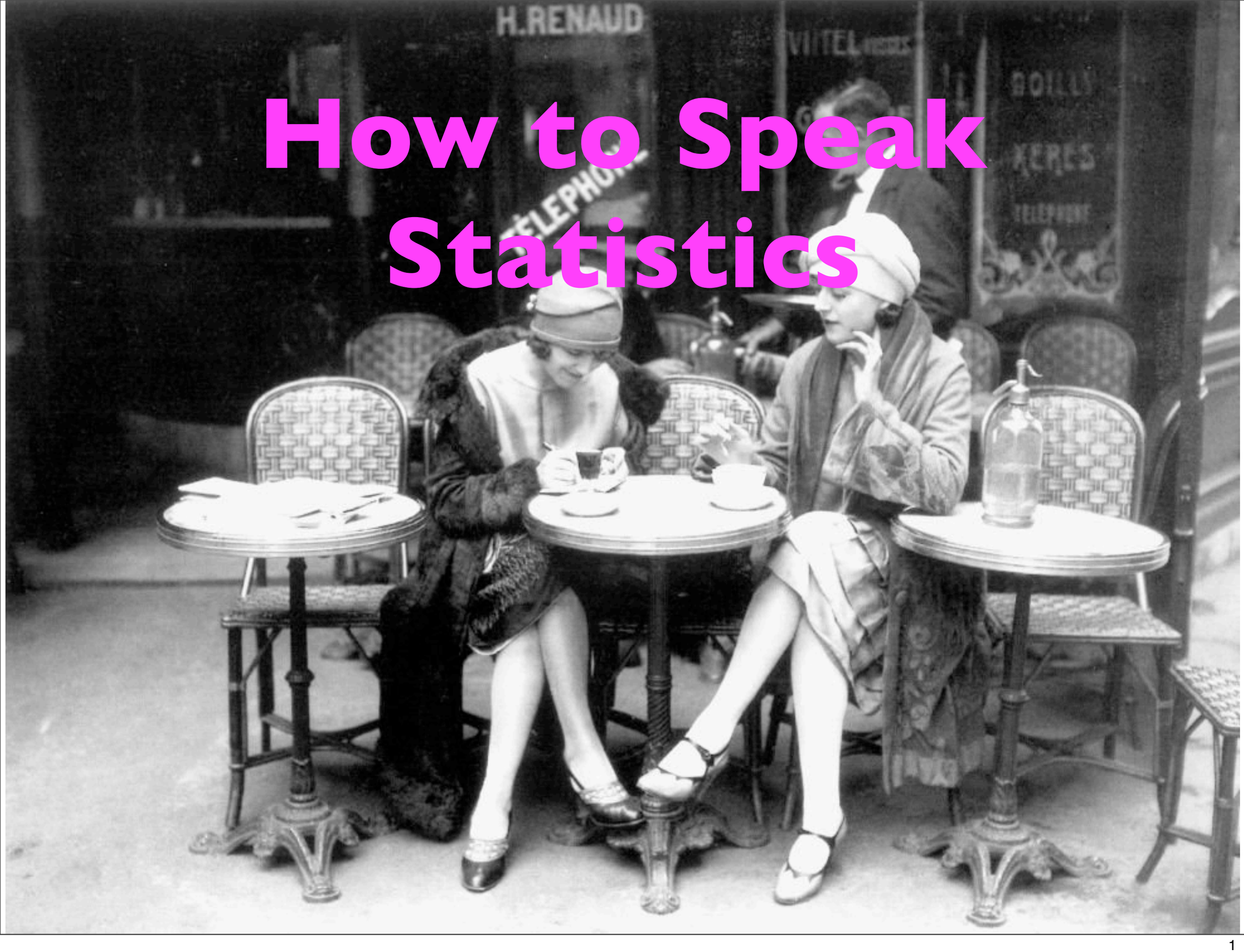
How to Speak Statistics



Hypotheses

- the null hypothesis - "null" means "nothing"
  - there is no difference between the means (or equivalently, the means are equal)
  - the correlation is zero
  - the slope (of the regression line) is zero
  - the observed frequencies and the expected frequencies are the same
- Hypotheses refer to the population!

Hypotheses

- appropriate decisions regarding the null
  - if p > .05 (or alpha), the null hypothesis is retained or we fail to reject the null (the experiment didn't work)
    - our results (not data) are consistent with the null hypothesis
  - if p < .05 the null hypothesis is rejected
  - things NOT to say about the null: proved, disproved, right, wrong, true, false, correct, incorrect (we don't know any of that!)

Hypotheses

- the alternative hypothesis - also called the experimental hypothesis
  - if $p > .05$, the alternative (experimental) hypothesis is not supported or not confirmed (the experiment didn't work)
  - if $p < .05$ the alternative (experimental) hypothesis is supported or confirmed
  - things NOT to say about the alternative: proved, disproved, right, wrong, true, false, correct, incorrect, retained, rejected

Interpreting a Significance Test

- significance tests we know: t-test, ANOVA, chi-square test (and various other tests on correlations and slopes, which are usually one of the above)
- after one of those tests is done:
  - if we are asked to make a decision regarding the null hypothesis, the correct answer is either "we reject the null" or "we retain (fail to reject) the null"
  - if we are asked about the alternative (experimental) hypothesis, the correct answer is either "the alternative hypothesis is supported" or "the alternative hypothesis is not supported"

What It Means

- if the null is retained (not rejected)
  - the experiment didn't work
  - true experiments: there is no effect of the IV on the DV - at least, we cannot be confident that the IV had an effect from the results we got
  - quasi-experiments: there is no relationship, association, correlation between the IV and the DV
    - the word "correlation" is normally used only when speaking of numeric variables

What It Means

- if the null is rejected (alternative supported)
  - the experiment worked - Yay! We get to publish and might get tenure!
  - true experiments: there is an effect of the IV on the DV - at least that's the way it appears
  - quasi-experiments: there is a relationship, association, correlation between the IV and the DV
    - the word "correlation" is normally used only when speaking of numeric variables

Statistical Significance

- if the null is rejected, we can say the <u>difference</u> (whatever it is) was statistically significant
- this is ALWAYS a statement about the sample
  - the sample means were significantly different
  - the sample correlation (or slope) was significantly different from zero
  - occasionally (around the water cooler): the effect is significant or the correlation is significant (but this is loose talk!)
- never say: the data/experiment are/is significant
- you are treading on thin ice if you say the result is significant (but it's ok with me)

Statistical Significance

- a statistically significant effect is not necessarily an important effect or an effect that anyone should care about
- it's not even necessarily a large effect - the effect could be so trivially small that no reasonable person would pay attention to it
- a statistically significant effect is one that probably did not occur by chance alone - it is probably not due to random error
- the p-value (roughly speaking) tells the approximate probability that this result or difference was due to nothing but chance

Effects (Relationships)

- when the null is rejected, we say there is an effect of the IV (on the DV)
- this does not necessarily imply a cause and effect relationship
- a safer word is relationship: There is a relationship between the IV and the DV.
- this means if we know the value of the IV, we can guess the value of the DV better than we can if we do not know the value of the IV (or better than chance)

- examples
    - if we are trying to guess a person's weight (without seeing him/her), does it help to know his/her gender? - if yes, then there is "an effect of gender on weight" or "a relationship between gender and weight"
    - 20 subjects are divided randomly into two groups. Each group is given a list of words and a set of instructions as to what to do with them. Later we ask them to recall the words. If we are trying to guess how many words a subject recalled, does it help to know the instructions she was given? If yes, then there is "an effect of instructions on recall."
    - or we can say "instructions affected recall." Effect is a noun (a thing). Affect is a verb (an action).
        - The word "effect" usually has the word "the" or "an" in front of it. The word "affect" never has the word "the" or "an" in front of it.

Effect Size

- p-values do not tell you about effect size!
- here are some effect size measures
  - contingency tables: phi, Cramer's V, likelihood ratio (relative risk), odds ratio
  - difference between two means: Cohen's d
  - difference among multiple means (ANOVA): eta squared, partial eta squared
  - regression relationships: R-squared (proportion of variability explained or PVE)

Effect Size

- small p-values are good - they mean the evidence is strong against the null
  - how really small p-values are written
    - in SPSS: p = .0000
    - in R: scientific notation (e.g., 7.71e-07)
    - in APA style: p < .001
- small p-values DO NOT mean the effect is large
- large values of effect size measures are good - they mean the effect is large

Confounding

- confounding occurs when a third variable is responsible for our effect (and not the IV)
- when the groups are different in some way other than the difference created by the treatment (the IV), suspect confounding
- when the groups are created by random assignment, confounding is less likely
- not everything that "goes wrong" with a study is a confound!
  - small sample sizes - not a confound!
  - unequal group sizes - not a confound!
  - biased sampling - may or may not create a confound

A Note On Group Size

- when we divide by N, we have "corrected" for differences in group size
  - the resulting statistic is "per subject"
- thus, when we are comparing means or proportions or percentages, group size is no longer an issue (e.g., means are not different because the group sizes are different)

Ratios

- when you divide two numbers, you have a ratio
- if those two numbers are variances, then you have an F-ratio (an example of a test statistic)
- if those two numbers are percent successes (or failures), then you have a likelihood ratio
- if those two numbers are odds of success (or failure), then you have an odds ratio
- if 45% of men and 30% of women were accepted into a certain grad program, then 45/30 = 1.5 is a likelihood ratio; the correct way to say it is "Men were 1.5 times more likely to be accepted than women."

Be Careful!

- if 45 men and 35 women are accepted, the likelihood ratio is...???
- percentages can be especially tricky, so be careful what conclusions you draw from them
- here is a table describing the results of a study on admissions to six UC Berkeley grad programs - was there gender bias in admissions?

	percent of applicants admitted						
gender	overall	dept. A	dept. B	dept. C	dept. D	dept. E	dept. F
male	44.5	62.1	63.0	36.9	33.1	27.7	5.9
female	30.4	82.4	68.0	34.1	34.9	23.9	7.0
	p<.001	p<.001	p=.77	p=.43	p=.64	p=.37	p=.64 ← p-values for differences between the percentages
LR (men/women):	1.46	0.75	0.93	1.08	0.95	1.16	0.84 ← likelihood rations

Here are the raw frequencies. What happened? Why does there appear to be a bias against women overall that doesn't appear in any of the individual departments?

```
> UCBAdmissions
, , Dept = A

          Gender
Admit       Male Female
  Admitted  512     89
  Rejected  313     19

, , Dept = B

          Gender
Admit       Male Female
  Admitted  353     17
  Rejected  207      8

, , Dept = C

          Gender
Admit       Male Female
  Admitted  120    202
  Rejected  205    391

, , Dept = D

          Gender
Admit       Male Female
  Admitted  138    131
  Rejected  279    244

, , Dept = E

          Gender
Admit       Male Female
  Admitted   53     94
  Rejected  138    299

, , Dept = F

          Gender
Admit       Male Female
  Admitted   22     24
  Rejected  351    317

, , Dept = Sum

          Gender
Admit       Male Female
  Admitted 1198    557
  Rejected 1493   1278
```

Advice

- redo any analysis I do in the handouts
- do the damn practice problems!!!
- if there is anything you don't understand, ASK
- DON'T FALL BEHIND!