Revised: 3/20/2021

Inference II

Michael C. Hackett

Assistant Professor, Computer Science

Community College of Philadelphia

Choosing a Sample Size

- We want a sample size large enough that the margin of error is sufficiently small enough so that the sample is useful.

A university newspaper is conducting a survey to determine what fraction of students support a $200 per year increase in fees to pay for a new football stadium. How big of a sample is required to ensure the margin of error is smaller than 0.04 using a 95% confidence level?

Choosing a Sample Size

A university newspaper is conducting a survey to determine what fraction of students support a $200 per year increase in fees to pay for a new football stadium. How big of a sample is required to ensure the margin of error is smaller than 0.04 using a 95% confidence level?

- The margin of error for a sample proportion:

$$z^* \sqrt{\frac{p(1-p)}{n}}$$

- The smallest sample size n so that this margin of error is smaller than 0.04.
  - For a 95% confidence level, the z^* is 1.96

$$1.96 \times \sqrt{\frac{p(1-p)}{n}} < 0.04$$

Choosing a Sample Size

- If no prior point estimate exists, we typically use this worst-case value which is when p is 0.5

$$1.96 \times \sqrt{\frac{0.5(1-0.5)}{n}} < 0.04$$

$$1.96^2 \times \frac{0.5(1-0.5)}{n} < 0.04^2$$

$$1.96^2 \times \frac{0.5(1-0.5)}{0.04^2} < n$$

$$600.25 < n$$

- A least 601 participants are needed to ensure a sample proportion within 0.04 of the true proportion with 95% confidence.

Choosing a Sample Size

Suppose we want to continually track the support of payday borrowers for regulation on lenders, where we would conduct a new poll every month. Running such frequent polls is expensive, so we decide a wider margin of error of 5% for each individual survey would be acceptable. Based on the original sample of borrowers where 70% supported some form of regulation, how big should our monthly sample be for a margin of error of 0.05 with 95% confidence?

$$1.96 \times \sqrt{\frac{0.7(1-0.7)}{n}} < 0.05$$

$$1.96^2 \times \frac{0.7(1-0.7)}{n} < 0.05^2$$

$$1.96^2 \times \frac{0.7(1-0.7)}{0.05^2} < n$$

$$322.7 < n \text{ (or more intuitively: } n > 322.7)$$

Difference of Two Proportions

- Confidence intervals and hypothesis tests can be applied to differences in population proportions
  - $p_1 - p_2$
  - A reasonable point estimate: $\hat{p}_1 - \hat{p}_2$

- The difference of two sample proportions can be modeled using a normal distribution, provided its independence and success-failure conditions are met

Difference of Two Proportions

- $\hat{p}_1 - \hat{p}_2$ can be modeled using a normal distribution when:

- *Independence*: The data are independent within and between the two groups.
  - Generally satisfied with data from two independent random samples or from a randomized experiment.
- *Success-failure*: Will holds for both groups, where we check successes and failures in each group separately.

$n_1\hat{p}_1 \geq 10$ and $n_1(1 - \hat{p}_1) \ \geq 10$

$n_2\hat{p}_2 \geq 10$ and $n_2(1 - \hat{p}_2) \ \geq 10$

- Standard Error:

$$SE = \sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}}$$

Difference of Two Proportions

- The same general formula for the confidence interval is used

$$\hat{p}_1 - \hat{p}_2 \pm z^* \times \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

- For the difference of two proportions, follow the same Prepare, Check, Calculate, Conclude steps for computing a confidence interval or completing a hypothesis test

Difference of Two Proportions

- *We consider an experiment for patients who underwent CPR for a heart attack and were subsequently admitted to a hospital. These patients were* ***randomly*** *divided into a treatment group where they received a blood thinner or the control group where they did not receive a blood thinner. The outcome variable of interest was whether the patients survived for at least 24 hours. Check whether we can model the difference in sample proportions using the normal distribution.*

	Survived	Died	Total
Control	11	39	50
Treatment	14	26	40
Total	25	65	90

Difference of Two Proportions

- *Independence*: Satisfied because it was a random experiment
- *Success-failure*: There are at least 10 in each experiment arm
  - 11, 14, 39, and 26

	Survived	Died	Total
Control	11	39	50
Treatment	14	26	40
Total	25	65	90

- Both conditions are satisfied; can be reasonably modeled using a normal distribution

Difference of Two Proportions

	Survived	Died	Total
Control	11	39	50
Treatment	14	26	40
Total	25	65	90

- *Create and interpret a 90% confidence interval of the difference for the survival rates in the CPR study.*
- Where p_t is the survival rate in the treatment group and p_c for the control group:

$$\hat{p}_t - \hat{p}_c = \frac{14}{40} - \frac{11}{50} = 0.35 - 0.22 = 0.13$$

$$SE = \sqrt{\frac{0.35(1 - 0.35)}{40} + \frac{0.22\ (1 - 0.22)}{50}} = 0.095$$

- 90% confidence interval = $0.13 \pm 1.65 \times 0.095 = (-0.027, 0.287)$

- 90% confident that blood thinners have a difference of -2.7% to +28.7% percentage point impact on survival rate for patients who are like those in the study.
  - Because 0% is contained in the interval, there is not enough information to say whether blood thinners help or harm heart attack patients who have been admitted after they have undergone CPR.

Difference of Two Proportions

A quadcopter company is considering a new manufacturer for rotor blades. The new manufacturer would be more expensive, but they claim their higher-quality blades are more reliable, with 3% more blades passing inspection than their competitor. Set up appropriate hypotheses for the test.

- H_0 : The higher-quality blades will pass inspection 3% more frequently than the standard-quality blades.

$$p_{highQuality} - p_{standard} = 0.03$$

- H_A : The higher-quality blades will pass inspection some amount different than 3% more often than the standard-quality blades.

$$p_{highQuality} - p_{standard} \neq 0.03$$

Difference of Two Proportions

- *The quality control engineer collects a sample of blades, examining 1000 blades from each company, and finds that 899 blades pass inspection from the current supplier and 958 pass inspection from the prospective supplier. Using these data, evaluate the hypotheses with a significance level of 5%.*

- Independence: Not necessarily random, so we assume it is reasonable that the blades are independent.
- Success-Failure: Condition holds (no need to do the math, 958 and 899 out of 1000 each)

Difference of Two Proportions

- *The quality control engineer collects a sample of blades, examining 1000 blades from each company, and finds that 899 blades pass inspection from the current supplier and 958 pass inspection from the prospective supplier. Using these data, evaluate the hypotheses with a significance level of 5%.*

$$\hat{p}_1 - \hat{p}_2 = 0.958 \; - 0.899 = 0.059$$

$$SE = \sqrt{\frac{0.958(1 - 0.958)}{1000} + \frac{0.889\,(1 - 0.889)}{1000}} = 0.0114$$

$$Z = \frac{0.059 - 0.03}{0.0114} = 2.54$$

```
> 2*(1-pnorm(2.54, mean=0.03))
[1] 0.01207312
```

Difference of Two Proportions

- *The quality control engineer collects a sample of blades, examining 1000 blades from each company, and finds that 899 blades pass inspection from the current supplier and 958 pass inspection from the prospective supplier. Using these data, evaluate the hypotheses with a significance level of 5%.*

- p-value 0.012 is less than the 0.05 significance level.
- Null hypothesis is rejected.

- Since we observed a larger-than-3% increase in blades that pass inspection (0.059), we have statistically significant evidence that the higher-quality blades pass inspection more than 3% as often as the currently used blades, exceeding the company's claims.

Chi-square test

- A method for assessing a null model when the data are categorical is a **chi-square test**.

- A chi-square test often used when:
  - Determining if, given a sample of cases that can be classified into several (more than two) groups, the sample is representative of the general population.
  - Evaluating whether data resemble a particular distribution, such as a normal or geometric distribution.

Chi-square test

- The hypothesis tests seen this far used the following for computing the Z score:

$$\frac{point\ estimate - null\ value}{SE\ of\ point\ estimate}$$

- This is based on…
  - Identifying the difference between a point estimate and an expected value if the null hypothesis was true
  - Standardizing that difference using the standard error of the point estimate

Chi-square test

- For categorized data, a Z score is computed for each category

$$\frac{observed\ count - expected\ count}{SE\ of\ observed\ count} = \frac{observed\ count - expected\ count}{\sqrt{expected\ count}}$$

- The Z scores are each squared and summed to arrive at the chi-square test statistic

$$X^2 = Z_1^2 + Z_2^2 + Z_3^2 + \cdots + Z_k^2$$

- X^2 summarizes how strongly the observed counts tend to deviate from the null counts.

Chi-square distribution

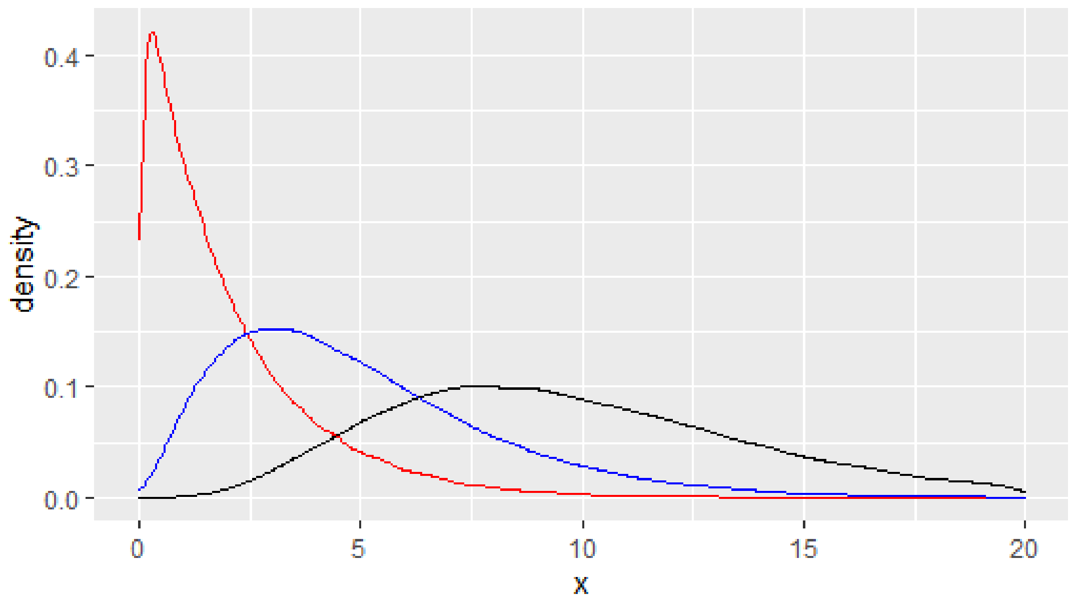
- Chi-square distributions characterize data sets and statistics that are always positive and usually right skewed.

- The chi-square distribution has only one parameter: **degrees of freedom** (**df**), which changes the shape, center, and spread of the distribution.
  - Recall a normal distribution had two parameters (mean and standard deviation) that could be used to describe its exact characteristics.

Chi-square distribution

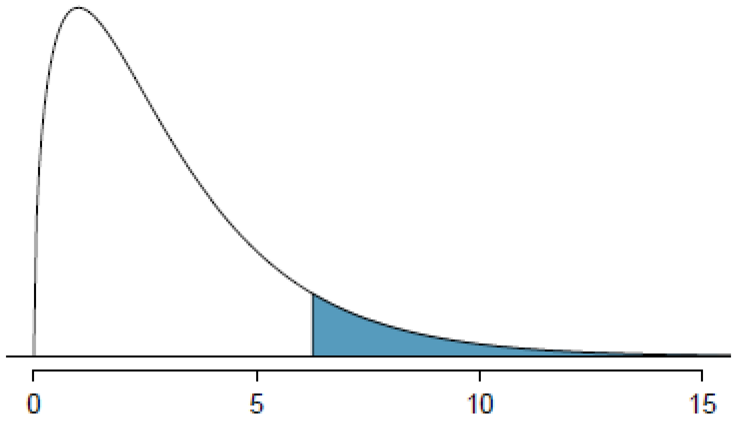
- As the degrees of freedom increase:
  - The center becomes larger
  - The variability increases
  - Becomes more symmetric

  - Red = 2 df
  - Blue = 5 df
  - Black = 10 df



Chi-square distribution

- The chi-square distribution below has 3 degrees of freedom and an upper shaded tail starting at 6.25

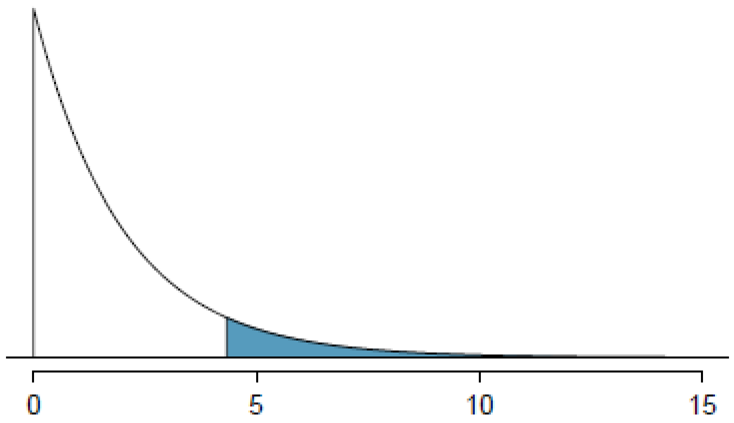


Area:

```
> pchisq(6.25, df=3, lower.tail=FALSE)
[1] 0.1000608
```

Chi-square distribution

- The chi-square distribution below has 2 degrees of freedom and an upper shaded tail starting at 4.3



Area:

```
> pchisq(4.3, df=2, lower.tail=FALSE)
[1] 0.1164842
```

Chi-square test

- A random sample of 275 jurors in a small county:

Race	White	Black	Hispanic	Other	Total
Representation in juries	205	26	25	19	275
Registered voters	0.72	0.07	0.12	0.09	1.00

- Jurors are identified their racial group
- *Are these jurors racially representative of the population?*
  - If the jury is representative of the population, then the proportions in the sample should roughly reflect the population of eligible jurors, i.e. registered voters.

Race	White	Black	Hispanic	Other	Total
Observed data	205	26	25	19	275
Expected counts	198	19.25	33	24.75	275

Chi-square test

Are the differences strong enough to provide convincing evidence that the jurors are not a random sample?

- $H_0 =$ Jurors were randomly sampled; there is no racial bias
- $H_A =$ Jurors were not randomly sampled; there is racial bias

- There is no sample mean like in the previous hypothesis tests
    - We perform a chi-square test to check the difference between the observed and the expected.

Chi-square test

Are the differences strong enough to provide convincing evidence that the jurors are not a random sample?

- $H_0 =$ Jurors were randomly sampled; there is no racial bias
- $H_A =$ Jurors were not randomly sampled; there is racial bias

- There is no sample mean like in the previous hypothesis tests
  - We perform a chi-square test to check the difference between the observed and the expected.

Chi-square test

$$\frac{observed\ count - null(expected)\ count}{SE\ of\ observed\ count}$$

White jurors:

$$Z_1 = \frac{205 - 198}{\sqrt{198}} = 0.50$$

Black jurors:

$$Z_2 = \frac{26 - 19.25}{\sqrt{19.25}} = 1.54$$

Hispanic jurors:

$$Z_3 = \frac{25 - 33}{\sqrt{33}} = -1.39$$

Hispanic jurors:

$$Z_4 = \frac{19 - 24.75}{\sqrt{24.75}} = -1.16$$

Chi-square test

$$X^2 = Z_1^2 + Z_2^2 + Z_3^2 + \ Z_4^2$$

$$X^2 = 0.5^2 + 1.54^2 + -1.39^2 + -1.16^2$$

$$X^2 = 0.25 + 2.37 + 1.93 + 1.35 = 5.9$$

Chi-square test

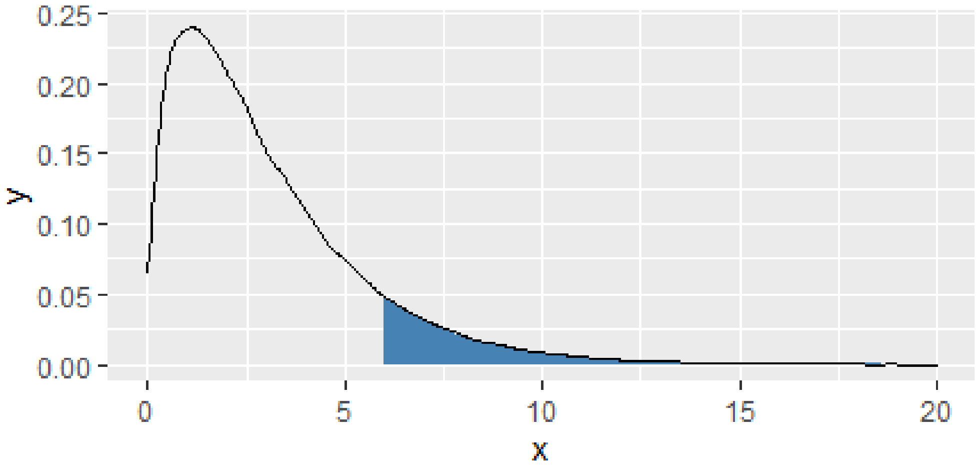
- A large X^2 value would suggest strong evidence favoring the alternative hypothesis: that there was racial bias.
  - We could not quantify what the chance was of observing such a large test statistic if the null hypothesis was true.

- If the null hypothesis was true and there was no racial bias, then X^2 would follow a chi-square distribution, with three degrees of freedom in this case.
  - X^2 follows a chi-square distribution with $k - 1$ degrees of freedom, where k is the number of bins.

Chi-square test

- The area is 0.117

```
> pchisq(5.9, df=3, lower.tail=FALSE)
[1] 0.1165781
```



- High p-value (over 10% significance level)
    - Fail to reject H_0
    - The data do not provide convincing evidence of racial bias in the juror selection

Testing two-way tables

- The previous example was a one-way table

Race	White	Black	Hispanic	Other	Total
Observed data	205	26	25	19	275
Expected counts	198	19.25	33	24.75	275

- The chi-square test also applies to two-way tables:

	Failure	Success	Total
`lifestyle`	109	125	234
`met`	120	112	232
`rosi`	90	143	233
Total	319	380	699

Testing two-way tables

- *This table summarizes the results of an experiment evaluating three treatments for Type 2 Diabetes in patients aged 10-17 who were being treated with metformin. The three treatments considered were continued treatment with metformin (met), treatment with metformin combined with rosiglitazone (rosi), or a lifestyle intervention program. Each patient had a primary outcome, which was either lacked glycemic control (failure) or did not lack that control (success).*

	Failure	Success	Total
`lifestyle`	109	125	234
`met`	120	112	232
`rosi`	90	143	233
Total	319	380	699

- *What are appropriate hypotheses to test?*

Testing two-way tables

- $H_0 =$ No difference in the effectiveness of the three treatments
- $H_A =$ There is some difference in the effectiveness of the three treatments

	Failure	Success	Total
lifestyle	109	125	234
met	120	112	232
rosi	90	143	233
Total	319	380	699

- *Compute the expected values for each of the six table cells.*
- *Compute the chi-square for the data*
- *Evaluate whether to reject the null hypothesis using a significance level of 0.05*

Testing two-way tables

- Computing expected counts in a two-way table:

$$Expected\ Count_{row\ i,\ col\ j} = \frac{row\ i\ total\ \times column\ j\ total}{table\ total}$$

	Failure	Success	Total
lifestyle	109	125	234
met	120	112	232
rosi	90	143	233
Total	319	380	699

$$\frac{234 \times 319}{699} = 106.8 \qquad \frac{234 \times 380}{699} = 127.2$$

$$\frac{232 \times 319}{699} = 105.9 \qquad \frac{234 \times 380}{699} = 126.1$$

$$\frac{233 \times 319}{699} = 106.3 \qquad \frac{234 \times 380}{699} = 126.7$$

Testing two-way tables

	Failure	Success	Total
lifestyle	109	125	234
met	120	112	232
rosi	90	143	233
Total	319	380	699

- Computing the chi-square:

$$\frac{(observed\ count - expected\ count)^2}{expected\ count}$$

$$\frac{234 \times 319}{699} = 106.8$$

$$\frac{232 \times 319}{699} = 105.9$$

$$\frac{233 \times 319}{699} = 106.3$$

$$\frac{234 \times 380}{699} = 127.2$$

$$\frac{234 \times 380}{699} = 126.1$$

$$\frac{234 \times 380}{699} = 126.7$$

$$Z_1 = \frac{(109 - 106.8)^2}{106.8} = 0.05$$

$$Z_2 = \frac{(120 - 105.9)^2}{105.9} = 1.87$$

$$Z_3 = \frac{(90 - 106.3)^2}{106.3} = 2.51$$

$$Z_4 = \frac{(125 - 127.2)^2}{127.2} = 0.04$$

$$Z_5 = \frac{(112 - 126.1)^2}{126.1} = 1.58$$

$$Z_6 = \frac{(143 - 126.7)^2}{126.7} = 2.10$$

$$X^2 = 0.05 + 1.87 + 2.51 + 0.04 + 1.58 + 2.10 = \mathbf{8.15}$$

Testing two-way tables

- Degrees of freedom in a two-way chi-squared test:

$$df = (number\ of\ rows - 1) \times (number\ of\ columns\ - 1)$$

$$df = (3 - 1) \times (2 - 1) = 2$$

- p-value: 0.016

```
> pchisq(8.16, df=2, lower.tail=FALSE)
[1] 0.01690747
```

- 0.016 < 0.05 : Reject H_0; There was some difference in the effectiveness of the three treatments