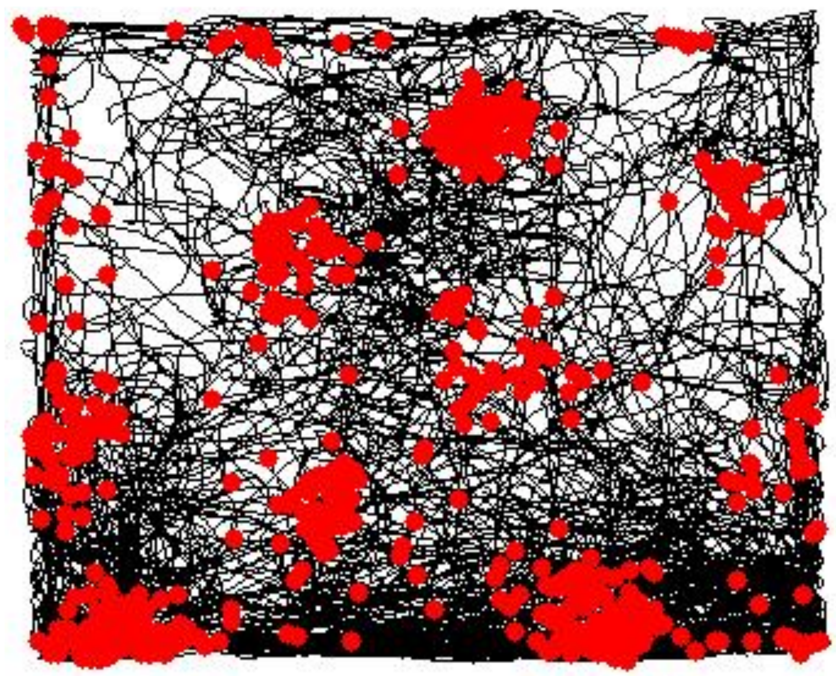
Can also compute t-distribution by computing t-statistic on every resampled population

http://www.stat.rutgers.edu/home/mxie/rcpapers/bootstrap.pdf

More nonparametric tests…

Shuffling tests:

How likely is this to occur by chance?



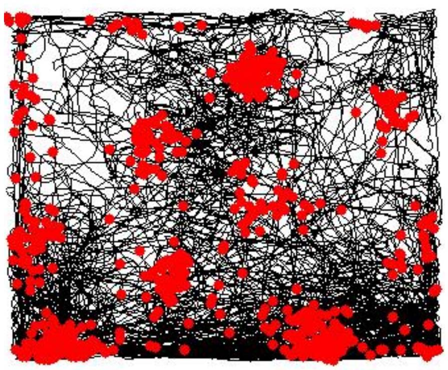
Ask this question while preserving spiking dynamics and behavior of the animal

Langston et al., 2010. Science.

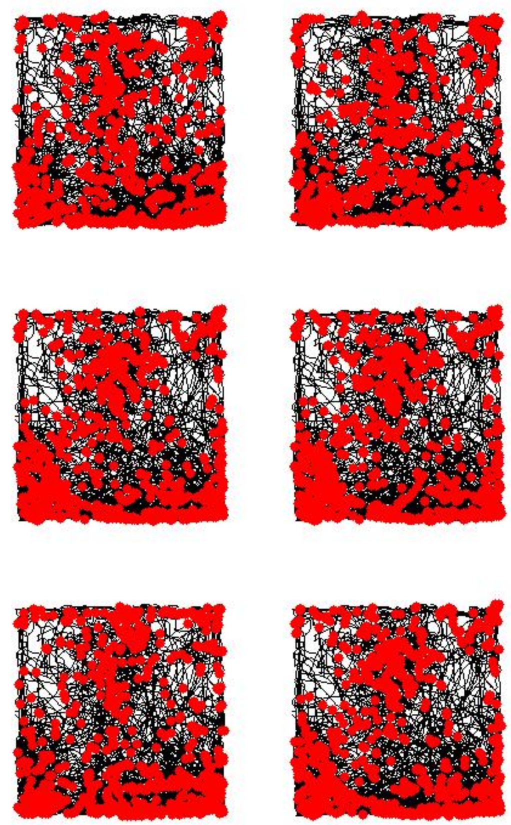
More nonparametric tests…

Shuffling tests:

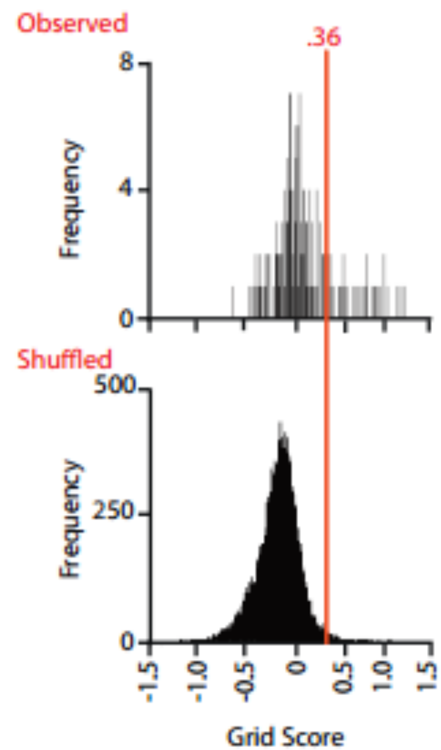
Quantify the “grid score”: degree of 60 degree symmetry



For all cells in dataset, randomly shift spike train along the path of the animal and re-compute the grid score



Analyze null distribution of grid scores



Giocomo et al., 2011. Cell.

So far, we have only dealt with testing one hypothesis at a time…

… but what if you have multiple?

Multiple comparisons

Efficacy of a new drug is related to the reduction of any one of a number of disease symptoms compared to an old drug

As the # of symptoms increases, the drug will appear to be an improvement because at least some symptoms will be reduced

With 5% chance of incorrectly rejecting the null hypothesis if it is really true, with 100 symptoms the chance of at least one incorrect rejection is **99.4%**

Look for an effect in individual mice




As the number of mice increases, so does the chance of finding a false positive

http://en.wikipedia.org/wiki/Multiple_comparisons_problem

Multiple comparisons: what to do?

Easiest thing to do:

Bonferroni correction! Divide significance level by the number of tests

Divide by number of symptoms

Divide by number of mice

ANOVA: analysis of variance

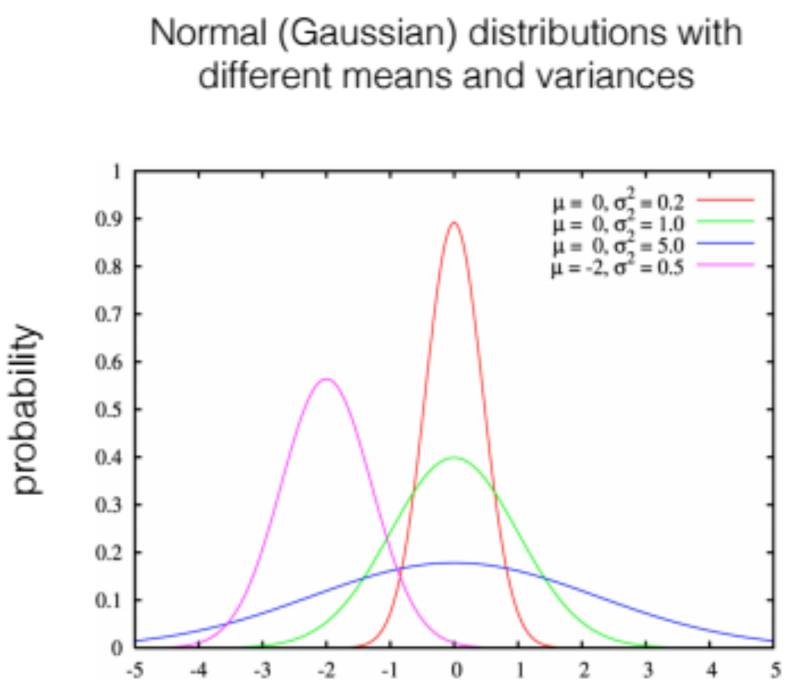
1-factor ANOVA: are all group means the same?

- Compare the variance of data across groups with the variance of data within groups
- If the across-group variance is sufficiently larger than the within group variance, the null hypothesis is rejected
- Must do post-hoc tests to determine which pair differs

So far, we have assumed that the samples within a group are independent…

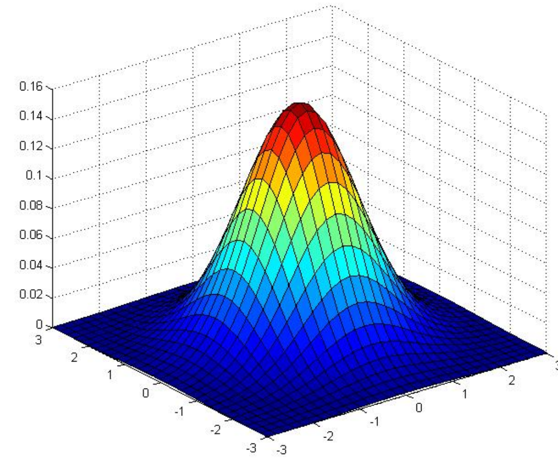
… but what if they aren't?

Probability distribution for 1 variable:

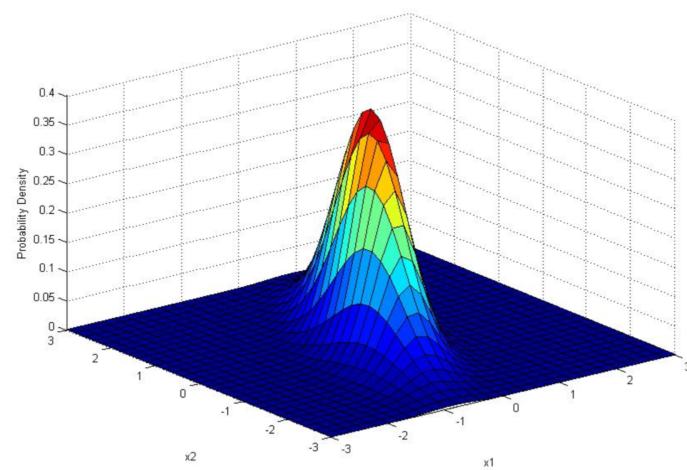


Probability distribution for 2 variables:

X and Y are **independent**:



X and Y are **dependent**:
If you know Y, that will change the probability distribution for X



Formalize independence and dependence

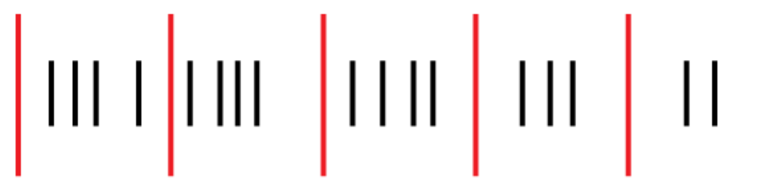
Conditional probability: $$\mathrm{P}(X|Y) = \frac{\mathrm{P}(X \cap Y)}{\mathrm{P}(Y)}$$

Events/variables are independent if: $$\mathrm{P}(X \cap Y) = \mathrm{P}(X)\mathrm{P}(Y)$$

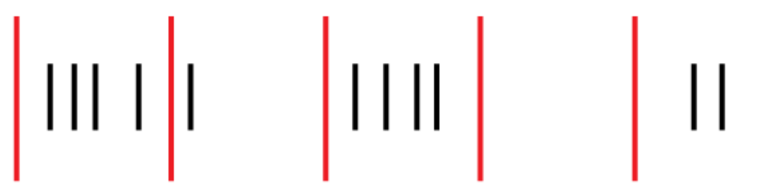
$$\mathrm{P}(X|Y) = \mathrm{P}(X)$$

Examples:

Independent spike trains:



Dependent spike trains:



Dependent variables: covariance and correlation

Describe linear relationship between two variables

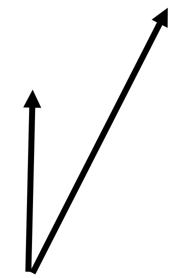
Covariance: how much two random variables change together

$$\sigma(X,Y) = \mathrm{E}[(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))]$$

Can be rewritten as: $\sigma(X,Y) = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$

Uncorrelated if: $\mathrm{E}[XY] = \mathrm{E}[X]\mathrm{E}[Y]$

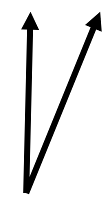
Related to dot product:



Correlation: Normalized version of covariance:

$$r = \frac{\sigma(X,Y)}{std(X)std(Y)}$$

Compare unit vectors:

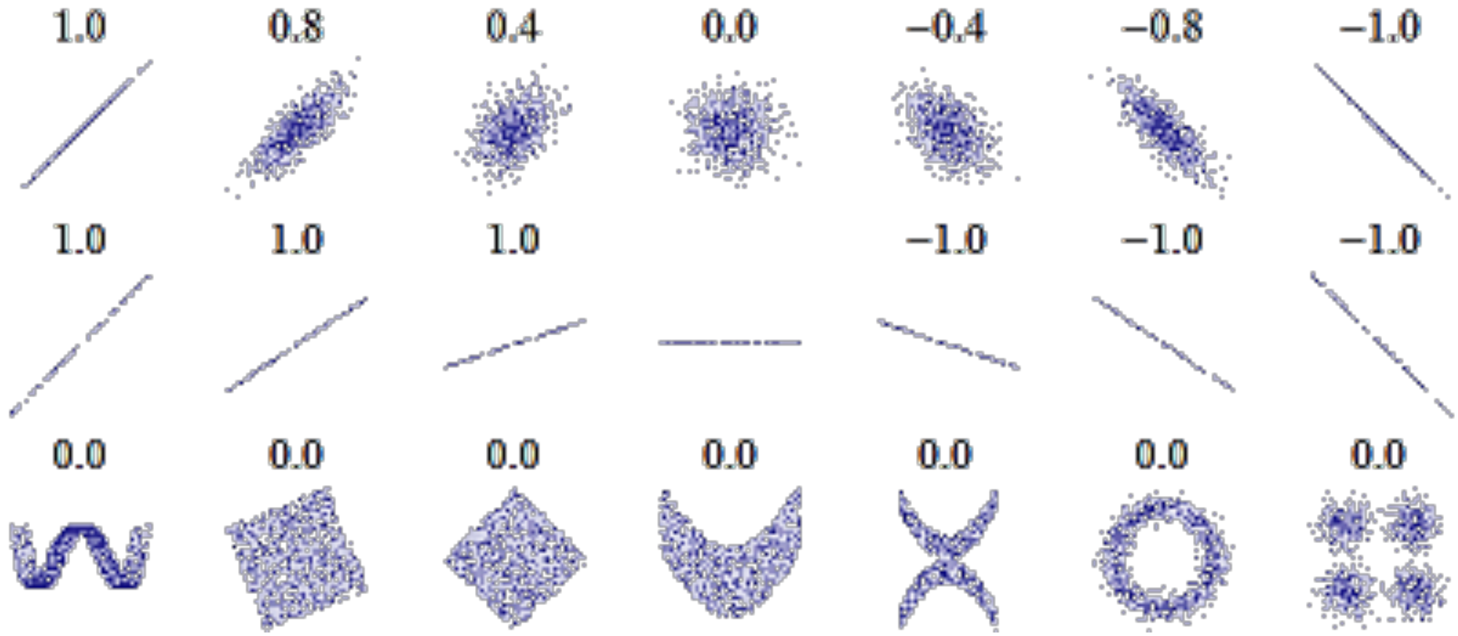


Another way to say it: z-score each variable (subtract off mean and normalize by the standard deviation) and then compare their means

R-value goes between -1 and 1
R-squared goes between 0 and 1, gives fraction of variance explained

MATLAB: cov(data)
corrcoeff(data)

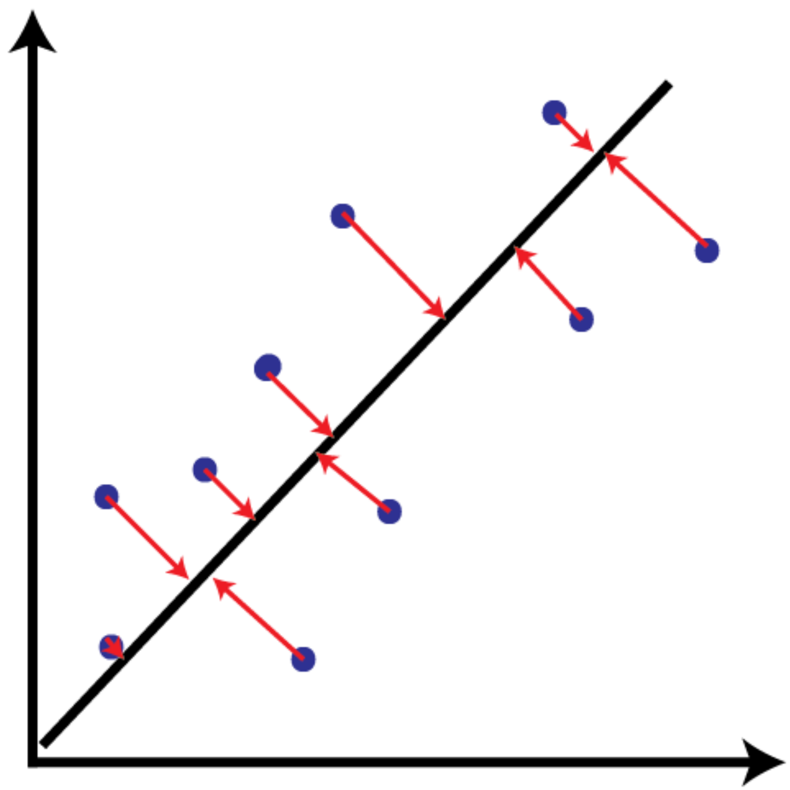
Correlation coefficients



If your data looks like the bottom row, use another method!
One option: mutual information (next lecture!)

Fitting lines to correlated variables

Linear regression through the method of least-squares:



Use correlation coefficient for strength of association

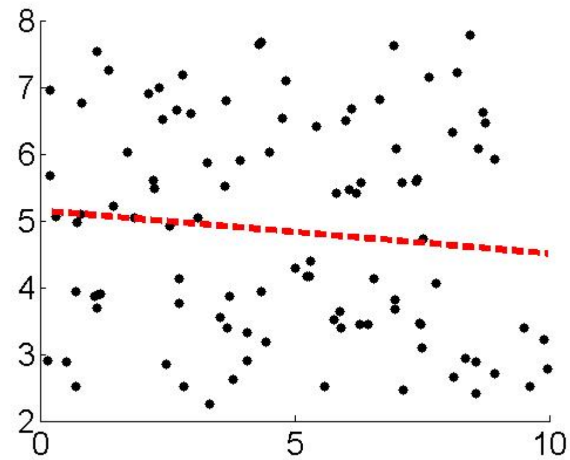
Use slope of regression line when considering X as a predictor for Y

Nonlinear - fit something other than a line

MATLAB: polyfit(data(:,1),data(:,2),1))

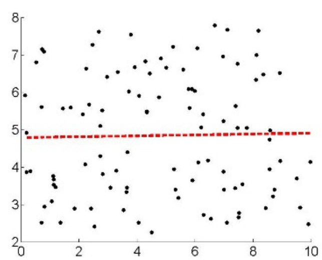
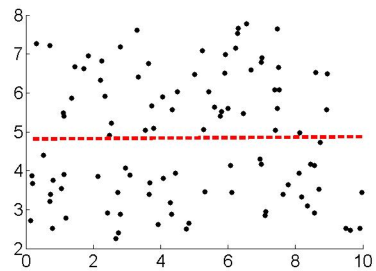
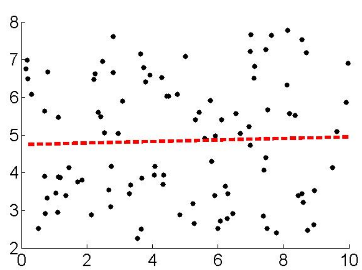
Shuffling + linear regression

Is this slope significantly different from 0?



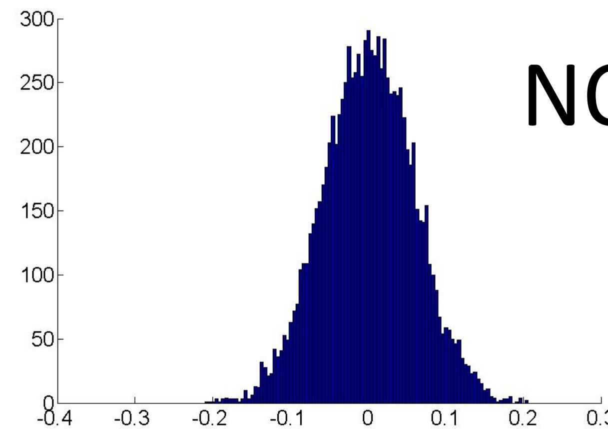
Analyze distribution of slope values!

Create null distribution of slopes by shuffling points along x-axis:

X 10,000

Null distribution

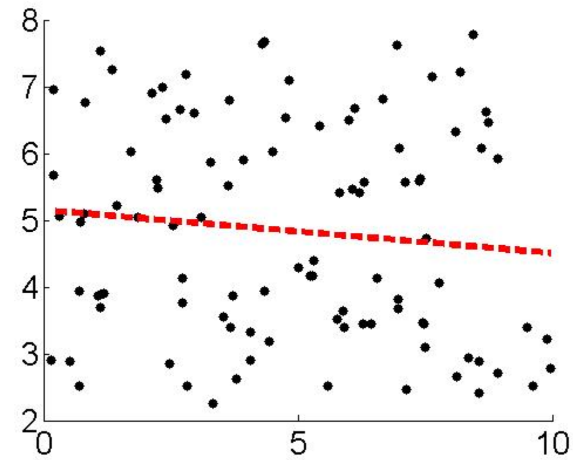


NO

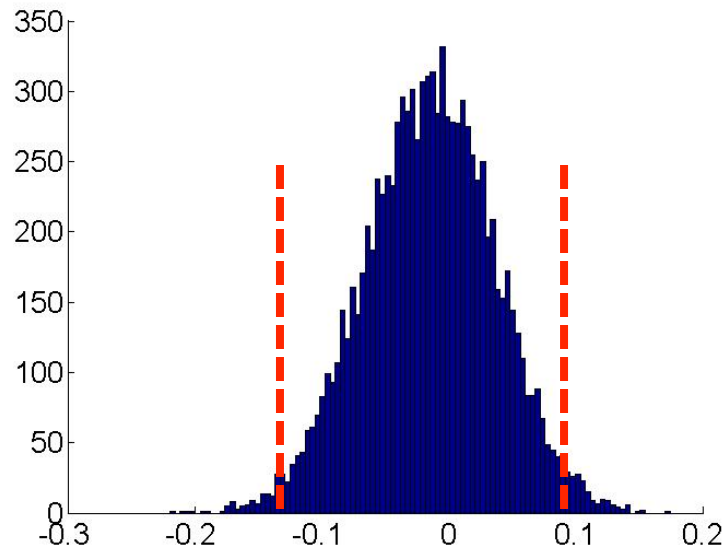
MATLAB: randperm(numel(data(:,1)))

Bootstrap + linear regression

Obtain a confidence interval on the slope



Resample from population and calculate the slope for each bootstrap iteration:



MATLAB: datasample(data(:,1),n)