```
> t.test(pre, post) #WRONG!

        Welch Two Sample t-test

data:  pre and post
t = 2.6242, df = 19.92, p-value = 0.01629
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  270.5633 2370.3458
sample estimates:
mean of x mean of y
 6753.636  5433.182
```

The number symbol (or "hash") # introduces a comment in R. The rest of the line is skipped.

It is seen that t has become considerably smaller, although still significant at the 5% level. The confidence interval has become almost four times wider than in the correct paired analysis. Both illustrate the loss of efficiency caused by not using the information that the "pre" and "post" measurements are from the same person. Alternatively, you could say that it demonstrates the gain in efficiency obtained by planning the experiment with two measurements on the same person, rather than having two independent groups of pre- and postmenstrual women.

5.7 The matched-pairs Wilcoxon test

The paired Wilcoxon test is the same as a one-sample Wilcoxon signed-rank test on the differences. The call is completely analogous to `t.test`:

```
> wilcox.test(pre, post, paired=T)
        Wilcoxon signed rank test with continuity correction

data:  pre and post
V = 66, p-value = 0.00384
alternative hypothesis: true location shift is not equal to 0

Warning message:
In wilcox.test.default(pre, post, paired = T) :
  cannot compute exact p-value with ties
```

The result does not show any material difference from that of the t test. The p-value is not quite so extreme, which is not too surprising since the Wilcoxon rank sum cannot get any larger than it does when all differences have the same sign, whereas the t statistic can become arbitrarily extreme.

Again, we have trouble with tied data invalidating the exact p calculations. This time it is the two identical differences of -1540.

In the present case it is actually very easy to calculate the exact p-value for the Wilcoxon test. It is the probability of 11 positive differences + the probability of 11 negative ones, $2 \times (1/2)^{11} = 1/1024 = 0.00098$, so the approximate p-value is almost four times too large.

5.8 Exercises

5.1 Do the values of the `react` data set (notice that this is a single vector, not a data frame) look reasonably normally distributed? Does the mean differ significantly from zero according to a t test?

5.2 In the data set `vitcap`, use a t test to compare the vital capacity for the two groups. Calculate a 99% confidence interval for the difference. The result of this comparison may be misleading. Why?

5.3 Perform the analyses of the `react` and `vitcap` data using nonparametric techniques.

5.4 Perform graphical checks of the assumptions for a paired t test in the `intake` data set.

5.5 The function `shapiro.test` computes a test of normality based on the degree of linearity of the Q–Q plot. Apply it to the `react` data. Does it help to remove the outliers?

5.6 The crossover trial in `ashina` can be analyzed for a drug effect in a simple way (how?) if you ignore a potential period effect. However, you can do better. Hint: Consider the intra-individual differences; if there were *only* a period effect present, how should the differences behave in the two groups? Compare the results of the simple method and the improved method.

5.7 Perform 10 one-sample t tests on simulated normally distributed data sets of 25 observations each. Repeat the experiment, but instead simulate samples from a different distribution; try the t distribution with 2 degrees of freedom and the exponential distribution (in the latter case, test for the mean being equal to 1). Can you find a way to automate this so that you can have a larger number of replications?