Today

- HW6 questions?
- Next reading presentation: Friday (R25)
- Statistical methods

Inferential statistics

Inferential statistics:

- Statistical tests you apply to quantitative data to determine the likelihood that the results you observe are due to chance, or conversely, whether they are *statistically significant*, meaning that they are not likely to have occurred due to chance.

Is it significant?

- You mostly look at the *p* statistic, which tells you the probability that the distribution (actually, the distribution or any less likely distribution) is due to chance. If p is less than 0.05 (i.e. 1 in 20 odds that it was chance), this indicates that the distribution is unlikely to have been produced by chance, and is usually taken as a significant result.

Inferential statistics

We will look at three classes of test (if time permits):

(1) t-tests and ANOVA
(2) regression
(3) chi-square and Fisher's exact tests

To pick a statistical method, you need to minimally know:

- your *dependent variable(s)*
- your *independent variable(s)*
- whether each of these is treated as *continuous* or *categorical*
- how many of each type of variable do you have,
- how many categories (levels) in each categorical variable.

t-test and ANOVA

t-test and ANOVA (ANalysis Of VAriance) are commonly used

- They tell you if there is a significant difference in means of a continuous dependent variable given the different levels of the categorical independent variable(s)
- Use them when you have
  - continuous dependent variable(s)
  - categorical independent variable(s)
- E.g.
  - Dependent variable is reaction time (continuous)
  - Independent variable is whether language is metaphorical or literal (categorical)

t-test and ANOVA

- With > 1 independent variable or an independent variable with > 2 levels perform ANOVA http://www.physics.csbsju.edu/stats/anova.html
- Otherwise, use a t-test. http://www.physics.csbsju.edu/stats/t-test.html
- You're looking for the p value once again.
- For any analysis in which at least one independent variable is within-subjects or items, use a Paired t-test or Repeated-Measures ANOVA.
- In other cases, use an unpaired t-test or Univariate (Factorial) ANOVA.

t-test and ANOVA

	1 Indep. Var.		2+ Indep. Vars.
Design	**2 levels**	**3+ levels**	
Within	Paired t-test	Repeated Measures ANOVA	
Between	Unpaired t-test	ANOVA	

t-test and ANOVA

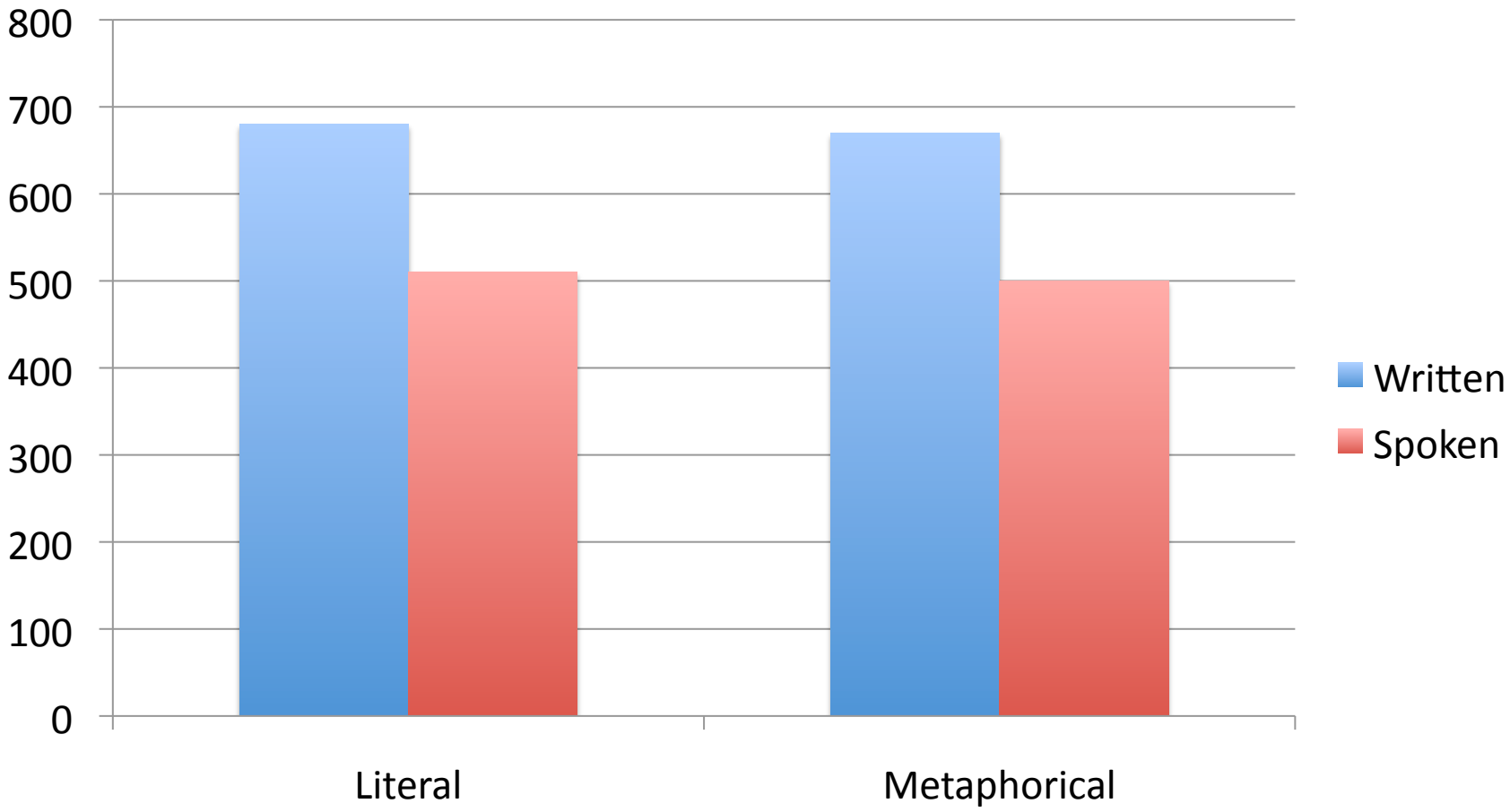
Analyses by participants and items

- For a participants analysis
  - calculate the mean values per condition per ppt
  - then run statistics (e.g. t-test, ANOVA) on these data
  - results tell you whether the outcome is significant across the sample of participants
- For an items analysis
  - calculate the mean values per condition per item
  - then run your statistical test on these data
  - tells you if outcome is significant across the sample of items
- You might need to use different tests for items and ppts analyses of the same data

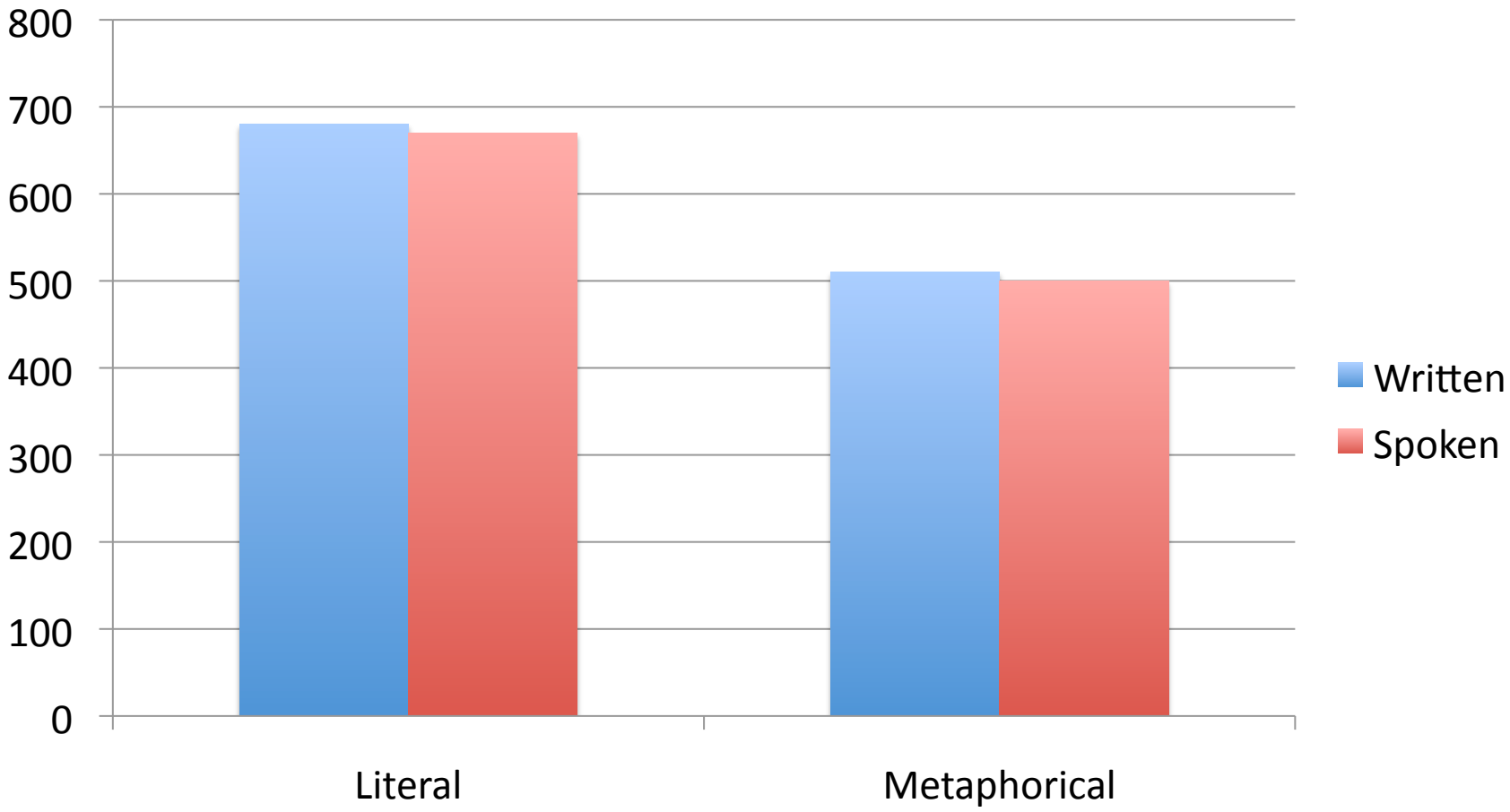
t-test and ANOVA

- With ANOVA, you can measure different types of effect
  - *Main* effects are effects produced by a single independent variable.
  - *Interaction* effects are effects produced by the interaction of independent variables

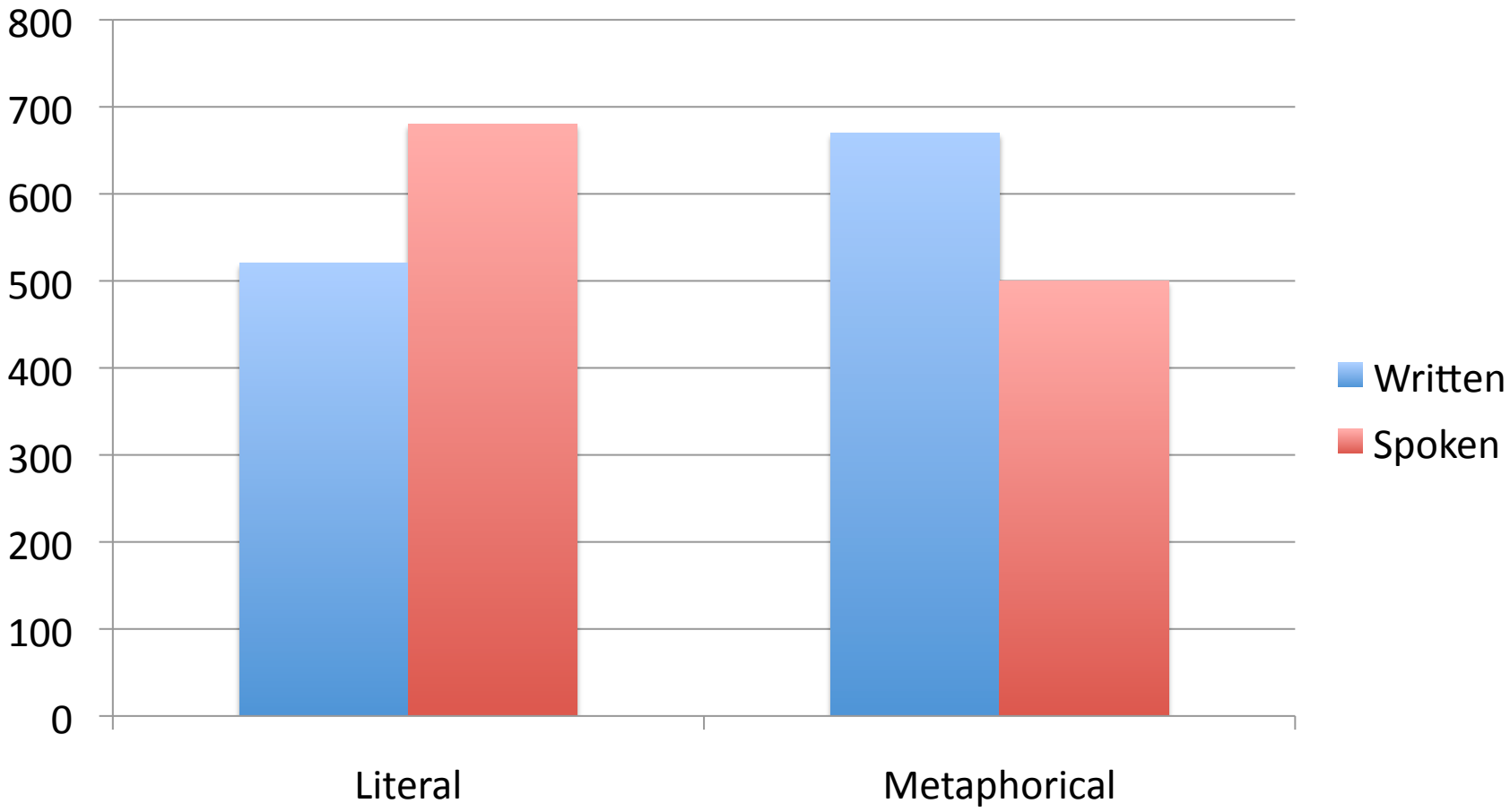
Main effect


800
700
600
500
400
300
200
100
0
Literal
Metaphorical
Written
Spoken

Main effect


800
700
600
500
400
300
200
100
0
Literal
Metaphorical
Written
Spoken

Interaction effect



Regression

- Use regression when the independent and dependent variables are continuous

- Linear regression attempts to explain the relationship between these two variables with a straight line fit to the data

- To get an intuitive idea of how regression works: http://www.mste.uiuc.edu/activity/regression/

Regression

- To perform regression, use any statistics program, or this: http://faculty.vassar.edu/lowry/VassarStats.html

- You're interested in the significance of the r^2 value, and the p value of the accompanying t-test

χ^2 or Fisher's Exact

χ^2 or Fisher's Exact

- Use when you have
  - a single categorical dependent variable
  - a single categorical independent variable
  - one observation per ppt.
  - average of ≥ 5 observations per cell (minimum)
- You're asking whether there is a significant difference in the category distributions of the dependent variable, across the levels of the independent variable.
- E.g. when people are drinking or searching (categorical independent variable), do they pick ”happiness" or ”joy"? (categorical dependent variable)?

χ^2 or Fisher's Exact

Use Fisher's exact test
http://www.quantitativeskills.com/sisa/statistics/fisher.htm

- if you have exactly two levels (categories) in both variables.
- You're interested in the p statistic for the two-tail p-value.
- It doesn't matter if you put the independent and dependent variables in rows or columns, just be consistent.

χ^2 or Fisher's Exact

Use chi-square
http://people.ku.edu/~preacher/chisq/chisq.htm

- if you have more than two levels in either variable.
- You're interested in the p-value