$$f(x) = P(X=x) = \begin{cases} \binom{x+r-1}{r-1} p^r q^x \\ 0 \text{ otherwise} \end{cases}$$

$x \sim NB(r, p)$

Mean $= \frac{rq}{p}$ Var $= \frac{rq}{p^2}$

X = no. of failure before the success
r = no. of success required
p = probability of success
q = probability of failure

16/1/2022

Test of Hypothesis

On the basis of experimental results, we might wish to accept or reject the theory that the me-meu (μ) has a particular value μ_0. The numerical method for the testing such as theory or hypothesis is called a significant test. One may wish to compile a new process with an existing process to see which one is better.

Notations & Definition

1) A hypothesis is a conjunction value of an unknown population

2) ~~The~~ statistical test uses the data obtained ~~of~~ from a sample to make a decision about whether or ~~not~~ not a non-hypothesis should be rejected.

3) The numerical value obtained from a statistical test is called a test value.

4) The level of significance is the maximum probability of computing a type I error.

5) In hypothesis testing situation there are

four possible outcomes.

	Ho True	Ho False
Reject Ho	Type 1 error	Correct decision
Accept Ho	Correct decision	Type 2 error

6/ A type 1 error occurs if one rejects the non-hypothesis when it is true.

7/ A type 2 error occurs when we do not reject the non-hypothesis when false.

8/ The critical value of valids seperates a critical region from a non critical region and is denoted by C.V. (Critical value)

Hypothesis Testing Common Phrase

1/ $>$ (Greater than; its above, higher than, increased etc...)

2/ $<$ (Less than, lower than, shorter than, decrease, reduce)

3/ $\geq$ (at Least, not Less than)

4/ $\leq$ (at most, not more than)

5) = (Equality, same as, have not change
6) ≠ (different from, as changed from, is not the same as)

The f/n solving hypothesis solving problem, the f/f steps are followed

1) State the hypothesis & identify the
2) Find the critical value or values
3) Compute the test value
4) Make the decision to accept or reject a non-hypothesis
5) Summarize the result

N.B ⇒ Reject Ho if calculated value or greater than the calculated value.

Eg 1:-

A sample of size 16 is taken from $N(\mu, 25)$ and mean is 25. Test the hypothesis $H_0: \mu = 47$ versus $H_1: \mu < 47$ at 4% level of critical.

Solution

Step 1: $H_0: \mu = 47$ versses $H_1: \mu < 47$

Step 2: $Z_{tab} = Z_\alpha = Z_{4\%} = Z_{0.04} = -1.75$

$\Rightarrow 0.04006 \Rightarrow -1.7$

$$\begin{array}{r} 0.05 \\ \hline -1.75 \end{array}$$

Step 3: $T.S = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{45 - 47}{5/\sqrt{16}} = \dfrac{-2}{5/4}$

$= -1.6$

Step 4: we Reject H_0 since calculated value (-1.6) is greater than tabulated value (-1.75).

Step 5: Population mean is less than 47.

(2) A sample of size 4 is taken from a standard deviation & S.D 3 gives mean 25. Test $H_0: \mu = 23$ & $H_1: \mu > 23$.

Solution

Step 1: $H_0: \mu = 23$ & $H_1: \mu > 23$

Step 2: $Z_{tab} = Z_{\alpha} = Z_{5\%} = Z_{0.05} = -1.64$

i.e -1.6

0.04

-1.64

Step 3: $T.S = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{25-23}{3/\sqrt{4}} = \dfrac{2}{3/2}$

$= 1.333$

Step 4: We do not reject H_0 because the calculated value (1.333) is less than the C.V (1.64).

(3) Suppose in eg 2 above we're interested in testing $H_0: \mu = 20$ & $H_1: \mu \neq 20$ at 4% level of significance.

Soln.

Step 1: $H_0: \mu = 20$ & $H_1: \mu \neq 20$

Step 2: $Z_{tab} = Z_\alpha = Z_{4\%} = Z_{0.04/2} = Z_{0.02}$

$\Rightarrow 0.02018$

$\Rightarrow 2.05$ ie $\; 2.0 + 0.05 = 2.05$

Step 3: $T.S = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{25 - 20}{3/\sqrt{4}} = \frac{5}{3/2} = 0.833$

(4) An educator estimates the dropout rate for seniors at highschool in USA to be 15% last yr. 38 seniors from a random sample of 200 USA seniors ~~students~~ reduce at $\alpha = 0.05$. Is there a enough evidence ~~non-evidence~~ to withdraw the educators claim?

Solution

$$\Rightarrow Z = \frac{P^n - P}{\sqrt{\frac{pq}{n}}}$$

Step 1: $H_0: P = 15\%$ vs $H_1: P \neq 15\%$
$= 0.15$

Step 2: $Z_{tab} = Z_\alpha = Z_{0.05/2} = Z_{0.025} = -1.96$

$\Rightarrow 0.02500 \Rightarrow -1.9$
$+ 0.06$
-1.96

Step 3: $\hat{P} = \frac{x}{n} = \frac{38}{200} = 0.19$

$P = 0.15$
$q = 1 - P = 1 - 0.15$

$= 0.85$

$n = 200$

$$z = \frac{0.19 - 0.15}{\sqrt{\frac{0.15 \times 0.85}{200}}} = 1.584 = 1.58$$

Step 4: We ~~don't~~ reject H_0 because the calculated (1.58) is greater than the cal val (−1.96).

Step 5: There is enough evidence to support the claim that the dropout rate for seniors at high school in USA is 15%.

(5) An Attorney claims that more than 25% of all lawyers advertise. A sample of 200 lawyers in a certain city shows that 63 had used some form of advertisement. At $\alpha = 0.05$ is there enough evidence to support the attorney's

claim.

Comparison of Sample variance or standard deviation with Specific Variance or S.D

Chi-square test for variance of standard deviation

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \text{ with } (n-1) \text{ d.f}$$

n = sample size $\quad$ σ = Standard deviation

S^2 = " variance

χ^2 = population variable

Eg:- An Instructor wishes to see whether the variation of his course of 23 students in the class is less than variance of the population. The variance of the class is 1.98. Is there enough evidence to support the claim of the variation of Student score.

$(\sigma^2 = 2.25, \alpha = 0.05)$.

Solution

$n=23$

Step 1: $H_0 : \sigma^2 = 2.25$ vs $\sigma^2 > 2.25$

Step 2: $Z_{tab} = \chi^2 = (n-1)\alpha = \chi^2(23)$

$\chi^2_{22}, 0.05 = 33.92$

Step 3: $Z_{cal} = \chi^2 = \dfrac{(n-1)S^2}{\sigma^2}$

$= \dfrac{(23-1)\,1.98}{2.25} = 19.36$

Step 4: We do not reject because the calculated value is less than the table value

A seat manufacturing company wishes to test the claim that the variance of nacotine content of its sig. is 0.644. The content is measured in mg (milligram) and assumed that it is normally distributed.

A sample of siggar has a standard deviation of 1.0mg at $\alpha = 0.05$.

Is there enough evidence to reject the

manufacturer's claim.

9/2/2022 Correlation And Regression

Correlation & Regression implies the relationship b/w two ~~and~~ or more than two attributes or traits values as ~~progesse~~ possessed or exhibited by individuals that make up a set sample of population.

Correlation or regression coefficient b/w two variables;

- It informs the researcher of students the extent of closeness or moving together of the values.

Note:- ~~Base~~ When we're trying to check the relationship b/w 2 variables, we use Correlation

- Cor

- When we have to check the relationship b/w more than 2 variables, we use Regression.

Correlation

It is concerned with relationship b/w variables. They are two methods for determining the correlation coefficient b/w two variables x & y; they are the
* Pearson (r)
* Spearman (ρ)

Their characteristics include;-
i. If $r = \pm 1.00 - \pm 0.80$, then x & y are highly correlated.
ii. If $r = \pm 0.80 - \pm 0.60$, then x & y are substantially correlated.
iii. If $r = \pm 0.60 - \pm 0.40$, then x & y are moderately correlated
iv. If $r = \pm 0.40 - \pm 0.20$; x & y are low correlated
v. If $r = \pm 0.20 - 0$; x & y are low correlated
vi. If $r = \pm 1$; x & y are perfectly correlated
vii. If $r = 0$;- there is no correlation.

A statistical measure which determines the amount of linear relationship b/w x & y is called coefficient of correlation.

Pearson's Coeff. of Correlation

$$r = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{[n\sum x_i^2 - (\sum x_i)^2][n\sum y_i^2 - (\sum y_i)^2]}}$$

r must be b/w $-1 \leq r \leq 1$

Spearman (ρ) Correlation coefficient

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

n = no. of paired variables

d_i = deviation b/w paired variable

Eg:-

Compute the value of Correlation of Coeff for a data obtained in a study of age and blood pressure measurement of 6 randomly selected objects as shown below.

Subject	Age (x)	Pressure (y)	xy	x^2	y^2
A	43	128	5504	1849	16384
B	48	120	5760	2304	14400
C	56	135	7560	3136	18225
D	61	145	8845	3721	21025
E	67	141	9447	4489	19881
F	70	152	10640	4900	23104
Total	345	821	47756	20399	118019

$\sum xy = 47756$ $\quad \sum x^2 = 20399$

$\sum x = 345$

$\sum y^2 = 118019$ $\quad \sum y = 821$

$n = 6$

$$r = \frac{6(47756) - (345 \times 821)}{\sqrt{[(6 \times 20399) - 345^2][6 \times 118019 - 321^2]}}$$

$\therefore r = 0.8884$

Condition:- It is highly correlated.

Eg 2:
Two students were asked to rate 8 diff. textbooks for a specific course on ascending scale from 0-20. Points were assigned to each or several categories such as reading level, use of illustration & a use of column. Determine the corellalation coeff. b/w the 2 students rating.

Textbook	Rating 1	Rating 2	r_x	r_y	d_i	d_i^2
A	4	4	2	1	1	1
B	10	6	5	2	3	9
C	18	20	7	8	-1	1
D	20	14	8	6	2	4
E	12	16	6	7	-1	1
F	2	8	1	4	-3	9
G	5	11	3	5	-2	4
H	9	7	4	3	1	1
Total						30

$\sum d_i^2 = 30 \quad n = 8$

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 30}{8(8^2-1)}$$

$$\rho = 0.6428$$

Conclusion: It is ~~positively~~ substantially correlated.

Linear Regression

Regression analysis is concerned with derivation of an equ. which defines a best fitting regression line. The linear regression line can be described in general by the expression of equation of straight line.

$$Y = a + bx$$

Y = dependent variable
x = independent " or regression variable

$a = \bar{y} - b\bar{x}$ where $\bar{x} = \frac{\Sigma x}{n}$, $\bar{y} = \frac{\Sigma y}{n}$

$$b = \frac{\Sigma xy - n\bar{x}\bar{y}}{\Sigma x^2 - n\bar{x}^2} \text{ or } b = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2}$$

Eg 3;- In forming data records the sales of heating fuel in 1000 liters for a distributor over 10 weeks along with average temp. for each week;-

1. ~~Compute~~ the best fitting regression equ- for sales (i.e y & temperature).
x = level of sales
2. ~~What will be~~ y = temperature

2. What will be the regression model estimate when:-
a) When temperature is 5°C
b) Temperature falls down to 3°C

Week	1	2	3	4	5	6	7	8	9	10
Fuel oil Sales (x)	26	17	7	12	80	40	20	15	10	5
Fuel (y) temp	4	10	14	12	4	5	8	11	13	15

$\bar{y} = 18.2$ $\quad \Sigma x = 96$ $\quad \Sigma xy = 1366$

$\bar{x} = 9.6$ $\quad \Sigma x^2 = 1076$ $\quad \Sigma y = 182$

$n = 10$

$$b = \frac{(10 \times 1366) - (182 \times 96)}{(10 \times 1076) - 96^2} = -2.4689$$

$a = \bar{y} - b\bar{x} = 18.2 - (-2.4689 \times 9.6)$

$a = 41.901$

a) $y = a + bx$

$= 41.901 + (-2.4689)x$

$y = 41.901 - 2.4689x$

b) i) $y = 41.901 - 2.4689(5)$

$= 41.901 - 12.3445$

$= (29.5565 \times 1000)$

$= 29556.5$ litre

ii) $y = 41.901 - 2.4689(3)$

$= 41.901 - 7.4067$

$= (34.4943 \times 1000)$

$= 34494.3$ litre

Correction

$$b = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$$

$$= \frac{10(1366) - (182)(96)}{10(4408) - (182)^2}$$

$$= \frac{-3812}{44080 - 33124} = -0.3479$$

$a = \bar{y} - b\bar{x}$

$= 9.6 - b(18.2)$

$= 9.6 - (-0.3479)\cdot(18.2)$

$= 9.6 - (-6.33178)$

$= 9.6 + 6.33178$

$a = 15.93178$

(a) $y = a + bx$

$= 15.93178 + (-0.3479)\cdot x$

$= 15.93178 - 0.3479x$

(b) if Temp. is $5°C$

$y = 15.93178 - 0.3479(5)$

$= 15.93178 - 1.7395$

$= (14.19228 \times 1000)$

$= 14192.28$ litres

At Temp @ 3°C

$\Rightarrow V = 15.93178 - 0.3479(3)$

$= 15.93178 - 1.0437$

$= 14.88808 \times 1000$

$= 14888.08$ litres