# The Student's t-Distribution and Statistical Inference

**t-Statistic:** It is defined as the ratio of standard normal variable and square root of $\chi^2$ chi-square random variable divided by its degree of freedom. In its simplest form is given as $t = \dfrac{\bar{X} - \mu}{s/\sqrt{n}}$ where $s^2 = \dfrac{\sum (x - \bar{x})^2}{n-1}$ unbiased estimate of population variance $\sigma^2$.

**t-distribution**

let "x" continuous distribution with interval $(-\infty, +\infty)$is said to be t-distribution with probability density function (pdf) given as

$$f(x) = \frac{1}{\sqrt{n}\,\beta\left(\frac{n}{2}, \frac{1}{2}\right)\left[1 + \frac{t^2}{n}\right]^{\frac{n+1}{2}}}$$

It has only one parameter "degree of freedom"

**Write down the assumption of "t-distribution".**

i) The populations from which small samples are drawn are normally distributed.
ii) The samples are drawn/selected randomly.
iii) The samples are drawn independently.
iv) The populations have same variance.

**Write down the assumption of paired t-test.**

i) The populations from which two samples are selected normally distributed.
ii) The two samples are drawn/selected randomly.
iii) The two samples are drawn dependently.
vi) The variance of two populations may or may not be equal.

**Write down the properties of t-distribution.**

i) The total area under curve is unity.
ii) It ranges from $-\infty to + \infty$.
iii) It is continuous distribution.
iv) It has only one parameter called degree of freedom "v.
v) It is symmetrical but flatter than the normal distribution. But as the degree of freedom increase, the "t-distribution" approximates the normal distribution.
vi) Its mean is zero and variance is $n/n-1$
vii) The "t-distribution" has mode at t=0 and mean, median and mode are identical on this point.
viii) As degree of freedom "n=1", then "t-distribution" tends to Cauchy distribution but at df "n=1" t-distribution does not tends to normality. Hence Cauchy distribution does not tends to normality as $n \to \infty$ only with "n" degree of freedom as $n \to \infty$. T distribution tends to normality.
ix) It does not posses the moment generating function.
x) It does not depend on population variance but it depends on degree of freedom "n". It is a chief merit of this distribution.

**Write down the application of t-distribution.**

i) Testing the equality of population mean $\mu = \mu_0$ for small samples size and $\sigma^2$ is unknown.
ii) Finding confidence interval for population mean $\mu$ when n<30 and $\sigma^2$ is unknown.
iii) Testing the equality of two population mean $\mu_1 = \mu_2$ for small samples sizes and $\sigma_1^2 = \sigma_2^2$ but unknown.
iv) t-distribution is also used in LSD-test.

**What is meant by degree of freedom?**

Ans: The degree of freedom may be defined as the total number of sample observations minus the number of population parameter estimated from the sample. It is denoted by $v = n-1$

**What are the parameters of t-distribution?**

Ans: The degree of freedom is the only one parameter of t-distribution. It is denoted by $v \text{ or } n$

**Under what conditions can we not use the normal distribution but can use the "t: distribution to find confidence intervals for the unknown population mean?**
When the sample size "n" is smaller than 30, and the population standard deviation "$\sigma$" is not known, we cannot use the normal distribution for determining confidence intervals for the unknown population mean but we can use the students "t" distribution
**What is the relationship between the "t" distribution and the standard normal distribution?**
The "t" distribution is bell-shaped and is symmetrical about its mean zero. But if flatter or platy kurtic than the standard normal distribution so that more of its area falls within the tails. While there is only one standard normal distribution, there is different (t) distribution for each sample size "n" However, as "n" becomes larger, the (t) distribution approaches the standard normal distribution until, when $n \geq 30$, they are approximately equal.
**When "Z-test and t-test are recommended in hypothesis testing or interval estimation?**
**a) Z-test can be applied in the following conditions**
**i)** If the variance of the population $\sigma^2$ is known "Z-test" is appropriate (irrespective of the normality of the population and the sample size).
ii) If the variance of the population $\sigma^2$ is not known and sample size is large (n>30) "Z-test" is appropriate because the estimate of the population variance ($s^2$) from large sample is a satisfactory (unbiased) estimate of the population variance $\sigma^2$.
**b) t-test is applicable only**, if the population variance is unknown and sample size is small $n \leq 30$ provided that the given population is normally distributed.
**What tests are based on "t" distribution?**
i) Testing the hypothesis about single population mean when $\sigma^2$ are unknown and "n<30"
ii) Testing the hypothesis about the difference of two population means
a) When $\sigma_1^2$ *and* $\sigma_2^2$ are unknown but equal $\sigma_1^2 = \sigma_2^2$ and $n_1$ *and* $n_2 < 30$
b) When $\sigma_1^2$ *and* $\sigma_2^2$ are unknown but equal $\sigma_1^2 \neq \sigma_2^2$ and $n_1$ *and* $n_2 < 30$
iii) Testing the hypothesis about the equality or difference of two population means when samples means are correlated. When $\sigma_1^2$ *and* $\sigma_2^2$ are unknown but $n_1$ *and* $n_2 < 30$
iv) Testing the hypothesis about the equality or differences of two population means when observations are paired. Also $n_1$ *and* $n_2 < 30$
v) Testing the hypothesis about population simple correlation coefficient
vi) Testing the hypothesis about population partial correlation coefficient
vii) Testing the hypothesis about population multiple correlation coefficient
viii) Testing the hypothesis about intercept in simple linear regression model $H_0 : \alpha = \alpha_0$
ix) Testing the hypothesis about slope (regression coefficient) in SLRM $H_0 : \beta = \beta_0$
x) Testing the hypothesis about mean and individual prediction in SLRM $H_0 : \mu_{Y.X} = \mu_0$
xi) Testing the hypothesis about the equality of two different slope (regression coefficient) in two different linear regression models $H_0 : \beta_1 = \beta_2$
xii) Testing the hypothesis about the variance of residuals in SLRM $H_0 : \sigma_u^2 = \sigma_{u0}^2$
xiii) Testing the hypothesis about the intercept in MLRM $H_0 : \beta_0 = \beta_{00}$
xiv) Testing the hypothesis about first slope in MLRM $H_0 : \beta_1 = \beta_{10}$
xv) Testing the hypothesis about first slope in MLRM $H_0 : \beta_2 = \beta_{20}$
**What is objective of pooled estimator?**
To estimate the unknown population parameter, we drawn specified number of samples of different sizes from the same population and calculate estimates respectively then pooled the estimator to arrive at the true value of population parameter, thus reduced the biasness in estimator.

Pooled estimator of population mean m $\quad \overline{X}_P = \dfrac{n_1\overline{X}_1 + n_2\overline{X}_2}{n_1 + n_2}$

Pooled estimator of population proportion $\quad \hat{P}_C = \dfrac{n_1\hat{P}_1 + n_2\hat{P}_2}{n_1 + n_2}$

Pooled estimator of population variance $\quad S_P^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

**Differentiate between dependent and independent samples**
If two samples are selected, one each of two population then
i) Two samples are independent if the selection of objects from one population is unrelated to the selection of objects from other population.
ii) Two samples are dependent if for each object selected from one population an object is chosen from other population to from a pair of similar objects.
**How do you recognize independent and dependent samples?**
i) Independent samples are always two different samples
ii) Dependent samples always consists of matched or pair observations
**Define large samples and small samples**
When $n \geq 30$ is large sample and $n < 30$ is small samples

**Important points**

| Null hypothesis $H_0$ | Alternative hypothesis $H_1$ |
|---|---|
| Coin is unbiased<br>Drug is ineffective in curing a particular disease<br>The difference b/w two teaching methods are null or zero<br>Not significantly different<br>Insignificant<br>At least $H_0 : \mu \geq \mu_0$<br>At most $H_0 : \mu \leq \mu_0$<br>Not more than $H_0 : \mu \leq \mu_0$<br>Not less than $H_0 : \mu \geq \mu_0$<br>No effect At least $H_0 : \mu = \mu_0$<br>Fair $H_0 : \mu = \mu_0$ | Coin is biased<br>Drug is effective<br>The difference b/w two teaching methods are null or zero<br>Significantly different<br>Significant<br>Better $H_1 : \mu > \mu_0$<br>Improved $H_1 : \mu > \mu_0$<br>Benifitted $H_1 : \mu > \mu_0$<br>Deterícted $H_1 : \mu < \mu_0$<br>Above $H_1 : \mu > \mu_0$<br>Exceeds $H_1 : \mu > \mu_0$<br>More than $H_1 : \mu > \mu_0$<br>Taller $H_1 : \mu > \mu_0$<br>Increased $H_1 : \mu > \mu_0$<br>Biased $H_1 : \mu \neq \mu_0$<br>Fair $H_1 : \mu \neq \mu_0$<br>Raise $H_1 : \mu > \mu_0$<br>Higher $H_1 : \mu > \mu_0$<br>Inferior $H_1 : \mu < \mu_0$<br>Reduce $H_1 : \mu < \mu_0$ |

Q.18.4: A random sample of size n is drawn from a normal population with mean 5 and variance $\sigma^2$

Solution:

Given that $\mu = 5$

i) If n=25, $\overline{X}$ =3 and s=2 , what is t?

**Solution:** $t = \dfrac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$

$$t = \frac{3-5}{\frac{2}{\sqrt{25}}} = \frac{-2}{0.4} = -5$$

ii) If n=9, $\overline{X}$ =2 and t=-2 , what is s?

**Solution:**

$$t = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$$

$$-2 = \frac{2-5}{\frac{s}{\sqrt{9}}} = \frac{-3}{\frac{s}{3}} = \frac{-3(3)}{s} = \frac{-9}{s}$$

$$-2s = -9 \rightarrow s = \frac{-9}{-2} \rightarrow s = 4.5$$

iii) If n=25, s=10 and t=2 , what is $\overline{X}$ ?

**Solution:**

$$t = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$$

$$2 = \frac{\overline{X} - 5}{\frac{10}{\sqrt{25}}} = \frac{\overline{X} - 5}{2} \qquad \overline{X} = 4 + 5 = 9$$

iv) If s=25, $\overline{X}$ =14 and t=13, what is n?

**Solution:**

$$t = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$$

$$3 = \frac{14-5}{\frac{15}{\sqrt{n}}} = \frac{9\sqrt{n}}{15} \qquad 9\sqrt{n} = 45 \qquad \sqrt{n} = 45/9 = 5 \qquad n = 25$$

Q.18.5 (a): Explain how a confidence interval is constructed for the mean of a normal population for small sample.

Solution:

Construction of confidence for population mean ( $\mu$ ) When (n<30) small and $\sigma^2$ are unknown

Let $X_1, X_2, X_3, ..., X_n$ be a random samples drawn from a normal population with mean $\mu$ and unknown variance $\sigma^2$. Here $s^2 = \dfrac{\sum\left(X - \overline{X}\right)}{n-1}$ is an unbiased estimate of $\sigma^2$. Then student "t-distribution" is $t = \dfrac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$ with "v=n-1" degree of freedom

Then $100(1-\alpha)\%$ confidence interval in form of probability statement for population mean $\mu$ is

$$P\left[-t_{\frac{\alpha}{2}}(v) \le t_c \le -t_{\frac{\alpha}{2}}(v)\right] = 1 - \alpha$$

Substituting the value of test-statistic

$$P\left[-t_{\frac{\alpha}{2}}(v) \le \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \le t_{\frac{\alpha}{2}}(v)\right] = 1-\alpha$$

Multiplying inside the bracket $\frac{s}{\sqrt{n}}$

$$P\left[-t_{\frac{\alpha}{2}}(v)\frac{s}{\sqrt{n}} \le \bar{X}-\mu \le t_{\frac{\alpha}{2}}(v)\frac{s}{\sqrt{n}}\right] = 1-\alpha$$

Subtracting the $\bar{X}$ inside the bracket

$$P\left[-\bar{X}-t_{\frac{\alpha}{2}}(v)\frac{s}{\sqrt{n}} \le \bar{X}-\mu-\bar{X} \le -\bar{X}+t_{\frac{\alpha}{2}}(v)\frac{s}{\sqrt{n}}\right] = 1-\alpha$$

$$P\left[-\bar{X}-t_{\frac{\alpha}{2}}(v)\frac{s}{\sqrt{n}} \le -\mu \le -\bar{X}+t_{\frac{\alpha}{2}}(v)\frac{s}{\sqrt{n}}\right] = 1-\alpha$$

Multiplying (-1) inside the bracket and sign of inequality will be change

$$P\left[\bar{X}+t_{\frac{\alpha}{2}}(v)\frac{s}{\sqrt{n}} \ge \mu \ge \bar{X}-t_{\frac{\alpha}{2}}(v)\frac{s}{\sqrt{n}}\right] = 1-\alpha$$

Or

$$P\left[\bar{X}-t_{\frac{\alpha}{2}}(v)\frac{s}{\sqrt{n}} \le \mu \le \bar{X}+t_{\frac{\alpha}{2}}(v)\frac{s}{\sqrt{n}}\right] = 1-\alpha$$

Or $\bar{X} \pm t_{\frac{\alpha}{2}}(v)\frac{s}{\sqrt{n}}$ Hence the required result

Q.18.5(b): The breaking strength of 10 specimens of 0.104 inches diameter hard-drawn copper wire are found to be 578,572,570,568,572,570,570,572,596,584,lbs.Calculatate the 95% confidence interval for mean breaking strength of this kind of wire, assuming that the measurements are normally distributed.

Solution:

The $(1-\alpha)$ % confidence interval for population mean " $\mu$ " when population variance $\sigma^2$ is unknown and sample size is small $n<30$

$\bar{X} \pm \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}}(v)$ Where $s=\sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}$ and $v=n-1$d.f

Or

$$\bar{X}-\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v) < \mu < \bar{X}+\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v)$$

Given that

$n=10$ $\bar{X}=\frac{\sum X}{n}=575.2$ $s=\sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}=8.70$ $(1-\alpha)=0.99$

$\alpha=1-0.95=0.05$ $v=n-1=9$ $t_{\frac{\alpha}{2}}(v)=t_{0.025}(9)=2.262$

Then 95 $(1-\alpha)$ % C.I for $\mu$

$$\bar{X}-\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v) < \mu < \bar{X}+\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v)$$

$$575.2-\frac{8.70}{\sqrt{10}}(2.262) < \mu < 575.2+\frac{8.70}{\sqrt{10}}(2.262)$$

$$575.2-6.22 < \mu < 575.2+6.22$$

$$568.98 < \mu < 581.42$$

Hence the required 95% confidence interval for population mean is (568.98, 581.42)

Q.18.6 (a): What is meant by the term" Confidence Interval"? Find a 90% confidence interval for the mean of a normal distribution if σ =2 and if a sample of size 8 gave the values 9, 14, 10, 12, 7, 13, 11, 12.What would be the confidence interval if $\sigma$ were unknown?

Solution:
i) When population standard deviation are known
The $(1-\alpha)$ % confidence interval for population mean " $\mu$ " when population variance $\sigma^2$ is known and sample size is small $n<30$

$$\bar{X} \pm \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}$$

Or

$$\bar{X} - \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}$$

Given that

$n=8$ $\qquad \bar{X} = \frac{\sum X}{n} = 11$ $\qquad (1-\alpha)=0.90$ $\quad \alpha = 0.10$ $\quad Z_{\frac{\alpha}{2}} = Z_{0.05} = 1.645$

Then 90 $(1-\alpha)$ % C.I for $\mu$

$$\bar{X} - \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}$$

$$11 - \frac{2}{\sqrt{8}}(1.645) < \mu < 11 + \frac{2}{\sqrt{8}}(1.645)$$

$$11 - 1.16 < \mu < 11 + 1.16$$

$$9.84 < \mu < 12.16$$

Hence the required 90% confidence interval for population mean are (9.84, 12.16)
ii) When population standard deviation are unknown and "n<30"
The $(1-\alpha)$ % confidence interval for population mean " $\mu$ " when population variance $\sigma^2$ is unknown and sample size is small $n<30$

$$\bar{X} \pm \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}}(v)$$

Where $s = \sqrt{\frac{\sum (X-\bar{X})^2}{n-1}}$ and $v = n-1$ d.f

Or

$$\bar{X} - \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}}(v) < \mu < \bar{X} + \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}}(v)$$

Given that

$n=8$ $\qquad \bar{X} = \frac{\sum X}{n} = 11$ $\qquad s = \sqrt{\frac{\sum (X-\bar{X})^2}{n-1}} = 2.27$ $\qquad (1-\alpha) = 0.90$

$\alpha = 0.10$ $\qquad v = n-1 = 7$ $\qquad t_{\frac{\alpha}{2}}(v) = t_{0.05}(7) = 1.895$

Then 90 $(1-\alpha)$ % C.I for $\mu$

$$\bar{X} - \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}}(v) < \mu < \bar{X} + \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}}(v)$$

$$11 - \frac{2.27}{\sqrt{8}}(1.895) < \mu < 11 + \frac{2.27}{\sqrt{8}}(1.895)$$

$$11 - 1.52 < \mu < 11 + 1.52$$

$$9.48 < \mu < 12.52$$

Hence the required 90% confidence interval for population mean are (9.48,12.52)
Q.18.6 (b): Find a 90% confidence interval for the mean of a normal distribution σ =3 given the sample as 2,3,-0.2,-0.4 and -0.9.How will these limits be affected if ∂ is not known?
Solution:
i) When population standard deviation are known
The $(1-\alpha)$ % confidence interval for population mean " $\mu$ " when population variance $\sigma^2$ is known and sample size is small $n<30$

$$\bar{X} \pm \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}$$

Or

$$\bar{X} - \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}$$

Given that

$n=4 \qquad \overline{X}=\frac{\sum X}{n}=0.2 \qquad (1-\alpha)=0.90 \quad \alpha=0.10 \qquad Z_{\frac{\alpha}{2}}=Z_{0.05}=1.645$

Then $90(1-\alpha)$ % C.I for $\mu$

$$\overline{X}-\frac{\sigma}{\sqrt{n}}Z_{\frac{\alpha}{2}}<\mu<\overline{X}+\frac{\sigma}{\sqrt{n}}Z_{\frac{\alpha}{2}}$$

$$0.2-\frac{3}{\sqrt{4}}(1.645)<\mu<0.2+\frac{3}{\sqrt{4}}(1.645)$$

$$0.2-2.47<\mu<0.2+2.47$$

$$-2.27<\mu<2.67$$

Hence the required 90% confidence interval for population mean are (-2.27,2.67)

ii) When population standard deviation are unknown and "n<30"

The $(1-\alpha)$ % confidence interval for population mean "$\mu$" when population variance $\sigma^2$ is unknown and sample size is small $n<30$

$\overline{X}\pm\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v) \qquad$ Where $s=\sqrt{\frac{\sum(X-\overline{X})^2}{n-1}}$ and $v=n-1$ d.f

Or

$$\overline{X}-\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v)<\mu<\overline{X}+\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v)$$

Given that

$n=8 \qquad \overline{X}=\frac{\sum X}{n}=11 \qquad s=\sqrt{\frac{\sum(X-\overline{X})^2}{n-1}}=1.43 \qquad (1-\alpha)=0.90$

$\alpha=0.10 \qquad v=n-1=3 \qquad t_{\frac{\alpha}{2}}(v)=t_{0.05}(3)=2.353$

Then $90(1-\alpha)$ % C.I for $\mu$

$$\overline{X}-\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v)<\mu<\overline{X}+\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v)$$

$$0.2-\frac{1.43}{\sqrt{4}}(2.353)<\mu<0.2+\frac{1.43}{\sqrt{4}}(2.353)$$

$$0.2-1.68<\mu<0.2+1.68$$

$$-1.48<\mu<1.88$$

Hence the required 90% confidence interval for population mean is (-1.48, 1.88)

Q.18.7(a): A random sample of 10 boys had the following : 70,120, 110,101,88,83,95,107,100,98. Find the 95%confidence interval for the population mean. Assume that 1Q'sare normally distributed and the variance is not known.

Solution:

The $(1-\alpha)$ % confidence interval for population mean "$\mu$" when population variance $\sigma^2$ is unknown and sample size is small $n<30$

$\overline{X}\pm\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v) \qquad$ Where $s=\sqrt{\frac{\sum(X-\overline{X})^2}{n-1}}$ and $v=n-1$ d.f

Or

$$\overline{X}-\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v)<\mu<\overline{X}+\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v)$$

Given that

$n=10 \qquad \overline{X}=\frac{\sum X}{n}=97.2 \qquad s=\sqrt{\frac{\sum(X-\overline{X})^2}{n-1}}=14.27 \qquad (1-\alpha)=0.95$

$\alpha=1-0.95=0.05 \qquad v=n-1=9 \qquad t_{\frac{\alpha}{2}}(v)=t_{0.025}(9)=2.262$

Then $95(1-\alpha)$ % C.I for $\mu$

$$\overline{X}-\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v)<\mu<\overline{X}+\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v)$$

$$97.2-\frac{14.27}{\sqrt{10}}(2.262)<\mu<97.2+\frac{14.27}{\sqrt{10}}(2.262)$$
$$97.2-10.2<\mu<97.2+10.2$$
$$87.0<\mu<107.4$$

Hence the required 95% confidence interval for population mean are (87.0,107.4)

Q.18.7 (b): A sample of size 16 from a normal population with unknown standard deviation, gave $\bar{x}=14.5$ and s=5.Find a 90% confidence interval on the Population mean

Solution:

The $(1-\alpha)$ % confidence interval for population mean "$\mu$" when population variance $\sigma^2$ is unknown and sample size is small $n<30$

$$\bar{X}\pm\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v) \qquad \text{Where } s=\sqrt{\frac{\sum\left(X-\bar{X}\right)^2}{n-1}} \text{ and } v=n-1\,\text{d.f}$$

Or

$$\bar{X}-\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v)<\mu<\bar{X}+\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v)$$

Given that

$$n=16 \qquad \bar{X}=14.5 \qquad s=5 \qquad (1-\alpha)=0.90 \qquad \alpha=1-0.90=0.10$$
$$v=n-1=15 \qquad t_{\frac{\alpha}{2}}(v)=t_{0.05}(15)=1.75$$

Then 95$(1-\alpha)$% C.I for $\mu$

$$\bar{X}-\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v)<\mu<\bar{X}+\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v)$$
$$14.5-\frac{5}{\sqrt{16}}(1.75)<\mu<14.5+\frac{5}{\sqrt{16}}(1.75)$$
$$14.5-2.2<\mu<14.5+2.2$$
$$12.3<\mu<16.7$$

Hence the required 90% confidence interval for population mean are (12.3,16.7)

Q.18.8: The masses, in grams, of twelve ball bearings taken at random from a batch are 31.4, 33.1, 35.9, 34.7, 33.4, 34.5, 35.0, 32.5, 36.9, 36.4, 35.8, 33.2. Calculate a 90% confidence interval for the mean mass of the population, supposed normal, from which these masses were drawn.

Solution:

The $(1-\alpha)$ % confidence interval for population mean "$\mu$" when population variance $\sigma^2$ is unknown and sample size is small $n<30$

$$\bar{X}\pm\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v) \qquad \text{Where } s=\sqrt{\frac{\sum\left(X-\bar{X}\right)^2}{n-1}} \text{ and } v=n-1\,\text{d.f}$$

Or

$$\bar{X}-\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v)<\mu<\bar{X}+\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v)$$

Given that

$$n=12 \qquad \bar{X}=\frac{\sum X}{n}=34.4 \qquad s=\sqrt{\frac{\sum\left(X-\bar{X}\right)^2}{n-1}}=1.70 \qquad (1-\alpha)=0.90$$
$$\alpha=1-0.9\backslash 0=0.10 \qquad v=n-1=11 \qquad t_{\frac{\alpha}{2}}(v)=t_{0.05}(11)=1.796$$

Then 95$(1-\alpha)$% C.I for $\mu$

$$\bar{X}-\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v)<\mu<\bar{X}+\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(v)$$
$$34.4-\frac{1.70}{\sqrt{12}}(1.796)<\mu<34.4+\frac{1.70}{\sqrt{16}}(1.796)$$
$$34.4-0.88<\mu<34.4+0.88$$
$$33.52<\mu<35.28$$

Hence the required 90% confidence interval for population mean are (33.52,35.28)

**Construction of confidence for population mean ( $\mu_1$ *and* $\mu_2$ ) When $\sigma_1^2$ *and* $\sigma_2^2$ are unknown but equal $\sigma_1^2 = \sigma_2^2$ and sample sizes small ( $n_1, n_2 < 30$ )**

Let $X_{11}, X_{12}, X_{13}, ..., X_{1n}$ and $X_{21}, X_{22}, X_{23}, ..., X_{2n}$ be a two normal population samples are drawn randomly and independently with population with mean " $\mu_1$ *and* $\mu_2$ "and unknown variance " $\sigma_1^2$ *and* $\sigma_2^2$ " but equal " $\sigma_1^2 = \sigma_2^2$ " and $n_1, n_2 < 30$. Here $s_1^2 = \frac{\sum\left(X_1 - \overline{X}_1\right)}{n_1 - 1}$ and

$$s_2^2 = \frac{\sum\left(X_2 - \overline{X}_2\right)}{n_2 - 1}$$

Or $s_p^2 = \frac{\sum\left(X_1 - \overline{X}_1\right) + \sum\left(X_2 - \overline{X}_2\right)}{n_1 + n_2 - 2} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

" $s_1^2$ *and* $s_2^2$ " be an unbiased estimate of " $\sigma_1^2$ *and* $\sigma_2^2$ ".

Then student "t-distribution" is $t = \frac{\overline{X}_1 - \overline{X}_1 - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ with " $v = n_1 + n_2 - 2$ " degree of freedom

Then $100(1-\alpha)\%$ confidence interval in form of probability statement for population mean $\mu_1$ *and* $\mu_2$ is

$$P\left[-t_{\frac{\alpha}{2}}(v) \le t_c \le -t_{\frac{\alpha}{2}}(v)\right] = 1 - \alpha$$

Substituting the value of test-statistic

$$P\left[-t_{\frac{\alpha}{2}}(v) \le \frac{\overline{X}_1 - \overline{X}_1 - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \le t_{\frac{\alpha}{2}}(v)\right] = 1 - \alpha$$

Multiplying inside the bracket $s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

$$P\left[-t_{\frac{\alpha}{2}}(v)s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2) \le t_{\frac{\alpha}{2}}(v)s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right] = 1 - \alpha$$

Subtracting the " $\overline{X}_1 - \overline{X}_2$ " inside the bracket

$$P\left[-(\overline{X}_1 - \overline{X}_2) - t_{\frac{\alpha}{2}}(v)s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le (\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2) - (\overline{X}_1 - \overline{X}_2) \le -(\overline{X}_1 - \overline{X}_2) + t_{\frac{\alpha}{2}}(v)s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right] = 1 - \alpha$$

$$P\left[-(\overline{X}_1 - \overline{X}_2) - t_{\frac{\alpha}{2}}(v)s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le -(\mu_1 - \mu_2) \le -(\overline{X}_1 - \overline{X}_2) + t_{\frac{\alpha}{2}}(v)s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right] = 1 - \alpha$$

Multiplying (-1) inside the bracket and sign of inequality will be change

$$P\left[(\overline{X}_1 - \overline{X}_2) + t_{\frac{\alpha}{2}}(v)s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \ge (\mu_1 - \mu_2) \ge (\overline{X}_1 - \overline{X}_2) - t_{\frac{\alpha}{2}}(v)s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right] = 1 - \alpha$$

Or

$$P\left[(\overline{X}_1 - \overline{X}_2) - t_{\frac{\alpha}{2}}(v)s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le (\mu_1 - \mu_2) \le (\overline{X}_1 - \overline{X}_2) + t_{\frac{\alpha}{2}}(v)s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right] = 1 - \alpha$$

Or $\overline{X}_1 - \overline{X}_2 \pm t_{\frac{\alpha}{2}}(v)s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ Hence the required result where $v = n_1 + n_2 - 2$

When unknown variance " $\sigma_1^2$ *and* $\sigma_2^2$ " but not equal " $\sigma_1^2 \ne \sigma_2^2$ " and $n_1, n_2 < 30$

$$\bar{X}_1 - \bar{X}_2 \pm t_{\frac{\alpha}{2}}(v)\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} \qquad \text{Where } v = \frac{\left[\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right]^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$$

Q.18.9 (a): Give two random samples of size $n_1=9$ and $n_2=16$, from two independent normal population, with $\bar{X}_1 = 64$, $\bar{X}_2 = 59$, $s_1=6$, and $s_2=5$, find 95% confidence interval for $\mu_1-\mu_2$, assuming that $\sigma_1=\sigma_2$.

Solution:

The $100(1-\alpha)\%$ confidence interval for the difference between two population means "$\mu_1-\mu_2$" when population variances $\sigma_1^2$ and $\sigma_2^2$ are unknown but equal $\sigma_1^2 = \sigma_2^2$ and sample sizes are small $n_1, n_2 < 30$.

$$\bar{X}_1 - \bar{X}_2 \pm S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\, t_{\frac{\alpha}{2}}(v)$$

Given that

$n_1 = 9 \quad \bar{X}_1 = 64 \quad s^2{}_1 = 36 \quad n_2 = 16 \quad \bar{X}_2 = 59 \quad s^2{}_2 = 25 \quad (1-\alpha)\% = 95\%$

$(1-\alpha) = 0.95 \quad \alpha = 0.05 \quad \frac{\alpha}{2} = 0.025 \quad v = 23 \quad t_{\frac{\alpha}{2}}(v) = t_{0.025}(23) = 2.069$

$$S_p == \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{(9-1)(36)+(16-1)(25)}{9+16-2}} = 5.37$$

Then $95(1-\alpha)\%$ C.I for $\mu_1-\mu_2$

$$(\bar{X}_1-\bar{X}_2) - S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\, t_{\frac{\alpha}{2}}(v) < \mu_1-\mu_2 < (\bar{X}_1-\bar{X}_2) + S_p\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\, t_{\frac{\alpha}{2}}(v)$$

$$(64-59) - 5.37\sqrt{\frac{1}{9}+\frac{1}{16}}(2.069) < \mu_1-\mu_2 < (64-59) - 5.37\sqrt{\frac{1}{9}+\frac{1}{16}}(2.069)$$

$$5-4.63 < \mu_1-\mu_2 < 5+4.63$$

$$0.37 < \mu_1-\mu_2 < 9.63$$

Hence the required 95% confidence interval for the difference of population means is (0.37, 9.63)

Q.18.9 (b): A sample from a normal population with unknown variance consists of the observations 34, 25, 43, 37, 45. A sample from a second normal population with the same unknown variance as the first consists if the observation 20, 31, 23, 35, 41, 29, 39. Find a 95% confidence interval on $\mu_1-\mu_2$.

Solution:

The $100(1-\alpha)\%$ confidence interval for the difference between two population means "$\mu_1-\mu_2$" when population variances $\sigma_1^2$ and $\sigma_2^2$ are unknown but equal $\sigma_1^2 = \sigma_2^2$ and sample sizes are small $n_1, n_2 < 30$.

$$\bar{X}_1 - \bar{X}_2 \pm S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\, t_{\frac{\alpha}{2}}(v)$$

Given that

$n_1 = 5 \quad \bar{X}_1 = 36.8 \quad s^2{}_1 = 63.2 \quad n_2 = 7 \quad \bar{X}_2 = 31.14 \quad s^2{}_2 = 61.48$

$(1-\alpha)\% = 95\% \quad \alpha = 0.05 \quad v = (n_1+n_2-2) = 10 \quad t_{\frac{\alpha}{2}}(v) = t_{0.025}(10) = 2.228$

$$S_p == \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}} = 7.88$$

Then $99(1-\alpha)\%$ C.I for $\mu_1-\mu_2$

$$(\bar{X}_1-\bar{X}_2) - S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\, t_{\frac{\alpha}{2}}(v) < \mu_1-\mu_2 < (\bar{X}_1-\bar{X}_2) + S_p\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\, t_{\frac{\alpha}{2}}(v)$$

$$(36.8-31.14)-7.88\sqrt{\frac{1}{5}+\frac{1}{7}}(2.228)<\mu_1-\mu_2<(36.8-31.14)-7.88\sqrt{\frac{1}{5}+\frac{1}{7}}(2.228)$$

$$5.66-10.28<\mu_1-\mu_2<5.66+10.28$$

$$-4.62<\mu_1-\mu_2<15.94$$

Hence the required 95% confidence interval for the difference of population means is (-4.62, 15.94)

Q.18.10: A course in mathematics is taught to 10 students by the conventional class room procedure .A second group of 12 students was given the same course by means of programmed materials. At the end of the semester, the same examination was given to each group. Their grades are given as follows:

| | |
|---|---|
| Groups 1 | 70 66 77 73 72 68 74 75 69 |
| Groups 2 | 77 83 92 85 82 84 80 78 91 93 80 87 |

Find a 90% confidence interval for the differences between the average grades of the two groups. Assume the population to be approximately normal with equal variances.

Solution:

The $100(1-\alpha)\%$ confidence interval for the difference between two population means "$\mu_1-\mu_2$" when population variances $\sigma_1{}^2$ and $\sigma_2{}^2$ are unknown but equal $\sigma_1{}^2=\sigma_2{}^2$ and sample sizes are small $n_1, n_2<30$.

$$\bar{X}_1-\bar{X}_2\pm S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\,t_{\frac{\alpha}{2}}(v)$$

Given that

$$n_1=10 \quad \bar{X}_1=72 \quad s^2{}_1=36.07 \quad n_2=12 \quad \bar{X}_2=84.3 \quad s^2{}_2=10.91$$

$$(1-\alpha)\%=90\% \quad (1-\alpha)=0.90 \quad \alpha=0.10 \quad \frac{\alpha}{2}=0.05 \quad v=(n_1+n_2-2)=20$$

$$t_{\frac{\alpha}{2}}(v)=t_{0.05}(20)=1.725$$

$$S_p==\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}=\sqrt{\frac{(10-1)(36.07)+(12-1)(10.91)}{10+12-2}}=4.72$$

Then $90(1-\alpha)\%$ C.I for $\mu_1-\mu_2$

$$(\bar{X}_1-\bar{X}_2)-S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\,t_{\frac{\alpha}{2}}(v)<\mu_1-\mu_2<(\bar{X}_1-\bar{X}_2)+S_p\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\,t_{\frac{\alpha}{2}}(v)$$

$$(72-84.3–)-4.72\sqrt{\frac{1}{10}+\frac{1}{20}}(1.725)<\mu_1-\mu_2<(72-84.3–)-4.72\sqrt{\frac{1}{10}+\frac{1}{20}}(1.725)$$

$$-12.3-3.5<\mu_1-\mu_2<-12.3+3.5$$

$$-15.8<\mu_1-\mu_2<-8.8 \qquad \text{Or} \qquad 8.8<\mu_2-\mu_1<15.8$$

Hence the required 90% confidence interval for the difference of population means is (8.8 and 15.8)

Q.18.11: From the area planted in one variety of guayule, 54 plants were selected at random. Of these plants, 15 were " Off types " and 12 were "Aberrant". The rubber per centages of these plants was:

| | |
|---|---|
| Off types | 4.47, 5.88, 6.21, 5.55, 6.09, 5.70, 5.82, 4.84, 5.59, 5.59, 5.22, 4.45, 6.74, 6.06, 6.04 |
| Aberrant | 6.48, 6.36, 4.28, 7.71, 6.40, 7.06, 5.51, 8.93, 7.71, 7.20, 7.37, 5.91 |

Compute a 90%confidence interval for the difference of two population means.

Solution:

The $100(1-\alpha)\%$ confidence interval for the difference between two population means "$\mu_1-\mu_2$" when population variances $\sigma_1{}^2$ and $\sigma_2{}^2$ are unknown but equal $\sigma_1{}^2=\sigma_2{}^2$ and sample sizes are small $n_1, n_2<30$ .

$$\bar{X}_1-\bar{X}_2\pm S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\,t_{\frac{\alpha}{2}}(v)$$

Given that

$n_1 = 12 \quad \overline{X}_1 = 6.74 \quad s^2{}_1 = 1.452 \quad n_2 = 15 \quad \overline{X}_2 = 5.62 \quad s^2{}_2 = 0.412$

$(1-\alpha) = 0.90 \quad \alpha = 0.10 \quad \frac{\alpha}{2} = 0.05 \quad v = (n_1 + n_2 - 2) = 25 \quad t_{\frac{\alpha}{2}}(v) = t_{0.05}(25) = 1.708$

$$S_p == \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{(12-1)(1.452)+(15-1)(0.412)}{12+15-2}} = 0.93$$

Then $90(1-\alpha)\%$ C.I for $\mu_1 - \mu_2$

$$(\overline{X}_1 - \overline{X}_2) - S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\, t_{\frac{\alpha}{2}}(v) < \mu_1 - \mu_2 < (\overline{X}_1 - \overline{X}_2) + S_p\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\, t_{\frac{\alpha}{2}}(v)$$

$$(6.74-5.62) - 0.93\sqrt{\frac{1}{12}+\frac{1}{15}}(1.708) < \mu_1 - \mu_2 < (6.74-5.62) + 0.93\sqrt{\frac{1}{12}+\frac{1}{15}}(1.708)$$

$$1.12 - 0.62 < \mu_1 - \mu_2 < 1.12 + 0.62$$

$$0.50 < \mu_1 - \mu_2 < 1.74$$

Hence the required 90% confidence interval for the difference of population means is (0.50 and 1.74)

Q.18.12 (b): Describe the procedure for testing hypothesis about mean of a normal population when standard deviation is unknown and sample size is small.

Solution:

**Describe the procedure for testing of hypothesis about the mean of a normal population when population standard deviation is unknown and the sample size is small.**

## Procedure:

i) We set up our null and alternative hypothesis

a) $H_0 : \mu = \mu_0$ b) $H_0 : \mu \geq \mu_0$ c) $H_0 : \mu \leq \mu_0$

a) $H_1 : \mu \neq \mu_0$ b) $H_1 : \mu < \mu_0$ c) $H_1 : \mu > \mu_0$

ii) Assumption: A sample is drawn randomly and independently from a normal population with population mean $\mu$ and unknown $\sigma$ and sample size is small (n<30)

iii) Level of significance

$\alpha = (Comonly \;\; uesd\ 5\%\ or\ 1\%)$

iv) Test-statistic

a) $t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}} = \frac{\overline{X} - \mu_0}{S.E(\overline{X})}$ When sampling done with replacement

b) $t = \frac{\overline{X} - \mu_0}{\frac{s}{\sqrt{n}}\sqrt{\left(\frac{N-n}{N-1}\right)}} = \frac{\overline{X} - \mu_0}{S.E(\overline{X})}$ Sampling done without replacement

If $H_0$ is true; it has t-distribution with v=n-1 degree of freedom

v) Critical region

It is naturally depend on alternative hypothesis

a) $H_1 : \mu \neq \mu_0$ We used two sided test

$|t_c| \geq t_{\frac{\alpha}{2}}(v)$

b) $H_1 : \mu < \mu_0$ We use one sided test

$t_c < -t_{\alpha}(v)$

C) $H_1 : \mu > \mu_0$ We use one sided test

$t_c > t_{\alpha}(v)$

vi) Calculation

In this step we calculate the value of “t” test statistic on the basis of sample data.

vii) Conclusion

If our calculated value does not fall’s in critical region then we accept $H_0$ other wise we reject it.

Q.18.13: Given the following information, what is your conclusion in testing each of the indicated null hypotheses? Assume the population is normal.

| | Sample size (n) | Sample mean ($\bar{X}$) | Estimate of Variance from Sample($s^2$) | Significance level (α) | Hypotheses $H_0$ | $H_1$ |
|---|---|---|---|---|---|---|
| a | 9 | 12 | 36 | 0.05 | μ=10 | μ>10 |
| b | 16 | 13 | 64 | 0.05 | μ=10 | μ≠10 |
| c | 16 | 11 | 81 | 0.01 | μ≤10 | μ>10 |
| d | 25 | 8 | 64 | 0.01 | μ≥10 | μ<10 |
| e | 25 | 9 | 49 | 0.02 | μ=10 | μ≠10 |

Solution☹a)
i) We set up our null and alternative hypothesis
$H_0: \mu = 10$
$H_1: \mu > 10$
ii) Assumption: A sample is drawn randomly and independently from a normal population with population mean $\mu$ and unknown $\sigma$ and sample size is small.
iii) Level of significance
$\alpha = 5\% = 0.05$
iv) Test-statistic
$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$
If $H_0$ is true; which has t-distribution with v=n-1 degree of freedom
v) Critical region
$t_c > t_\alpha(v)$
$t_c > t_{0.05}(8)$ $\qquad v = n-1 = 9-1 = 8$
$t_c > 1.860$
vi) Calculation
$n = 9 \qquad \mu = 10 \qquad \bar{X} = 12 \qquad \sigma^2 = 36 \qquad \sigma = 6$
$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{12-10}{6/\sqrt{9}} = 1.0$$
vii) Conclusion
Since our calculated value does not fall's in critical region, so we accept $H_0$ and conclude that population mean is 10 at 5% level of significance.
b)
i) We set up our null and alternative hypothesis
$H_0: \mu = 10$
$H_1: \mu \neq 10$
ii) Assumption: A sample is drawn randomly and independently from a normal population with population mean $\mu$ and unknown $\sigma$ and sample size is small.
iii) Level of significance
$\alpha = 5\% = 0.05$
iv) Test-Statistic
$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$
If $H_0$ is true; which has t-distribution with v=n-1 degree of freedom
v) Critical region
$|t_c| \geq t_{\frac{\alpha}{2}}(v)$
$|t_c| \geq t_{0.025}(15)$ $\qquad v = n-1 = 16-1 = 15$
$|t_c| \geq 1.753$
vi) Calculation

$n = 16 \qquad \mu = 10 \qquad \overline{X} = 13 \qquad s^2 = 64 \qquad s = 8$

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}} = \frac{13-10}{8/\sqrt{16}} = 1.5$$

vii) Conclusion

Since our calculated value does not fall's in critical region, so we accept $H_0$ and conclude that population mean is 10 at 5% level of significance.

c) Do yourself same as Q.13 (a)

d)

i) We state our null and alternative hypothesis

$H_0 : \mu \geq 10$

$H_1 : \mu < 10$

ii) Assumption: A sample is drawn randomly and independently from a normal population with population mean $\mu$ and unknown $\sigma$ and sample size is small.

iii) Level of significance

$\alpha = 1\% = 0.01$

iv) Test-Statistic

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

If $H_0$ is true; which has t-distribution with v=n-1 degree of freedom

v) Critical region

$t_c < -t_\alpha(v)$

$t_c < -t_{0.01}(24) \qquad\qquad v = n - 1 = 25 - 1 = 24$

$t_c < 2.492$

vi) Calculation

$n = 25 \qquad \mu = 10 \qquad \overline{X} = 8 \qquad s^2 = 64 \qquad s = 8$

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}} = \frac{8-10}{8/\sqrt{25}} = -1.25$$

vii) Conclusion

Since our calculated value does not fall's in critical region, so we accept $H_0$ and conclude that population mean is 10 at 1% level of significance.

e) Do your self same as Q.13 (b)

Q.18.14: A Sample o 12 jars of peanut butter was taken from a lot, each jar being labeled "8 ounces net weight." The individual weights in ounces are: 8.2, 8.0, 7.6, 7.6, 7.7, 7.5, 7.3, 7.4, 7.5, 8.0, 7.4, and 7.5. Test whether these values are consistent with a population mean of 8. Assume that the weights are normally distributed.

Solution:

i) We set up our null and alternative hypothesis

$H_0 : \mu = 8$

$H_1 : \mu \neq 8$

ii) Assumption: A sample is drawn randomly and independently from a normal population with population mean $\mu$ and unknown $\sigma$ and sample size is small.

iii) Level of significance

$\alpha = 5\% = 0.05$

iv) Test-Statistic

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

If $H_0$ is true; which has t-distribution with v=n-1 degree of freedom

v) Critical region

$|t_c| \geq t_{\frac{\alpha}{2}}(v)$

$|t_c| \geq t_{0.025}(11) \qquad\qquad v = n - 1 = 12 - 1 = 11$

$|t_c| \geq 2.201$

vi) Calculation

$n = 12 \qquad \mu = 8 \qquad \bar{X} = \dfrac{\sum X}{n} = 7.74 \qquad\qquad s = \sqrt{\dfrac{\sum\left(X - \bar{X}\right)^2}{n-1}} = 0.28$

$t = \dfrac{\bar{X} - \mu}{s/\sqrt{n}} = \dfrac{7.64 - 8}{0.28/\sqrt{12}} = -4.45$

vii) Conclusion

Since our calculated value fall in critical region, so we reject $H_0$ and conclude that population mean 8 is inconsistent at 5% level of significance.

Q.18.15 (a): The nine items of a sample had the following values:

45, 47, 50, 52, 48, 47, 49, 53, 51.

Does the mean of the nine items differ significantly from an assumed normal population mean of 47.5 at α=0.05?

Solution:

i) We set up our null and alternative hypothesis

$H_0 : \mu = 47.5$

$H_1 : \mu \neq 47.5$

ii) Assumption: A sample is drawn randomly and independently from a normal population with population mean $\mu$ and unknown $\sigma$ and sample size is small.

iii) Level of significance

$\alpha = 5\% = 0.05$

iv) Test-Statistic

$t = \dfrac{\bar{X} - \mu}{s/\sqrt{n}}$

If $H_0$ is true; which has t-distribution with v=n-1 degree of freedom

v) Critical region

$|t_c| \geq t_{\frac{\alpha}{2}}(v)$

$|t_c| \geq t_{0.025}(8) \qquad\qquad v = n - 1 = 9 - 1 = 8$

$|t_c| \geq 2.305$

vi) Calculation

$n = 9 \qquad \mu = 47.5 \qquad \bar{X} = \dfrac{\sum X}{n} = 49.11 \qquad\qquad s = \sqrt{\dfrac{\sum\left(X - \bar{X}\right)^2}{n-1}} = 2.62$

$t = \dfrac{\bar{X} - \mu}{s/\sqrt{n}} = \dfrac{49.11 - 47.5}{2.62/\sqrt{9}} = 1.84$

vii) Conclusion

Since our calculated value does not fall's in critical region, so we accept $H_0$ and conclude that population mean is 47.5 at 5% level of significance.

Q.18.15 (b): Ten cartons are taken at random from an automatic filling machine. The mean net weight of the 10 cartons is 15.90 oz., and the sum of squared deviations is 0.27 Does the sample mean differ significantly from the intended weight of 16 Oz.?

Solution:

i) We set up our null and alternative hypothesis

$H_0 : \mu = 16$

$H_1 : \mu \neq 16$

ii) Assumption: A sample is drawn randomly and independently from a normal population with population mean $\mu$ and unknown $\sigma$ and sample size is small.

iii) Level of significance

$\alpha = 5\% = 0.05$

iv) Test-Statistic

$t = \dfrac{\bar{X} - \mu}{s/\sqrt{n}}$

If $H_0$ is true; which has t-distribution with v=n-1 degree of freedom

v) Critical region

$|t_c| \geq t_{\frac{\alpha}{2}}(v)$

$|t_c| \geq t_{0.025}(9)$ $\qquad v = n-1 = 10-1 = 9$

$|t_c| \geq 2.262$

vi) Calculation

$n = 10 \qquad \mu = 16 \qquad \overline{X} = 15.90 \qquad s = \sqrt{\dfrac{\sum\left(X-\overline{X}\right)^2}{n-1}} = \sqrt{\dfrac{0.276}{9}} = 0.175$

$$t = \frac{\overline{X}-\mu}{s/\sqrt{n}} = \frac{15.90-16}{0.175/\sqrt{10}} = -1.81$$

vii) Conclusion

Since our calculated value does not fall's in critical region, so we accept $H_0$ and conclude that population mean is 16 at 5% level of significance.

Q.18.16: A random sample of 16 values from a normal population showed a mean of 41.5 inches and a sum of squares of deviation from this mean equal to 135 (inches)$^2$ .Show that the assumption of a mean of 43.5 inches for the population is not reasonable and that the 95% confidence limits for the mean are 39.9 and 43.1 inches.

Solution:

i) We set up our null and alternative hypothesis

$H_0 : \mu = 43.5$

$H_1 : \mu \neq 43.5$

ii) Assumption: A sample is drawn randomly and independently from a normal population with population mean $\mu$ and unknown $\sigma$ and sample size is small.

iii) Level of significance

$\alpha = 5\% = 0.05$

iv) Test-Statistic

$$t = \frac{\overline{X}-\mu}{s/\sqrt{n}}$$

If $H_0$ is true; which has t-distribution with v=n-1 degree of freedom

v) Critical region

$|t_c| \geq t_{\frac{\alpha}{2}}(v)$

$|t_c| \geq t_{0.025}(15)$ $\qquad v = n-1 = 16-1 = 15$

$|t_c| \geq 2.131$

vi) Calculation

$n = 16 \qquad \mu = 43.5 \qquad \overline{X} = 41.5 \qquad s = \sqrt{\dfrac{\sum\left(X-\overline{X}\right)^2}{n-1}} = \sqrt{\dfrac{135}{15}} = 3$

$$t = \frac{\overline{X}-\mu}{s/\sqrt{n}} = \frac{41.5-43.5}{3/\sqrt{16}} = -2.66$$

vii) Conclusion

Since our calculated value fall in critical region, so we reject $H_0$ and conclude that population mean 43.5 is not reasonable at 5% level of significance.

Now

**ii) do your self**

Q.18.17: In the past a machine has produced washers having a thickness of 0.050 inches. To determine whether the machine is in proper working order, a sample of 10 washers is chosen for which the mean thickness is 0.053 inches and the standard deviation is 0.003 inches. Test the hypothesis that the machine is in proper working order, using a level of significance of 0.05 and 0.01.

Solution:

i) We set up our null and alternative hypothesis

$H_0 : \mu = 0.05$

$H_1 : \mu \neq 0.05$

ii) Assumption: A sample is drawn randomly and independently from a normal population with population mean $\mu$ and unknown $\sigma$ and sample size is small.
iii) Level of significance
i) $\alpha = 5\% = 0.05$ ii) $\alpha = 1\% = 0.01$
iv) Test-Statistic
$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$
If $H_0$ is true; which has t-distribution with v=n-1 degree of freedom
v) Critical region
$|t_c| \geq t_{\frac{\alpha}{2}}(v)$
i) $\alpha = 5\% = 0.05$
$|t_c| \geq t_{0.025}(9)$ $v = n-1 = 10-1 = 9$
$|t_c| \geq 2.262$
ii) $\alpha = 1\% = 0.01$
$|t_c| \geq t_{0.005}(9)$ $v = n-1 = 10-1 = 9$
$|t_c| \geq 3.250$

vi) Calculation
$n = 10$ $\mu = 0.05$ $\bar{X} = 0.053$ $s = 0.003$
$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{0.053 - 0.050}{0.003/\sqrt{10}} = 3.162$$
vii) Conclusion
i) Since our calculated value fall in critical region, so we reject $H_0$ and conclude that machine not proper working at 5% level of significance.
ii) Since our calculated value does not fall's in critical region, so we accept $H_0$ and conclude that machine is proper working at 1% level of significance.
Q.18.18: A manufacturer claims that his light bulbs have an average lifetime of 1500 hours. A purchaser decides to check this claim and finds that for six bulbs the lifetimes are 1472, 1486, 1401, 1350, 1610, 1590, hours. Does this evidence support the manufacturer's claim? Assume that the lifetime of the light bulbs are normally distributed.
Solution:
i) We set up our null and alternative hypothesis
$H_0 : \mu \geq 1500$
$H_1 : \mu < 1500$
ii) Assumption: A sample is drawn randomly and independently from a normal population with population mean $\mu$ and unknown $\sigma$ and sample size is small.
iii) Level of significance
$\alpha = 5\% = 0.05$
iv) Test-Statistic
$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$
If $H_0$ is true; which has t-distribution with v=n-1 degree of freedom
v) Critical region
$t_c < -t_{\alpha}(v)$
$t_c < -t_{0.05}(5)$ $v = n-1 = 6-1 = 5$
$t_c < -2.015$
vi) Calculation
$n = 6$ $\mu = 1500$ $\bar{X} = \frac{\sum X}{n} = 1484.83$ $s = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}} = 102.08$
$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{1484.83 - 1500}{102.08/\sqrt{6}} = -0.36$$

vii) Conclusion

Since our calculated value does not fall's in critical region, so we accept $H_0$ and conclude that population mean is 1500 at 5% level of significance.

Q.18.19 (a): Describe the procedure for testing hypotheses about the equality of mean of two normal populations for small samples.

Solution: **Describe the procedure for testing equality of means of two normal populations when $\sigma_1 = \sigma_2$ but unknown for small sample sizes.**

## Procedure:

i) We state our null and alternative hypothesis

a) $H_0 : \mu_1 - \mu_2 = \Delta$  b) $H_0 : \mu_1 - \mu_2 \leq \Delta$  c) $H_0 : \mu_1 - \mu_2 \geq \Delta$

a) $H_1 : \mu_1 - \mu_2 \neq \Delta$  b) $H_1 : \mu_1 - \mu_2 > \Delta$  c) $H_1 : \mu_1 - \mu_2 < \Delta$

ii) Assumption: The two samples of sizes $n_1$ *and* $n_2$ are randomly and independently drawn from two normal populations with population mean $\mu_1$ *and* $\mu_2$ when population variances are $\sigma_1^2$ *and* $\sigma_2^2$.unknown but $\sigma_1^2 = \sigma_2^2$ and sample sizes are small.

iii) Level of significance

$\alpha = (Comonly\ \ uesd\ 5\%\ or\ 1\%)$

iv) Test-statistic

$$t_c = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

If $H_0$ is true; which hast-distribution with $(v = n_1 + n_2 - 2)$ degree of freedom

v) Critical region

It is naturally depend on alternative hypothesis

a) $H_1 : \mu_1 - \mu_2 \neq \Delta$  We used two sided test

$|t_c| \geq t_{\frac{\alpha}{2}}(v)$

b) $H_1 : \mu_1 - \mu_2 > \Delta$  We use one sided test

$t_c > t_\alpha(v)$

C) $H_1 : \mu_1 - \mu_2 < \Delta$  We use one sided test

$t_c < -t_\alpha(v)$

vi) Calculation

In this step we calculate the value of "t" test statistic on the basis of sample data.

Where $s_1^2 = \frac{1}{n_1 - 1}\left(\sum X_1^2 - \frac{\left(\sum X_1\right)^2}{n_2}\right) = \frac{\sum\left(X_1 - \bar{X}_1\right)^2}{n_1 - 1}$

And

$$s_2^2 = \frac{1}{n_2 - 1}\left(\sum X_2^2 - \frac{\left(\sum X_2\right)^2}{n_2}\right) = \frac{\sum\left(X_2 - \bar{X}_2\right)^2}{n_2 - 1}$$

$$S_p = \sqrt{\frac{\sum\left(X_1 - \bar{X}_1\right)^2 + \sum\left(X_1 - \bar{X}_1\right)^2}{n_1 + n_2 - 2}} \qquad \text{Or} \qquad S_P = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

vii) Conclusion

If our calculated value does not fall's in critical region then we accept $H_0$ other wise we reject it.

Q.18.19 (b): Two random samples taken independently from normal population with an identical variance yield the following results:

| | n | $\bar{X}$ | $s^2$ |
|---|---|---|---|
| Sample 1 | 12 | 10 | 1200 |
| Sample 2 | 18 | 25 | 900 |

Test the hypothesis that the true difference between the population means is 10, that is, that $\mu_2 - \mu_1 = 10$ against the alternative the $\mu_2 - \mu_1 > 10$ at the 5 per cent level of significance.

Solution:

i) We state our null and alternative hypothesis

a) $H_0 : \mu_2 - \mu_1 = 10$

a) $H_1 : \mu_2 - \mu_1 > 10$

ii) Assumption: The two samples of sizes $n_1$ *and* $n_2$ are randomly and independently drawn from two normal populations with population mean $\mu_1$ *and* $\mu_2$ when population variances are $\sigma_1^2$ *and* $\sigma_2^2$.unknown but $\sigma_1^2 = \sigma_2^2$ and sample sizes are small.

iii) Level of significance

$\alpha = 5\% = 0.05$

iv) Test-statistic

$$t_c = \frac{(\bar{X}_2 - \bar{X}_1) - (\mu_2 - \mu_1)}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

If $H_0$ is true; which has t-distribution with $(V = n_1 + n_2 - 2)$ degree of freedom

v) Critical region

$t_c > t_{0.05}(28)$ $\qquad v = 12 + 18 - 2 = 28$

$t_c > 1.701$

vi) Calculation

$n_1 = 12$ $\qquad \bar{X}_1 = 10$ $\qquad s_1^2 = 1200$

$n_2 = 18$ $\qquad \bar{X}_2 = 25$ $\qquad s_1^2 = 900$

$$S_P = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(12-1)1200 + (18-1)900}{12 + 18 - 2}} = 31.904$$

$$t_c = \frac{(\bar{X}_2 - \bar{X}_1) - (\mu_2 - \mu_1)}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(25-10) - (10)}{31.904\sqrt{\frac{1}{12} + \frac{1}{18}}} = 0.42$$

vii) Conclusion

Since our calculated value does not fall's in critical region then we accept $H_0$. And we conclude that the $H_0 : \mu_2 - \mu_1 = 10$ at 5% level of significance.

Q.18.20: The weights in grams of 10 male and 10 female juvenile ring-necked pheasants are:

Males: 1293, 1380, 1614, 1497, 1340, 1340, 1643, 1643, 1466, 1627, 1383, 1711,

Females: 1061, 1065, 1092, 1017, 1021, 1138, 1143, 1094, 1270, 1028

Test the hypothesis of difference of 350 grams between population means in favour of males against the alternative of a greater difference, using a 0.05 level of significance Assume that the weights are normally distributed.

Solution:

i) We state our null and alternative hypothesis

a) $H_0 : \mu_1 - \mu_2 = 350$

a) $H_1 : \mu_1 - \mu_2 > 350$

ii) Assumption: The two samples of sizes $n_1$ *and* $n_2$ are randomly and independently drawn from two normal populations with population mean $\mu_1$ *and* $\mu_2$ when population variances are $\sigma_1^2$ *and* $\sigma_2^2$.unknown but $\sigma_1^2 = \sigma_2^2$ and sample sizes are small.

iii) Level of significance

$\alpha = 5\% = 0.05$

iv) Test-statistic

$$t_c = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

If $H_0$ is true; which has t-distribution with $(V = n_1 + n_2 - 2)$ degree of freedom

v) Critical region

$t_c > t_\alpha(v)$

$t_c > t_{0.05}(28)$ $\qquad\qquad$ $v = 12 + 18 - 2 = 28$

$t_c > 1.701$

vi) Calculation

$n_1 = 10 \qquad n_2 = 10$

| $X_1$ | $(X_1 - \bar{X}_1)^2$ | $X_2$ | $(X_2 - \bar{X}_2)^2$ |
|---|---|---|---|
| 1293<br>1380<br>1614<br>1497<br>1340<br>1340<br>1643<br>1643<br>1466<br>1627<br>1383<br>1711 | | 1061<br>1065<br>1092<br>1017<br>1021<br>1138<br>1143<br>1094<br>1270<br>1028 | |
| 14954 | $\sum(X_1 - \bar{X}_1)^2 = 191586.4$ | 10929 | $\sum(X_2 - \bar{X}_2)^2 = 52848.9$ |

$\bar{X}_1 = \frac{\sum X}{n_1} = \frac{14954}{10} = 1495.4$ $\qquad\qquad$ $\bar{X}_2 = \frac{\sum X}{n_2} = \frac{10929}{10} = 1092.9$

$$S_p = \sqrt{\frac{\sum(X_1 - \bar{X}_1)^2 + \sum(X_1 - \bar{X}_1)^2}{n_1 + n_2 - 2}}$$

$$S_P = \sqrt{\frac{191586.4 + 52848.9}{10 + 10 - 2}} = 116.53$$

$$t_c = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(1495.4 - 1092.9) - (350)}{116.53\sqrt{\frac{1}{10} + \frac{1}{10}}} = 1.007$$

vii) Conclusion

Since our calculated value does not fall's in critical region then we accept $H_0$. And we conclude that the $H_0 : \mu_1 - \mu_2 = 350$ at 5% level of significance.

Q.18.21 (a): The heights of six randomly selected sailors are in inches:63, 65, 68, 69, 71, and 72. Those of ten randomly selected soldiers are 61,62,65,66,69,69,70, 71, 72 and 73. Discuss in the light of these data that the soldiers are on the average taller than sailors. Assume that the heights are normally distributed.

Solution:

i) We state our null and alternative hypothesis

a) $H_0 : \mu_1 \geq \mu_2$

a) $H_1 : \mu_1 < \mu_2$

ii) Assumption: The two samples of sizes $n_1$ *and* $n_2$ are randomly and independently drawn from two normal populations with population mean $\mu_1$ *and* $\mu_2$ when population variances are $\sigma_1^2$ *and* $\sigma_2^2$ .unknown but $\sigma_1^2 = \sigma_2^2$ and sample sizes are small.

iii) Level of significance

$\alpha = 5\% = 0.05$

iv) Test-statistic

$$t_c = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

If $H_0$ is true; which has t-distribution with $(V = n_1 + n_2 - 2)$ degree of freedom

v) Critical region

$t_c < -t_\alpha(v)$

$t_c < -t_{0.05}(14)$ $\qquad\qquad$ $v = 6 + 10 - 2 = 14$

$t_c < -1.761$

vi) Calculation

| $X_1$ | $(X_1 - \bar{X}_1)^2$ | $X_2$ | $(X_2 - \bar{X}_2)^2$ |
|---|---|---|---|
| 63<br>65<br>68<br>69<br>71<br>72 | | 61<br>62<br>65<br>66<br>69<br>69<br>70<br>71<br>72<br>73 | |
| 408 | $\sum(X_1 - \bar{X}_1)^2 = 60$ | 678 | $\sum(X_2 - \bar{X}_2)^2 = 153.60$ |

$\bar{X}_1 = \frac{\sum X}{n_1} = \frac{408}{6} = 68$ $\quad\quad$ $\bar{X}_2 = \frac{\sum X}{n_2} = \frac{678}{10} = 67.8$

$$S_p = \sqrt{\frac{\sum(X_1 - \bar{X}_1)^2 + \sum(X_1 - \bar{X}_1)^2}{n_1 + n_2 - 2}}$$

$$S_P = \sqrt{\frac{60 + 153.6}{6 + 10 - 2}} = 3.91$$

$$t_c = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(68 - 67.8) - (0)}{3.91\sqrt{\frac{1}{6} + \frac{1}{10}}} = 0.099$$

vii) Conclusion

Since our calculated value does not fall's in critical region then we reject $H_0$. And we conclude that the $H_1 : \mu_1 < \mu_2$ at 5% level of significance.

Q.18.21 (b): Eight pots, growing three barley plants each, were exposed to a high tension discharge while nine similar pots were enclosed in an earthed wire cage. The number of tillers (shoots) in each pot was as follows:

Caged: 17, 27, 18, 25, 27, 29, 27, 23, 17.

Electrified: 16, 16, 20, 16, 21, 17, 15, 20.

Discuss whether electrification exercises any real effect on tillering.

Solution:

i) We state our null and alternative hypothesis

a) $H_0 : \mu_1 = \mu_2$

a) $H_1 : \mu_1 \neq \mu_2$

ii) Assumption: The two samples of sizes $n_1$ *and* $n_2$ are randomly and independently drawn from two normal populations with population mean $\mu_1$ *and* $\mu_2$ when population variances are $\sigma_1^2$ *and* $\sigma_2^2$ .unknown but $\sigma_1^2 = \sigma_2^2$ and sample sizes are small.

iii) Level of significance

$\alpha = 5\% = 0.05$

iv) Test-statistic

$$t_c = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

If $H_0$ is true; which has t-distribution with $(V = n_1 + n_2 - 2)$ degree of freedom

v) Critical region

$|t_c| \geq t_{\frac{\alpha}{2}}(v)$

$|t_c| \geq t_{0.025}(15)$ $\quad\quad$ $v = 9 + 8 - 2 = 15$

$|t_c| \geq 2.13$

vi) Calculation

| $X_1$ | $(X_1-\overline{X}_1)^2$ | $X_2$ | $(X_2-\overline{X}_2)^2$ |
|---|---|---|---|
| 17<br>27<br>18<br>25<br>27<br>29<br>27<br>23<br>17 | | 16<br>16<br>20<br>16<br>21<br>17<br>15<br>20 | |
| 210 | $\sum\left(X_1-\overline{X}_1\right)^2=184$ | 141 | $\sum\left(X_2-\overline{X}_2\right)^2=37.88$ |

$n_1=9$ $\quad$ $n_2=8$

$\overline{X}_1=\frac{\sum X}{n_1}=\frac{210}{9}=23.33$ $\quad$ $\overline{X}_2=\frac{\sum X}{n_2}=\frac{141}{8}=17.63$

$$S_p=\sqrt{\frac{\sum\left(X_1-\overline{X}_1\right)^2+\sum\left(X_1-\overline{X}_1\right)^2}{n_1+n_2-2}}$$

$$S_P=\sqrt{\frac{184+37.88}{9+8-2}}=3.85$$

$$t_c=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}=\frac{(23.33-17.63)-(0)}{3.85\sqrt{\frac{1}{9}+\frac{1}{8}}}=3.05$$

vii) Conclusion

Since our calculated value fall in critical region then we reject $H_0$. And we conclude that the $H_1:\mu_1\neq\mu_2$ at 5% level of significance.

Q.18.22: Twelve hogs were fed on diet A, 15 on diet B. The gains in weights for the individual hogs (in pounds) were as shown:

A: 25, 30, 28, 34, 24, 25, 13, 32, 24, 30, 31, 35

B: 44, 34, 22, 8, 47, 31, 40, 30, 32, 35, 18, 21, 35, 29, 22.

What conclusions may be drawn from this experiment?

Solution:

i) We state our null and alternative hypothesis

$H_0:\mu_1=\mu_2$

$H_1:\mu_1\neq\mu_2$

ii) Assumption: The two samples of sizes $n_1$ *and* $n_2$ are randomly and independently drawn from two normal populations with population mean $\mu_1$ *and* $\mu_2$ when population variances are $\sigma_1^2$ *and* $\sigma_2^2$.unknown but $\sigma_1^2=\sigma_2^2$ and sample sizes are small.

iii) Level of significance

$\alpha=5\%=0.05$

iv) Test-statistic

$$t_c=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

If $H_0$ is true; which has t-distribution with $\left(V=n_1+n_2-2\right)$ degree of freedom

v) Critical region

$|t_c|\geq t_{\frac{\alpha}{2}}(v)$

$|t_c|\geq t_{0.025}(25)$ $\quad$ $v=12+15-2=25$

$|t_c|\geq 2.060$

vi) Calculation

| $X_1$ | $(X_1-\bar{X}_1)^2$ | $X_2$ | $(X_2-\bar{X}_2)^2$ |
|---|---|---|---|
| 25<br>30<br>28<br>34<br>24<br>25<br>13<br>32<br>24<br>30<br>31<br>35 | | 44<br>34<br>22<br>8<br>47<br>31<br>40<br>30<br>32<br>35<br>18<br>21<br>35<br>29<br>22 | |
| 331 | $\sum(X_1-\bar{X}_1)^2=390.92$ | 448 | $\sum(X_2-\bar{X}_2)^2=1493.73$ |

$\bar{X}_1=\frac{\sum X}{n_1}=\frac{331}{12}=27.58$ $\bar{X}_2=\frac{\sum X}{n_2}=\frac{448}{15}=29.87$

$$S_p=\sqrt{\frac{\sum(X_1-\bar{X}_1)^2+\sum(X_1-\bar{X}_1)^2}{n_1+n_2-2}}$$

$S_P=8.68$

$$t_c=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}=\frac{(27.58-29.87)-(0)}{8.68\sqrt{\frac{1}{12}+\frac{1}{15}}}=-0.68$$

vii) Conclusion

Since our calculated value does not fall's in critical region then we accept $H_0$. And we conclude that the $H_0:\mu_1=\mu_2$ at 5% level of significance.

Q.18.23 (b): A group of 12 children are found to have the following intelligence quotients:112, 109, 125, 113, 116, 131, 112, 123, 108, 113, 132, and 128.Is it reasonable to suppose that these children have come from a large population whose average 1Q is 115?

Solution:

i) We state our null and alternative hypothesis

a) $H_0:\mu=115$

a) $H_1:\mu\neq115$

ii) Assumption: The samples of size $n$ are randomly and independently drawn from the normal population with population mean $\mu$ when population variances are $\sigma^2$ unknown but sample size is small.

iii) Level of significance

$\alpha=5\%=0.05$

iv) Test-statistic

$$t_c=\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}$$

If $H_0$ is true; which has t-distribution with $(V=n-1)$ degree of freedom

v) Critical region

$|t_c|\geq t_{\frac{\alpha}{2}}(v)$

$|t_c|\geq t_{0.025}(11)$ $v=12-1=11$

$|t_c|\geq2.201$

vi) Calculation

| $X$ | $(X-\bar{X})^2$ |
|---|---|
| 112<br>109<br>125<br>113<br>116<br>131<br>112<br>123<br>108<br>113<br>132<br>128 | |
| 1422 | $\sum(X-\bar{X})^2=841$ |

$$\bar{X}=\frac{\sum X}{n}=\frac{1422}{12}=118.5$$

$$s=\sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}=8.74$$

$$t_c=\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}=\frac{118.5-115}{\frac{8.74}{\sqrt{12}}}=1.39$$

vii) Conclusion

Since our calculated value does not fall's in critical region then we accept $H_0$. And we conclude that the $H_0:\mu=115$ at 5% level of significance.

Q.18.23©: A second group of 10 children is tested, resulting in the following 1Q's:
117, 110, 106, 109, 116, 119, 107, 106, 105 and 108.
Is this group significantly different from the first group?

Solution:

i) We state our null and alternative hypothesis

$H_0:\mu_1=\mu_2$

$H_1:\mu_1\neq\mu_2$

ii) Assumption: The two samples of sizes $n_1$ *and* $n_2$ are randomly and independently drawn from two normal populations with population mean $\mu_1$ *and* $\mu_2$ when population variances are $\sigma_1^2$ *and* $\sigma_2^2$ .unknown but $\sigma_1^2=\sigma_2^2$ and sample sizes are small.

iii) Level of significance

$\alpha=5\%=0.05$

iv) Test-statistic

$$t_c=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

If $H_0$ is true; which has t-distribution with $(V=n_1+n_2-2)$ degree of freedom

v) Critical region

$|t_c|\geq t_{\frac{\alpha}{2}}(v)$ $\qquad v=12+10-2=20$

$|t_c|\geq 2.086$

vi) Calculation

| $X_1$ | $(X_1-\overline{X}_1)^2$ | $X_2$ | $(X_2-\overline{X}_2)^2$ |
|---|---|---|---|
| 112<br>109<br>125<br>113<br>116<br>131<br>112<br>123<br>108<br>113<br>132<br>128 | | 117<br>110<br>106<br>109<br>116<br>119<br>107<br>106<br>105<br>108 | |
| 1422 | $\sum(X_1-\overline{X}_1)^2=841$ | 1103 | $\sum(X_2-\overline{X}_2)^2=236.1$ |

$$\overline{X}_1=\frac{\sum X}{n_1}=\frac{1422}{12}=118.5 \qquad \overline{X}_2=\frac{\sum X}{n_2}=\frac{1103}{10}=110.3$$

$$S_p=\sqrt{\frac{\sum(X_1-\overline{X}_1)^2+\sum(X_1-\overline{X}_1)^2}{n_1+n_2-2}}$$

$$S_P=7.35$$

$$t_c=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}=\frac{(118.5-110.3)-(0)}{7.35\sqrt{\frac{1}{12}+\frac{1}{10}}}=2.61$$

vii) Conclusion

Since our calculated value fall in critical region then we reject $H_0$. And we conclude that the $H_1:\mu_1\neq\mu_2$ at 5% level of significance.

Q.18.24 (a): Two separate groups of subjects were tested. The experimental group (group E) had 10 subjects; the control group (group c) had 9 subjects. The data are given below; the scores are assumed to be normally distribution.

Group E: 12, 13, 16, 14, 15, 12, 15, 14, 13 and 16.

Group C: 10, 13, 14, 12, 15, 16, 13, 14 and 11.

Determine whether the means of the two groups differ significantly at the 0.05 level of significance.

Solution:

i) We state our null and alternative hypothesis

$H_0:\mu_1=\mu_2$

$H_1:\mu_1\neq\mu_2$

ii) Assumption: The two samples of sizes $n_1$ *and* $n_2$ are randomly and independently drawn from two normal populations with population mean $\mu_1$ *and* $\mu_2$ when population variances are $\sigma_1^2$ *and* $\sigma_2^2$ .unknown but $\sigma_1^2=\sigma_2^2$ and sample sizes are small.

iii) Level of significance

$\alpha=5\%=0.05$

iv) Test-statistic

$$t_c=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

If $H_0$ is true; which has t-distribution with $(V=n_1+n_2-2)$ degree of freedom

v) Critical region

$|t_c|\geq t_{\frac{0.05}{2}}(18)$ v=18

$|t_c|\geq 2.11$

vi) Calculation

| $X_1$ | $(X_1-\overline{X}_1)^2$ | $X_2$ | $(X_2-\overline{X}_2)^2$ |
|---|---|---|---|
| 12<br>13<br>16<br>14<br>15<br>12<br>15<br>14<br>13<br>16 | | 10<br>13<br>14<br>12<br>15<br>16<br>12<br>14<br>11 | |
| 140 | $\sum(X_1-\overline{X}_1)^2=20$ | 117 | $\sum(X_2-\overline{X}_2)^2=30$ |

$\overline{X}_1=\frac{\sum X}{n_1}=\frac{140}{10}=14$ $\overline{X}_2=\frac{\sum X}{n_2}=\frac{117}{9}=13$

$$S_p=\sqrt{\frac{\sum(X_1-\overline{X}_1)^2+\sum(X_1-\overline{X}_1)^2}{n_1+n_2-2}}$$

$S_P=1.72$

$$t_c=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}=\frac{(14-13)-(0)}{1.72\sqrt{\frac{1}{10}+\frac{1}{9}}}=1.27$$

vii) Conclusion

Since our calculated value does not fall's in critical region then we accept $H_0$. And we conclude that the $H_0:\mu_1=\mu_2$ at 5% level of significance.

Q.18.24 (b): the strength of ropes made out cotton yarn and coir gave on measurement the following values:

Cotton Yarn: 7.5, 5.4, 10.6, 9.0, 6.1, 10.2, 7.9, 9.7, 7.1, 8.5

Coir: 8.3, 6.1, 9.6, 10.4, 6.4, 10.0, 7.9, 8.9, 7.5, 9.7

Test whether there is a significant difference in the strength of the two types of ropes at 0.05 level of significance.

Solution:

i) We state our null and alternative hypothesis

$H_0:\mu_1=\mu_2$

$H_1:\mu_1\neq\mu_2$

ii) Assumption: The two samples of sizes $n_1$ *and* $n_2$ are randomly and independently drawn from two normal populations with population mean $\mu_1$ *and* $\mu_2$ when population variances are $\sigma_1^2$ *and* $\sigma_2^2$.unknown but $\sigma_1^2=\sigma_2^2$ and sample sizes are small.

iii) Level of significance

$\alpha=5\%=0.05$

iv) Test-statistic

$$t_c=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

If $H_0$ is true; which has t-distribution with $(V=n_1+n_2-2)$ degree of freedom

v) Critical region

$|t_c|\geq t_{\frac{0.05}{2}}(18)$ $v=10+10-2=18$

$|t_c|\geq 2.101$

vi) Calculation

| $X_1$ | $(X_1-\bar{X}_1)^2$ | $X_2$ | $(X_2-\bar{X}_2)^2$ |
|---|---|---|---|
| 7.5<br>5.4<br>10.6<br>9.0<br>6.1<br>10.2<br>7.9<br>9.7<br>7.1<br>8.5 | | 8.3<br>6.1<br>9.6<br>10.4<br>6.4<br>10.0<br>7.9<br>8.9<br>7.5<br>9.7 | |
| 82 | $\sum\left(X_1-\bar{X}_1\right)^2=26.78$ | 84.8 | $\sum\left(X_2-\bar{X}_2\right)^2=20.236$ |

$\bar{X}_1=\frac{\sum X}{n_1}=\frac{82}{10}=8.2$ $\bar{X}_2=\frac{\sum X}{n_2}=\frac{84.8}{10}=8.48$

$$S_p=\sqrt{\frac{\sum\left(X_1-\bar{X}_1\right)^2+\sum\left(X_1-\bar{X}_1\right)^2}{n_1+n_2-2}}$$

$S_P=1.62$

$$t_c=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}=\frac{(8.2-8.48)-(0)}{1.62\sqrt{\frac{1}{10}+\frac{1}{10}}}=-0.39$$

vii) Conclusion

Since our calculated value does not fall's in critical region then we accept $H_0$. And we conclude that the $H_0:\mu_1=\mu_2$ at 5% level of significance.

Q.18.25: Out of 20 children, 10 selected at random were given a ration of orange juice each day and the other 10, a ration of milk. Their gains in weights after a certain period were found to be as follows:

$1^{st}$ group: 2.6, 1.5, 4, 1, 3.5, 3.4, 2.5, 3, 4, 3.5

$2^{nd}$ group: 3.5, 2.5, 1.5, 2.5, 3, 2, 3, 2, 1.5, 2.5

Use t-test to determine whether the means differ significantly, assuming the standard deviation for each group is the same.

Solution:

i) We state our null and alternative hypothesis

$H_0:\mu_1=\mu_2$

$H_1:\mu_1\neq\mu_2$

ii) Assumption: The two samples of sizes $n_1$ *and* $n_2$ are randomly and independently drawn from two normal populations with population mean $\mu_1$ *and* $\mu_2$ when population variances are $\sigma_1^2$ *and* $\sigma_2^2$.unknown but $\sigma_1^2=\sigma_2^2$ and sample sizes are small.

iii) Level of significance

$\alpha=5\%=0.05$

iv) Test-statistic

$$t_c=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

If $H_0$ is true; which has t-distribution with $(V=n_1+n_2-2)$ degree of freedom

v) Critical region

$|t_c|\geq t_{\frac{0.05}{2}}(18)$ $v=10+10-2=18$

$|t_c|\geq 2.101$

vi) Calculation

| $X_1$ | $(X_1-\bar{X}_1)^2$ | $X_2$ | $(X_2-\bar{X}_2)^2$ |
|---|---|---|---|
| 2.6<br>1.5<br>4<br>1<br>3.5<br>3.4<br>2.5<br>3<br>4<br>3.5 | | 3.5<br>2.5<br>1.5<br>2.5<br>3<br>2<br>3<br>2<br>1.5<br>2.5 | |
| 29 | $\sum(X_1-\bar{X}_1)^2=9.22$ | 24 | $\sum(X_2-\bar{X}_2)^2=3.90$ |

$\bar{X}_1=\dfrac{\sum X}{n_1}=\dfrac{29}{10}=2.9$ $\bar{X}_2=\dfrac{\sum X}{n_2}=\dfrac{24}{10}=2.4$

$$S_p=\sqrt{\frac{\sum(X_1-\bar{X}_1)^2+\sum(X_1-\bar{X}_1)^2}{n_1+n_2-2}}$$

$S_P=0.85$

$$t_c=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}=\frac{(2.9-2.4)-(0)}{0.85\sqrt{\frac{1}{10}+\frac{1}{10}}}=1.32$$

vii) Conclusion

Since our calculated value does not fall's in critical region then we accept $H_0$. And we conclude that the $H_0:\mu_1=\mu_2$ at 5% level of significance.

Q.18.26: Given two random samples of size $n_1=9$ and $n_2=10$, from two independent normal population, with $\bar{X}_1=75$, $\bar{X}_2=60$, $s_1=13.61$ and $s_2=11.05$,test the hypothesis at the 0.05 level of significance that $\mu_1$-$\mu_2$ against the alternative that $\mu_1>\mu_2$ .Assume that the populations have unequal variances.

Solution:

i) We state our null and alternative hypothesis

$H_0:\mu_1=\mu_2$

$H_1:\mu_1>\mu_2$

ii) Assumption: The two samples of sizes $n_1$ *and* $n_2$ are randomly and independently drawn from two normal populations with population mean $\mu_1$ *and* $\mu_2$ when population variances are $\sigma_1^2$ *and* $\sigma_2^2$.unknown but $\sigma_1^2\neq\sigma_2^2$ and sample sizes are small.

iii) Level of significance

$\alpha=5\%=0.05$

iv) Test-statistic

$$t_c=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

If $H_0$ is true; which has t-distribution with $v=\dfrac{\left[\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right]^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$ degree of freedom

v) Critical region

$t_c>t_\alpha(v)$ $t_c>t_{0.05}(14)$

$$v=\frac{\left[\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right]^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}=\frac{\left[\frac{(13.61)^2}{9}+\frac{(11.05)^2}{16}\right]^2}{\frac{\left(\frac{(13.61)^2}{9}\right)^2}{8}+\frac{\left(\frac{(11.05)^2}{16}\right)^2}{15}}=14$$

$t_c > 1.76$

vi) Calculation

$n_1 = 9 \quad n_2 = 16 \qquad \bar{X}_1 = 75, \qquad \bar{X}_2 = 60, \qquad s_1 = 13.61 \qquad s_2 = 11.05$

$$t_c=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}=\frac{75-60}{\sqrt{\frac{(13.61)^2}{9}+\frac{(11.05)^2}{16}}}=2.82$$

vii) Conclusion

Since our calculated value fall in critical region then we reject $H_0$. And we conclude that the $H_1 : \mu_1 > \mu_2$ at 5% level of significance.

Q.18.27 (a): Distinguish between situations requiring a two – sample t-test and paired-sample t-test .What distributional assumptions and made in each use?

Solution:

**Describe the procedure for testing of hypothesis about two means with paired observations.**

When observations from two samples are paired either by naturally or by design testing of hypothesis two means are done as under.

Let we drawn the random sample of size “n” paired observation from a normal population with mean $\mu_D = \mu_1 - \mu_2$ and standard deviation $\sigma_d$ but unknown. Then we use test-statistic $t_c = \dfrac{\bar{d}-\mu_D}{\frac{s_d}{\sqrt{n}}}$

## Procedure:

i) We state our null and alternative hypothesis

a) $H_0 : \mu_D = 0$ b) $H_0 : \mu_D \le 0$ c) $H_0 : \mu_D \ge 0$

a) $H_1 : \mu_D \ne 0$ b) $H_1 : \mu_D > 0$ c) $H_1 : \mu_D < 0$

ii) Assumption: Sample of paired observations is random independent and drawn from normal population. When unknown population variance and n<30

iii) Level of significance

$\alpha = (Comonly \;\; uesd\; 5\% \; or \; 1\%)$

iv) Test-statistic

$t_c = \dfrac{\bar{d}-\mu_D}{\frac{s_d}{\sqrt{n}}}$ When sampling done with replacement

If $H_0$ is true; which hast-distribution with $(V = n-1)$ degree of freedom

v) Critical region

It is naturally depend on alternative hypothesis

a) $H_1 : \mu_D \ne 0$ We used two sided test

$|t_c| \ge t_{\frac{\alpha}{2}}(v)$

b) $H_1 : \mu_D > 0$ We use one sided test

$t_c > t_\alpha(v)$

C) $H_1 : \mu_D < 0$ We use one sided test

$t_c < -t_\alpha(v)$

vi) Calculation

In this step we calculate the value of “t” test statistic on the basis of sample data.

Where $s_d^2 = \dfrac{1}{n-1}\left(\sum d^2 - \dfrac{d^2}{n}\right) = \dfrac{\sum\left(d-\bar{d}\right)^2}{n-1}$

vii) Conclusion
If our calculated value does not fall's in critical region then we accept $H_0$ other wise we reject it.
Solution:
Construction of confidence for population mean ($\mu_D$) of paired observation When (n<30) small and $\sigma^2$ are unknown
Then $100(1-\alpha)\%$ confidence interval in form of probability statement for population mean $\mu$ is

$$P\left[-t_{\frac{\alpha}{2}}(v) \le t_c \le -t_{\frac{\alpha}{2}}(v)\right] = 1-\alpha$$

Substituting the value of test-statistic $t_c = \dfrac{\bar{d}-\mu_D}{\frac{s_d}{\sqrt{n}}}$

$$P\left[-t_{\frac{\alpha}{2}}(v) \le \frac{\bar{d}-\mu_D}{\frac{s_d}{\sqrt{n}}} \le t_{\frac{\alpha}{2}}(v)\right] = 1-\alpha$$

Multiplying inside the bracket $\dfrac{s_d}{\sqrt{n}}$

$$P\left[-t_{\frac{\alpha}{2}}(v)\frac{s_d}{\sqrt{n}} \le \bar{d}-\mu_D \le t_{\frac{\alpha}{2}}(v)\frac{s_d}{\sqrt{n}}\right] = 1-\alpha$$

Subtracting the $\bar{X}$ inside the bracket

$$P\left[-\bar{d}-t_{\frac{\alpha}{2}}(v)\frac{s_d}{\sqrt{n}} \le \bar{d}-\mu_D-\bar{d} \le -\bar{d}+t_{\frac{\alpha}{2}}(v)\frac{s_d}{\sqrt{n}}\right] = 1-\alpha$$

$$P\left[-\bar{d}-t_{\frac{\alpha}{2}}(v)\frac{s_d}{\sqrt{n}} \le -\mu_D \le -\bar{d}+t_{\frac{\alpha}{2}}(v)\frac{s_d}{\sqrt{n}}\right] = 1-\alpha$$

Multiplying (-1) inside the bracket and sign of inequality will be change

$$P\left[\bar{d}+t_{\frac{\alpha}{2}}(v)\frac{s_d}{\sqrt{n}} \ge \mu_D \ge \bar{d}-t_{\frac{\alpha}{2}}(v)\frac{s_d}{\sqrt{n}}\right] = 1-\alpha$$

Or

$$P\left[\bar{d}-t_{\frac{\alpha}{2}}(v)\frac{s_d}{\sqrt{n}} \le \mu_D \le \bar{d}+t_{\frac{\alpha}{2}}(v)\frac{s_d}{\sqrt{n}}\right] = 1-\alpha$$

Or $\bar{d} \pm t_{\frac{\alpha}{2}}(v)\dfrac{s_d}{\sqrt{n}}$ Hence the required result

Q.18.27 (b): The weights of 4 persons before they stopped smoking and 5 weeks after they stopped smoking are as follows:

| Person | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Before | 148 | 176 | 153 | 116 |
| After | 154 | 176 | 151 | 121 |

Use the t-test for paired observations to test the hypothesis at the 0.05 level of significance, that giving up smoking has no effect on a person' weight.
Solution:
i) We set up our null and alternative hypothesis

$H_0 : \mu_D = 0$

$H_1 : \mu_D \neq 0$

ii) Assumption: Sample of paired observations is random independent and drawn from normal population. When unknown population variance and n<30.
iii) Level of significance
$\alpha = 0.05$

iv) Test-statistic

$$t_c = \frac{\bar{d} - \mu_D}{\frac{s_d}{\sqrt{n}}}$$

If $H_0$ is true; which has t-distribution with $(V = n-1)$ degree of freedom

v) Critical region

$$|t_c| \geq t_{\frac{\alpha}{2}}(v)$$

$$|t_c| \geq t_{0.025}(3)$$

$$|t_c| \geq 3.18$$

vi) Calculation

| Person | 1 | 2 | 3 | 4 | Total |
|---|---|---|---|---|---|
| Before | 148 | 176 | 153 | 116 | |
| After | 154 | 176 | 151 | 121 | |
| d=Y-X | 6 | 0 | -2 | 5 | 9 |
| $d^2$ | 36 | 0 | 4 | 25 | 65 |

$$\bar{d} = \frac{\sum d}{n} = \frac{9}{4} = 2.25$$

$$s_d = \sqrt{\frac{1}{n-1}\left(\sum d^2 - \frac{\left(\sum d\right)^2}{n}\right)} = \sqrt{\frac{1}{3}\left(65 - \frac{(9)^2}{4}\right)} = 3.86$$

$$t_c = \frac{\bar{d} - \mu_D}{\frac{s_d}{\sqrt{n}}} = \frac{2.25 - 0}{3.86/\sqrt{4}} = 1.17$$

vii) Conclusion

Since our calculated value does not fall's in critical region then we accept $H_0$. And we conclude that the smoking has no effect on person's weights.

Q.18.28: To verify whether a course in statistics improved performance, a similar test was given to 12 participants both before and after the course. The original grades recorded in alphabetical order of the Participants were ,the grades were 44, 40, 61, 52, 32, 44, 70, 41, 67, 72, 53, and 72.After the course, the grades were in the same order 53, 38, 69, 57, 46, 39, 73, 48, 73, 74, 60 and 78.

a) Was the course useful, as measured by performance on the test? Consider these 12 participants as a sample from a population.

Solution:

i) We set up our null and alternative hypothesis

$$H_0 : \mu_D \leq 0$$

$$H_1 : \mu_D > 0$$

ii) Assumption: Sample of paired observations is random independent and drawn from normal population. When unknown population variance and n<30.

iii) Level of significance

$$\alpha = 0.05$$

iv) Test-statistic

$$t_c = \frac{\bar{d} - \mu_D}{\frac{s_d}{\sqrt{n}}}$$

If $H_0$ is true; which has t-distribution with $(V = n-1)$ degree of freedom

v) Critical region

$$t_c > t_{\alpha}(v)$$

$$t_c > t_{0.05}(11)$$

$$t_c > 1.796$$

vi) Calculation

| X | Y | $d = Y - X$ | $d^2$ |
|---|---|---|---|
| 44 | 53 | 9 | 81 |
| 40 | 38 | -2 | 4 |
| 61 | 69 | 8 | 64 |
| 52 | 57 | 5 | 25 |
| 32 | 46 | 14 | 196 |
| 44 | 39 | -5 | 25 |
| 70 | 73 | 3 | 9 |
| 41 | 48 | 7 | 49 |
| 67 | 73 | 6 | 36 |
| 72 | 74 | 2 | 4 |
| 53 | 60 | 7 | 49 |
| 72 | 78 | 6 | 36 |
| 648 | 708 | d=60 | $d^2$ =578 |

$$\bar{d} = \frac{\sum d}{n} = \frac{60}{12} = 5.0$$

$$s_d = \sqrt{\frac{1}{n-1}\left(\sum d^2 - \frac{\left(\sum d\right)^2}{n}\right)} = 5.03$$

$$t_c = \frac{\bar{d} - \mu_D}{\frac{s_d}{\sqrt{n}}} = \frac{5-0}{5.03/\sqrt{12}} = 3.44$$

vii) Conclusion

Since our calculated value fall in critical region then we reject $H_0$ . And we conclude that the performance has been improved.

b) Would the same conclusion be reached if tests were not considered paired? Use 5% level of significance in both cases.

Solution:

i) We state our null and alternative hypothesis

$H_0 : \mu_1 \le \mu_2$

$H_1 : \mu_1 > \mu_2$

ii) Assumption: The two samples of sizes $n_1$ *and* $n_2$ are randomly and independently drawn from two normal populations with population mean $\mu_1$ *and* $\mu_2$ when population variances are $\sigma_1^2$ *and* $\sigma_2^2$ .unknown but $\sigma_1^2 = \sigma_2^2$ and sample sizes are small.

iii) Level of significance

$\alpha = 5\% = 0.05$

iv) Test-statistic

$$t_c = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

If $H_0$ is true; which has t-distribution with $(V = n_1 + n_2 - 2)$ degree of freedom

v) Critical region

$t_c > t_{0.05}(24)$ $\qquad v = 12 + 12 - 2 = 22$

$t_c > 1.717$

vi) Calculation

| $X_1$ | $(X_1-\bar{X}_1)^2$ | $X_2$ | $(X_2-\bar{X}_2)^2$ |
|---|---|---|---|
| 44<br>40<br>61<br>52<br>32<br>44<br>70<br>41<br>67<br>72<br>53<br>72 | | 53<br>38<br>69<br>57<br>46<br>39<br>73<br>48<br>73<br>74<br>60<br>78 | |
| 648 | $\sum\left(X_1-\bar{X}_1\right)^2=2176$ | 708 | $\sum\left(X_2-\bar{X}_2\right)^2=2250$ |

$\bar{X}_1=\frac{\sum X}{n_1}=\frac{648}{12}=54$ $\bar{X}_2=\frac{\sum X}{n_2}=\frac{708}{12}=59$

$$S_p=\sqrt{\frac{\sum\left(X_1-\bar{X}_1\right)^2+\sum\left(X_1-\bar{X}_1\right)^2}{n_1+n_2-2}}$$

$S_P=14.18$

$$t_c=\frac{(\bar{X}_2-\bar{X}_1)-(\mu_2-\mu_1)}{s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}=\frac{(59-54)-(0)}{14.18\sqrt{\frac{1}{12}+\frac{1}{12}}}=0.86$$

vii) Conclusion

Since our calculated value does not fall's in critical region then we accept $H_0$.

Note: Here we cannot get the same conclusion in both cases

Q.18.29: In a certain experiment to compare two types of sheep food A and B, the following result of increase in weights were observed:

| Sheep No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Food A | 49 | 53 | 51 | 52 | 47 | 50 | 52 | 53 |
| Food B | 52 | 55 | 52 | 53 | 50 | 54 | 54 | 53 |

a) Assuming that the two samples of sheep are independent, can we conclude that food B is better than food A?

b) Examine the case when the same set of eight sheep was used in both the foods.

a) Solution:

i) We state our null and alternative hypothesis

$H_0:\mu_1\geq\mu_2$

$H_1:\mu_1<\mu_2$

ii) Assumption: The two samples of sizes $n_1$ *and* $n_2$ are randomly and independently drawn from two normal populations with population mean $\mu_1$ *and* $\mu_2$ when population variances are $\sigma_1^2$ *and* $\sigma_2^2$.unknown but $\sigma_1^2=\sigma_2^2$ and sample sizes are small.

iii) Level of significance

$\alpha=5\%=0.05$

iv) Test-statistic

$$t_c=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

If $H_0$ is true; which has t-distribution with $(V=n_1+n_2-2)$ degree of freedom

v) Critical region

$t_c<-t_{0.05}(14)$ $v=8+8-2=14$

$t_c<-1.76$

vi) Calculation

| $X_1$ | $(X_1-\bar{X}_1)^2$ | $X_2$ | $(X_2-\bar{X}_2)^2$ |
|---|---|---|---|
| 49<br>53<br>51<br>52<br>47<br>50<br>52<br>53 | | 52<br>55<br>52<br>53<br>50<br>54<br>54<br>53 | |
| 407 | $\sum(X_1-\bar{X}_1)^2=30.875$ | 423 | $\sum(X_2-\bar{X}_2)^2=16.875$ |

$$\bar{X}_1=\frac{\sum X}{n_1}=\frac{407}{8}=50.875 \qquad \bar{X}_2=\frac{\sum X}{n_2}=\frac{423}{8}=52.875$$

$$S_p=\sqrt{\frac{\sum(X_1-\bar{X}_1)^2+\sum(X_1-\bar{X}_1)^2}{n_1+n_2-2}}$$

$$S_P=1.85$$

$$t_c=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}=\frac{(50.875-52.875)-(0)}{1.85\sqrt{\frac{1}{8}+\frac{1}{8}}}=-2.16$$

vii) Conclusion

Since our calculated value fall in critical region then we reject $H_0$ and we conclude that food “B” is better than food “A”

b)

Solution:

i) We set up our null and alternative hypothesis

$H_0:\mu_D\le 0$

$H_1:\mu_D>0$

ii) Assumption: Sample of paired observations is random independent and drawn from normal population. When unknown population variance and n<30.

iii) Level of significance

$\alpha=0.05$

iv) Test-statistic

$$t_c=\frac{\bar{d}-\mu_D}{\frac{s_d}{\sqrt{n}}}$$

If $H_0$ is true; which has t-distribution with $(V=n-1)$degree of freedom

v) Critical region

$t_c>t_{0.05}(7)$

$t_c>1.895$

vi) Calculation

| X | Y | $d=Y-X$ | $d^2$ |
|---|---|---|---|
| 49<br>53<br>51<br>52<br>47<br>50<br>52<br>53 | 52<br>55<br>52<br>53<br>50<br>54<br>54<br>53 | 3<br>2<br>1<br>1<br>3<br>4<br>2<br>0 | 9<br>4<br>1<br>1<br>9<br>16<br>4<br>0 |
| 407 | 423 | d=16 | $d^2$=44 |

$$\bar{d}=\frac{\sum d}{n}=\frac{16}{8}=2.0$$

$$s_d = \sqrt{\frac{1}{n-1}\left(\sum d^2 - \frac{\left(\sum d\right)^2}{n}\right)} = 1.31$$

$$t_c = \frac{\bar{d} - \mu_D}{\frac{s_d}{\sqrt{n}}} = \frac{2-0}{1.31/\sqrt{8}} = 4.32$$

vii) Conclusion

Since our calculated value fall in critical region then we reject $H_0$. And we conclude that the food "B" better than food "A".

Q.18.30: The government awarded grants to nine different experimental stations of the agricultural department to test the yield capabilities of two varieties of wheat. Five acres of each variety are planted at each station and the yields in mounds per acre , recorded as follows:

| Station | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Variety 1 | 38 | 23 | 35 | 41 | 44 | 29 | 37 | 31 | 38 |
| Varity 2 | 45 | 25 | 31 | 38 | 50 | 33 | 35 | 40 | 43 |

Test the hypothesis, at the 0.05 level of significance, that the average yields of the two varieties of wheat are equal against the alternative hypothesis that they are unequal, assuming the distribution of yields to be approximately normal. Explain why pairing is necessary in this problem.

Solution:

i) We set up our null and alternative hypothesis

$H_0 : \mu_D = 0$

$H_1 : \mu_D \neq 0$

ii) Assumption: Sample of paired observations is random independent and drawn from normal population. When unknown population variance and n<30.

iii) Level of significance

$\alpha = 0.05$

iv) Test-statistic

$$t_c = \frac{\bar{d} - \mu_D}{\frac{s_d}{\sqrt{n}}}$$

If $H_0$ is true; which has t-distribution with $(V = n-1)$ degree of freedom

v) Critical region

$|t_c| \geq t_{\frac{0.05}{2}}(8)$ $\qquad v = 9-1 = 8$

$|t_c| \geq 2.31$

vi) Calculation

| X | Y | $d = Y - X$ | $d^2$ |
|---|---|---|---|
| 38 | 45 | | |
| 23 | 25 | | |
| 35 | 31 | | |
| 41 | 38 | | |
| 44 | 50 | | |
| 29 | 33 | | |
| 37 | 35 | | |
| 31 | 40 | | |
| 38 | 43 | | |
| | | 24 | 240 |

$$\bar{d} = \frac{\sum d}{n} = \frac{24}{9} = 2.67$$

$$s_d = \sqrt{\frac{1}{n-1}\left(\sum d^2 - \frac{\left(\sum d\right)^2}{n}\right)} = 4.69$$

$$t_c = \frac{\bar{d} - \mu_D}{\frac{s_d}{\sqrt{n}}} = \frac{2.67 - 0}{4.69/\sqrt{9}} = 1.71$$

vii) Conclusion

Since our calculated value does not fall's in critical region then we accept $H_0$. And we conclude that the average yields of two varieties of wheat are equal at 0.05.

Q.18.31: A taxi company is trying to decide whether the use of radial tires instead of regular belted tires improves fuel economy. Twelve cars were equipped with radial tires and driven over a prescribed test course. Without changing drivers, the same cars were then equipped with regular belted tires and driven once again over the test course .The gasoline consumption in km per liter, was recorded as follows:

| | |
|---|---|
| Radial Tires | 4.2, 4.7, 6.6, 7.0, 6.7, 4.5, 5.7, 6.0, 7.4, 4.9, 6.1, 5.2, |
| Belted Tires | 4.1, 4.9, 6.2, 6.9, 6.8, 4.4, 5.7, 5.8, 6.9, 4.7, 6.0, 4.9 |

At the 0.025 level of significance, can we conclude that cars equipped with radial tires give better fuel economy than those equipped with belted tires? Assume the population to be normally distributed.

Solution:

i) We set up our null and alternative hypothesis

$H_0 : \mu_D \le 0$

$H_1 : \mu_D > 0$

ii) Assumption: Sample of paired observations is random independent and drawn from normal population. When unknown population variance and n<30.

iii) Level of significance

$\alpha = 0.025$

iv) Test-statistic

$$t_c = \frac{\bar{d} - \mu_D}{\frac{s_d}{\sqrt{n}}}$$

If $H_0$ is true; it hast-distribution with $(V = n-1)$degree of freedom

v) Critical region

$t_c > t_\alpha(v)$

$t_c > t_{0.025}(11)$

$t_c > 2.201$

vi) Calculation

| X | Y | $d = X - Y$ | $d^2$ |
|---|---|---|---|
| 4.2 | 4.1 | | |
| 4.7 | 4.9 | | |
| 6.6 | 6.2 | | |
| 7.0 | 6.9 | | |
| 6.7 | 6.8 | | |
| 4.5 | 4.4 | | |
| 5.7 | 5.7 | | |
| 6.0 | 5.8 | | |
| 7.4 | 6.9 | | |
| 4.9 | 4.7 | | |
| 6.1 | 6.0 | | |
| 5.2 | 4.9 | | |
| 154 | 176 | 1.7 | 0.67 |

$$\bar{d} = \frac{\sum d}{n} = \frac{1.7}{12} = 0.1417$$

$$s_d = \sqrt{\frac{1}{n-1}\left(\sum d^2 - \frac{\left(\sum d\right)^2}{n}\right)} = 0.1975$$

$$t_c = \frac{\bar{d} - \mu_D}{\frac{s_d}{\sqrt{n}}} = \frac{0.1417 - 0}{0.1975/\sqrt{12}} = 2.5$$

vii) Conclusion

Since our calculated value fall in critical region then we reject $H_0$. And we conclude that the redial tires give better fuel then belt tires at 0.025.

Q.18.32: In the experiment on the effectiveness of a teaching machine, a machine instructed group of students was compared with a teacher instructed group on an achievement test. The following scores were obtained:

| Score | 40-49 | 50-59 | 60-69 | 70-79 | 80-89 | 90-99 | Total |
|---|---|---|---|---|---|---|---|
| Teacher instructed | 21 | 40 | 55 | 38 | 10 | 2 | 166 |
| Machine instructed | 18 | 35 | 42 | 46 | 19 | 4 | 164 |
| Total | 39 | 75 | 97 | 84 | 29 | 6 | 330 |

Apply t-test to determine whether there is significant difference in achievement of the two groups.

Solution:

i) We state our null and alternative hypothesis

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

ii) Assumption: The two samples of sizes $n_1$ *and* $n_2$ are randomly and independently drawn from two normal populations with population mean $\mu_1$ *and* $\mu_2$ when population variances are $\sigma_1^2$ *and* $\sigma_2^2$.unknown but $\sigma_1^2 = \sigma_2^2$ and sample sizes are large the t-test may be thereore correctly applied.

iii) Level of significance

$\alpha = 5\% = 0.05$

iv) Test-statistic

$$t_c = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

If $H_0$ is true; which has t-distribution with $(V = n_1 + n_2 - 2)$ degree of freedom

v) Critical region

$|t_c| \geq t_{\frac{\alpha}{2}}(v)$ $\qquad v = 166 + 164 - 2 = 328$

$|t_c| \geq t_{0.025}(328)$

$|t_c| \geq 1.96$

vi) Calculation

| $X$ | $f_1$ | $f_2$ | $f_1X$ | $f_1X^2$ | $f_2X$ | $f_2X^2$ |
|---|---|---|---|---|---|---|
| 44.5 | 21 | 18 | | | | |
| 54.5 | 45 | 35 | | | | |
| 64.5 | 55 | 42 | | | | |
| 74.5 | 38 | 46 | | | | |
| 84.5 | 10 | 19 | | | | |
| 94.5 | 2 | 4 | | | | |
| | $\sum f_1 = 166 = n_1$ | $\sum f_2 = 164 = n_2$ | 10527 | 689381.5 | 10828 | 741031 |

$\bar{X}_1 = 63.42$ $\qquad \bar{X}_2 = 66.02$

$$S_1^2 = \frac{\sum f_1 X^2}{\sum f_1} - \left(\frac{\sum f_1 X}{\sum f_1}\right)^2 \quad \text{And} \quad S_2^2 = \frac{\sum f_2 X^2}{\sum f_2} - \left(\frac{\sum f_2 X}{\sum f_2}\right)^2$$

$$S_p = \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}}$$

$S_P = 12.09$

$$t_c = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(63.42 - 66.02) - (0)}{12.09\sqrt{\frac{1}{166} + \frac{1}{164}}} = -1.95$$

vii) Conclusion

Since our calculated value does not fall's in critical region then we accept $H_0$ and we conclude that there is no significant difference between the achievement of the two groups at 5% level of significance.