

# BIOS 600: Principles of Statistical Inference

## Case Studies in Probability

Fall 2012

# Breast Cancer Survivors Study

The Susan G. Komen Foundation funded a small study of the relationship between diet beliefs and behaviors among breast cancer survivors. In the study,

- ▶ 49% of survivors ate 5+ servings of fruits and/or vegetables a day
- ▶ 56% of survivors agreed with the statement “Eating 5+ servings of fruits and/or vegetables a day will decrease the risk of breast cancer recurrence”
- ▶ 72% of survivors either ate 5+ servings/day, agreed that doing so would decrease breast cancer recurrence risk, or both

# Breast Cancer Survivors Study

1. What is the probability that a survivor both agrees with the statement and eats 5+ servings of fruits and vegetables a day?
2. Among survivors who agree with the statement, what is the probability of eating 5+ servings of fruits and/or vegetables per day?
3. Are agreeing with the statement and eating 5+ servings/day independent events? Cite the data in explaining your answer.

# Breast Cancer Survivors Study

1. What is the probability that a survivor both agrees with the statement and eats 5+ servings of fruits and vegetables a day?

$$\begin{aligned}Pr(\text{believe} \cup \text{eat}) &= Pr(\text{believe}) + Pr(\text{eat}) - Pr(\text{believe} \cap \text{eat}) \\0.72 &= 0.56 + 0.49 - Pr(\text{believe} \cap \text{eat}) \\0.33 &= Pr(\text{believe} \cap \text{eat})\end{aligned}$$

# Breast Cancer Survivors Study

2. Among survivors who agree with the statement, what is the probability of eating 5+ servings of fruits and/or vegetables per day?

$$Pr(\text{eat} \mid \text{believe}) = \frac{Pr(\text{eat and believe})}{Pr(\text{believe})} = \frac{0.33}{0.56} = 0.59$$

# Breast Cancer Survivors Study

3. Are agreeing with the statement and eating 5+ servings/day independent events? Cite the data in explaining your answer.

$$Pr(\text{eat}) = 0.49$$

$$Pr(\text{eat} \mid \text{believe}) = 0.59$$

## Type II Diabetes Study

You are designing a pilot study and plan to recruit 50 high-risk non-diabetic adults in a study of dietary and behavioral risk factors for type II diabetes. According to the best available figures, 10% of high-risk adults should get type II diabetes during the study period.

- ▶ How many study subjects do you expect to develop the disease?
- ▶ What is the probability that *exactly* 10% of your study subjects get type II diabetes during the study?
- ▶ What is the probability that at most 2 study subjects get diabetes?
- ▶ If you want to have 90% probability that at least 5 patients will get diabetes, do you need to recruit more patients?



## Type II Diabetes Study

You are designing a pilot study and plan to recruit 50 high-risk non-diabetic adults in a study of dietary and behavioral risk factors for type II diabetes. According to the best available figures, 10% of high-risk adults should get type II diabetes during the study period.

- ▶ How many study subjects do you expect to develop the disease?

- ▶  $np = 50(.10) = 5$

- ▶ What is the probability that *exactly* 10% of your study subjects get type II diabetes during the study?

- ▶ 10% is 5 patients, so take

$$\binom{50}{5} 0.1^5 0.9^{45} = \frac{50(49)(48)(47)(46)(45!)}{5(4)(3)(2)(1)(45!)} (.1^5)(.9^{45}) = 0.1849$$



## Type II Diabetes Study

You are designing a pilot study and plan to recruit 50 high-risk non-diabetic adults in a study of dietary and behavioral risk factors for type II diabetes. According to the best available figures, 10% of high-risk adults should get type II diabetes during the study period.

- ▶ What is the probability that at most 2 study subjects get diabetes?

$$\begin{aligned} & Pr(0) + Pr(1) + Pr(2) \\ &= \binom{50}{0} (.1^0)(.9^{50}) + \binom{50}{1} (.1^1)(.9^{49}) + \binom{50}{2} (.1^2)(.9^{48}) \\ &= 0.0052 + 0.0286 + 0.0779 = 0.1117 \end{aligned}$$

Stata: **bitesti 50 2 .1**

## Type II Diabetes Study

You are designing a pilot study and plan to recruit 50 high-risk non-diabetic adults in a study of dietary and behavioral risk factors for type II diabetes. According to the best available figures, 10% of high-risk adults should get type II diabetes during the study period.

- ▶ If you want to have 90% probability that at least 5 patients will get diabetes, do you need to recruit more patients?

$$\begin{aligned} & 1 - Pr(0) - Pr(1) - Pr(2) - Pr(3) - Pr(4) \\ &= 1 - 0.1117 - Pr(3) - Pr(4) \\ &= 0.8883 - Pr(3) - Pr(4) \\ &= 0.8883 - \binom{50}{3} (.1^3)(.9^{47}) - \binom{50}{4} (.1^4)(.9^{46}) \\ &= 0.8883 - 0.1386 - 0.1809 = 0.5688 \end{aligned}$$

You do need to recruit more patients

How many do you need? `display 1-binomial(n,4,.1)`

78

# Parking on Campus

Suppose UNC writes parking tickets approximately twice per week (2 of 5 weekdays) and that you will always get a parking ticket if you park on campus on one of those days. You want to park on campus on Tuesdays and Thursdays during the fall term during your favorite class. A student parking permit costs \$227, and a parking fine is \$30.

- ▶ Assuming BIOS 600 (which meets 30 times) is your favorite class, are you better off financially buying a permit or risking the parking tickets?
- ▶ How many days of parking does it take to justify purchasing the parking permit?

| PARKING VIOLATION                                                                                                                                                                               |                                                      |                  |                 |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------|------------------|-----------------|
| <b>NOTICE</b> This vehicle is improperly parked. Violations are as follows:                                                                                                                     |                                                      |                  |                 |
| <input type="checkbox"/> Vehicle has no valid Parking Permit                                                                                                                                    | <input type="checkbox"/> Blocking Driveway or Access |                  |                 |
| <input type="checkbox"/> Parked in No Parking Area / Space                                                                                                                                      | <input type="checkbox"/> Blocking other Vehicle      |                  |                 |
| <input type="checkbox"/> Parked in Fire Lane                                                                                                                                                    | <input type="checkbox"/> Parked in 2 Spaces          |                  |                 |
| <input type="checkbox"/> Parked in Handicap Space                                                                                                                                               | <input type="checkbox"/> Other _____                 |                  |                 |
| <input type="checkbox"/> Parked in Reserved or Assigned Space                                                                                                                                   |                                                      |                  |                 |
| This vehicle's description has been permanently recorded. Any additional infractions of our regulations could result in towing of vehicle owner's expense and revocation of parking privileges. |                                                      |                  |                 |
| License No. _____                                                                                                                                                                               | Date _____                                           | Permit No. _____ | Date _____      |
| Vehicle Make / Model _____                                                                                                                                                                      | Color _____                                          | Time _____       |                 |
| Driver's Name / License _____                                                                                                                                                                   | Location _____                                       |                  | Issued By _____ |
|                                                                                                                                                                                                 |                                                      |                  |                 |
| Violation _____                                                                                                                                                                                 |                                                      |                  |                 |
| License Plate No. _____                                                                                                                                                                         | Date _____                                           | Permit No. _____ |                 |
| Vehicle Make/Model _____                                                                                                                                                                        | Color _____                                          |                  |                 |
| Date _____                                                                                                                                                                                      | Time _____                                           | Location _____   | Issued By _____ |

## Parking on Campus

Suppose UNC writes parking tickets approximately twice per week (2 of 5 weekdays) and that you will always get a parking ticket if you park on campus on one of those days. You want to park on campus on Tuesdays and Thursdays during the fall term during your favorite class. A student parking permit costs \$227, and a parking fine is \$30.

- ▶ Assuming BIOS 600 (which meets 30 times) is your favorite class, are you better off financially buying a permit or risking the parking tickets?
  - ▶ Number of parking tickets is binomial with  $n = 30$  and  $p = 0.4$ , so the expected number of tickets is  $np = 30(.4) = 12$
  - ▶ Thus the expected cost is  $12(\$30) = \$360$
  - ▶ Buy the permit!

## Parking on Campus

Suppose UNC writes parking tickets approximately twice per week (2 of 5 weekdays) and that you will always get a parking ticket if you park on campus on one of those days. You want to park on campus on Tuesdays and Thursdays during the fall term during your favorite class. A student parking permit costs \$227, and a parking fine is \$30.

- ▶ How many days of parking does it take to justify purchasing the parking permit?

$$30(n)(0.4) = 227$$

$$12n = 227$$

$$n = 18.91$$

- ▶ You need to park 19 times to justify purchasing the permit

# Screening for Down syndrome

In 2011 researchers at Brown University described a prenatal screening test for Down syndrome based on DNA sequencing of maternal plasma, which has sensitivity 0.986 and specificity 0.998.

- ▶ If you are pregnant and are given the screening test with a positive result, what is the probability that you are in fact carrying a baby with Down syndrome? Use the table on the next slide (from JAMA) to determine your age-specific risk based on Down syndrome incidence at 16 weeks' gestation.

## Screening for Down syndrome

| Age   | Risk             | Age       | Risk            |
|-------|------------------|-----------|-----------------|
| 15-19 | $\frac{1}{1250}$ | 37        | $\frac{1}{150}$ |
| 20-24 | $\frac{1}{1400}$ | 38        | $\frac{1}{120}$ |
| 25-29 | $\frac{1}{1100}$ | 39        | $\frac{1}{100}$ |
| 30-31 | $\frac{1}{900}$  | 40        | $\frac{1}{75}$  |
| 32    | $\frac{1}{750}$  | 41        | $\frac{1}{60}$  |
| 33    | $\frac{1}{420}$  | 42        | $\frac{1}{45}$  |
| 34    | $\frac{1}{325}$  | 43        | $\frac{1}{35}$  |
| 35    | $\frac{1}{250}$  | 44        | $\frac{1}{30}$  |
| 36    | $\frac{1}{200}$  | $\geq 45$ | $\frac{1}{20}$  |

## Screening for Down syndrome

| Age   | PPV  | Age       | PPV  |
|-------|------|-----------|------|
| 15-19 | 0.28 | 37        | 0.77 |
| 20-24 | 0.26 | 38        | 0.81 |
| 25-29 | 0.31 | 39        | 0.83 |
| 30-31 | 0.35 | 40        | 0.87 |
| 32    | 0.40 | 41        | 0.89 |
| 33    | 0.54 | 42        | 0.92 |
| 34    | 0.60 | 43        | 0.94 |
| 35    | 0.66 | 44        | 0.94 |
| 36    | 0.71 | $\geq 45$ | 0.96 |



## Case Study: TV Drama/Comedy Viewers and Health Information

The CDC conducted a large population-representative survey asking respondents if they had learned something about a health issue or disease from a TV drama or comedy show in the previous 6 months (call the event  $L$  the event the respondent did learn something new, and call the event  $F$  that the respondent is female).

Data from the survey were used to estimate the following probabilities:  $Pr(L) = 0.58$ ,  $Pr(F) = 0.50$ ,  $Pr(L \cap F) = 0.31$ .

Create a hypothetical 1000 table and calculate  $Pr(L \cup F)$ ,  $Pr(L | F)$ , and  $Pr(L | \bar{F})$ .

## Case Study: TV Drama/Comedy Viewers and Health Information

|           | $L$ | $\bar{L}$ | Total |
|-----------|-----|-----------|-------|
| $F$       | 310 | 190       | 500   |
| $\bar{F}$ | 270 | 230       | 500   |
| Total     | 580 | 420       | 1000  |

$$Pr(L \cup F) = \frac{310+270+190}{1000} = 0.77$$

$$Pr(L | F) = \frac{Pr(L \cap F)}{Pr(F)} = \frac{0.31}{0.50} = 0.62$$

$$Pr(L | \bar{F}) = \frac{Pr(L \cap \bar{F})}{Pr(\bar{F})} = \frac{0.27}{0.50} = 0.54$$