Comparing Two Groups
Matched Pairs

Up until now, we have learned inferential procedures for what we called the one sample case; that is, we have been interested in a single proportion, such as the percentage of people who support the mayor, or a single mean, such as the mean body temperature for healthy adults. However, many questions that we might want to study require us to have two sets of data, or to look at the same measurement taken under two different conditions. So now we are going to look at expanding our statistical toolbox to include tests and confidence intervals to see if two measurements, or two groups, differ.

EXAMPLE 1
We are going to go back now to our Anthropology data where we were interested in looking at the error in age estimate, that is, the difference between estimated age and actual age at death.

We have looked at the estimates given by the DiGangi method, in which age estimates were made by examining the first rib. Previously we learned that the DiGangi method tends to underestimate the age of death, with the errors in age estimation negative for more than 75% of the observations. This is graphically visible in Figure **??**. And we can see that on average the DiGangi method underestimates the age of death by more than 14 years.

When we first introduced this data set, we also discussed the Suchey-Brooks method. This method estimates the age from the pubic bone, a weight bearing part of the body, so it may perform differently than the DiGangi method. Looking at some summary statistics and the boxplot for the error in age estimation from the Suchey-Brooks method, shown in Figure **??**, we can also see that it tends to underestimate. We see that 75% of our observations are zero or less and the average estimated error in age estimation is more than 7 years. So both methods underestimate the actual age.

Figure 1: Summary statistics and box plots for error in age estimation
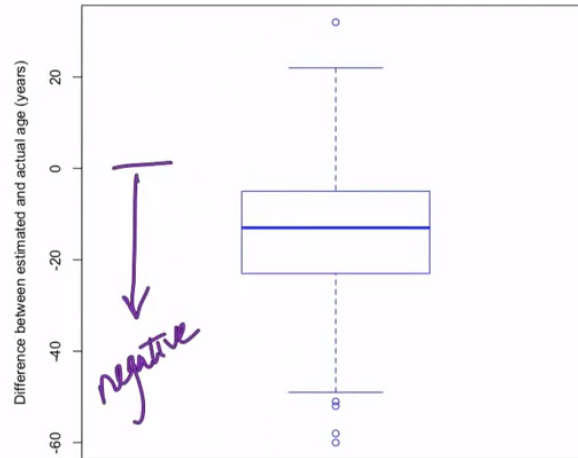
(a) DiGangi method

(b) Suchey-Brooks method

But maybe the Suchey-Brooks method performs a little bit better. If we look at the box plots for the two methods side-by-side in Figure ??, we can see that the values for the Suchey-Brooks method seem a little closer 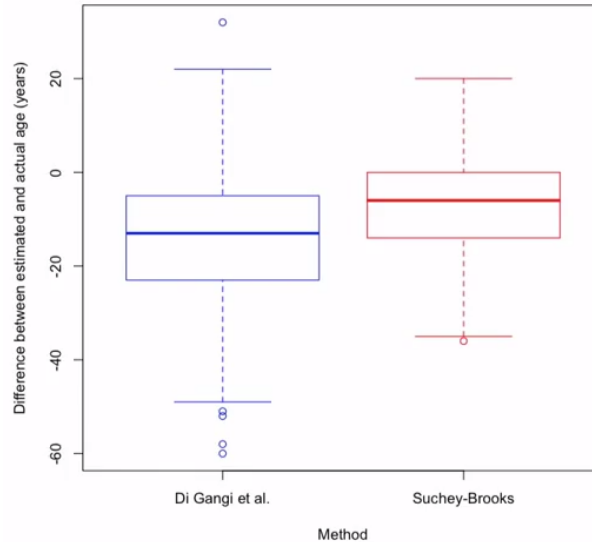to zero, which is what we would get if the actual and estimated age were the same. But there is a lot of overlap in these box plots, so you

2

may wonder if the difference between the two methods is significant.

Figure 2: Side-by-side box plots of error in age estimation for the DiGangi and Suchey-Brooks method

Before answering that question, there is something fundamentally wrong with the representation in the box plots in Figure **??**. The estimates by each method were taken on the same skeletons, but we don't know this from the plot, in which we cannot tell how the measurements on each skeleton coincide. For example, our largest underestimate using the DiGangi method was a whopping 60 years below the actual age. Using the Suchey-Brooks method, the largest underestimate was only 36 years. But are these even the same skeleton? Because these side-by-side box blots don't match up measurements on the same skeleton, this is not the best representation of our data.

This is an example of **matched pairs**. Matched pairs are useful when we want to compare two conditions or treatments or methods, on the same or similar observational or experimental units. The units could be the same subject or skeleton, or perhaps something like twins or siblings. Some examples of matched pairs include

- two age estimates using different methods on the same skeleton,
- pre and post-test scores on the same person,
- measurements in pairs at the same time or place, and
- compare an outcome with and without a treatment on the same subject.

The last example is referred to as a **crossover study**: an *experimental design* using matched pairs. In a crossover study it is important to randomize which experimental unit in the pair gets which treatment, or the treatment order if both treatments are being compared on

the same experimental unit. Measurements on a matched pair are *not* independent, so the analysis must reflect the pairing. Using matched pairs allows the researcher to control for the variables that differ among independent observational or experimental units. As a result, there is less variability, or greater precision, and thus more power in our tests and narrower confidence intervals in our estimate of the difference between the treatments or conditions being compared.

In carrying out the analysis to compare the DiGangi and Suchey-Brooks methods, the general approach is to treat each pair as one observation. For each skeleton, we calculate the difference in the error in each estimation between the Suchey-Brooks and the DiGangi methods, and then we carry out the analysis on these *differences*. Since there will only be one observation per skeleton now, it simplifies the analysis because we already know methods for the one sample case. And since the skeletons are independent, so are the single observations. In Figure **??**, we see the new dataset which includes the new variable *Differences*, calculated as the difference between the Suchey-Brooks method and the DiGangi method.

```
Observation  DiGangiError  SucheyBrooksError  Difference
     1           -34            -18              16
     2           -12             -9               3
     3           -40            -11              29
     4           -15             2               17
     5           -28            -14              14
     6            -9             1               10
     7           -18            -15               3
     8           -23            -12              11
     9           -31            -24               7
    10           -16            -14               2
    11           -26            -23               3
    12           -29             0               29
    13            -8             -4               4
    14           -23            -21               2
    15           -16            -14               2
    16           -18             -7              11
    17           -23             3               26
    18           -42            -33               9
```

Figure 3: Partial skeleton dataset including new variable *Difference*

For example, we can calculate for the first observation

$$\text{Difference} = \text{SucheyBrooksError} - \text{DiGangiError}$$
$$= -18 - (-34)$$
$$= 16$$

Before we proceed, there is one small item we need to point out about these data. For two of the observations the Suchey-Brooks estimate is missing, displayed in Figure **??**. Different statistical packages use different symbols for missing data and in R the symbol is NA. We originally choose the sample of 400 skeletons that we have been working with in order to have a DiGangi estimate for each, so the skeletons all have their first rib intact. But that does not mean that all other bones were intact, and indeed two of our 400 skeletons had damaged or missing pubic bones, so the Suchey-Brooks estimate was not possible. But we

don't expect that a missing bone is related to age of death or its estimate, so we don't expect that leaving out missing observations will bias the results. We'll continue with the 398 skeletons for which we have both estimates for both methods so that we can compare them.

| Observation	DiGangiError	SucheyBrooksError	Difference
.	.	.	.
.	.	.	.
.	.	.	.
225	−19	−5	14
226	−38	−9	29
227	−11	7	18
228	−14	NA	NA
229	−29	−27	2
230	5	15	10
231	−4	−19	−15
232	−10	−6	4
233	−36	−22	14
234	3	−8	−11
.	.	.	.
.	.	.	.
.	.	.	.
335	−11	5	16
336	4	−5	−9
337	−1	3	4
338	−29	NA	NA
339	−60	−25	35
340	−38	−24	14
341	−52	−17	35
342	−12	5	17

Figure 4: Partial skeleton dataset with missing observations highlighted

Let's have a look at these 398 differences in Figure **??**. From the plots, we can see that their distribution is fairly symmetric and it is centered above zero. Indeed the average difference is 6.854, so the difference in the error in the two methods are on average almost seven years. Since both methods tend to underestimate age at death, so their errors are negative, and in this case we are subtracting the DiGangi errors from the Suchey-Brooks errors, a positive mean is an indication of higher error estimates in the Suchey-Brooks method. This means that they are closer to zero on average, so it seems we have an indication that the Suchey-Brooks method is more accurate. One more statistic to note from these differences is the standard deviation of them which is 11.056 years.

Figure 5: Plots of differences in age estimation between DiGangi and Suchey-Brooks method
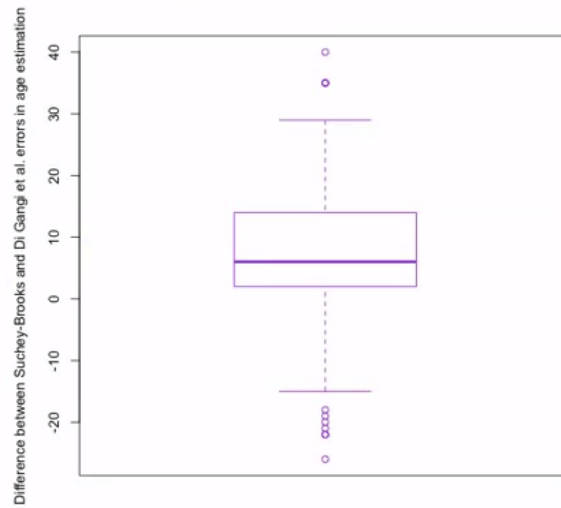
(a) Boxplot of differences
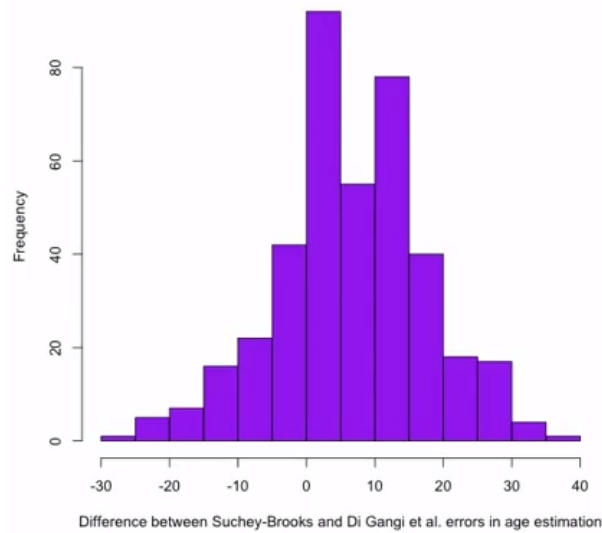
(b) Histogram of differences

We will define d to be

$$d \equiv \text{Error in age estimate for Suchey-Brooks method} - \text{Error in DiGangi method}$$

and henceforth, any statistic or parameter with a subscript d will indicate that it is calculated from the difference.

To test whether the mean error in age estimation differs between these two methods, we will specify our null and alternative hypothesis as

$$H_0 : \mu_d = 0 \text{ vs } H_a : \mu_d \neq 0$$

6

The statistics we will need are $\bar{x}_d = 6.854$, $s_d = 11.056$, and $n = 398$.

We can then calculate our test statistic

$$
\begin{aligned}
\text{Test statistic} &= \frac{\bar{x}_d - \mu_d}{s_d/\sqrt{n}} \\
&= \frac{6.854 - 0}{11.056/\sqrt{398}} \\
&\doteq 12.37
\end{aligned}
$$

If the null hypothesis is true, this test statistic should be a value from a t-distribution with $df = 397$. With such a high degree of freedom, this will be quite close to a standard normal distribution, and in a standard normal distribution, we know that 99.7% of the data values are within 3 standard deviations of the mean. So a value greater than 12.37 or less than negative 12.37 is very unlikely. Indeed, the p-value in this case is very, very small. With such an extremely small p-value, we can conclude that we have extremely strong evidence to reject the null hypothesis. That is, the mean of the difference between the methods in the errors in age estimation is not zero. So the Suchey-Brooks method and the DiGangi method don't give the same mean error in age estimation.

To provide a range of plausible values for the difference in the error in age estimation between the Suchey-Brooks and the DiGangi methods, we can construct a 95% confidence interval

$$
\begin{aligned}
&\bar{x}_d \pm t_{397,0.025} s_d/\sqrt{n} \\
&= 6.854 \pm 1.966 \times 11.056/\sqrt{398} \\
&\doteq [5.76, 7.94]
\end{aligned}
$$

We estimate that the true mean difference in the error in age estimation between the Suchey-Brooks and DiGangi methods is possibly as low as 5.76 and as high as almost 8 years. From our statistical test, we had very strong evidence that the mean of the differences was not 0, and correspondingly, we don't see 0 within this confidence interval.

For all of the inferential procedures that we have learned, we need independent observations. In matched pairs designs, this isn't the case. If we have two measurements on the same skeleton, we expect that those two measurements will be related to each other. However, we can combine them into a single measurement typically by taking a difference and then we have one set of independent observations as long as the original observations are independent. Reducing the problem means we can now analyze the data using the one sample procedures that we have already learned.