7.4 Inferences Using Paired Data

In a *matched pairs* experiment, choose subjects _____________ to minimize __________________________. Then randomly assign the treatment to ___________________. Since differences within pairs _____________ treatment have been minimized, differences in pairs _____________ treatment should be due mostly to ___________________. Ideal examples include

- Testing a drug on pairs of ________________________________ to get rid of variability across subjects due to age, sex, genetics, etc. Treatments are assigned randomly within each pair.
- Testing _______________ subjects twice, ______________________________, to reduce effect of variability ___________________. The order of tests is randomly chosen for each subject.

Notation:

- $\{(X_1, Y_1), \cdots, (X_n, Y_n)\}$: raw _____________ data
- $\{D_i = X_i - Y_i\}$: a random sample from a population of _____________
- $\mu_D = \mu_X - \mu_Y$: unknown mean _____________; its point estimate is _____________
- σ_D: unknown standard deviation of differences; its point estimate is _____________

Apply one-sample procedures (______ for large n, ______ for small n) to the differences, $\{D_i\}$.

Studying the differences within matched pairs is the ___________________ of this section. Everything that follows, below, is review.

Matched-Pairs Confidence Intervals

Suppose $D_1, \cdots, D_n$ is a random sample from a population of differences within pairs. Recall:

A $(100\%)(1-\alpha)$ confidence interval for μ_D contains μ_D for a proportion $1-\alpha$ of random samples.

- (§5.2) For large n ($n > 30$), the interval is

- (§5.4) For small n and an $\approx$ _____________ population of differences, the interval is

Matched-Pairs Hypothesis Tests

Suppose $D_1, \cdots, D_n$ is a random sample from a population of differences within pairs. Recall:

To test $H_0 : \mu_D = \mu_0$:

1. State null and alternative hypotheses, H_0 and H_1
2. Check assumptions: Are $\{D_i\}$ a random sample? For small n, are $\{D_i\} \approx$ normal?
3. Find the test statistic depending on n:
   - (§6.1) For large n, use $Z = \dfrac{\bar{D} - \mu_0}{s_D/\sqrt{n}}$, which is $\sim$_____________ ($\approx$) under H_0
   - (§6.4) For small n, use $T = \dfrac{\bar{D} - \mu_0}{s_D/\sqrt{n}}$, which is $\sim$_____________ under H_0
4. Find the P-value as usual and draw a conclusion

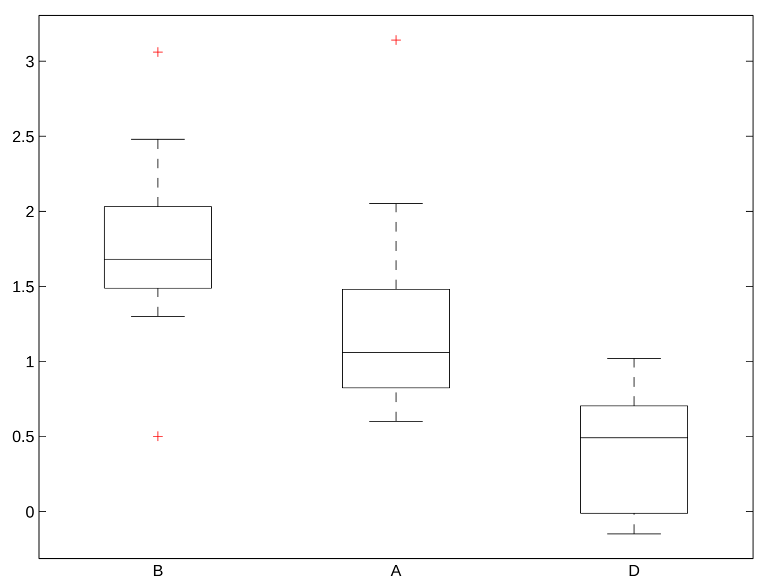
Examples

e.g. (Hollander & Wolfe, 1973; thanks to Guilherme Ludwig) Here are measurements from the Hamilton depression scale for 9 patients before (B) and after (A) taking a tranquilizer:

B	1.83	.50	1.62	2.48	1.68	1.88	1.55	3.06	1.30
A	.88	.65	.60	2.05	1.06	1.29	1.06	3.14	1.29

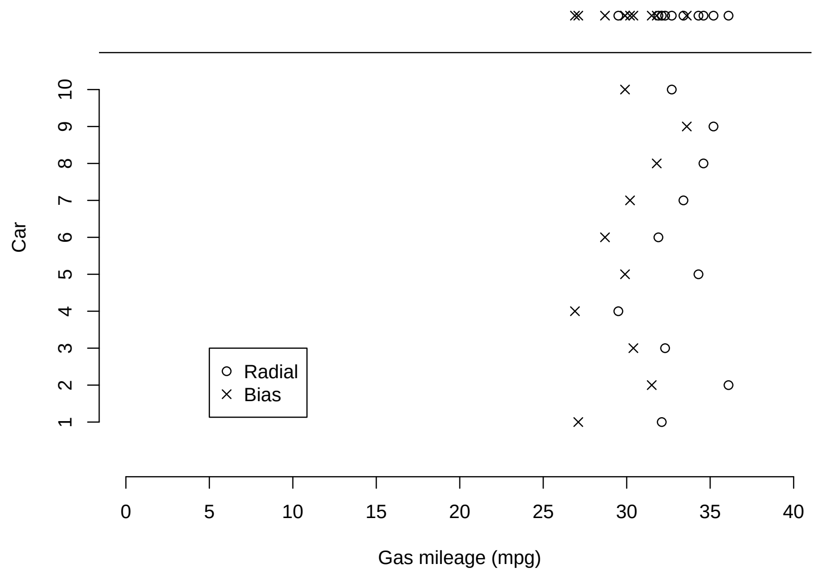
- Make boxplots of B and A. Give two reasons why the §7.3 small-sample test for the difference of two means is a poor choice.

- Plot the differences, $D = B - A$. Is a matched pairs t test of $H_0 : \mu_D = 0$ vs. $H_1 : \mu_D > 0$ reasonable? If so, do it.



e.g. (p. 302 #17) A taxicab company is trying to decide if it should switch from bias tires to radial tires to improve fuel economy. Each of 10 taxis was equipped with one of the two tire types and driven on a test course. Without changing drivers, tires were then switched to the other type and the test course was repeated. The fuel economy (in mpg) for the 10 cars is as follows:

Car	Radial	Bias	D
1	32.1	27.1	5.0
2	36.1	31.5	4.6
3	32.3	30.4	1.9
4	29.5	26.9	2.6
5	34.3	29.9	4.4
6	31.9	28.7	3.2
7	33.4	30.2	3.2
8	34.6	31.8	2.8
9	35.2	33.6	1.6
10	32.7	29.9	2.8
			$\bar{d} = 3.21$
			$s_D = 1.13$



a. Because switching tires on the taxi fleet is expensive, management does not want to switch unless a hypothesis test provides strong evidence that the mileage will be improved. State the appropriate null and alternate hypotheses, and find the P-value.

b. It will be profitable to switch to radial tires if the mean mileage improvement is greater than 2 mpg. Should the switch be made?