Statistics for Data Science-2

Week 11 Solve with us

Table of contents

1. In the past the standard deviation of weights (in grams) of salt packages filled by a machine was 40. To decrease the standard deviation of weights, a new machine is introduced. A random sample of 25 packages from the new machine showed a standard deviation of 32. Based on this data, what is your conclusion, at a significance level of 0.05, on the standard deviation of weights of salt packages filled by the new machine?
(Use $F^{-1}_{\chi^2_{24}}(0.05) = 13.85$)
   (a) Standard deviation of weights is 40.
   (b) Standard deviation of weights is less than 40.

Solution:

As per given information, the null and alternative hypothesis are given by

$$H_0 : \sigma = 40, \qquad H_A : \sigma < 40$$

Define a test statistic T as $T = S^2$.

We know that $\dfrac{(n-1)S^2}{\sigma^2} = \dfrac{24S^2}{40^2} \sim \chi^2_{24}$.

Test: reject the null hypothesis if $S^2 < c^2$.

If the significance level of the test is 0.05, then

$$
\begin{aligned}
& P(S^2 < c^2) = 0.05 \\
\Rightarrow & P\left(\frac{24S^2}{40^2} < \frac{24c^2}{40^2}\right) = 0.05 \\
\Rightarrow & P\left(\chi^2_{24} < \frac{24c^2}{40^2}\right) = 0.05 \\
\Rightarrow & \frac{24c^2}{40^2} = 13.85 \\
\Rightarrow & c^2 = \frac{1600 \times 13.85}{24} = 923.33
\end{aligned}
$$

Since $S^2 = 32^2 = 1024 > 923.33$, we will not reject the null hypothesis.
Therefore, standard deviation of weights is 40.

2. A sociologist focusing on popular culture and media believes that the average number of hours per week (hrs/week) spent on social media is different for men and women. The researcher knows that the standard deviations of amount of time spent on social media are 8 hrs/week and 10 hrs/week for men and women, respectively. Examining two independent random samples of 100 individuals each, if the average number of hrs/week spent on social media for the sample of men is 3 hours greater than that for the sample of women, what conclusion can be made from a hypothesis test where, $H_0 : \mu_M = \mu_W$ and $H_A : \mu_M \neq \mu_W$? Take $\alpha = 0.01$. Use $F_Z^{-1}(0.005) = -2.57$

   a) Reject H_0
   b) Accept H_0

Solution:

Let X_i and Y_i represent the average number of hrs/week spent on social media by men and women respectively.

$X_1, X_2, \ldots, X_{100} \sim \text{N}(\mu_1, 8^2)$ and $Y_1, Y_2, \ldots, Y_{100} \sim \text{N}(\mu_2, 10^2)$

$|\overline{X} - \overline{Y}| = 3$

Consider, $H_0 : \mu_1 = \mu_2, H_A : \mu_1 \neq \mu_2$

$T = \overline{X} - \overline{Y} \sim \text{N}(\mu_1 - \mu_2, \frac{64}{100} + \frac{100}{100})$ i.e. $\text{N}(\mu_1 - \mu_2, \frac{164}{100})$

Test: Reject H_0 if $|T| > c$.

$$\alpha = P(|T| > c \mid H_0) = P\left(|\frac{T}{\sqrt{164/100}}| > \frac{c}{\sqrt{164/100}}\right)$$

$$= P\left(|Z| > \frac{c}{\sqrt{164/100}}\right) = 2F_Z\left(\frac{-c}{\sqrt{164/100}}\right)$$

$$\Rightarrow c = -\sqrt{\frac{164}{100}} F_Z^{-1}(\alpha/2)$$

$$\Rightarrow c = -\sqrt{\frac{164}{100}} F_Z^{-1}(0.005)$$

$$\Rightarrow c = -\sqrt{\frac{164}{100}} \times (-2.57) = 3.29$$

Since, $|\overline{X} - \overline{Y}| = 3 < 3.29$

Therefore, we will accept H_0.

3. The manufacturer of a new car claims that a typical car gets a mileage of 36 kilometres per litre. We think that the mileage is less. To test our suspicion, we perform the hypothesis test with $H_0 : \mu = 36$ and $H_A : \mu < 36$. Suppose we take a random sample of 100 new cars and find that their average mileage is 36.4 kilometres per litre and sample standard deviation is 4, what does a t-test say about a null hypothesis with a significance level of 0.05?

    a) Reject H_0
    b) Accept H_0

Hint: Use $F_{t_{99}}^{-1}(0.05) = -1.66$

Solution:

Null hypothesis, $H_0 : \mu = 36$

Alternate hypothesis, $H_A : \mu < 36$

Test: Reject H_0 if $\overline{X} < c$

Given, $\alpha = 0.05$ and $\overline{X} = 36.8$
In this problem, we do not know the population variance, σ^2.
The sample variance $S^2 = 4^2$

$$\alpha = P(\overline{X} < c | \mu = 36)$$

$$\alpha = P\left(\frac{\overline{X} - 36}{\sqrt{S^2/n}} < \frac{c - 36}{\sqrt{S^2/n}}\right)$$

$$\alpha = P\left(\frac{\overline{X} - 36}{\sqrt{16/100}} < \frac{c - 36}{\sqrt{16/100}}\right)$$

$$\alpha = F_{t_{99}}\left(\frac{c-36}{\sqrt{16/100}}\right)$$

$$0.05 = F_{t_{99}}\left(\frac{c-36}{\sqrt{16/100}}\right)$$

$$c = 36 + \sqrt{\frac{16}{100}}F_{t_{99}}^{-1}(0.05)$$

$$c = 35.336$$

Since, $\overline{X} = 36.4 > c$, accept H_0.

4. The average life expectancy of a particular breed of animal is expected to be 5 years. The life length (in years) of a random sample of 8 animals of that breed are
$5.6, 4.8, 4.4, 5.2, 4.2, 5.8, 4, 5$.
Use a 0.05 level of significance to check the hypothesis that $\mu = 5$ years against the alternative that $\mu \neq 5$ years. Assume the life expectancy to be normal.
   a) Reject H_0
   b) Accept H_0

Use $s = 0.442, F_{t_7}^{-1}(0.025) = -2.36$

Solution:

The null and alternative hypothesis are defined by

$$H_0 : \mu = 5, \quad H_A : \mu \neq 5$$

Define a test statistic T as $T = \overline{X} - 5$.
Test: reject H_0 if $\mid \overline{X} - 5 \mid > c$

Notice that when H_0 is true, $\dfrac{\overline{X} - 5}{\frac{0.442}{\sqrt{8}}} \sim t_7$

Given that $\alpha = 0.05$, therefore,

$$\alpha = P(\mid \overline{X} - \mu \mid > c \mid \mu = 5)$$

$$\implies 0.05 = P\left(\mid \frac{\overline{X} - 5}{\frac{0.442}{\sqrt{8}}} \mid > \frac{c}{\frac{0.442}{\sqrt{8}}} \right)$$

$$\implies 0.05 = 2P(t_7 < \frac{-c}{\frac{0.442}{\sqrt{8}}})$$

$$\implies 0.025 = F_{t_7}(\frac{-c}{\frac{0.442}{\sqrt{8}}})$$

$$\implies -2.36 = \frac{-c}{\frac{0.442}{\sqrt{8}}}$$

$$\implies c = 0.368$$

Now, $\mid \overline{X} - 10 \mid = \mid 4.875 - 5 \mid = 0.125 < 0.368$.
Therefore, accept the null hypothesis.

5. An IITM instructor conducts two live sessions for two different classes, call it A and B, in Statistics. Session A had 25 students attending while session B had 50 students. The instructor conducted a test for the two sessions. Although there was no significant difference in mean grades, session A had a standard deviation of 12 while session B had a standard deviation of 15. Can we conclude at the 0.05 level of significance that the variability in marks of class B is greater than that of A?

   a) Yes
   b) No

   Hint: Use $F^{-1}_{F_{(49,24)}}(0.95) = 1.86$

Solution:

$H_0 : \sigma_1 = \sigma_2, H_A : \sigma_1 < \sigma_2$

Test: Reject H_0 if $\dfrac{S_B^2}{S_A^2} > 1 + c_R$

We know that, $\dfrac{S_B^2}{S_A^2} \sim F(n_2 - 1, n_1 - 1)$

$n_1 = 25, n_2 = 50$

$\Rightarrow \dfrac{S_B^2}{S_A^2} \sim F(49, 24)$

Therefore,

$$\alpha = 1 - F_{F(49,24)}(1 + c_R)$$

$$\Rightarrow 1 + c_R = F^{-1}_{F(49,24)}(1 - \alpha) = F^{-1}_{F(49,24)}(0.95)$$

$$\Rightarrow 1 + c_R = 1.86$$

Since, $\dfrac{S_B^2}{S_A^2} = \frac{15^2}{12^2} = 1.5625 < 1.86$
Therefore, we will accept H_0.
This implies that at the 0.05 level of significance the variability in marks of class B is not greater than that of A.

6. A die is tossed 150 times with the following results:

X	1	2	3	4	5	6
Observed count	30	22	20	32	23	23

Can we conclude that the die is fair at a significance level of 0.05? (use $F^{-1}_{\chi^2_5}(0.95) = 11.07$)

(a) Yes
(b) No

Solution:

If die is fair, then each outcome is equally likely to occur. That is expected counts for each outcome will be $\frac{1}{6} \times 150 = 25$.

Therefore

x	1	2	3	4	5	6
Observed count	30	22	20	32	23	23
Expected count	25	25	25	25	25	25

Define the null and alternative hypothesis as

H_0 : Samples are i.i.d. Uniform$\{1, 2, 3, 4, 5, 6\}$,

H_A : Samples are not i.i.d. Uniform$\{1, 2, 3, 4, 5, 6\}$

Value of the test statistic T is given by

$$T = \frac{(30-25)^2}{25} + \frac{(22-25)^2}{25} + \frac{(20-25)^2}{25} + \frac{(32-25)^2}{25} + \frac{(23-25)^2}{25} + \frac{(23-25)^2}{25}$$

$$= 4.64$$

We will reject the null hypothesis if $T > c$.
At a significance level of 0.05, we have

$$\begin{aligned} 0.05 &= P(T > c) \\ \Rightarrow 0.05 &= 1 - P(T \leq c) \\ \Rightarrow P(T \leq c) &= 0.95 \\ \Rightarrow F_{\chi^2_5}(c) &= 0.95 \\ \Rightarrow c &= 11.07 \end{aligned}$$

Since $T = 4.64 < 11.07$, we will accept the null hypothesis.
It implies that die is fair.