Lecture 13
Tests for counting

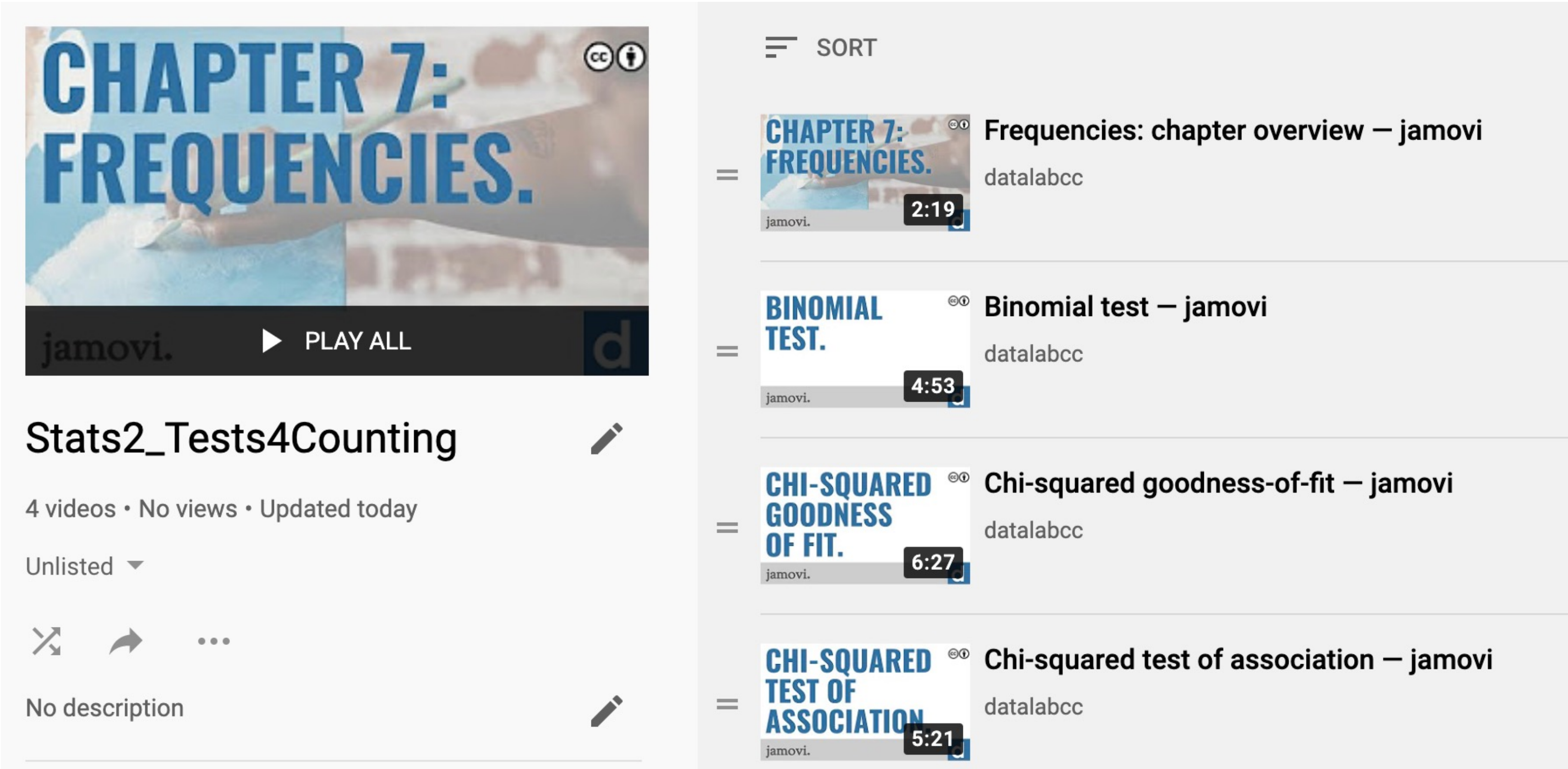
CHOONG-WAN WOO | COCOAN lab | http://cocoanlab.github.io

What we learned…

- Sampling distribution
- Standard error of the mean
- Confidence interval
- One-sample t-test
- Paired t-test
- Independent samples t-test
- Resampling (bootstrap, permutation tests)

Tests for counting

https://www.youtube.com/playlist?list=PLXCuLG6zw7mKzyUxZDs06SOtWP8fjJF98



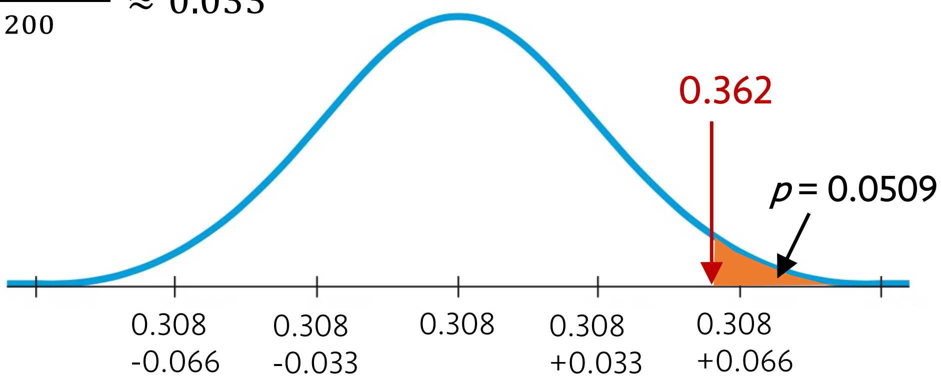
Binomial tests

- **Hypothesis**: questions like, has the Facebook users who update their status daily increased since last month?

- **Null hypothesis:** null because it assumes no changes, thus p = 30.8%
- **Alternative hypothesis**: H_A: p > 30.8%

- We observed a new $\hat{p}$ from 200 respondents.

- Based on the null hypothesis, $SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.308 \times 0.692}{200}} \approx 0.033$
- Let's say the observed $\hat{p}$ = 36.2%
- Then, $z = \frac{0.362 - 0.308}{0.033} = 1.6364$
- p = 0.0509 (one-tail)



Chi-square tests

Births	Sign
23	Aries
20	Taurus
18	Gemini
23	Cancer
20	Leo
19	Virgo
18	Libra
21	Scorpio
19	Sagittarius
22	Capricorn
24	Aquarius
29	Pisces

Birth totals by sign for 256 Fortune 400 executives.

Example: zodiac signs of 256 heads of the largest 400 companies

- If the zodiac signs cannot predict the future, we should expect 1/12 counts for each category.
- How closely do the observed numbers of births per sign fit this simple “null” model?
- **“Goodness-of-fit”** test

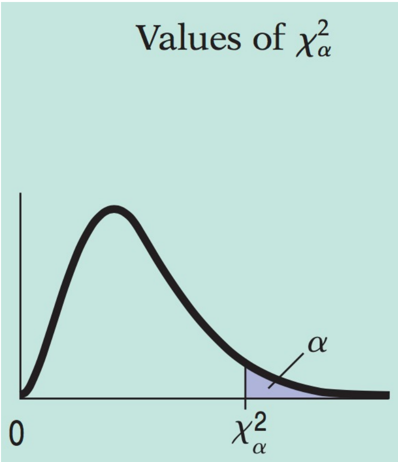
Goodness-of-fit tests

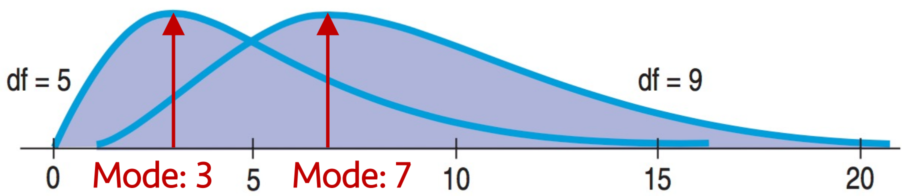
- Procedure:
  - First, observed value minus expected value for each cell: similar to residuals
  - The residual values can be positive and negative, so we need to square them.
  - We divide the residuals by the expected counts.
  - $\sum \frac{(Obs - Exp)^2}{Exp}$
  - How well the theory (expected values) fits the data: **goodness-of-fit**

- It follows the chi-square (χ^2) distribution.
  - $\chi^2 = \sum \frac{(Obs - Exp)^2}{Exp}$
  - This family of models also depends on the degrees of freedom.
  - In the chi-square test, *df* = *n* − 1, where *n* is the number of categories, not the sample size.

Chi-Square P-values

- The chi-square should be used *only* for testing hypotheses, *not* for constructing confidence intervals.
- We can do only *one-sided* test (by squaring the differences, we made all the deviations positive).
- There's *no* direction to the rejection of the null model. All we know is that it doesn't fit.
- It is testing all of the cells together. There are many ways the null hypothesis can be wrong (*many-sided* in some sense)
- Chi-square models are skewed.
  - The mode is at $\chi^2 = df - 2$, and its mean is at *df*.





Trouble with Goodness-of-fit tests

- Goodness-of-fit: **How well does the theory fit the data?**
- **The only null hypothesis available** (H_0: the theory is true)
  - We can only reject or fail to reject the null hypothesis.
  - We can never confirm the theory is true.
- **It is also difficult to know what is the alternative.**
  - The theory can be wrong in many ways.
- **Thus, there is no way to prove that a favored model is true, with goodness-of-fit tests.**
  - Alternative: model comparison

Chi-square test of Homogeneity

- Testing whether the proportions are same across multiple groups
- two-way table

Post-graduation activities of the class of 2006 for several colleges of a large university

	Agriculture	Arts & Sciences	Engineering	Social Science	Total
Employed	379	305	243	125	**1052**
Grad School	186	238	202	96	**722**
Other	104	123	37	58	**322**
Total	**669**	**666**	**482**	**279**	**2096**

Percentage

	Agriculture	Arts & Sciences	Engineering	Social Science	Total
Employed	56.7%	45.8%	50.4%	44.8%	**50.2**
Grad School	27.8	35.7	41.9	34.4	**34.4**
Other	15.5	18.5	7.7	20.8	**15.4**
Total	**100**	**100**	**100**	**100**	**100**

Expected values for the '06 graduates

	Agriculture	Arts & Sciences	Engineering	Social Science	Total
Employed	335.777	334.271	241.920	140.032	**1052**
Grad School	230.448	229.414	166.032	96.106	**722**
Other	102.776	102.315	74.048	42.862	**322**
Total	**669**	**666**	**482**	**279**	**2096**

Chi-square test of Homogeneity

- $\chi^2 = \sum \frac{(Obs - Exp)^2}{Exp}$
- The example of the agriculture school $\frac{(Obs - Exp)^2}{Exp} = \frac{(379 - 335.777)^2}{335.777} = 5.564$
- And summing these across all the schools,
  - $\chi^2 = \sum \frac{(Obs - Exp)^2}{Exp} = 54.51$
- Degrees of freedom:
  - $(R-1)(C-1)$, where **R**: the number of rows, **C**: the number of columns

Chi-square test of independence

Race effects on police vehicle search

		Race			
		Black	White	Other	Total
Search	No	787	594	27	**1408**
	Yes	813	293	19	**1125**
	Total	**1600**	**887**	**46**	**2533**

- Are police search and race independent? or have relationship?
- Contingency table
- From L09:
  - “**Independence:** the occurrence of A does not change the probability of B, $P(\mathbf{B}|\mathbf{A}) = P(\mathbf{B})$”

- The calculation is identical to the homogeneity test.

- What’s different?
  - Independence test: Two categorical variables measured on a single population
    - Homogeneity test: a single categorical variable independently measured on two or more populations
  - Independence test’s question: “Are the variables independent?”
    - Homogeneity test: “Are the groups homogeneous?”