“You can, for example, never fortell what any one man will do, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant. So says the statistician.”

Sherlock Holmes, *The Sign of Four*

From the BBC website (9/25/09) comes the following report:

From the BBC website (9/25/09) comes the following report:

An experimental HIV vaccine has for the first time cut the risk of infection, researchers say. The vaccine - a combination of two earlier experimental vaccines - was given to 16,000 people in Thailand, in the largest ever such vaccine trial. Researchers found that it reduced by nearly a third the risk of contracting HIV, the virus that leads to Aids. It has been hailed as a significant, scientific breakthrough, but a global vaccine is still some way off. The study was carried out by the US army and the Thai government over seven years on volunteers - all HIV-negative men and women aged between 18 and 30 - in parts of Thailand.

$\approx$

The vaccine was a combination of two older vaccines that on their own had not cut infection rates. The vaccine is based on B and E strains of HIV that most commonly circulate in Thailand not the C strain which predominates in Africa.

The vaccine was a combination of two older vaccines that on their own had not cut infection rates. The vaccine is based on B and E strains of HIV that most commonly circulate in Thailand not the C strain which predominates in Africa.

Half of the volunteers were given the vaccine, while the other half were given a placebo - and all were given counselling on HIV/Aids prevention. Participants were tested for HIV infection every six months for three years.

$\approx$

Question 1: This study is what is called a Phase III Clinical Trial, but in terms of what you have learned in MA 2611, what kind of study is this?

Question 1: This study is what is called a Phase III Clinical Trial, but in terms of what you have learned in MA 2611, what kind of study is this?

Answer: A controlled experiment.

$\approx$

The data are:

Group	Infected	Uninfected	Total
Placebo	74	8124	8198
Vaccine	51	8146	8197

$\approx$

Question 2: How would you analyze these data using the estimation techniques you've learned in MA 2611?

$\approx$

Let's use the 5-step method:

1. **The Scientific Goal:**

Let's use the 5-step method:

1. **The Scientific Goal:** Determine if the vaccine is effective.
2. **The Statistical Model:**

Let's use the 5-step method:

1. **The Scientific Goal:** Determine if the vaccine is effective.
2. **The Statistical Model:** Two population binomial: $b(8198, p_p)$ and $b(8197, p_v)$.
3. **The Model Parameters To Be Estimated:**

Let's use the 5-step method:

1. **The Scientific Goal:** Determine if the vaccine is effective.
2. **The Statistical Model:** Two population binomial: $b(8198, p_p)$ and $b(8197, p_v)$.
3. **The Model Parameters To Be Estimated:** p_p, the population proportion that would get HIV infection if given a placebo, and p_v, the population proportion that would get HIV infection if given the vaccine.

Let's use the 5-step method:

1. **The Scientific Goal:** Determine if the vaccine is effective.
2. **The Statistical Model:** Two population binomial: $b(8198, p_p)$ and $b(8197, p_v)$.
3. **The Model Parameters To Be Estimated:** p_p, the population proportion that would get HIV infection if given a placebo, and p_v, the population proportion that would get HIV infection if given the vaccine.

   More specifically, since we are trying to see if the vaccine is effective, we would like to estimate their difference, $p_p - p_v$.

$\approx$

4. **Point and Interval Estimates:**

4. **Point and Interval Estimates:**
The point estimate of $p_p - p_v$ is

4. **Point and Interval Estimates:**
The point estimate of $p_p - p_v$ is
$\hat{p}_p - \hat{p}_v = 74/8198 - 51/8197 = 0.00903 - 0.00622 = 0.00281.$

4. **Point and Interval Estimates:**
The point estimate of $p_p - p_v$ is
$\hat{p}_p - \hat{p}_v = 74/8198 - 51/8197 = 0.00903 - 0.00622 = 0.00281.$

As the interval estimate, let's construct a 95% confidence interval for $p_p - p_v$.

4. **Point and Interval Estimates:**
The point estimate of $p_p - p_v$ is
$\hat{p}_p - \hat{p}_v = 74/8198 - 51/8197 = 0.00903 - 0.00622 = 0.00281.$

As the interval estimate, let's construct a 95% confidence interval for $p_p - p_v$.

Note that we could use the large sample interval, since the numbers of HIV cases and non-cases are large in both groups, but I will construct an approximate score confidence interval, since it works for both large and small samples and is what I have asked you to use for all your MA 2611 work.

$\approx$

For the confidence interval, we compute (recalling that $z_{0.975} = 1.96$),

$$\tilde{n}_p = 8198 + 0.5 \times 1.96^2 = 8199.92,$$

For the confidence interval, we compute (recalling that $z_{0.975} = 1.96$),

$$\tilde{n}_p = 8198 + 0.5 \times 1.96^2 = 8199.92,$$

$$\tilde{p}_p = (74 + 0.25 \times 1.96^2)/\tilde{n}_p = 0.00914,$$

For the confidence interval, we compute (recalling that $z_{0.975} = 1.96$),

$$\tilde{n}_p = 8198 + 0.5 \times 1.96^2 = 8199.92,$$

$$\tilde{p}_p = (74 + 0.25 \times 1.96^2)/\tilde{n}_p = 0.00914,$$

$$\tilde{n}_v = 8197.92 + 0.5 \times 1.96^2 = 8198.92,$$

For the confidence interval, we compute (recalling that $z_{0.975} = 1.96$),

$$\tilde{n}_p = 8198 + 0.5 \times 1.96^2 = 8199.92,$$

$$\tilde{p}_p = (74 + 0.25 \times 1.96^2)/\tilde{n}_p = 0.00914,$$

$$\tilde{n}_v = 8197.92 + 0.5 \times 1.96^2 = 8198.92,$$

$$\tilde{p}_v = (51 + 0.25 \times 1.96^2)/\tilde{n}_v = 0.00634,$$

And the CI is

$$\tilde{p}_p - \tilde{p}_v \pm z_{0.975}\sqrt{\frac{\tilde{p}_p(1-\tilde{p}_p)}{\tilde{n}_p} + \frac{\tilde{p}_v(1-\tilde{p}_v)}{\tilde{n}_v}}$$

$$= 0.00914 - 0.00634$$

$$\pm 1.96\sqrt{\frac{0.00914(1-0.00914)}{8199.92} + \frac{0.00634(1-0.00634)}{8198.92}}$$

$$= (0.00012, 0.00548)$$

$\approx$

5. **Results and Interpretation:**

5. **Results and Interpretation:** The confidence interval, (0.00012, 0.00548), suggests that the reduction in the HIV infection rate is in the range 0.012%-0.548%. No big deal!

5. **Results and Interpretation:** The confidence interval, (0.00012, 0.00548), suggests that the reduction in the HIV infection rate is in the range 0.012%-0.548%. No big deal! But consider this:

5. **Results and Interpretation:** The confidence interval, (0.00012, 0.00548), suggests that the reduction in the HIV infection rate is in the range 0.012%-0.548%. No big deal! But consider this:

   There are roughly one billion people aged 18-30.

5. **Results and Interpretation:** The confidence interval, (0.00012, 0.00548), suggests that the reduction in the HIV infection rate is in the range 0.012%-0.548%. No big deal! But consider this:

   There are roughly one billion people aged 18-30.

   This means the vaccine could potentially prevent the infection of between $1 \times 10^9 \times 0.00012 = 120,000$ and $1 \times 10^9 \times 0.00549 = 5,490,000$ individuals over a seven year span.

5. **Results and Interpretation:** The confidence interval, (0.00012, 0.00548), suggests that the reduction in the HIV infection rate is in the range 0.012%-0.548%. No big deal! But consider this:

   There are roughly one billion people aged 18-30.

   This means the vaccine could potentially prevent the infection of between $1 \times 10^9 \times 0.00012 = 120,000$ and $1 \times 10^9 \times 0.00549 = 5,490,000$ individuals over a seven year span.

   That would be quite an achievement.

$\approx$