Section B

The Paired t-Test; the Hypothesis Testing Component

Hypothesis Testing Approach

- Want to draw a conclusion about a population parameter
  - In a population of women who use oral contraceptives, is the average (expected) change in blood pressure (after-before) 0 or not?

- Sometimes the term *expected* is used for the population average

- μ is the expected (population) mean change in blood pressure

- Hypothesis testing approach allows us to choose between two competing possibilities for μ using a single imperfect (paired) sample

Hypothesis Testing

- Two, mutually exclusive, exhaustive possibilities for "truth" about mean change
  - Null hypothesis: represented by H_o: ("h-knot" or "h-oh")
    - H_o: $\mu = 0$
  - Alternative hypothesis
    - H_A: $\mu \neq 0$

- We will use the results from our study to choose between the null and alternative hypotheses

The Null Hypothesis, H_0

- **Null:** typically represents the hypothesis that there is “no association” or “no difference”
  - For example, there is no association between oral contraceptive use and blood pressure
    - H_o: $\mu = 0$

- **Alternative:** the very general complement to the null
  - For example, there is an association between blood pressure and oral contraceptive use
    - H_A: $\mu \neq 0$

Hypothesis Testing

- We are testing both hypotheses at the same time
  - Our result will allow us to either:
    - "Reject H_0"

      or
    - "Fail to reject H_o"

- We start by assuming the null (H_o) is true, and asking:
  - How likely is the result we got from our sample if H_o is the truth —i.e., no change in mean blood pressure after taking OCs?
  - $\overline{x}$ would have to be far from zero to claim H_A is true
    - But is $\overline{x}$ = 4.8 mmHg big enough to choose H_A ?

Hypothesis Testing Question

- Is our sample result “unlikely” when H_o is true—and therefore we should H_0 in favor of H_A?
  - We need some measure of how probable the result from our sample is, if the null hypothesis is true
  - Need the probability of having gotten such an extreme sample mean as 4.8 if the null hypothesis (H_0: $\mu = 0$) was true?
    - This probability is called the *p-value*

Hypothesis Testing Question

- Does our sample result allow us to reject H_0 in favor H_A?
  - If that probability (p-value) is small, it suggests the observed result cannot be easily explained by chance

- So, what can we turn to to evaluate how unusual our sample statistic is when the null is true?
  - We need a mechanism that will explain the behavior of the sample mean across many different random samples of 10 women—when the truth is that oral contraceptives do not affect blood pressure
  - Luckily, we've already defined this mechanism: it's the *sampling distribution of the sample mean!*

Sampling Distribution

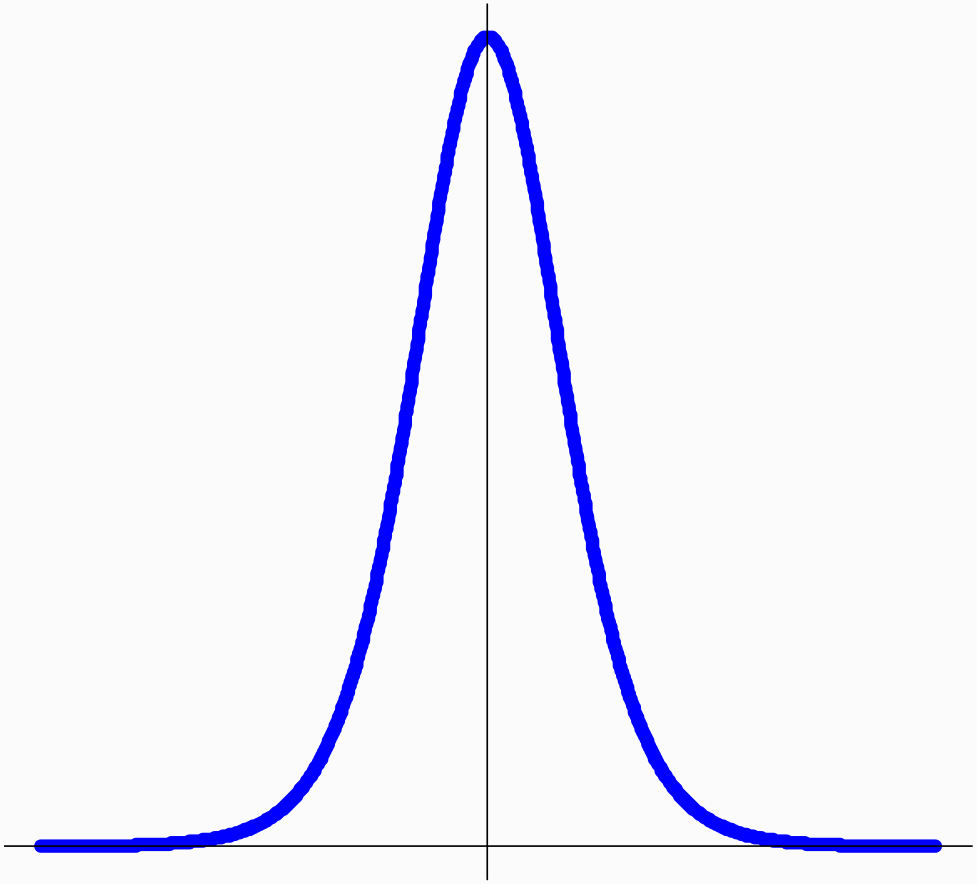
- *Sampling distribution of the sample mean* is the (theoretical) distribution of all possible values of $\bar{x}$ from samples of same size, *n*

- For BP example, theory tells us it is a t_9 *distribution*

- Recall, the sampling distribution is centered at the "truth," the underlying value of the population mean, μ
  - In hypothesis testing, we start under the assumption that H_0 is true—so the sampling distribution under this assumption will be centered at μ_0, the null mean

Sampling Distribution

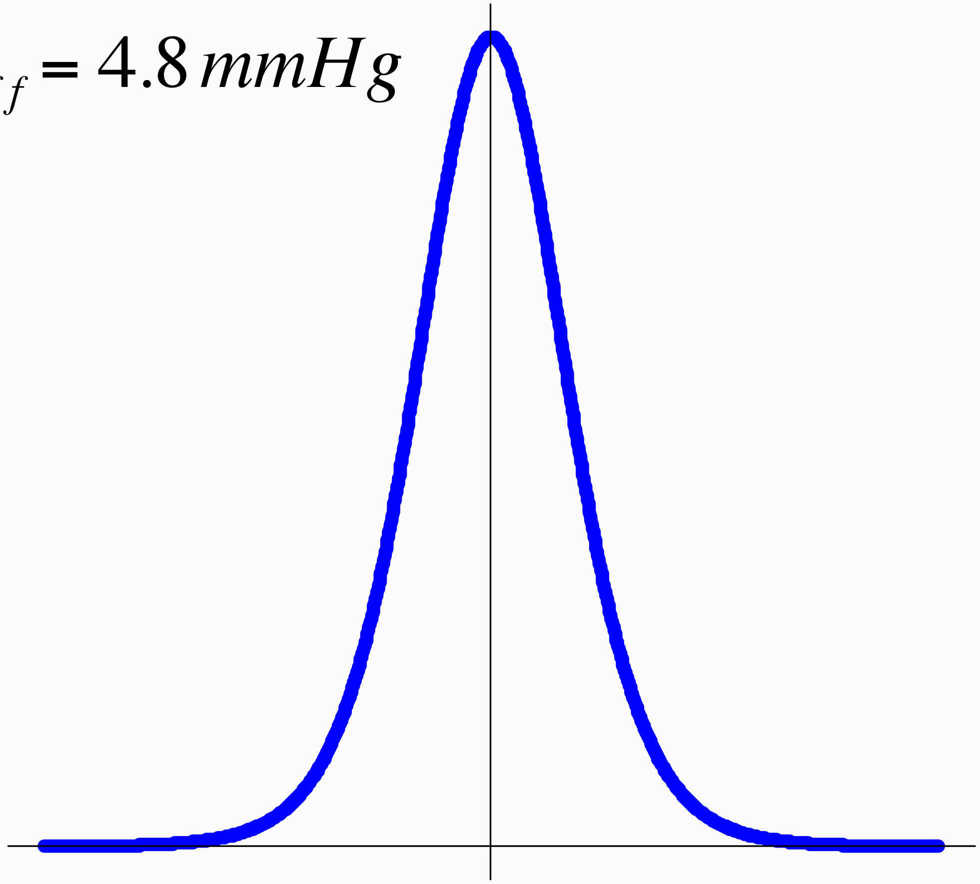
- Sampling distribution of sample mean differences (after–before) in BP, from samples of size n=10



Getting a p-Value

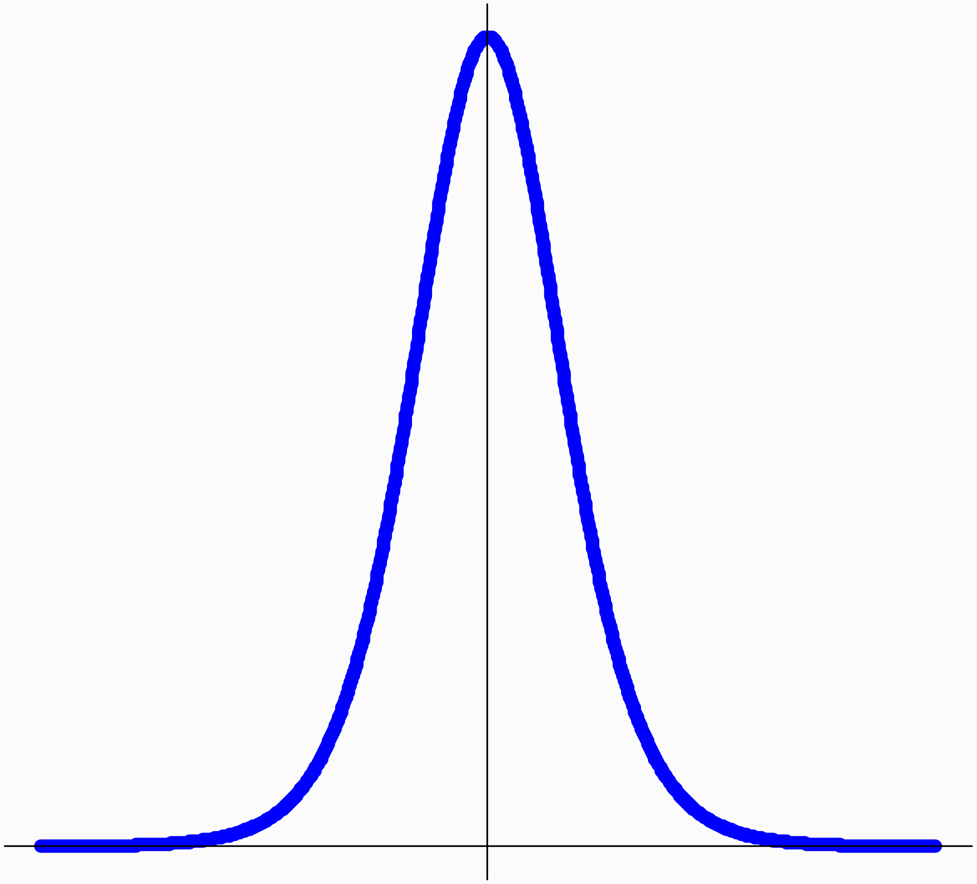
- To compute a p-value, we need to find our value of $\bar{x}_{diff}$ on the graph and figure out how "unusual" it is

- Recall: $\bar{x}_{diff} = 4.8\,mmHg$



Getting a p-Value

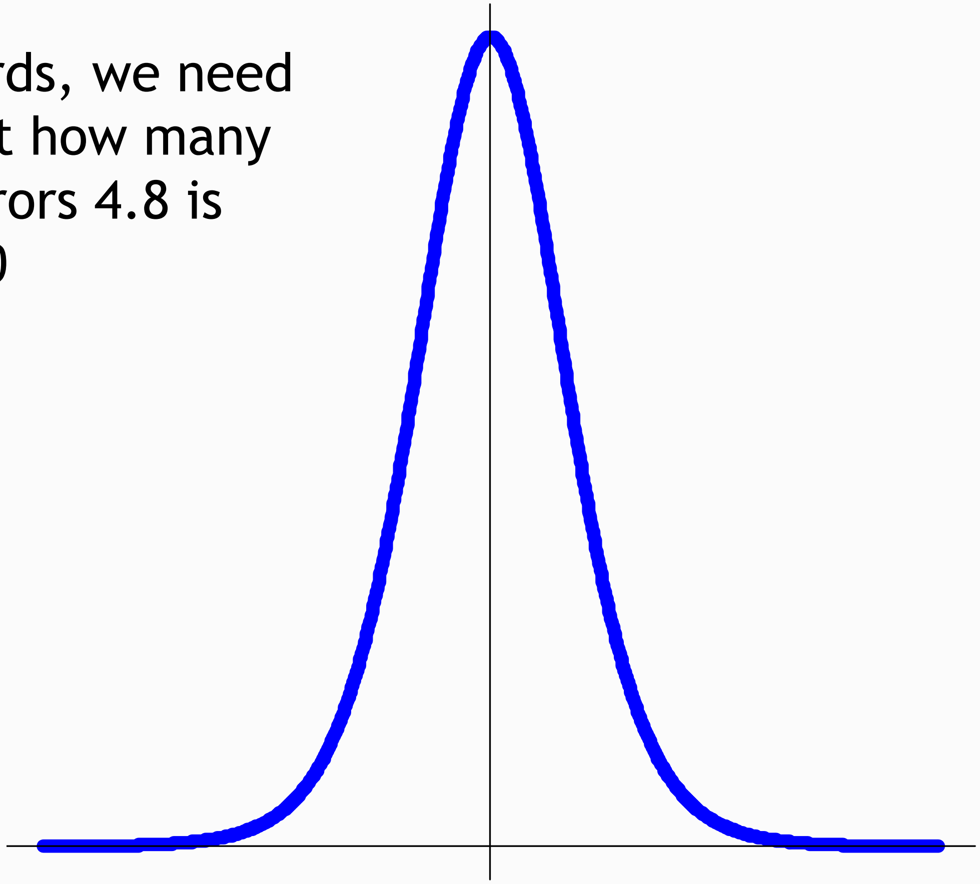
- Where is $\bar{x}_{diff} = 4.8\, mmHg$ under the curve?



Getting a p-Value

- We need to figure out how "far" our result—4.8—is from 0, in "standard statistical units"

- In other words, we need to figure out how many standard errors 4.8 is away from 0



How Are p-Values Calculated?

- Calculate the distance in standard errors
  - Called a *t-statistic,* but synonymous with *z-score,* normal score, etc.—think of it as a distance

$$t = \frac{\bar{x}_{diff} - 0}{S\hat{E}(\bar{x})}$$

$$t = \frac{4.8 - 0}{4.6/\sqrt{10}} = \frac{4.8}{1.45} \approx 3.3$$

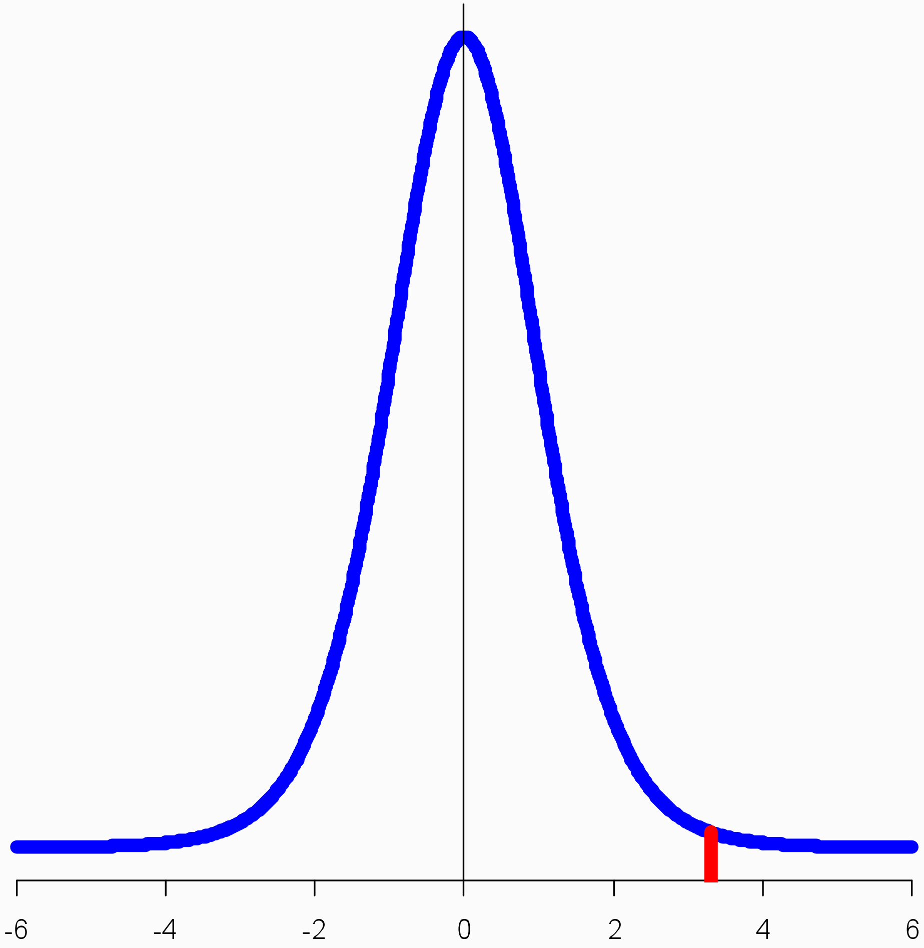
How Are p-Values Calculated?

- We observed a sample mean that was 3.3 standard errors of the mean (SEM) away from what we would have expected the mean to be if OC use were not associated with blood pressure

- Is a result 3.3 standard errors above its mean unusual?
  - Lets see where it falls on the sampling distribution

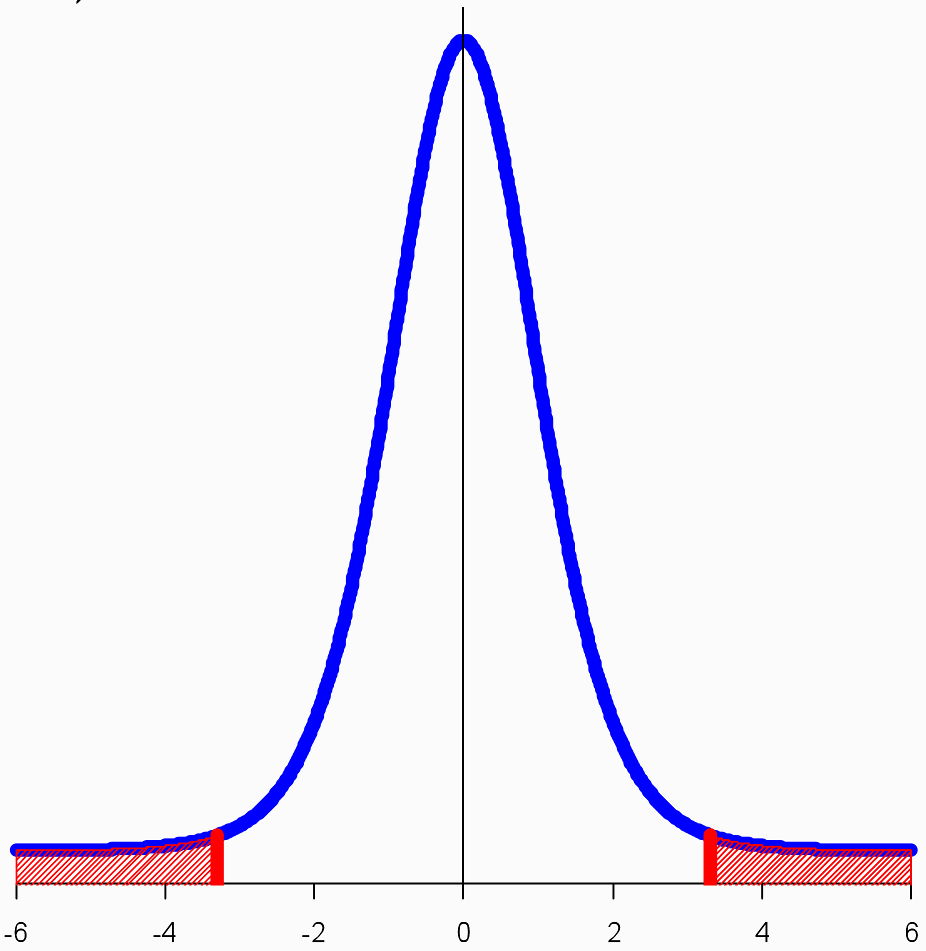
How Are p-Values Calculated?

- 3.3 on the sampling distribution (t_9)



How Are p-Values Calculated?

- The p-value is the probability of getting a sample result as (or more) extreme than what you observed (3.3) away from $\mu_0 = 0$ (in either direction from 0)



How Are p-Values Calculated?

- We could look this up in a t-table . . .
- Better option—let Stata do the work for us!

How to Use STATA to Perform a Paired t-Test

- At the command line:

`ttesti` n s $\bar{x}_{diff}$ $\mu 0$

- For the BP-OC data:

```
ttesti  10  4.8  4.6  0
```

Stata Output

- Using *ttesti*

```
. ttesti 10 4.8 4.6 0

One-sample t test
------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |      10         4.8    1.454648         4.6    1.509358    8.090642
------------------------------------------------------------------------------
    mean = mean(x)                                                t =   3.2998
Ho: mean = 0                                     degrees of freedom =        9

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.9954         Pr(|T| > |t|) = 0.0092          Pr(T > t) = 0.0046
```

Stata Output

- 95% CI

```
. ttesti 10 4.8 4.6 0

One-sample t test
------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |      10         4.8    1.454648         4.6    1.509358    8.090642
------------------------------------------------------------------------------
    mean = mean(x)                                                t =   3.2998
Ho: mean = 0                                     degrees of freedom =        9

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.9954         Pr(|T| > |t|) = 0.0092          Pr(T > t) = 0.0046
```

Stata Output

- p-value

```
. ttesti 10 4.8 4.6 0

One-sample t test
------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |      10         4.8    1.454648         4.6    1.509358    8.090642
------------------------------------------------------------------------------
    mean = mean(x)                                                t =   3.2998
Ho: mean = 0                                     degrees of freedom =        9

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.9954         Pr(|T| > |t|) = 0.0092          Pr(T > t) = 0.0046
```

Interpreting the p-Value

- The p-value in the blood pressure/OC example is **.0092**
  - Interpretation: if the true before OC/after OC blood pressure difference is 0 among all women taking OCs, then the chance of seeing a mean difference as extreme/more extreme as 4.8 in a sample of 10 women is **.0092**

Using the p-Value to Make a Decision

- We now need to use the p-value to choose a course of action: either reject H_0, or fail to reject H_0
  - We need to decide if our sample result is unlikely enough to have occurred by chance if the null was true
    - Our measure of this “unlikeliness” is $p = 0.0092$

Using the p-Value to Make a Decision

- Establishing a cutoff
  - In general, to make a decision about what p-value constitutes "unusual" results, there needs to be a cutoff, such that all p-values less than the cutoff result in rejection of the null
  - Standard cutoff is .05—this is an arbitrary value
  - Cut off is called *alpha-level* of the test

Using the p-Value to Make a Decision

- Establishing a cutoff
  - Frequently, the result of a hypothesis test with a p-value less than .05 (or some other arbitrary cutoff) is called *statistically significant*
  - At the .05 level, we have a statistically significant blood pressure difference in the BP/OC example

Blood Pressure: Oral Contraceptive Example

- Statistical method
  - The changes in blood pressures after oral contraceptive use were calculated for 10 women
  - A paired t-test was used to determine if there was a statistically significant change in blood pressure, and a 95% confidence was calculated for the mean blood pressure change (after-before)

Blood Pressure: Oral Contraceptive Example

- Result
  - Blood pressure measurements increased on average 4.8 mmHg with standard deviation 4.6 mmHg
  - The 95% confidence interval for the mean change was 1.5 mmHg–8.1 mmHg
  - The blood pressure measurements after oral contraceptive use were statistically significantly higher than before oral contraceptive use (p=.009)

Blood Pressure: Oral Contraceptive Example

- Discussion
  - A limitation of this study is that there was no comparison group of women who did not use oral contraceptives
  - We do not know if blood pressures may have risen without oral contraceptive usage