T-Test

This test compares the means of two variables, computes the difference between the two variables for each row and tests to see if the average difference is significantly different from zero

Details / Restrictions:

- Value of t will be zero if there is no difference in the data

- One Sample t-test. To test whether a mean may be different from a non-zero value, subtract that value from each data value i.e. To test the mean number of visits to shop is equal to 3, test whether the mean of (number of visits to shop − 3) is equal to 0.

- To make a paired test between two data values, test whether the difference between them is zero.

SE = difference in the two means $\sqrt{\frac{var_1}{n_1} + \frac{var_2}{n_2}}$

Variance = $(\text{standard deviation})^2$

Hypothesis $H_0 = \mu_{new} - \mu_{usual} = 0$ (no difference in means)
H_1 $\mu_{new} - \mu_{usual} \neq 0$ (difference in means).

T test formula: $\frac{\bar{x}_{1_{new}} - \bar{x}_{2_{usual}}}{SE\ (\bar{x}_{new} - \bar{x}_{usual})}$

$Df = n_1 + n_2 - 2$

Using df and $\alpha = 0.05$ (95% confidence) create the confidence interval. If t is in this interval → no evidence to reject H_0

P if P is less than α (0.05), we have evidence against H_0 (i.e. it is outside 95% interval)

95% CI for difference: long term mean for n_1 between 2 values

Example Stats 2012 q3

no fraud vs fraud notes (measure of lenght)

fraud	N	mean	st dev	SE Mean
no	100	124.720	0.355	0.036
yes	100	130.193	0.298	0.030

Difference = $\bar{x}_1$ mu(no) vs $\bar{x}_2$ mu(yes)

Estimate for difference -0.47330 $\bar{x}_1 - \bar{x}_2$

95% CI for difference (-0.5645, -0.3815) long term CI for "no"

T test of difference 0 (vs not =): T-value = -10.20

P-Value < 0.001 DF = 198

Both used pooled St Dev = 0.3284

H_0 mean(no) - mean(yes) = 0

H_1 mean(no) - mean(yes) $\neq 0$.

$$T\ test = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{variance_1}{n_1} + \frac{var_2}{n_2}}} \quad \frac{-0.473}{\sqrt{\frac{(0.3284)^2}{100} + \frac{(0.3284)^2}{100}}} = -10.1969 = \underline{-10.2}$$

95% CI df = 198 $\alpha = 0.05$ $-1.97 < t < 1.97$

T outside interval, evidence against H_0

P-value: P (< 0.001) is < 0.05, evidence against H_0

95% CI for difference: (-0.5645, -0.38151)

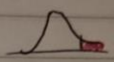
Long term mean for lenght of not fraud is between (-0.5645 and -0.3815) smaller than fraud notes

One Sided T-test.

Is the population (long term mean bounce height) from the new process higher than the population (long term mean bounce height) from the usual process?

H_0 $\mu_{new} - \mu_{usual} \leq 0$ Focusing on right tail

H_1 $\mu_{new} - \mu_{usual} > 0$



Test statistic = $\dfrac{(\bar{x}_{new} - \bar{x}_{usual}) - \text{hypothesised value}}{SE(\bar{x}_{new} - \bar{x}_{usual})}$

hypothesised value = use the worst case which is value = 0

DF = $n_{new} + n_{usual} - 2$

Need 5% area to right: $\alpha = 0.05$ do $\alpha * 2 = 0.10$ gives 0.05 each side

If t is outside interval, evidence against H_0, average of new is higher than average of usual

Example:

Difference mu(new) - mu(normal)

Estimate for difference: 2.016 difference in means

95% lower bound for difference: 0.779

T-Test for difference = 0 (vs >): T-Value = 2.64

P value = 0.004 DF = 198

Both used Pooled StDev = 5.2926

95% lower bound = Sample difference - 1.65 * SE (difference)

2.016 - 1.65 * (0.75) = 0.778

1.65 = upper bound figure

Test Statistic

like doing a 95% confidence interval

$$z = \frac{\mu_0 - \text{Sample (mean)}}{\frac{sd}{\sqrt{n}}} \qquad \frac{\text{population mean} - \text{sample mean}}{\text{standard error of sample}}$$

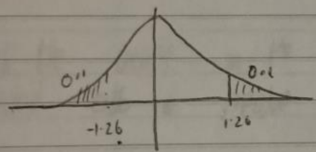
Eg what Sample mean of 248.40

Population-mean = 250

sd = 8.48

n = 50

$$SE = \frac{8.48}{\sqrt{50}} = 1.27 \qquad \frac{250 - 248.4}{\frac{8.48}{\sqrt{50}}} = 1.26 \text{ z}$$



$$2 * (1 - 0.9) = 0.2$$

- Probability of 0.2 of getting our data or a more extreme value if the null hypothesis is true

- Greater than 0.5 → not enough evidence against H_0