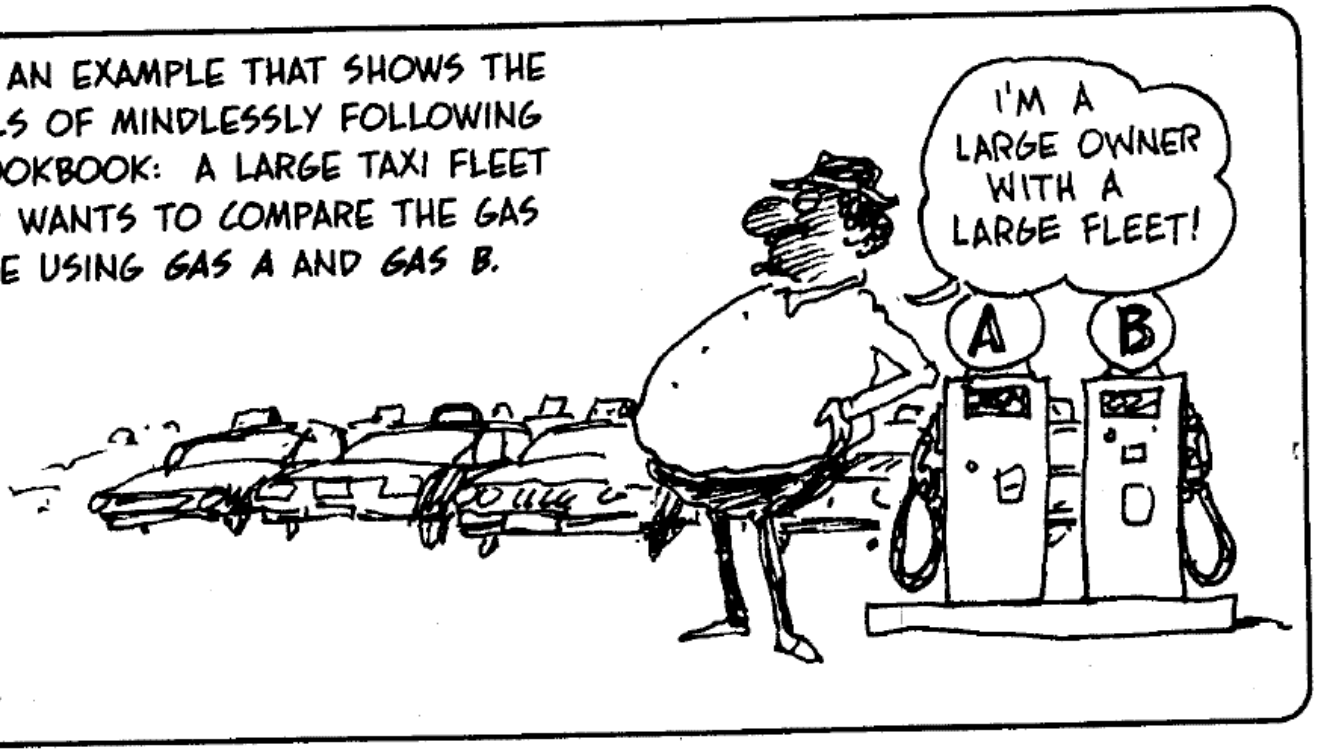
THE CARTOON GUIDE TO STATISTICS

LARRY GONICK
Author of *The Cartoon History of the Universe*
& WOOLLCOTT SMITH

HERE'S AN EXAMPLE THAT SHOWS THE PITFALLS OF MINDLESSLY FOLLOWING THE COOKBOOK: A LARGE TAXI FLEET OWNER WANTS TO COMPARE THE GAS MILEAGE USING *GAS A* AND *GAS B*.



STARTING WITH 100 CABS, HE RANDOMLY ASSIGNS 50 TO EACH GASOLINE, AND, AFTER A DAY'S DRIVING, DETERMINES

	SAMPLE SIZE	MEAN MILEAGE	STANDARD DEVIATION
A	50	25	5.00
B	50	26	4.00

To be strict, should have been t_{98}, but for large *df*, close enough to *N(0,1)*.

$$\hat{\sigma} = \sqrt{\frac{(50-1)5^2 + (50-1)4^2}{50+50-2}} = 4.5277$$

$$4.5277\sqrt{\frac{1}{50}+\frac{1}{50}} = 0.9055$$

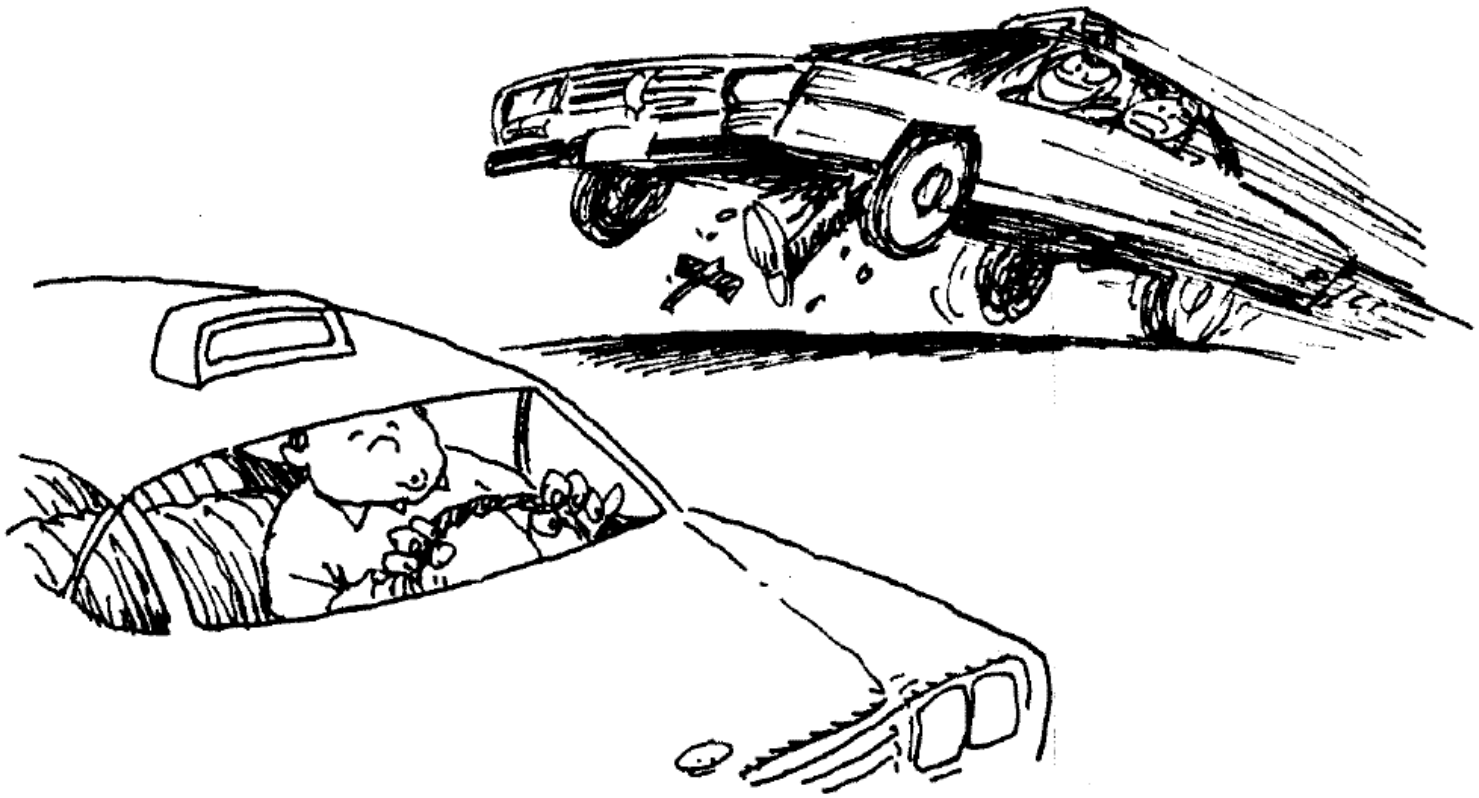
AT THE 95% CONFIDENCE LEVEL, WE HAVE

$$\mu_1 - \mu_2 = \bar{x}_1 - \bar{x}_2 \pm z_{.025}(.905)$$

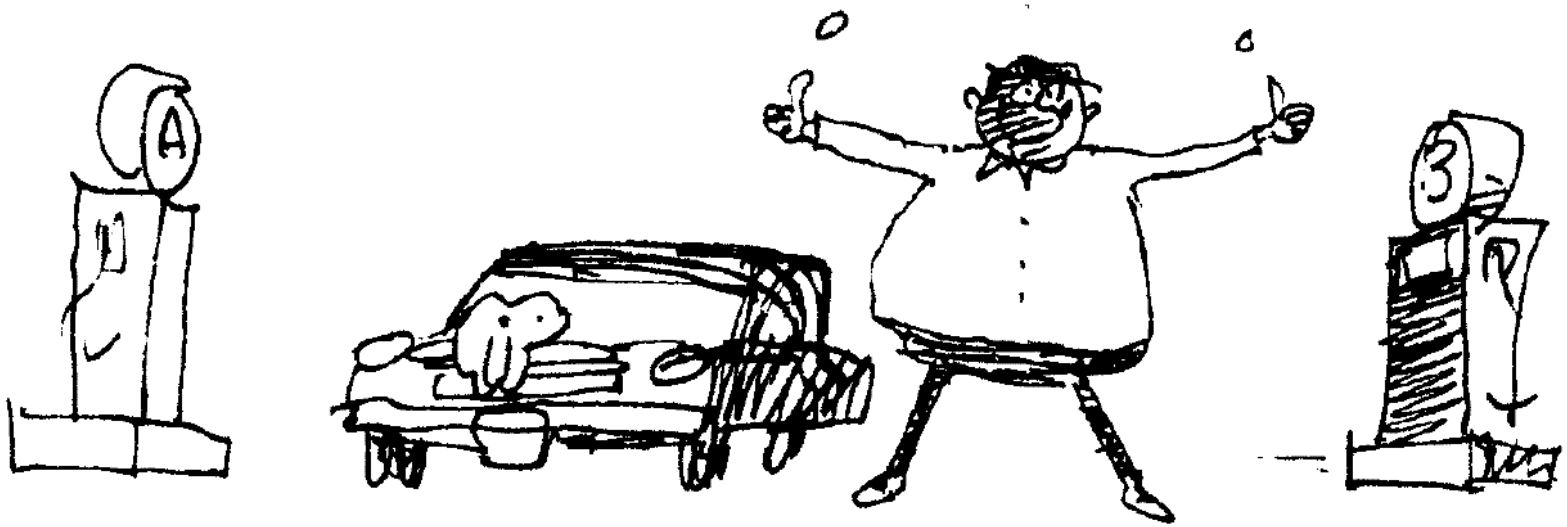
$$= -1 \pm (1.96)(.905)$$

$$= -1 \pm 1.774$$

ALTHOUGH GAS B APPEARS TO BE SLIGHTLY BETTER THAN GAS A, THE CONFIDENCE INTERVAL WAS WIDE BECAUSE OF THE LARGE STANDARD DEVIATIONS—I.E., THE ***MILEAGES VARIED WIDELY FROM ONE CAB TO THE NEXT.*** WHY SUCH HIGH VARIABILITY? BECAUSE CABS—AND CABBIES—HAVE DIFFERENT ***PERSONALITIES!***



A FAR BETTER WAY TO DO THIS STUDY IS TO ASSIGN GAS A AND GAS B TO THE *SAME CAB ON DIFFERENT DAYS.*



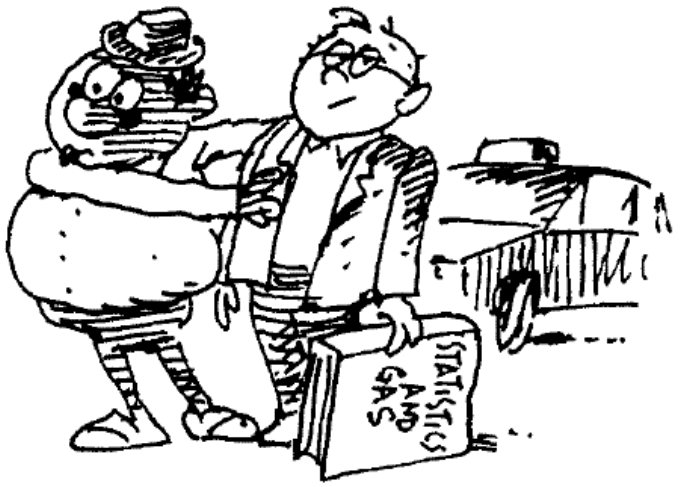
WE STILL RANDOMIZE THE TREATMENT BY FLIPPING A COIN TO DECIDE WHETHER TO USE GAS A ON TUESDAY OR WEDNESDAY. WE CAN ALSO CUT THE EXPERIMENT DOWN TO 10 CABS, SAVING THE OWNER A LOT OF TIME AND MONEY!

CAB	GAS A	GAS B	DIFFERENCE
1	27.01	26.95	0.06
2	20.00	20.44	−0.44
3	23.41	25.05	−1.64
4	25.22	26.32	−1.10
5	30.11	29.56	0.55
6	25.55	26.60	−1.05
7	22.23	22.93	−0.70
8	19.78	20.23	−0.45
9	33.45	33.95	−0.50
10	25.22	26.01	−0.79
MEAN	25.20	25.80	−0.60
STANDARD DEVIATION	4.27	4.10	0.61

NOTE THAT THE MEANS AND STANDARD DEVIATIONS OF GAS A AND GAS B ARE ABOUT THE SAME. THAT'S TO BE EXPECTED, SINCE THEY HAVE THE SAME SOURCE OF VARIABILITY AS IN THE UNPAIRED EXPERIMENT. BUT NOW THE *DIFFERENCE COLUMN* HAS A VERY *SMALL* STANDARD DEVIATION. THE DIFFERENCE COLUMN, BY COMPARING GAS PERFORMANCE WITHIN A SINGLE CAR, ELIMINATES VARIABILITY BETWEEN TAXIS.

THE DIFFERENCES d_i PROVIDE A ***SINGLE MEASURE OF DIFFERENCE FOR EACH TAXI,*** AND WE CAN USE IT TO MAKE A SMALL-SAMPLE t TEST STATISTIC:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$



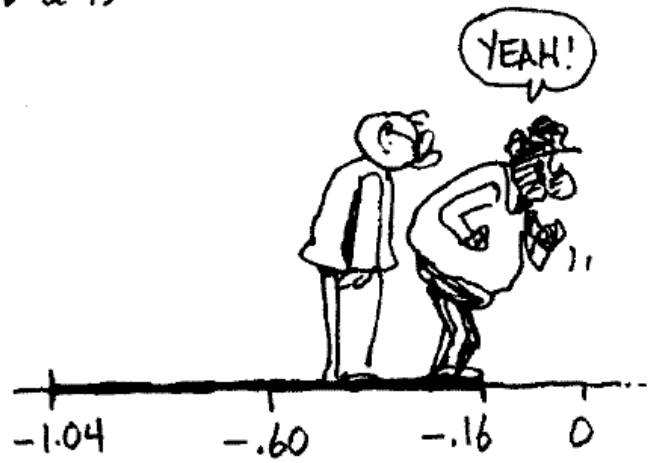
THE 95% CONFIDENCE INTERVAL AROUND $\bar{d}$ IS

$$\mu_d = \bar{d} \pm t_{.025}\left(s_d/\sqrt{n}\right)$$

SAMPLE MEAN — CRITICAL VALUE — STANDARD ERROR

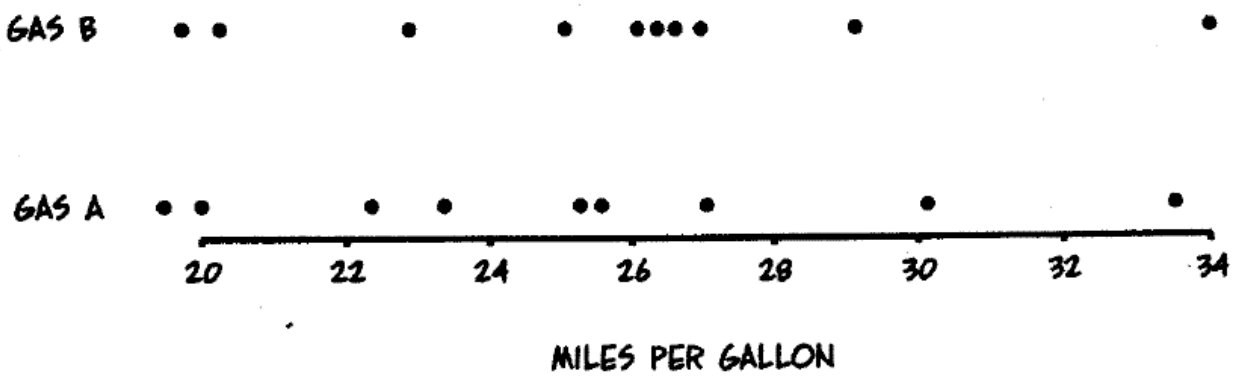
$$= -.6 \pm (2.26)\left(\frac{.61}{\sqrt{10}}\right)$$

$$= -.60 \pm .44$$

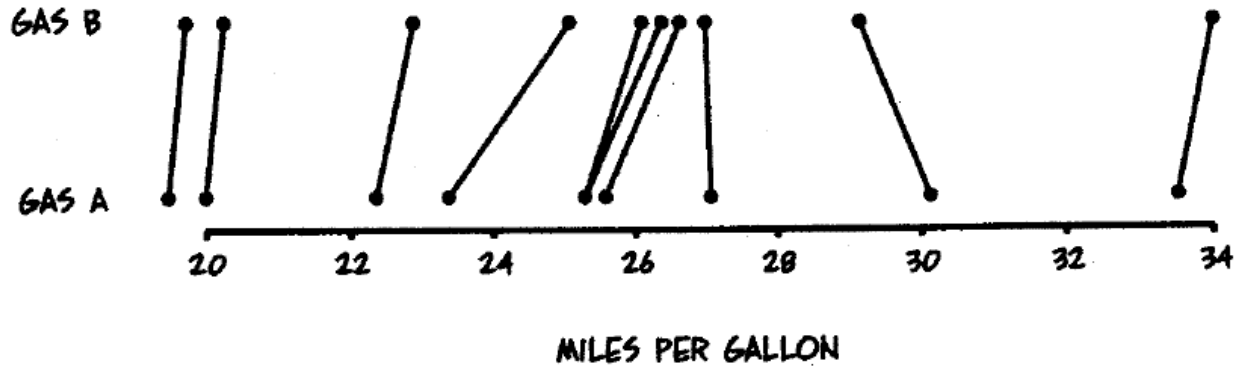


SO WE HAVE $-1.04 \leq \mu_d \leq -.16$ WITH 95% CONFIDENCE, GOOD EVIDENCE THAT GAS B REALLY IS BETTER.

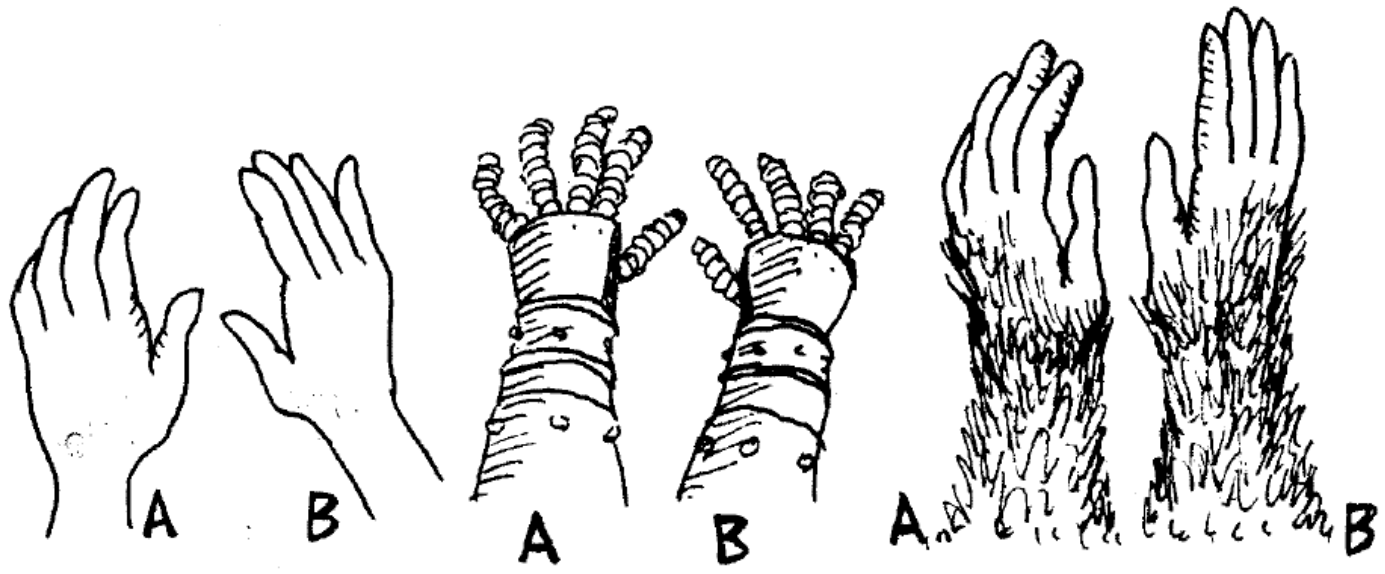
HERE ARE PLOTS OF THE GAS MILEAGE DATA: THE FIRST ONE SHOWS THE MILEAGES UNPAIRED:



AND HERE'S THE SAME DATA PAIRED BY TAXICAB.



A PAIRED COMPARISON EXPERIMENT IS ONE OF THE MOST EFFECTIVE WAYS TO REDUCE NATURAL VARIABILITY WHILE COMPARING TREATMENTS. FOR EXAMPLE, IN COMPARING HAND CREAMS, THE TWO BRANDS ARE RANDOMLY ASSIGNED TO EACH SUBJECT'S RIGHT OR LEFT HANDS. THIS ELIMINATES VARIABILITY DUE TO SKIN DIFFERENCES.



OR, IN COMPARING TWO BREAKFAST CEREALS, EACH TASTER RATES BOTH CEREALS (IN RANDOM ORDER). THE PAIRED COMPARISON REMOVES THE NATURAL BIAS OF THE TASTER FOR OR AGAINST CEREAL IN GENERAL.