1. Define sampling distribution with examples. If X has a chi-square distribution with n d.f., find m.g.f. $M_x(t)$ Hence under usual notations show that (i) $k_1 k_3 = 2k_2^2$ and (ii) $2\beta_2 - 3\beta_1 - 6 = 0$.

Solve:

A sampling distribution describes the data chosen for a sample from among a larger population.

Suppose, we take multiple random samples of 30 women from a given population and calculate the mean height for each sample. The distribution of these sample means is the sampling distribution of the sample mean.

The square of a standard normal variate is known as a chi-square variate with 1 degree of freedom (d.f)

population

Thus, if $X \sim N(\mu, \sigma^2)$ then

$$z = \frac{X-\mu}{\sigma} \sim N(0,1) \text{ and}$$

$$z^2 = \left(\frac{X-\mu}{\sigma}\right)^2 \text{ is a chi-square variate with 1 d.f.}$$

In general if $x_i, (i = 1, 2 \dots n)$ are n independent normal variates with means μ_i and variance $\sigma_i^2 (i = 1, 2 \dots n)$ then

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2 \text{ a chi-square variate with } n \text{ d.f.}$$

$X \sim \chi^2_{(n)}$ (chi-square distribution with n d.f) and then. the probability density function (p.d.f) is :-

$$f(x) = \frac{1}{2^{n/2}\Gamma\frac{n}{2}} \cdot e^{-x/2} \cdot x^{n/2 - 1} \quad ; \quad 0 \leq x < \infty$$

$$0 \quad ; \text{ otherwise.}$$

if $X \sim \chi^2_{(n)}$, then the moment generating function (m.g.f) is given by,

$$M_x(t) = E(e^{tx})$$

$$= \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$$

$$= \int_0^{\infty} e^{tx} \frac{1}{2^{n/2}\Gamma\frac{n}{2}} \cdot x^{n/2-1} \cdot e^{-x/2} \cdot dx.$$

$$= \frac{1}{2^{n/2}\Gamma\frac{n}{2}} \cdot \int_0^{\infty} e^{-x\left(\frac{1}{2}-t\right)} \cdot x^{n/2-1} \cdot dx \quad \left[\because \int_0^{\infty} e^{-ax} x^{n-1} dx = \frac{\Gamma n}{a^n}\right]$$

$$= \frac{1}{2^{n/2}\Gamma\frac{n}{2}} \cdot \frac{\Gamma\frac{n}{2}}{\left(\frac{1}{2}-t\right)^{n/2}}$$

$$= \frac{1}{2^{n/2}} \cdot \frac{1}{\left(\frac{1}{2}\right)^{n/2}(1-2t)^{n/2}}$$

$$= \frac{1}{(1-2t)^{n/2}}$$

$$= (1-2t)^{-n/2}$$

$\therefore$ The m.g.f of a chi-square distribution with n.d.f is $(1-2t)^{-n/2}$

Again,

If $X \sim \chi^2_{(n)}$ then.

$$K_x(t) = \log M_x(t)$$
$$= \log(1-2t)^{-n/2}$$
$$= -\frac{n}{2}\log(1-2t)$$
$$= -\frac{n}{2}\left(-2t - \frac{4t^2}{2} - \frac{8t^3}{3} - \frac{16t^4}{4} - \cdots\cdots\right)$$
$$= nt + nt^2 + \frac{4}{3}nt^3 + 2nt^4 + \cdots$$

We know,

$$k_n = \text{coefficient of } \frac{t^n}{n!} \text{ in } k_x(t)$$
$$\therefore k_1 = \text{coefficient of } \frac{t}{1!} \text{ in } k_x(t)$$
$$k_1 = n$$

Similarly,

$$k_2 = n \cdot 2! = 2n$$
$$k_3 = \frac{4}{3}n \cdot 3! = 8n$$
$$k_4 = 2n \cdot 4! = 48n.$$

Hence,

$$\text{Mean} = \mu = k_1 = n$$
$$\text{variance} = \mu_2 = k_2 = 2n$$
$$\mu_3 = k_3 = 8n$$
$$\mu_4 = k_4 + 3k_2^2 = 48n + 12n^2$$

Now.

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(8n)^2}{(2n)^3} = \frac{8}{n}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{48n + 12n^2}{(2n)^2}$$
$$= \frac{12}{n} + 3.$$

(i) Given Left Hand side is –

$$k_1 k_3 = n \cdot 8n = 8n^2$$

And the Right Hand side is –

$$2k_2^2 = 2 \cdot (2n)^2 = 8n^2$$

$$\therefore k_1 k_3 = 2k_2^2 \text{ [showed]}$$

(ii). the given equation is –

$$2\beta_2 - 3\beta_1 - 6$$

$$\Rightarrow 2 \cdot \left(\frac{12}{n} + 3\right) - 3 \cdot \frac{8}{n} - 6$$

$$\Rightarrow \frac{24}{n} + 6 - \frac{24}{n} - 6$$

$$= 0$$

$$\therefore 2\beta_2 - 3\beta_1 - 6 = 0 \text{ [showed]}$$

2. Define with examples (i) Parameter and Statistic, (ii) Null and Alternative hypotheses, (iii) Critical value and critical Region. Disscuss about the test procedure of hypothesis testing.

Solve:

(i) A parameter is a number describing a whole population.

Examples: Mean, mode, median, variance are examples of parameter.

Suppose, we want to know the average height of all adult women in a city. Now the calculated average height would be the population mean (μ). This mean is a parameter because it describes a characteristics of the entire population.

A statistic is a number describing a sample taken from the population.

Example: Sample mean, Sample variance. are example of statistic.

Suppose, we select 50 women from the city and measure their heights. The average height of these 50 women is the sample mean ($\bar{x}$). This mean is a statistic because it describes a characteristic of the sample.

(ii) Null hypothesis is the claim that there's no effect in the population. If the sample provide enough evidence against the claim that there's no effect in the population then we can reject the null hypothesis. Otherwise we fail to reject the null hypothesis. It is denoted by H_0.

Alternative hypothesis is the claim that there's an effect in the population. Alternative hypothesis accept when we reject null hypothesis. It is denoted by H_1.

Example:

Suppose a 'X' companies manager whose monthly income is 113,000 dollars.

Then the Null and alternative hypothesis will be:-

Null hypothesis: Manager's income is 113,000 dollars

Alternative hypothesis: Manager's income is not 113,000 dollar.

(iii) A critical value is the value that defines the boundary of the critical region.

The critical region is the set of all values of the test statistic that leads to the rejection of the null hypothesis.

Let's assume, for 5% significance level the value of z distribution is 1.645 (for one-tailed test). And it is the critical value. And the critical region consists of values greater than 1.645. If the test statistic calculate from sample data falls into this region, the null hypothesis will be rejected.

Test procedure of hypothesis testing:

1. State Null hypothesis and Alternative hypothesis.
2. Chose the significance level. It is the probability of rejecting the null hypothesis when it is actually true. Common choices are. 0.05, 0.01 and 0.10.
3. Select the appropriate test statistic depends on the given type of data. Common test statistic include the z-test, t-test, chi-square and F-statistic.
4. Determine the critical value. from the chosen significance level and the distribution of the test statistic.
5. Use the sample data to calculate the value of the test statistic.
6. Compare the calculate test statistic to the critical value
7. Decide whether to reject or fail to reject the null hypothesis.

CamScanner

3. What is power of a test? What is BCR stand for? Explain the test procedure for testing the (i) $H_0: \mu = \mu_0$ vs $H_1: \mu < \mu_0$ (For large sample case), (ii) $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$ (for small and unknown variance) and (iii) $H_0: \rho = 0$ vs $H_1: \rho \neq 0$.

Solve:

power of a test is the probability of rejecting the null hypothesis (H_0) when the alternative hypothesis (H_1) is true.

The power of a test is given by:

$$\text{Power} = 1 - \beta.$$

where β is the probability of making a Type II error.

BCR stands for Bayes Classification Rate. It is the highest possible classification accuracy that can be achieved by any classifier, assuming that the true underlying distributions of the data are known.

(i) $H_0: \mu = \mu_0$ vs $H_1: \mu < \mu_0$

Test procedure:

1. State the hypothesis.

   $H_0: \mu = \mu_0$ (Null hypothesis)

   $H_1: \mu < \mu_0$ (Alternative hypothesis)

2. Chose the Significance Level (Common choices are 0.05, 0.01 or 0.10).

3. Calculate the test statistic. As given hypothesis is for large sample case the test will be z-test statistic. The z-test statistic is calculate using the following formula:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

where $\bar{x}$ is the mean, μ_0 is the population mean, σ is the population standard deviation and n is the sample size.

4. Determine the critical value. The given hypothesis is left tail test so the critical value $Z_{-\alpha}$ such that $P(Z < Z_{-\alpha}) = \alpha$

5. Make the decision.

For left tail test reject H_0 if $Z < Z_{-\alpha}$

6. Based on the comparison draw a conclusion.

(ii) $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$ (for small and unknown variance)

<u>Test procedure:</u>

1. State the hypothesis.

$H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$

2. Choose the significance level.

3. Calculate the test statistic using the following formula.

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$ (As sample size small and unknown variance)

4. Determine the degrees of freedom (df) and Critical value. As it is two tailed test, df is (n-1) and the critical value $t_{\alpha/2}$ such that the area in each tail is $\alpha/2$.

5. Compare the calculate test statistic t to the critical value $\pm t_{\alpha/2}$. Reject Ho if t is less than $-t_{\alpha/2}$ or greater than $t_{\alpha/2}$.

6. Draw a conclusion

(iii). $H_0 : \rho = 0$ vs $H_1 : \rho \neq 0$

<u>Test procedure:</u>

1. State the hypothesis.

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$

2. choose the significance level.
3. Determine the degrees of freedom. $(n-2)$ for this case as it is correlation test.

4. Calculate the test statistic using the following formula:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

5. Determine the critical value.

6. Compare the calculated test statistic to the critical value.

$$|t| > t_{\alpha/2; df}$$

If t is greater than the critical value, reject the null hypothesis.

7. Draw a conclusion.

5. What are dichotomous and manifold classifications? State the uses of Yate's correction. Find the value of chi-square for 2×2 contingency table with Yates correction.

Solve:

An attribute is said to be dichotomously classified is it possessed only two catagories. In other words, the binary type qualitative data is also termed as dichotomously.

An attribute is said to be manifold if it possess more than two catagories.

In a 2×2 contingency table, the number of d.f is $(2-1)(2-1)=1$. If any one of the theoretical cell frequencies is less than 5, then use of pooling method for χ^2-test results in χ^2 with 0 d.f. which is meaningless. In this case we apply a correction which is known as "Yate's Correction for Continuity".

Yate's correction is used primarily to adjust the chi-square test for continuity when dealing with small sample size in a 2×2 contingency table. It improves accuracy and helps maintain the correct Type I error rate.

For the 2×2 table,

Here, $N = a+b+c+d$.

a	b
c	d

a	b	a+b
c	d	c+d
a+c	b+d	N

Under the hypothesis of independent of attributes,

$$E(a) = \frac{(a+b)(a+c)}{N}$$

$$E(b) = \frac{(a+b)(b+d)}{N}$$

$$E(c) = \frac{(a+c)(c+d)}{N}$$

$$E(d) = \frac{(b+d)(c+d)}{N}.$$

For 2×2 contingency table, the Yate's corrected chi-square statistic is calculated by.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Now,

$$\chi^2 = \frac{\{a-E(a)\}^2}{E(a)} + \frac{\{b-E(b)\}^2}{E(b)} + \frac{\{c-E(c)\}^2}{E(c)} + \frac{\{d-E(d)\}^2}{E(d)}$$

Here,

$$a - E(a) = a - \frac{(a+b)(a+c)}{N} = \frac{a(a+b+c+d) - (a^2 - ac + ab + bc)}{N}$$

$$= \frac{ad-bc}{N}$$

Similarly we get,

$$b - E(b) = -\frac{ad-bc}{N}$$

$$c - E(c) = -\frac{ad-bc}{N}$$

$$d - E(d) = \frac{ad-bc}{N}.$$

$$\chi^2 = \frac{(ad-bc)^2}{N^2}\left[\frac{1}{E(a)}+\frac{1}{E(b)}+\frac{1}{E(c)}+\frac{1}{E(d)}\right]$$

$$= \frac{(ad-bc)^2}{N}\left[\left\{\frac{1}{(a+b)(a+c)}+\frac{1}{(a+b)(b+d)}\right\}+\left\{\frac{1}{(a+c)(c+d)}+\frac{1}{(b+d)(c+d)}\right\}\right]$$

$$= \frac{(ad-bc)^2}{N}\left[\frac{b+d+a+c}{(a+b)(a+c)(b+d)}+\frac{b+d+a+c}{(a+c)(c+d)(b+d)}\right]$$

$$= (ad-bc)^2\left[\frac{c+d+a+b}{(a+b)(a+c)(b+d)(c+d)}\right]$$

$$= \frac{N(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)} \quad \text{——} (1)$$

According to Yate's correction, we subtract $\frac{1}{2}$ from a and d and add $\frac{1}{2}$ to b and c so that the marginal totals are not disturbed at all. Thus corrected value of χ^2 is given as:-

$$\chi^2 = \frac{N\left[\left(a-\frac{1}{2}\right)\left(d-\frac{1}{2}\right)-\left(b+\frac{1}{2}\right)\left(c+\frac{1}{2}\right)\right]^2}{(a+c)(b+d)(a+b)(c+d)}$$

$$= \frac{N\left[(ad-bc)-\frac{1}{2}(a+b+c+d)\right]^2}{(a+c)(b+d)(a+b)(c+d)}$$

$$= \frac{N\left[(ad-bc)-\frac{N}{2}\right]^2}{(a+c)(b+d)(a+b)(c+d)}$$

which is required chi-square value. for 2x2 contingency table with Yate's correction.

6. What is point estimation? State the different methods of point estimation. Discuss the principle of MLE. Find the MLE of θ for $f(x/\theta) = \frac{e^{-\theta}\theta^{x}}{x!}$; $\theta = 0, 1, 2 \dots \alpha$.

Solve:

In statistic, point estimation is a process of finding an approximate value of some parameters such as mean(μ) of a population from the random sample of the population.

Methods of point estimation:

There are different types of methods of point estimation. Some are-

1. Method of maximum Likelihood Estimation (MLE)
2. Method of minimum variance.
3. Method of moment.
4. Method of least squares.
5. Method of minimum chi-square.
6. Method of inverse probability.

Given

$$f(x/\theta) = \frac{e^{-\theta}\theta^{x}}{x!} \; ; \; \theta = 0, 1, 2 \dots \alpha .$$

The likelihood function

$$L(x;\theta) = f(x_1;\theta) \cdot f(x_2;\theta) \dots f(x_n;\theta)$$

$$= \frac{e^{-\theta} \cdot \theta^{x_1}}{x_1!} \cdot \frac{e^{-\theta} \cdot \theta^{x_2}}{x_2!} \dots \frac{e^{-\theta} \cdot \theta^{x_n}}{x_n!}$$

$$= e^{-n\theta} \left\{ \frac{\theta^{x_1+x_2+\dots x_n}}{\prod_{i=1}^{n} (x_i)!} \right\}$$

$$\log L = -n\theta + \sum_{i=1}^{n} x_i \log\theta - \sum_{i=1}^{n} \log (x_i!)$$

The likelihood equation for estimating θ is

$$\frac{\partial}{\partial\theta} \log L = 0$$

$$\Rightarrow -n + \frac{\sum_{i=1}^{n} x_i}{\theta} = 0$$

$$\Rightarrow -n + \frac{n\bar{x}}{\theta} = 0$$

$$\therefore \theta = \bar{x}$$

Thus the M.L.E for θ is the sample mean. $\bar{x}$.

Principle of MLE:

Let $x_1, x_2 \cdots x_n$ be a random sample of size n from a population with density function $f(x;\theta)$ Then the likelihood function of their sample values $x_1, x_2 \cdots x_n$ usually denoted by –

$$L(x/\theta) = L(x;\theta) = f(x_1;\theta).f(x_2;\theta) \cdots\cdots f(x_n;\theta)$$
$$= \prod_{i=1}^{n} f(x_i;\theta) \quad \text{— (I)}$$

To maximize (I), we have to get.

$$\frac{\partial}{\partial\theta} L(x;\theta) = \frac{\partial}{\partial\theta}\left\{\prod_{i=1}^{n} f(x_i,\theta)\right\} \quad \text{— (II)}$$

and $$\frac{\partial^2}{\partial\theta^2} L(x;\theta) = \frac{\partial^2}{\partial\theta^2}\left\{\prod_{i=1}^{n} f(x_i,\theta)\right\} \text{— (III)}$$

if $\frac{\partial L}{\partial\theta} = 0$ and $\frac{\partial^2 L}{\partial\theta^2} < 0$ then the value of θ obtaine from (II) is the maximum value of θ.

which is required MLE for θ.