$$E(X) = np \quad Var(X) = npq = np(1-p)$$

★ **The Poisson Distribution**

$$\boxed{f(X) = \frac{e^{-\lambda}\lambda^x}{x!}} \quad x = 0, 1, 2, \ldots, \lambda > 0.$$

$$E(X) = var(X) = \lambda.$$

• STATISTICAL INFERENCE: ESTIMATION

★ **Point Estimation**

$$\hat{\theta} = f(x_1, x_2, \ldots, x_n)$$

$\hat{\theta} \Rightarrow$ statistic, estimator

Particular numerical value of $\hat{\theta}$ is known as an estimate.

Let

$$\hat{\theta} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n) = \overline{X} \rightarrow \text{sample mean}$$

$\overline{X}$ is an estimator of the true mean value.

The estimator $\hat{\theta}$ obtained is known as a point estimator because it provides only a single (point) estimate of θ.

★ **Interval Estimation**

If X is normally distributed,

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

Sample Mean,

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$\theta_1, \theta_2 \rightarrow$ estimators

$$P(\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2) = 1 - \alpha, \quad 0 < \alpha < 1$$

i.e., the probability is $1-\alpha$ that the interval from $\hat{\theta}_1$ to $\hat{\theta}_2$ contains the true θ.

This interval is known as a confidence interval of size $1-\alpha$ for θ, $1-\alpha$ being known as the confidence coefficient.

- If $\alpha = 0.05$, then $1-\alpha = 0.95$, meaning that if we construct a confidence interval with a confidence coefficient of 0.95, then in repeated such constructions resulting from repeated sampling we shall be right in 95 out of 100 cases if we maintain that the interval contains the true θ.

$\alpha \rightarrow$ level of significance
(probability of ~~comit~~ committing a Type-1 error)

★ Methods of Estimation

1. Least squares (LS)
2. Maximum likelihood
3. Method of Moments (MOM)

- Maximum Likelihood

likelihood function:

$g(x_1, x_2, \ldots, x_n, \theta) \rightarrow$ joint pdf of n X values

$X \sim f(X, \theta)$

$$g(x_1, x_2, \ldots, x_n; \theta) = f(x_1;\theta) f(x_2;\theta) \ldots f(x_n;\theta)$$

If θ is known, g is the joint probability of observing the given sample values,

$g \rightarrow$ function of θ given the values of $x_1, x_2, \ldots, x_n$.

Likelihood func^n (func^n of θ given $x_1, x_2, \ldots, x_n$)

$$L(\theta; x_1, x_2, \ldots, x_n) = f(x_1;\theta) f(x_2;\theta) \ldots f(x_n;\theta)$$

The ML estimator of θ is that value of θ that maximizes the (sample) likelihood function, L.

* Small Sample Properties

- Unbiasedness

$$E(\hat{\theta}) = \theta$$

Biased :

$$Bias(\hat{\theta}) = E(\hat{\theta}) - \theta$$

Unbiasedness is a property of repeated sampling, not of any given sample: Keeping the sample size fixed, we draw several samples, each time obtaining an estimate of the unknown parameter. The average value of these estimates is expected to be equal to the true value if the estimator is to be unbiased.

- Best Unbiased, or Efficient, Estimator

Let $\hat{\theta}_1, \hat{\theta}_2 \Rightarrow$ unbiased estimators of θ.

If $var(\hat{\theta}_1) \leq var(\hat{\theta}_2)$ then

$\hat{\theta}_1 \Rightarrow$ minimum-variance unbiased, best unbiased or efficient estimator.

- Linearity

An estimator $\hat{\theta}$ is said to be a linear estimator of θ if it is a linear function of the sample observations.

$$\bar{X} = \frac{1}{n}\sum X_i = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$$

- Best Linear Unbiased Estimator

$\hat{\theta} \Rightarrow$ Linear, unbiased, minimum variance

in the class of all linear unbiased estimators of θ.

- Minimum Mean-Square-Error (MSE) Estimator

$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 \Rightarrow$ Dispersion around the true value.

$var(\hat{\theta}) = E[\hat{\theta} - E(\hat{\theta})]^2 \rightarrow$ dispersion around mean or average.

$$MSE(\hat{\theta}) = E(\hat{\theta}-\theta)^2$$
$$= E[\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta]^2$$
$$= E[\hat{\theta} - E(\hat{\theta})]^2 + E[E(\hat{\theta}) - \theta]^2 + 2E[\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta]$$

$$\cancel{= E[\hat{\theta} - E(\hat{\theta})]^2}$$

$$2E[\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta] = 2\{[E(\hat{\theta})]^2 - [E(\hat{\theta})]^2 - \theta E(\hat{\theta}) + \theta E(\hat{\theta})\}$$
$$= 0.$$

$$\therefore MSE(\hat{\theta}) = E[\hat{\theta} - E(\hat{\theta})]^2 + E[E(\hat{\theta}) - \theta]^2$$
$$= var(\hat{\theta}) + bias(\hat{\theta})^2$$
= variance of $\hat{\theta}$ plus square bias.

$$\boxed{MSE(\hat{\theta}) = var(\hat{\theta}) + bias(\hat{\theta})^2}$$

Minimum MSE criterion is used when the best unbiased criterion is incapable of producing estimators with smaller variances.

★ Large-Sample Properties

• Asymptotic Unbiasedness

$$\boxed{\lim_{n\to\infty} E(\hat{\theta}_n) = \theta}$$

E.g. $S^2 = \frac{\sum(x_i - \bar{x})^2}{n}$ $\quad E(S^2) = \sigma^2\left(1 - \frac{1}{n}\right)$

$$\lim_{n\to\infty} E(S^2) = \sigma^2$$

• Consistency

$\hat{\theta}$ is said to be a consistent estimator if it approaches the true value θ as the sample size gets larger and larger.

If in the limit (i.e. when n increases indefinitely) the distribution of $\hat{\theta}$ collapses to the single point θ, i.e. the

~~As~~ distribution of $\hat{\theta}$ has zero spread, or variance, we say that $\hat{\theta}$ is a consistent estimator of θ.

$$\boxed{\lim_{n\to\infty} P\{|\hat{\theta}-\theta| < \delta\} = 1} \quad \delta > 0$$

$$\underset{n\to\infty}{\text{plim}}\ \hat{\theta} = \theta$$

Sufficient condition:

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta, \quad \lim_{n\to\infty} var(\hat{\theta}_n) = 0$$

Invariance (Slutsky property)

If $\hat{\theta} \Rightarrow$ consistent estimator of θ

$h(\hat{\theta}) \Rightarrow$ continuous function of $\hat{\theta}$

$$\underset{n\to\infty}{\text{plim}}\ h(\hat{\theta}) = h(\theta)$$

- **Asymptotic Efficiency**

If $\hat{\theta}$ is consistent and its asymptotic variance is smaller than the asymptotic variance of all other consistent estimators of θ, $\hat{\theta}$ is called asymptotically efficient.

- **Asymptotic Normality**

An estimator $\hat{\theta}$ is said to be asymptotically normally distributed if its sampling distribution tends to approach the normal distribution as the sample size n increases indefinitely.

* Statistical Inference: Hypothesis Testing

Q. Could our sample have come from the PDF

$$f(x;\theta) = \theta^*$$

Null hypothesis: $\theta = \theta^*$ (H_0)

Alternative hypothesis: $\theta \neq \theta^*$ (H_1)

Simple hypothesis:

$H_0: \mu = 15$ and $\sigma = 2$ (Values are specified)

Composite hypothesis:

$H_0: \mu = 15$ and $\sigma > 2$ (Values are not specified)

<u>Example:</u>

$$X_i \sim N(\mu, \sigma^2) = N(\mu, 2.5^2)$$

$$\bar{X} = 67 \quad n = 100$$

$$H_0: \mu = \mu^* = 69$$

$$H_1: \mu \neq 69.$$

The Confidence Interval Approach

$$X_i \sim N(\mu, \sigma^2)$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

If $\alpha = 0.05$, we will have a 95% confidence interval and if ~~to~~ this confidence interval includes μ^*, we may not reject the null hypothesis - 95 out of 100 intervals thus established are likely to include μ^*.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$Z_i = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$P(-1.96 \leq Z_i \leq 1.96) = 0.95$$

$$P\left[-1.96 \leq \frac{\bar{X}-\mu}{2.5/10} \leq 1.96\right] = 0.95$$

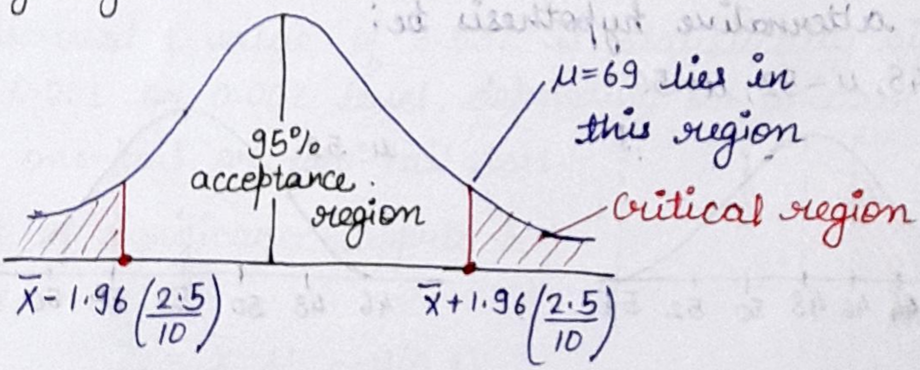
$$\Rightarrow P\left[(\bar{X}-1.96)0.25 \leq \mu \leq (\bar{X}+1.96)0.25\right] = 0.95$$

$$\Rightarrow P(66.51 \leq \mu \leq 67.49$$

$\mu = 69$ does not lie in the confidence interval, thus we reject the null hypothesis.

The confidence interval that we have established is called the acceptance region and the area(s) outside the acceptance region is (are) called the critical region(s), or regions of rejection of the null hypothesis. The lower and upper limits of the acceptance region are called the **critical values**.

If the hypothesised value falls inside the acceptance region, one may not reject the null hypothesis; otherwise one may reject it.



Errors:

Type I error: Rejection of the null hypothesis when it is true.

e.g. $\bar{X} = 67$ could have come from the population with a mean of 69.

Type II error: Not rejecting the null-hypothesis when it is false.

$\alpha \Rightarrow$ Probability of Type I error
(level of significance)

$\beta \Rightarrow$ Probability of Not committing a Type II error
(Power of the test)

The Power of a test is its ability to reject a false Null Hypothesis.

Example: $X \sim N(\mu, 100)$

$H_0 : \mu = 50 \qquad n = 25$

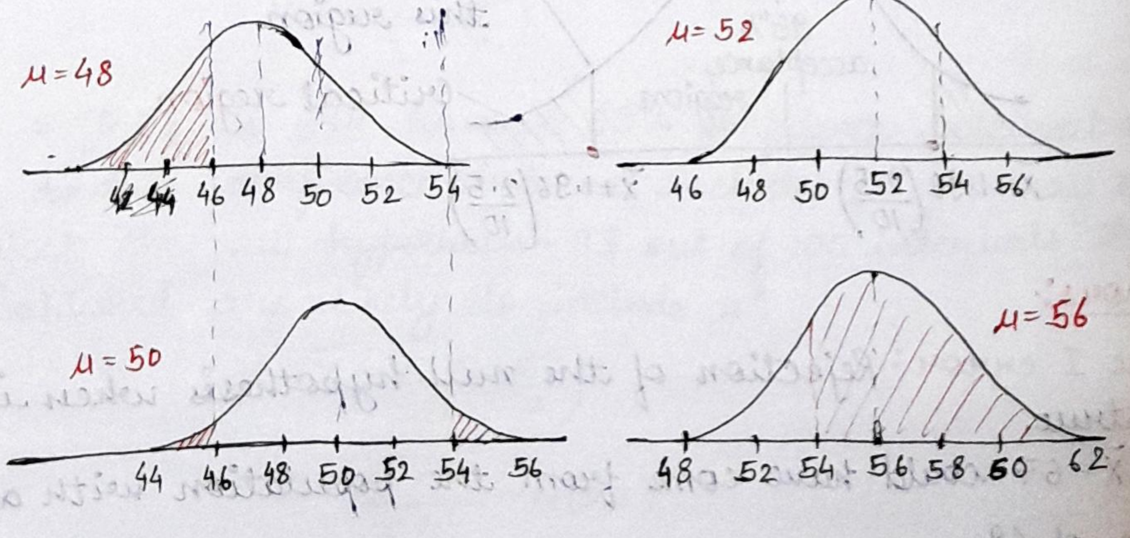
$\bar{X} \sim N(\mu, 100/25)$

Confidence interval: $\mu \pm 1.96\sqrt{4} \Rightarrow \mu \pm 3.92$ i.e.
$(46.08, 53.92)$

Power: what is the probability that $\bar{X}$ will lie in the preceding critical region(s) if the true μ has a value different from 50?

Let the alternative hypothesis be:

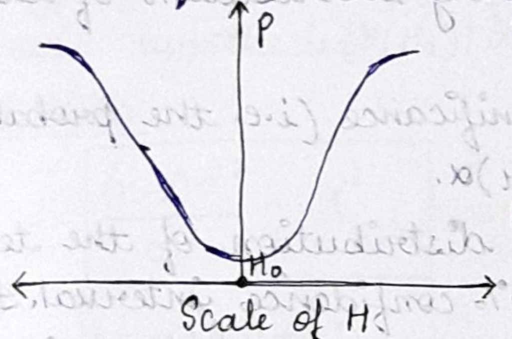
$\mu = 48, \mu = 52, \mu = 56$



The shaded area indicates the probability that $\bar{X}$ will fall into the critical region. This probability is:

0.17 if $\mu = 48$ 0.17 if $\mu = 52$

0.05 if $\mu = 50$ 0.85 if $\mu = 56$



P → probability of rejecting H.

Power function of test of hypothesis

★ The p-Value (Exact Level of Significance)

The p-value is defined as the lowest significance level at which a ~~nt~~ null hypothesis can be rejected.

Example:

Suppose that in an application involving 20 df we obtain a t value of 3.552. Now the p-value, or the exact probability of obtaining a t-value of 3.552 or greater is 0.001 (one-tailed) or 0.002 (two-tailed).

∴ The observed t value of 3.552 is statistically significant at the 0.001 or 0.002 level, depending on whether we are using a one-tail or two-tail test.

★ The test of significance approach

Example:

$$Z_i = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

If the probability of obtaining the computed Z_i value from the normal distribution is less than 1% or 5%, we can reject the null hypothesis - if the hypothesis were true, the chances of obtaining the particular Z value should be very high.

Step 1: State the null hypothesis H_0 and the alternative hypothesis H_1.

Step 2: Select the test static (e.g. $\bar{X}$)

Step 3: Determine the probability distribution of the test statistic (e.g. $\bar{X} \sim N(\mu, \sigma^2/n)$)

Step 4: Choose the level of significance (i.e. the probability of committing a type I error) α.

Step 5: Using the probability distribution of the test statistic, establish a $100(1-\alpha)\%$ confidence interval. If the value of the parameter under the null hypothesis (e.g. $\mu = \mu^* = a$) lies in the confidence region, the region of acceptance, do not reject the null hypothesis. But if it falls ~~out~~ outside this interval (i.e., it falls into the region of rejection), you may reject the null hypothesis. In not rejecting or rejecting a null hypothesis ~~you are~~ you are taking a chance of being wrong $\alpha\%$ of the time.

- **Null Hypothesis versus Alternative Hypothesis**

Model: $Y_i = \alpha + \beta X_i + u_i$

for the overall model:

$$\cancel{H_0 = \beta} \quad H_0: \beta = 0 \quad H_a: \beta \neq 0$$

For the individual coefficients:

Two-tailed test:

(a) Null hypothesis (H_0): $\beta = 0$

Alternative hypothesis (H_a): $\beta \neq 0$

(b) Null hypothesis (H_0): $\alpha = 0$

Alternative hypothesis (H_a): $\alpha \neq 0$.

One - tailed test:

(a) Null hypothesis (H_0): $\beta = 0$

Alternative Hypothesis (H_a): $\beta > 0$ or $\beta < 0$

(b) Null hypothesis (H_0): $\alpha = 0$

Alternative hypothesis (H_a): $\alpha > 0$ or $\alpha < 0$.

* Test Statistic

$$\boxed{\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2}}$$

$$Y_i = \alpha + \beta X_i + u_i$$

$$\left.\begin{array}{l} E(Y_i) = \alpha + \beta X_i \\ var(Y_i) = \sigma^2 \end{array}\right\} \boxed{Y_i \sim N(\alpha + \beta X_i, \sigma^2)}$$

Y_i's are ~~ider id~~ independently and normally distributed but not identically because the mean for each Y_i is different.

If in repeated sampling confidence intervals like it are constructed a great many times on the $1-\alpha$ probability basis, then, in the long run, on the average, such intervals will enclose in $1-\alpha$ of the cases the true value of the parameter.

* Test Statistic

For the overall model:

$$F = \frac{ESS/df_1}{RSS/df_2} = \frac{ESS/k-1}{RSS/n-k} = \frac{ESS/1}{RSS/n-2} \sim F_{(1,n-2)}$$

~~for the overall, model~~:

This F statistic is trying to compare the explained part of the model to the unexplained or residual part. As the ESS or explained part goes up, RSS goes down & F stat goes up and there will be more chances of rejecting the null hypothesis.

for the individual coefficients:

for $\hat{\beta}$: $\boxed{t_{\hat{\beta}} = \frac{\hat{\beta} - E(\hat{\beta})}{SE(\hat{\beta})}} = \frac{\hat{\beta} - \beta}{SE(\hat{\beta})} = \frac{\hat{\beta}}{SE(\hat{\beta})} \sim t_{(n-2)}$

$$SE(\hat{\beta}) = \sqrt{\frac{\hat{\sigma}^2}{\cancel{(n-2)}}} \sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2}} = \sqrt{\frac{\frac{\hat{u}_i^2}{n-2}}{\sum x_i^2}}$$

$$t = \frac{(\hat{\beta} - \beta)\sqrt{\sum x_i^2}}{\hat{\sigma}}$$ ~~Estimated~~

$$= \frac{\text{Estimator} - \text{Parameter}}{\text{Estimated standard error of estimator}}$$

$$P[-t_{\alpha/2} \leqslant t \leqslant t_{\alpha/2}] = 1 - \alpha$$

$t_{\alpha/2}$ ⇒ critical value of t at $\alpha/2$ level of significance

$$P\left[-t_{\alpha/2} \leqslant \frac{\hat{\beta} - \beta}{Se(\hat{\beta})} \leqslant t_{\alpha/2}\right] = 1 - \alpha$$

$$P[\hat{\beta} - t_{\alpha/2} Se(\hat{\beta}) \leqslant \beta \leqslant \hat{\beta} + t_{\alpha/2} Se(\hat{\beta})] = 1 - \alpha$$

∴ $100(1-\alpha)\%$ confidence interval for β:

$$\hat{\beta} \pm t_{\alpha/2} Se(\hat{\beta})$$

The width of the confidence interval is proportional to the standard error of the estimator. That is, the larger the standard error, the larger is the width of the confidence interval. The larger the standard error of the estimator, the greater is the uncertainty of estimating the true value of the unknown parameter. The standard error of an estimator is often described as a measure of the precision of the estimator (i.e.

how precisely the estimator measures the true population value).

For $\hat{\alpha}$: $\boxed{t_{\hat{\alpha}} = \frac{\hat{\alpha} - E(\hat{\alpha})}{SE(\hat{\alpha})}} = \frac{\hat{\alpha} - \alpha}{SE(\hat{\alpha})} = \frac{\hat{\alpha}}{SE(\hat{\alpha})} \sim t_{(n-2)}$

$$SE(\hat{\alpha}) = \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right)} = \sqrt{\frac{\hat{u}_i^2}{n-2}\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right)}$$

Q. For the model $Y_i = \alpha + \beta X_i + u_i$, if the scale of Y_i and X_i are changed by w_1 and w_2 respectively, i.e.,

$$Y_i^* = w_1 Y_i \ , \ X_i^* = w_2 X_i \ ,$$

what will be the impact on the test statistics?

Sol^n^:

$$\beta^* = \frac{\sum x_i^* y_i^*}{\sum x_i^{*2}} = \frac{w_1 w_2 \sum x_i y_i}{w_2^2 \sum x_i^2} = \left(\frac{w_1}{w_2}\right)\hat{\beta}$$

$$\alpha^* = \bar{Y}^* - \hat{\beta}^* \bar{X}^* = \cancel{\bar{Y}} - \cancel{\frac{w_2}{w_1}\hat{\beta}\bar{X}}$$

$$= w_1 \bar{Y} - \left(\frac{w_1}{w_2}\right)\hat{\beta}\, w_2 \bar{X} = w_1(\bar{Y} - \hat{\beta}\bar{X}) = w_1 \hat{\alpha}$$

$$\sigma^{*2} = \frac{\sum \hat{u}_i^{*2}}{n-2} = \frac{\sum \left\{ w_1 Y_i - w_1\hat{\alpha} - \left(\frac{w_1}{w_2}\right)\hat{\beta} w_2 X_i \right\}^2}{n-2}$$

$$= \frac{w_1^2 \sum \hat{u}_i^2}{n-2} = w_1^2 \hat{\sigma}^2 \qquad \sigma^{*2} = w_1^2 \hat{\sigma}^2$$

$$\text{var}(\beta^*) = \frac{w_1^2 \hat{\sigma}^2}{w_2^2 \sum x_i^2} = \left(\frac{w_1}{w_2}\right)^2 \text{var}(\hat{\beta}) \qquad SE(\beta^*) = \frac{w_1}{w_2} SE(\hat{\beta})$$

$$\text{var}(\alpha^*) = \sigma^{*2}\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right) = w_1^2 \hat{\sigma}^2 \left(\frac{1}{n} + \frac{\cancel{w_2^2}\bar{X}^2}{\cancel{w_2^2}\sum x_i^2}\right) = w_1^2 \text{var}(\hat{\alpha})$$

$$SE(\alpha^*) = w_1 SE(\hat{\alpha}) \qquad \boxed{SE(\beta^*) = \frac{w_1}{w_2} SE(\hat{\beta})}$$

$$\boxed{SE(\alpha^*) = w_1 SE(\hat{\alpha})}$$

$$t_{\beta^*} = \frac{\beta^*}{SE(\beta^*)} = \frac{\left(\frac{w_1}{w_2}\right)\hat{\beta}}{\left(\frac{w_1}{w_2}\right)SE(\hat{\beta})} = \frac{\hat{\beta}}{SE(\hat{\beta})} = t_{\hat{\beta}}$$

$$t_{\alpha^*} = \frac{\alpha^*}{SE(\alpha^*)} = \frac{w_1\hat{\alpha}}{w_1 SE(\hat{\alpha})} = \frac{\hat{\alpha}}{SE(\hat{\alpha})} = t_{\hat{\alpha}}$$

$$F^* = \frac{ESS^*/df_1}{RSS^*/df_2} = \frac{(\beta^{*2}\sum x_i^{*2})/df_1}{(\sum \hat{u}_i^{*2})/df_2} = \frac{\left\{\left(\frac{w_1}{w_2}\right)^2 w_2^2 \sum x_i^2\right\}/df_1}{(w_1^2 \sum \hat{u}_i^2)/df_2}$$

$$= F$$

Thus, changes in scale of measurement of the variables do not affect the test statistics.

★ Statistical Distribution of the test statistics

$$t_{\hat{\beta}} = \frac{\hat{\beta} - E(\hat{\beta})}{\sqrt{var(\hat{\beta})}} = \frac{\hat{\beta}-\beta}{\sqrt{\frac{\sigma^2}{\sum x_i^2}}} = \frac{(\hat{\beta}-\beta)\sqrt{\sum x_i^2}}{\sqrt{\frac{\sum \hat{u}_i^2}{n-2}}}$$

$$= \frac{\left(\frac{\hat{\beta}-\beta}{\sigma}\right)\sqrt{\sum x_i^2}}{\sqrt{\frac{\sum\left(\frac{\hat{u}_i^2}{\sigma^2}\right)}{n-2}}} = \frac{\left(\frac{\hat{\beta}-\beta}{\sigma}\right)\sqrt{\sum x_i^2}}{\sqrt{\frac{\sum\left(\frac{\hat{u}_i - \bar{\hat{u}}}{\sigma}\right)^2}{n-2}}}$$

$$t_{\hat{\beta}} = \frac{Z_1}{\sqrt{\frac{Z_2}{k_2}}} \sim t_{k_2} \qquad Z_1 \sim N(0,1)$$

$$Z_2 \sim \chi^2_{k_2},\ k_2 = n-2$$

$$t_{\hat{\alpha}} = \frac{\hat{\alpha} - E(\hat{\alpha})}{\sqrt{var(\hat{\alpha})}} = \frac{\hat{\alpha}-\alpha}{\sqrt{\sigma^2\left(\frac{1}{n}+\frac{\bar{X}^2}{\sum x_i^2}\right)}} = \frac{(\hat{\alpha}-\alpha)\Big/\sqrt{\left(\frac{1}{n}+\frac{\bar{X}^2}{\sum x_i^2}\right)}}{\sqrt{\sigma^2}}$$

$$= \frac{(\hat{\alpha}-\alpha)\Big/\sqrt{\left(\frac{1}{n}+\frac{\bar{X}^2}{\sum x_i^2}\right)}}{\sqrt{\frac{\sum \hat{u}_i^2}{n-2}}}$$

$$= \frac{\dfrac{(\hat{\alpha}-\alpha)}{\sigma} \Big/ \sqrt{\left(\dfrac{1}{n}+\dfrac{\bar{X}^2}{\sum x_i^2}\right)}}{\sqrt{\dfrac{\sum\left(\dfrac{\hat{u}_i^2}{\sigma^2}\right)}{n-2}}} = \frac{Z_1}{\sqrt{\dfrac{Z_2}{k_2}}} \sim t_{k_2}$$

$Z_1 \sim N(0,1)$

$Z_2 \sim \chi^2_{k_2}$

$k_2 = n-2$.

$$F = \frac{ESS/df_1}{RSS/df_2} = \frac{\dfrac{\hat{\beta}^2}{\sigma^2}\left(\sum x_i^2\right)/df_1}{\sum\left(\dfrac{\hat{u}_i^2}{\sigma^2}\right)/df_2} = \frac{\left(\dfrac{\hat{\beta}-\beta}{\sigma}\right)^2 \sum x_i^2/1}{\sum\left(\dfrac{\hat{u}_i^2}{\sigma^2}\right)/n-2} = \frac{Z_1/k_1}{Z_2/k_2}$$

$$= F_{k_1,k_2}.$$

$Z_1 \sim \chi^2_{k_1},\ k_1 = 1$

$Z_2 \sim \chi^2_{k_2},\ k_2 = n-2$

- **Relationship between the R^2 and the F statistic in the Regression Model.**

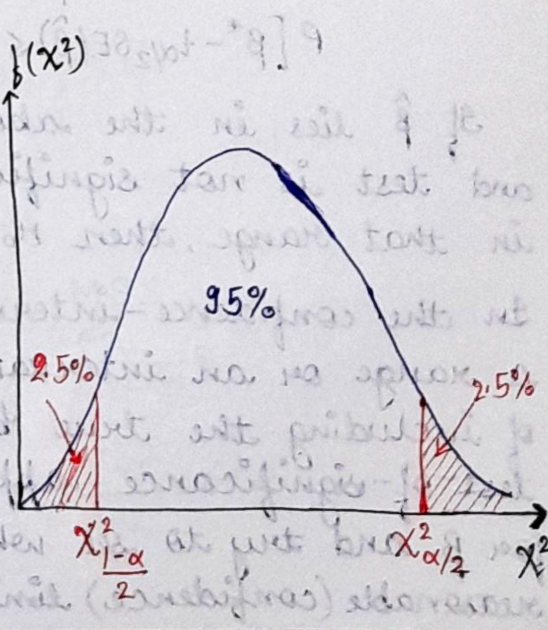
$$F = \frac{ESS/df_1}{RSS/df_2} = \frac{ESS/df_1}{(TSS-ESS)/df_2} = \frac{\left(\dfrac{ESS}{TSS}\right)/df_1}{\left(\dfrac{TSS-ESS}{TSS}\right)/df_2} = \left(\frac{R^2}{1-R^2}\right)\left(\frac{n-k}{k-1}\right)$$

$$\boxed{F = \frac{R^2}{1-R^2}\left(\frac{n-k}{k-1}\right)}$$

$$\chi^2 = (n-2)\frac{\hat{\sigma}^2}{\sigma^2}$$

$$\Pr\left(\chi^2_{1-\alpha/2} \leq \chi^2 \leq \chi^2_{\alpha/2}\right) = 1-\alpha$$

$$P\left[(n-2)\frac{\hat{\sigma}^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq (n-2)\frac{\hat{\sigma}^2}{\chi^2_{\frac{1-\alpha}{2}}}\right] = 1-\alpha$$



★ Two-sided hypothesis

A two-sided alternative hypothesis reflects the fact that we do not have a strong apriori or theoretical expectation about the direction in which the alternative hypothesis should move from the null hypothesis.

In statistics, when we reject the null hypothesis, we ~~say~~ say that our finding is statistically ~~sign~~ significant. On the other hand, when we do not reject the null hypothesis, we say that our finding is not ~~statiscal~~ statistically significant.

★ One-sided or One-Tail Test

Sometimes we have a strong apriori or theoretical expectation (or expectations based on some previous empirical work) that the alternative hypothesis is one-sided or unidirectional rather than two-sided.

- The test-of-significance approach

$$t = \frac{\hat{\beta} - \beta}{SE(\hat{\beta})}$$

$$P\left[-t_{\alpha/2} \leq \frac{\hat{\beta} - \beta}{SE(\hat{\beta})} \leq t_{\alpha/2}\right] = 1 - \alpha$$

Let $H_0 : \beta = \beta^*$

$H_1 : \beta \neq \beta^*$

$$P\left[\beta^* - t_{\alpha/2} SE(\hat{\beta}) \leq \hat{\beta} \leq \beta^* + t_{\alpha/2} SE(\hat{\beta})\right] = 1 - \alpha$$

If $\hat{\beta}$ lies in the above range then H_0 is not rejected and test is not significant, whereas if it does not lie in that range, then H_0 is rejected and test is significant.

In the confidence-interval approach, we try to establish a range or an interval that has a certain probability of including the true but unknown β, whereas in the test-of-significance approach we hypothesize some value for β and try to see whether the computed $\hat{\beta}$ lies within reasonable (confidence) limits around the hypothesized value

Type of Hypothesis	H_0: The Null Hypothesis	H_1: The Alternative Hypothesis	Decision Rule: Reject H_0 if
Two-Tail	$\beta = \beta^*$	$\beta \neq \beta^*$	$\lvert t \rvert > t_{\alpha/2, df}$
Right-Tail	$\beta \leq \beta^*$	$\beta > \beta^*$	$t > t_{\alpha, df}$
Left-Tail	$\beta \geq \beta^*$	$\beta < \beta^*$	$t < -t_{\alpha, df}$

$$\chi^2 = (n-2)\frac{\hat{\sigma}^2}{\sigma^2}$$

H_0: The Null Hypothesis	H_1: The Alternative Hypothesis	Critical Region: Reject H_0 if
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\frac{df(\hat{\sigma}^2)}{\sigma_0^2} > \chi^2_{\alpha, df}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\frac{df(\hat{\sigma}^2)}{\sigma_0^2} < \chi^2_{(1-\alpha), df}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\frac{df(\hat{\sigma}^2)}{\sigma_0^2} > \chi^2_{\alpha/2, df}$ or $< \chi^2_{(1-\alpha)/2, df}$

★ p-value

Exact probability of committing a Type I error. The lowest significance level at which a null hypothesis can be rejected.

• Analysis of Variance

$$\boxed{\underbrace{\sum y_i^2}_{TSS} = \underbrace{\sum \hat{y}_i^2}_{ESS} + \underbrace{\sum \hat{u}_i^2}_{RSS} = \hat{\beta}^2 \sum x_i^2 + \sum \hat{u}_i^2}$$

Source of variation	SS	df	MSS
Due to regression (ESS)	$\sum \hat{y}_i^2 = \hat{\beta}^2 \sum x_i^2$	$k-1$	$\hat{\beta}^2 \sum x_i^2$
Due to residuals (RSS)	$\sum \hat{u}_i^2$	$n-2$	$\frac{\sum u_i^2}{n-2} = \hat{\sigma}^2$
TSS	$\sum y_i^2$	$n-1$	

$$\cancel{F} = \frac{ESS/\cancel{k-1}}{RSS/\cancel{n-k}}$$

$$F = \frac{ESS/k-1}{RSS/n-k} = \frac{\hat{\beta}^2 \sum x_i^2}{\sum \hat{u}_i^2/(n-2)} = \frac{\hat{\beta}^2 \sum x_i^2}{\hat{\sigma}^2}$$

$$E(\hat{\beta}^2 \sum x_i^2) = \sigma^2 + \beta^2 \sum x_i^2$$

$$E\left(\frac{\sum u_i^2}{n-2}\right) = E(\hat{\sigma}^2) = \sigma^2$$

If $\beta = 0$, then both ~~estimates~~ provide us with identical estimates of true σ^2. In this ~~case~~ situation, the explanatory variable X has no linear influence on Y whatsoever and the entire variation in Y is explained by the random disturbances u_i.

$\therefore$ The ~~ratio~~ F ratio provides a test of the null hypothesis $H_0: \beta = 0$

- Mean Prediction

Prediction of $E(Y|X_0)$

$$\hat{Y}_0 = \hat{\alpha} + \hat{\beta} X_0 \text{ (BLUE)}$$

$$var(\hat{Y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right]$$

$$E(Y_0|X_0) = \cancel{\beta_1 + \beta}\ \alpha + \beta X_0$$

$$\hat{Y}_0 = \hat{\alpha} + \hat{\beta} X_0$$

$$E(\hat{Y}_0) = E(\hat{\alpha}) + X_0 E(\hat{\beta}) = \alpha + \beta X_0$$

$$\therefore E(\hat{Y}_0) = E(Y_0|X_0) = \alpha + \beta X_0$$

$$var(\hat{Y}_0) = var(\hat{\alpha}) + var(\hat{\beta}) X_0^2 + 2X_0 cov(\hat{\alpha}, \hat{\beta})$$

$$= \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right] + X_0^2 \frac{\sigma^2}{\sum x_i^2} \cancel{+ 2X_0 (-\bar{X})} - 2X_0 \bar{X} \frac{\sigma^2}{\sum x_i^2}$$

$$= \frac{\sigma^2}{n} + \frac{\sigma^2}{\sum x_i^2}\left[X_0^2 + \bar{X}^2 - 2X_0\bar{X}\right]$$

$$\text{var} = \frac{\sigma^2}{n} + \frac{\sigma^2}{\sum x_i^2}(X_0 - \bar{X})^2$$

$$\boxed{\text{var}(\hat{Y}_0) = \sigma^2\left[\frac{1}{n} + \frac{(X_0-\bar{X})^2}{\sum x_i^2}\right]}$$

$$t = \frac{\hat{Y}_0 - (\alpha + \beta X_0)}{SE(\hat{Y}_0)} \sim t_{(n-2)}$$

$$P\left[\hat{\alpha} + \hat{\beta}X_0 - t_{\alpha/2}SE(\hat{Y}_0) \leqslant \alpha + \beta X_0 \leqslant \hat{\alpha} + \hat{\beta}X_0 + t_{\alpha/2}SE(\hat{Y}_0)\right] = 1-\alpha.$$

- Individual Prediction

Predicting an individual value Y_0 given X_0.

$$Y_0 = \alpha + \beta X_0 + u_0$$

$$\hat{Y}_0 = \hat{\alpha} + \hat{\beta}X_0$$

Prediction error: $Y_0 - \hat{Y}_0$

$$Y_0 - \hat{Y}_0 = \alpha + \beta X_0 + u_0 - (\hat{\alpha} + \hat{\beta}X_0)$$

$$= (\alpha - \hat{\alpha}) + X_0(\beta - \hat{\beta}) + u_0$$

$$E(Y_0 - \hat{Y}_0) = E(\alpha - \hat{\alpha}) + X_0 E(\beta - \hat{\beta}) + E(u_0)$$

$$= 0$$

$$\cancel{(Y_0 - \hat{Y}_0)^2 = (\alpha - \hat{\alpha})^2 + X_0^2(\beta - \hat{\beta})^2 + \hat{u}_0^2}$$

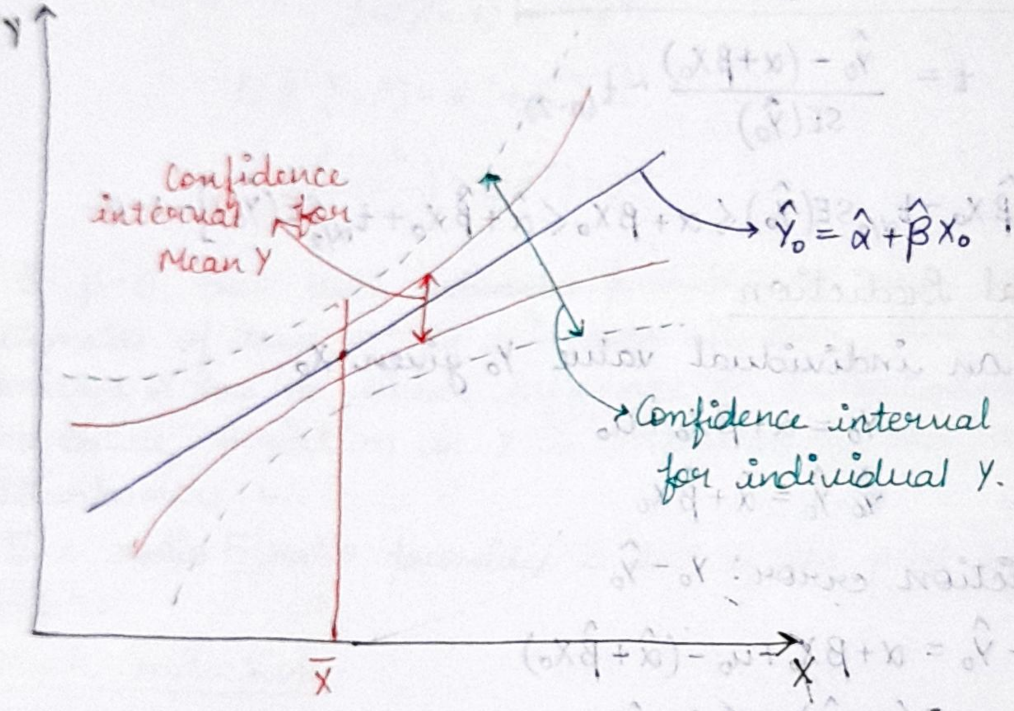
$$\text{var}(Y_0 - \hat{Y}_0) = \text{var}(\hat{\alpha}) + X_0^2\,\text{var}(\hat{\beta}) + 2X_0\,\text{cov}(\hat{\alpha}, \hat{\beta}) + \text{var}(u_0)$$

$$= \sigma^2\left[\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right] + X_0^2\frac{\sigma^2}{\sum x_i^2} + 2X_0\left[-\bar{X}\frac{\sigma^2}{\sum x_i^2}\right] + \sigma^2$$

$$\boxed{\text{var}(Y_0 - \hat{Y}_0) = \sigma^2\left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right]}$$

$$t = \frac{Y_0 - \hat{Y}_0}{SE(Y_0 - \hat{Y}_0)}$$

By computing confidence intervals for the above two tests, ~~we can~~ for all X values, a band will be formed around the regression line.



The width of these bands is smallest when $X_0 = \bar{X}$. However, the width widens sharply as X_0 moves away from $\bar{X}$. This change would suggest that the predictive ability of the historical sample regression line falls markedly as X_0 departs progressively from $\bar{X}$.

- Jarque - Bera (JB) Test of Normality

$$\boxed{JB = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right]} \sim \chi^2_2$$

$n \rightarrow$ Sample size

$S \rightarrow$ Skewness coefficient

$K \rightarrow$ Kurtosis coefficient

H_0: Residuals are normally distributed

★ REGRESSION THROUGH THE ORIGIN :

$$PRF: Y_i = \beta X_i + u_i$$

Example :

1. $(ER_i - r_f) = \beta_i (ER_m - r_f)$ (CAPM)

$ER_i \rightarrow$ Expected rate of return on security i

$ER_m \rightarrow$ expected rate of return on the market portfolio

$r_f \rightarrow$ Risk-free rate.

$\beta_i \rightarrow$ Beta coefficient, a measure of systematic risk, i.e. the risk that cannot be eliminated through diversification. Extent to which the ith security's rate of return moves with the market.

$R_i - R_f = \alpha + \beta(R_m - r_f) + u_i$ (Market Model)

2. Milton Friedman's permanent income hypothesis
Permanent consumption $\propto$ permanent income.

3. Cost-Analysis Theory
Variable cost of production $\propto$ Output.

4. Rate of change of price $\propto$ Rate of change of money supply.
(Inflation)

Model:

$$SRF: Y_i = \hat{\beta} X_i + \hat{u}_i$$

$$\boxed{\hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2}}$$

$$Min: \sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta} X_i)^2$$

$$\frac{d\sum \hat{u}_i^2}{d\hat{\beta}} = 2\sum(Y_i - \hat{\beta} X_i)(-X_i) \rightarrow \hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2}$$

$$\hat{\beta} = \frac{\sum X_i(\beta X_i + u_i)}{\sum X_i^2} = \beta + \frac{\sum X_i u_i}{\sum X_i^2}$$

$$E(\hat{\beta}) = \beta \ , \ \therefore \ E(\hat{\beta} - \beta)^2 = E\left(\frac{\sum X_i u_i}{\sum X_i^2}\right)^2$$

$$E(\hat{\beta} - \beta)^2 = E\Big\{ \cancel{X_i^2 \sum} \ \frac{\sum X_i^2 E(u_i^2)}{(\sum X_i^2)^2} = \frac{\sigma^2 \cancel{\sum X_i^2}}{(\sum X_i^2)^{\cancel{2}}}$$

$$E(\hat{\beta} - \beta)^2 = \frac{\sigma^2}{\sum X_i^2} \qquad \text{Also, } \boxed{\sum \hat{u}_i X_i = 0}$$

$$\boxed{var(\hat{\beta}) = \frac{\sigma^2}{\sum X_i^2}}$$

If we impose, $\sum \hat{u}_i = 0$

$$\sum Y_i = \hat{\beta} \sum X_i + \sum \hat{u}_i$$

$$= \hat{\beta} \sum X_i$$

$$\hat{\beta} = \frac{\sum Y_i}{\sum X_i} = \frac{\bar{Y}}{\bar{X}} \ \{\text{Biased estimator of } \beta\}$$

$\therefore$ In this case, both $\sum \hat{u}_i X_i = 0$ and $\sum \hat{u}_i = 0$ cannot be satisfied.

$$Y_i = \hat{Y}_i + \hat{u}_i$$

$$\bar{Y} = \bar{\hat{Y}} + \bar{\hat{u}} \quad \{\bar{\hat{u}} \neq 0\}$$

$$\therefore \ \boxed{\bar{Y} \neq \bar{\hat{Y}}}$$

Now, $k = 1$

$$\therefore \ \boxed{\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-1}}$$

$$\boxed{\text{raw } r^2 = \frac{(\sum x_i y_i)^2}{\sum x_i^2 \sum y_i^2}}$$

$$r^2 = 1 - \cancel{ESS}\ \frac{RSS}{TSS} = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2}$$

$$RSS = \sum \hat{u}_i^2 = \sum y_i^2 - \beta^2 \sum x_i^2 \leq \sum y_i^2 \text{ if } \beta \neq 0$$

{for the intercept model}

But in this case,

$$RSS = \sum \hat{u}_i^2 = \sum Y_i^2 - \hat{\beta}^2 \sum X_i^2$$

$$TSS = \sum y_i^2 = \sum Y_i^2 - N\bar{Y}^2$$

Now, it is not necessary that $RSS \leq TSS$.

$\therefore$ r^2 could also come out to be -ve.

$\therefore$ r^2 is defined differently as:

$$\text{raw } r^2 = \frac{(\sum X_i Y_i)^2}{\sum X_i^2 \sum Y_i^2}$$

$\therefore$ Model with intercept has the following advantages:

1) If the intercept term is included in the model, but it turns out to be statistically not significant, for all practical purposes we have a regression through origin.

2) ~~If~~ If in fact, there is an intercept in the model but we insist on fitting a regression through the origin, we would be committing a specification error.

- **Regression on Standardized Variables**

$$Y_i^* = \frac{Y_i - \bar{Y}}{S_Y} \qquad X_i^* = \frac{X_i - \bar{X}}{S_X}$$

$$Y_i^* = \alpha^* + \beta^* X_i^* + u_i^*$$
$$= \beta^* X_i^* + u_i^*$$

If the (standardized) X^* increases by one standard deviation (SE), on average, the ~~std~~ (std.) Y^* increases $\hat{\beta}^*$ times SE units.

$$\alpha^* = \bar{Y}^* - \hat{\beta}^* \bar{X}^* = 0$$

$$\boxed{\hat{\beta}^* = \hat{\beta}\frac{\sigma_X}{\sigma_Y}}$$

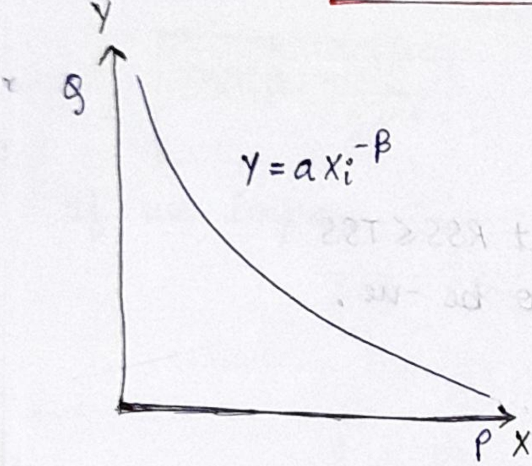
★ Log-Linear Model

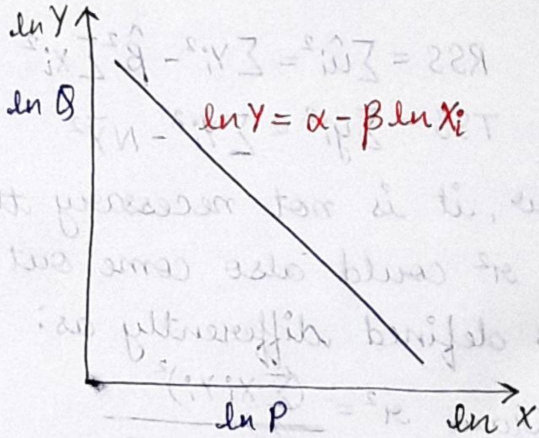
Exponential Regression Model

$$Y_i = \cancel{\beta} a X_i^{\beta} e^{u_i}$$

$$\ln Y_i = \ln a + \beta \ln X_i + u_i$$

$$\boxed{\ln Y_i = \alpha + \beta \ln X_i + u_i}$$





In this model, β measures the elasticity of Y w.r.t. X, i.e. the percentage change in Y for a given (small) percentage change in X.

Thus, if Y represents the quantity of a commodity demanded and X its unit price, β measures the price elasticity of demand.

The model assumes that the elasticity coefficient b/w Y and X, β remains constant throughout, hence the alternative name constant elasticity model.

$\hat{\alpha}$ and $\hat{\beta}$ are unbiased estimators for a and β.

but $a = \text{antilog}(\hat{\alpha})$ is biased.

$$\text{Elasticity} = \left(\frac{dY}{dX}\right)\left(\frac{X}{Y}\right)$$

$$\ln Y_i = \alpha + \beta \ln X_i + u_i$$

diff. w.r.t. X_i

$$\frac{1}{Y_i}\frac{\partial Y_i}{\partial X_i} = \beta \frac{1}{X_i}$$

$$\beta = \left(\frac{\partial Y}{\partial X}\right)\left(\frac{X}{Y}\right)$$

$$\boxed{\ln X_t - \ln X_{t-1} = \frac{(X_t - X_{t-1})}{X_{t-1}}}$$

* constant elasticity model will give a constant total revenue change for a given percentage change in price regardless of the absolute level of price.

* Log-Lin Model (Semilog Models)

$$Y_t = Y_0(1+r)^t$$

$Y_t \rightarrow$ real expenditure on services at time t.

$Y_0 \rightarrow$ initial value of the expenditure on services.

$$\ln Y_t = \ln Y_0 + t \ln(1+r)$$

$$\alpha = \ln Y_0$$
$$\beta = \ln(1+r)$$

$$\ln Y_t = \alpha + \beta t + u_t$$

The slope coefficient measures the constant proportional or relative change in Y for a given absolute change in the value of the regressor.

$$\frac{1}{Y_t}\frac{\partial Y_t}{\partial t} = \beta \Rightarrow \beta = \frac{\partial Y_t / Y_t}{\partial t}$$

$$\beta = \frac{\text{Relative change in regressand}}{\text{Absolute change in regressor.}}$$

Absolute change: $X_t - X_{t-1}$

Relative or proportional change: $\dfrac{X_t - X_{t-1}}{X_{t-1}} = \dfrac{X_t}{X_{t-1}} - 1$

Percentage change $= \left(\dfrac{X_t - X_{t-1}}{X_{t-1}}\right) 100$.

★ **Linear Trend Model**

$$Y_t = \alpha + \beta t + u_t$$

$\beta \rightarrow$ Absolute change in Y_t wrt change in t.

★ **Lin-Log Model**

$$Y_i = \alpha + \beta \ln X_i + u_i$$

$\beta \rightarrow$ Absolute change in Y wrt percentage change in X.

$$\frac{\partial Y_i}{\partial X_i} = \frac{\beta}{X_i} \rightarrow \beta = \frac{\partial Y_i}{\partial X_i / X_i}$$

$$\beta = \frac{\text{Change in } Y}{\text{Relative change in } X}$$

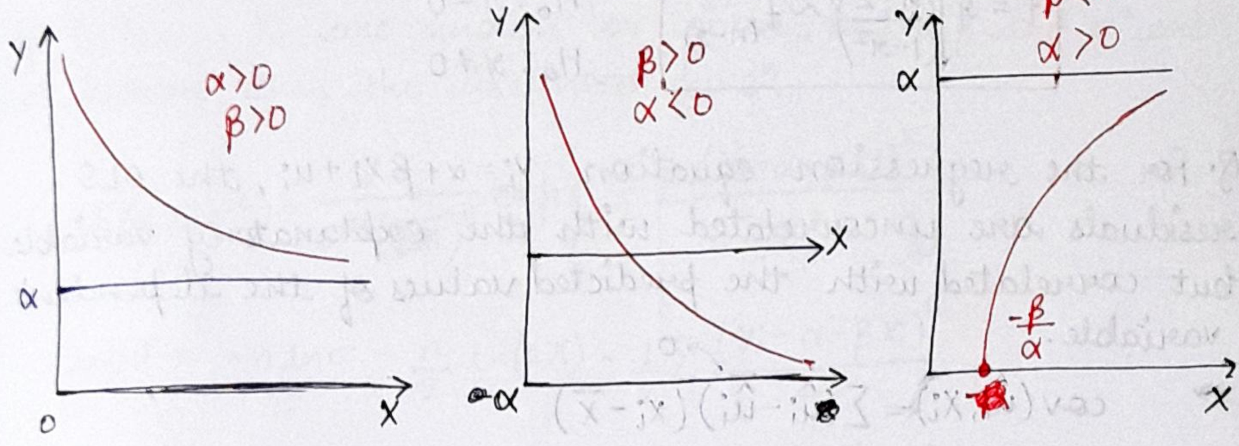
Example:

Engel Expenditure Model:

The total expenditure that is devoted to food tends to increase in arithmetic progression as total expenditure increases in geometric progression.

* Reciprocal Models

$$Y_i = \alpha + \beta\left(\frac{1}{X_i}\right) + u_i$$

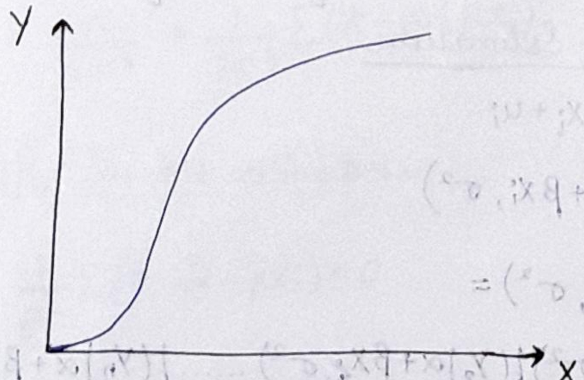


$$\frac{\partial Y_i}{\partial X_i} = \frac{-\beta}{X_i^2} \Rightarrow \beta > 0 \text{ if } \frac{\partial Y_i}{\partial X_i} < 0$$

$$\beta < 0 \text{ if } \frac{\partial Y_i}{\partial X_i} > 0$$

* Log Hyperbola Model

$$\ln Y_i = \alpha - \beta\left(\frac{1}{X_i}\right) + u_i$$



$$\frac{1}{Y_i}\frac{\partial Y_i}{\partial X_i} = \frac{\beta}{X_i^2}$$

$$\beta = \frac{\partial Y_i}{\partial X_i}\left(\frac{X_i^2}{Y_i}\right)$$

★ Statistical Test of the Pair-wise Correlation Coefficient

X, Y → pair-wise correlations coefficient, r

$$\boxed{t = r\sqrt{\left(\frac{n-2}{1-r^2}\right)} \sim t_{(n-2)}}$$

$H_0 : r = 0$

$H_a : r \neq 0$

Q. for the regression equation $Y_i = \alpha + \beta X_i + u_i$, the OLS residuals are uncorrelated with the explanatory variable but correlated with the predicted values of the dependent variable.

$$\text{cov}(\hat{u}_i, X_i) = \sum(\hat{u}_i - \overline{\hat{u}_i}^{\,\to 0})(X_i - \bar{X})$$

$$= \underbrace{\sum \hat{u}_i X_i}_{0} - \bar{X}\underbrace{\sum \hat{u}_i}_{0} = 0.$$

$$\text{cov}(\hat{u}_i, \hat{Y}_i) = \sum(\hat{u}_i - \underbrace{\bar{\hat{u}}_i}_{0})(\hat{Y}_i - \bar{\hat{Y}}_i) = \sum \hat{u}_i(\cancel{\hat{\alpha}} + \hat{\beta} X_i - \cancel{\hat{\alpha}} - \hat{\beta}\bar{X})$$

$$= \sum \hat{u}_i \hat{\beta} X_i - \hat{\beta} \sum \hat{u}_i \bar{X}$$

$$= \hat{\beta}\underbrace{\sum \hat{u}_i X_i}_{0} - \hat{\beta}\bar{X}\underbrace{\sum \hat{u}_i}_{0} = 0.$$

★ Maximum-Likelihood Estimation

$$Y_i = \alpha + \beta X_i + u_i$$

$$Y_i \sim N(\alpha + \beta X_i, \sigma^2)$$

$$f(Y_1, Y_2, Y_3, \ldots\ldots, Y_n | \alpha + \beta X_i, \sigma^2) =$$

$$f(Y_1 | \alpha + \beta X_1, \sigma^2) f(Y_2 | \alpha + \beta X_2, \sigma^2) \ldots\ldots f(Y_n | \alpha + \beta X_n, \sigma^2)$$

{as $Y_1, Y_2, \ldots, Y_n$ are independent}

$$f(Y_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-1}{2}\frac{(Y_i - \alpha - \beta X_i)^2}{\sigma^2}\right\}$$

$$f(Y_1, Y_2, \ldots, Y_n \mid \alpha+\beta X_i, \sigma^2) = \frac{1}{\sigma^n (\sqrt{2\pi})^n} \exp\left\{ \frac{-1}{2} \sum \frac{(Y_i - \alpha - \beta X_i)^2}{\sigma^2} \right\}.$$

If $Y_1, Y_2, \ldots, Y_n$ are known or given, but α, β and σ^2 are not known, then the likelihood func^n,

$$LF(\alpha, \beta, \sigma^2) = \frac{1}{\sigma^n (\sqrt{2\pi})^n} \exp\left\{ \frac{-1}{2} \sum \frac{(Y_i - \alpha - \beta X_i)^2}{\sigma^2} \right\}$$

$$\ln LF = -n \ln \sigma - \frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum \frac{(Y_i - \alpha - \beta X_i)^2}{\sigma^2}$$

$$= -\frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln(2\pi) - \frac{1}{2} \frac{\sum (Y_i - \alpha - \beta X_i)^2}{\sigma^2}$$

$$\frac{\partial \ln LF}{\partial \alpha} = \frac{-1}{\sigma^2} \sum (Y_i - \alpha - \beta X_i)(-1) = \frac{1}{\sigma^2} \sum (Y_i - \alpha - \beta X_i)$$

$$\frac{\partial \ln LF}{\partial \beta} = \frac{1}{\sigma^2} \sum X_i (Y_i - \alpha - \beta X_i)$$

$$\frac{\partial \ln LF}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum (Y_i - \alpha - \beta X_i)^2$$

Let $\tilde{\alpha}, \tilde{\beta}, \tilde{\sigma}^2 \rightarrow$ ML estimators

$$\Rightarrow \frac{1}{\tilde{\sigma}^2} \sum (Y_i - \tilde{\alpha} - \tilde{\beta} X_i) = 0$$

$$\Rightarrow \frac{1}{\tilde{\sigma}^2} \sum X_i (Y_i - \tilde{\alpha} - \tilde{\beta} X_i) = 0$$

$$\Rightarrow -\frac{n}{2\tilde{\sigma}^2} + \frac{1}{2\tilde{\sigma}^4} \sum (Y_i - \tilde{\alpha} - \tilde{\beta} X_i)^2 = 0$$

$$\sum Y_i = n\tilde{\alpha} + \tilde{\beta} \sum X_i$$

$$\sum Y_i X_i = \tilde{\alpha} \sum X_i + \tilde{\beta} \sum X_i^2$$

$\therefore$ We get same normal equations as OLS estimators,

$\therefore$ same values of α & β.

$$\tilde{\sigma}^2 = \frac{1}{n}\sum(Y_i - \tilde{\alpha} - \tilde{\beta}X_i)^2$$

$$= \frac{1}{n}\sum(Y_i - \hat{\alpha} - \hat{\beta}X_i)^2 = \frac{1}{n}\sum\hat{u}_i^2$$

$$\boxed{\tilde{\sigma}^2 = \frac{1}{n}\sum\hat{u}_i^2}$$

$$E(\tilde{\sigma}^2) = \frac{1}{n}E\left(\sum\hat{u}_i^2\right) = \frac{n-2}{n}\sigma^2 = \sigma^2 - \frac{2}{n}\sigma^2$$

$\therefore$ It underestimates true σ^2 by $\frac{2}{n}\sigma^2$.

$\therefore$ $\tilde{\sigma}^2$ is biased downward.

If $n \to \infty$ $E(\tilde{\sigma}^2) \to \sigma^2$

$\therefore$ $\tilde{\sigma}^2$ is asymptotically unbiased.

Bivariate Normal Probability Density (Function)

$$f(Y_1, Y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)} * \right.$$

$$\left.\left[\left(\frac{Y_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{(Y_1-\mu_1)(Y_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{Y_2-\mu_2}{\sigma_2}\right)^2\right]\right\}$$