STAT 401 - Statistical Methods for Research Workers

Two-sample t-test

Jarad Niemi

Iowa State University

6 September 2013

Do Japanese cars get better mileage than American cars?

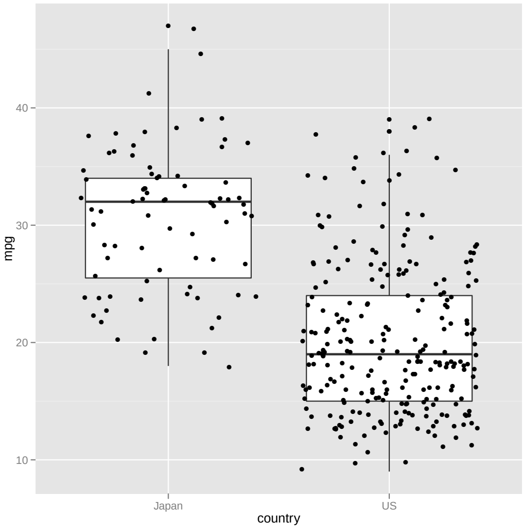
- Statistical hypothesis:

  H_0: Mean mpg of Japanese cars is the same as mean mpg of American cars.
  H_1: Mean mpg of Japanese cars is different than mean mpg of American cars.

- Statistical question:

  What is the difference in mean mpg between Japanese and American cars?

- Data collection:
  - Collect a random sample of Japan/American cars


40
30
20
10
mpg
Japan
US
country

Assumptions

Let

- Y_{1j} represent the jth Japanese car
- Y_{2j} represent the jth American car

Assume

$$Y_{1j} \stackrel{iid}{\sim} N(\mu_1, \sigma^2) \qquad Y_{2j} \stackrel{iid}{\sim} N(\mu_2, \sigma^2)$$

Restate the hypotheses using this notation

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 \neq \mu_2$

Alternatively

H_0: $\mu_1 - \mu_2 = 0$

H_1: $\mu_1 - \mu_2 \neq 0$

Test statistic

The test statistic we use here is

$$\frac{\overline{Y}_1 - \overline{Y}_2 - (\mu_1 - \mu_2)}{SE(\overline{Y}_1 - \overline{Y}_2)}$$

where

- $\overline{Y}_1$ is the sample average mpg of the Japanese cars
- $\overline{Y}_2$ is the sample average mpg of the American cars

and

$$SE(\overline{Y}_1 - \overline{Y}_2) = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \qquad s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}}$$

where

- s_1 is the sample standard deviation of the mpg of the Japanese cars
- s_2 is the sample standard deviation of the mpg of the American cars

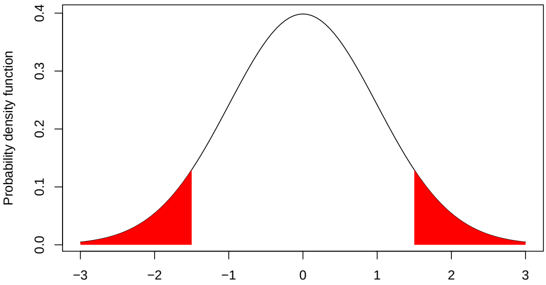
Pvalue

If H_0 is true, then $\mu_1 = \mu_2$ and the test statistic

$$t = \frac{\overline{Y}_1 - \overline{Y}_2 - (\mu_1 - \mu_2)}{SE(\overline{Y}_1 - \overline{Y}_2)} \sim t_{n_1+n_2-2}$$

where t_{df} is a t-distribution with df degrees of freedom.

Pvalue is $P(|t_{n_1+n_2-2}| > |t|) = P(t_{n_1+n_2-2} > |t|) + P(t_{n_1+n_2-2} < -|t|)$ or as a picture



Hand calculation

To calculate the quantity by hand, we need 6 numbers:

Car	N	Mean	SD
Japanese	79	30.5	6.11
American	249	20.1	6.41

Calculate

$$\begin{aligned} s_p &= \sqrt{\frac{(79-1)\cdot 6.11^2+(249-1)\cdot 6.41^2}{79+249-2}} && = 6.34 \\ SE(\overline{Y}_1 - \overline{Y}_2) &= 6.34\sqrt{\frac{1}{279}+\frac{1}{249}} && = 0.82 \\ t &= \frac{30.5-20.1}{0.82} && = 12.6 \end{aligned}$$

Finally, we are interested in finding $P(|t_{326}| > |12.6|) < 0.0001$ which is found using a table or software.

Confidence interval

Alternatively, we can construct a 100(1-α)% confidence interval. The formula is

$$\overline{Y}_1 - \overline{Y}_2 \pm t_{n_1+n_2-2}(1-\alpha/2) SE(\overline{Y}_1 - \overline{Y}_2)$$

where $\pm$ indicates plus and minus and $t_{df}(1-\alpha/2)$ is the value such that $P(t_{df} < t_{df}(1-\alpha/2)) = 1-\alpha/2$. If $\alpha = 0.05$ and $df = 326$, then $t_{df}(1-\alpha/2) = 1.97$.

The 95% confidence interval is

$$30.5 - 20.1 \pm 1.97 \cdot 0.82 = (8.73, 11.9)$$

We are 95% confident that, on average, Japanese cars get between 8.73 and 11.9 more mpg than American cars.

SAS code for two-sample t-test

```
DATA mpg;
    INFILE 'mpg.csv' DELIMITER=',' FIRSTOBS=2;
    INPUT mpg country $;

PROC TTEST DATA=mpg;
    CLASS country;
    VAR mpg;
    RUN;
```

```
                                 The TTEST Procedure

                                   Variable:  mpg

      country            N        Mean     Std Dev     Std Err     Minimum     Maximum

      Japan             79     30.4810      6.1077      0.6872     18.0000     47.0000
      US               249     20.1446      6.4147      0.4065      9.0000     39.0000
      Diff (1-2)               10.3364      6.3426      0.8190

 country       Method               Mean       95% CL Mean        Std Dev      95% CL Std Dev

 Japan                           30.4810     29.1130  31.8491      6.1077      5.2814   7.2429
 US                              20.1446     19.3439  20.9452      6.4147      5.8964   7.0336
 Diff (1-2)    Pooled            10.3364      8.7252  11.9477      6.3426      5.8909   6.8699
 Diff (1-2)    Satterthwaite     10.3364      8.7576  11.9152

                 Method           Variances        DF    t Value    Pr > |t|

                 Pooled           Equal           326      12.62      <.0001
                 Satterthwaite    Unequal      136.87      12.95      <.0001

                                Equality of Variances

                     Method        Num DF     Den DF    F Value    Pr > F

                     Folded F         248         78       1.10    0.6194
```

Conclusion

Mean miles per gallon of Japanese cars is significantly different than mean miles per gallon of American cars (two-sample t-test t=12.62, $p < 0.0001$). Japanese cars get an average of 10.3 [95% CI (8.7,11.9)] more miles per gallon than American cars.