1 Simple Regression:

Model: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ where $i = 1, 2, ...n$ and $\varepsilon_i \in \mathcal{N}(0, \sigma)$ independent.

$\overline{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$, $\overline{y} = \dfrac{\sum_{i=1}^{n} y_i}{n}$

$S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2$, $S_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})y_i$ and $S_{yy} = \sum_{i=1}^{n}(y_i - \overline{y})^2$

Point estimators of slope and intercept: $\widehat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}}$, $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1\overline{x}$, (unbiased, jointly Normal).

Coefficient of Determination: $\dfrac{S_{xy}^2}{S_{xx}S_{yy}}$ (dimensionless).

Residuals: $e_i = y_i - \widehat{y}_i$ where $\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i$

Sum of squares of the residuals: $SS_E = S_{yy} - \dfrac{S_{xy}^2}{S_{xx}}$

Mean square residual: $MS_E = \dfrac{SS_E}{n-2}$

Point estimator of σ^2: $\widehat{\sigma}^2 = MS_E \Rightarrow \widehat{\sigma} = \sqrt{MS_E}$, where $\dfrac{MS_E}{\sigma^2} \in \chi^2_{n-2}$, independent of $\widehat{\beta}_1$ and $\widehat{\beta}_0$.

Standard deviation of $\widehat{\beta}_1$and $\widehat{\beta}_0$: $\dfrac{\sigma}{\sqrt{S_{xx}}}$ and $\sigma\sqrt{\dfrac{1}{n} + \dfrac{\overline{x}^2}{S_{xx}}}$ respectively.

Their covariance: $-\overline{x}\dfrac{\sigma^2}{S_{xx}}$.

Confidence interval for slope: $\widehat{\beta}_1 \pm \mathsf{t}_{\frac{\alpha}{2},n-2}\sqrt{\dfrac{MS_e}{S_{xx}}}$

Confidence interval for intercept: $\widehat{\beta}_0 \pm \mathsf{t}_{\frac{\alpha}{2},n-2}\sqrt{MS_e\left(\dfrac{1}{n} + \dfrac{\overline{x}^2}{S_{xx}}\right)}$

Confidence ellipse for both: $\dfrac{(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T\mathbb{X}^T\mathbb{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{MS_e} = F_{2,n-2}(1-\alpha)$

Confidence interval for σ^2: $\left(\dfrac{SS_E}{\chi^2_{1-\frac{\alpha}{2},n-2}}, \dfrac{SS_E}{\chi^2_{\frac{\alpha}{2},n-2}}\right)$ (for σ, take square root).

Confidence interval for $\mathbb{E}(y|x_0) \equiv \beta_0 + \beta_1 x_0$:

$\widehat{\beta}_0 + \widehat{\beta}_1 x_0 \pm \mathsf{t}_{\frac{\alpha}{2},n-2}\sqrt{MS_e\left(\dfrac{1}{n} + \dfrac{(x_0 - \overline{x})^2}{S_{xx}}\right)}$

Prediction interval for a new y yet to be taken at x_0, i.e. for $\beta_0 + \beta_1 x_0 + \varepsilon_0$:

$\widehat{\beta}_0 + \widehat{\beta}_1 x_0 \pm \mathsf{t}_{\frac{\alpha}{2},n-2}\sqrt{MS_e\left(1 + \dfrac{1}{n} + \dfrac{(x_0 - \overline{x})^2}{S_{xx}}\right)}$

Testing H_0: $\beta_1 = 0$ against H_A: $\beta_1 \neq 0$, use $T = \dfrac{\widehat{\beta}_1}{\sqrt{\dfrac{MS_e}{S_{xx}}}}$ as test statistic,

check against critical values of $\pm t_{\frac{\alpha}{2},n-2}$ (α is now level of significance).
When H_A: $\beta_1 > 0$, check against $+t_{\alpha,n-2}$ ($\beta_1 < 0$, $-t_{\alpha,n-2}$).
Testing H_0: $\beta_0 = 0$ .... , use $T = \dfrac{\widehat{\beta}_0}{\sqrt{MS_e\left(\dfrac{1}{n} + \dfrac{\overline{x}^2}{S_{xx}}\right)}}$.
To test normality, display **sorted** residuals against $F^{-1}(\frac{i}{n+1})$,
$i = 1, ,2, \ldots\ n$, where F^{-1} is the $\mathcal{N}(0,1)$ distribution-function inverse;
should get a straight-line pattern.

– –

Lack-of-fit Test (several y observations at each x):

Group means: $\overline{y}_i \equiv \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$ and overall (grand) mean: $\overline{\overline{y}} = \frac{\sum_{i=1}^{m} n_i\,\overline{y}_i}{\sum_{i=1}^{m} n_i}$
$\overline{x} \equiv \frac{\sum_{i=1}^{m} n_i\,x_i}{\sum_{i=1}^{m} n_i}$, $S_{xx} \equiv \sum_{i=1}^{m} n_i(x_i - \overline{x})^2$ and $S_{xy} = \sum_{i=1}^{m} n_i(x_i - \overline{x})\,\overline{y}_i$
After these modifications, the rest of the formulas (for $\widehat{\beta}_1$and $\widehat{\beta}_0$, $\widehat{\sigma}$,
CI, PI, tests of significance etc.) remain identical.
To test whether the true model is linear:
Test statistic: $T = \dfrac{SS_{LOF}}{SS_{PE}} \cdot \dfrac{n-m}{m-2}$,
check against the critical value of $F_{m-2,n-m}$,
where $SS_{LOF} = \sum\limits_{i=1}^{m} n_i(\overline{y}_i - \widehat{y}_i)^2$ and $SS_{PE} = \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n_i}(y_{ij} - \overline{y}_i)^2$.
Note that the last two add up to SS_e.

– –

Weighted Regression

With each x_i and y_i, a positive weight w_i is also given.
Model changes only in $\varepsilon_i \in \mathcal{N}(0, \frac{\sigma}{\sqrt{w_i}})$.

$$\overline{x} = \frac{\sum\limits_{i=1}^{n} w_i\,x_i}{\sum\limits_{i=1}^{n} w_i},\ \overline{y} = \frac{\sum\limits_{i=1}^{n} w_i\,y_i}{\sum\limits_{i=1}^{n} w_i}$$

$$S_{xx} = \sum_{i=1}^{n} w_i(x_i - \overline{x})^2,\ S_{xy} = \sum_{i=1}^{n} w_i(x_i - \overline{x})y_i \text{ and } S_{yy} = \sum_{i=1}^{n} w_i(y_i - \overline{y})^2$$

Standard error of $\widehat{\beta}_0$ changes to $\sigma\sqrt{\frac{1}{\sum_{i=1}^{n} w_i} + \dfrac{\overline{x}^2}{S_{xx}}}$

that of $\mathbb{E}[y\,|\,x = x_0]$ to $\sigma\sqrt{\frac{1}{\sum_{i=1}^{n} w_i} + \dfrac{(x_0 - \overline{x})^2}{S_{xx}}}$.

with the corresponding change in the CI and PI construction.
The $\widehat{\sigma}^2$ estimator is still the same $MS_E = \dfrac{S_{yy} - \frac{S_{xy}^2}{S_{xx}}}{n-2}$.

Bivariate Model

x and y are now generated from a bivariate normal distribution having the usual 5 parameters: μ_x, σ_x, μ_y, σ_y and ρ.
We know that the conditional distribution of $y \,|\, x$ is:
$\mathcal{N}(\mu_y + \rho\, \sigma_y \frac{x-\mu_x}{\sigma_x}, \sigma_y \sqrt{1-\rho^2})$.
The 'best' regression line for predicting y based on x is still the old $y = \widehat{\beta}_0 + \widehat{\beta}_1 x$ (same formulas), but the distribution of the regression-coefficient estimators is now much more complicated. Using large n theory:

$$\mathbb{E}\left(\widehat{\beta}_1\right) = \beta_1 \left[1 + \frac{0}{n} + ...\right] \text{ and } \operatorname{Var}\left(\widehat{\beta}_1\right) = \frac{\sigma_y^2\,(1-\rho^2)}{n\,\sigma_x^2} + ...$$

One could thus still construct an approximate CI for β_1.
Similarly, on can show that the ρ estimator, namely $r = \dfrac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$, has

$$\mathbb{E}\,(r) = \rho - \frac{\rho(1-\rho^2)}{2n} + ... \text{ and } \operatorname{Var}(r) = \frac{(1-\rho^2)^2}{n} + ...$$

Arctanh(r) has a lot better 'behaviour' (especially in terms of skewness), its expected value and variance is $\operatorname{arctanh}(\rho) + \dfrac{\rho}{2n} + ...$ and $\dfrac{1}{n-3} + ...$ respectively. This is utilized in constructing CI for ρ.

2 Multivariate Linear Regression:

Multivariate Normal Distribution:

Defined by $\mathbf{X} = \mathbb{A}\mathbf{z} + \boldsymbol{\mu}$, where $\mathbf{z}$ are independent, standardized Normal.
Variance-covariance matrix: $\mathbb{V} = \mathbb{A}\mathbb{A}^{\mathrm{T}}$
(be able to solve for $\mathbb{A}$ - not unique).
Probability density function: $\frac{1}{\sqrt{(2\pi)^k \det(\mathbb{V})}} \cdot \exp\left(-\frac{(\mathbf{x}-\boldsymbol{\mu})^T \mathbb{V}^{-1} (\mathbf{x}-\boldsymbol{\mu})}{2}\right)$
Moment generating function: $\exp\left(\mathbf{t}^T \boldsymbol{\mu} + \frac{\mathbf{t}^T \mathbb{V}\,\mathbf{t}}{2}\right)$
$\mathbb{V}$ matrix estimator: $\widehat{\mathbb{V}}_{ij} = \dfrac{S_{ij}}{n-1}$, where $S_{ij} = \sum_{\ell=1}^{n} (x_{\ell i} - \overline{x}_i) \cdot (x_{\ell j} - \overline{x}_j)$
Hidden extrapolation: Compare $\sum_{i=1}^{k} \sum_{j=1}^{k} (x_i^{new} - \overline{x}_i)(\mathbb{V}^{-1})_{ij} (x_j^{new} - \overline{x}_j)$ with the largest of $\sum_{i=1}^{k} \sum_{j=1}^{k} (x_{\ell i} - \overline{x}_i)(\mathbb{V}^{-1})_{ij} (x_{\ell j} - \overline{x}_j)$, where $\ell = 1, 2, ...n$
(n is the number of observations, k is the number of the x-variables).
Correlation-matrix estimator: $r_{ij} = \dfrac{S_{ij}}{\sqrt{S_{ii}S_{jj}}}$
Partial correlation coefficient: $\rho_{ij\,|\,k} = \dfrac{\rho_{ij} - \rho_{ik}\rho_{jk}}{\sqrt{1-\rho_{ik}^2} \cdot \sqrt{1-\rho_{jk}^2}}$
(can be estimated by substituting r for ρ).

Multivariate Regression - Main Results

(covers polynomial and dummy-variable regression):

Model: $\mathbf{y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon}$ all from $\mathcal{N}(0, \sigma)$, independent.

Regression coefficient estimators: $\widehat{\boldsymbol{\beta}} = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbf{y}$

Their variance-covariance matrix: $\sigma^2(\mathbb{X}^T\mathbb{X})^{-1} \simeq MS_E \cdot (\mathbb{X}^T\mathbb{X})^{-1}$

Fitted values: $\widehat{\mathbf{y}} = \mathbb{X}\widehat{\boldsymbol{\beta}}$

Residuals: $\mathbf{e} = \mathbf{y} - \widehat{\mathbf{y}}$

$SS_E = \mathbf{y}^T\mathbf{y} - \widehat{\boldsymbol{\beta}}^T\mathbb{X}^T\mathbf{y}$, $MS_E = \dfrac{SS_E}{n-(k+1)}$

CI for β_i: $\widehat{\boldsymbol{\beta}}_i \pm \mathtt{t}_{\frac{\alpha}{2}, n-k-1} \cdot \sqrt{MS_E \cdot [(\mathbb{X}^T\mathbb{X})^{-1}]_{i,i}}$

CI for σ^2: $\left(\dfrac{SS_E}{\chi^2_{1-\frac{\alpha}{2}}}, \dfrac{SS_E}{\chi^2_{\frac{\alpha}{2}}}\right)$, use χ^2 with $n-(k+1)$ df.

CI for $\mathbb{E}(y_0)$: $\widehat{\boldsymbol{\beta}}^T\mathbf{x}_0 \pm \mathtt{t}_{\frac{\alpha}{2}, n-k-1} \cdot \sqrt{\mathbf{x}_0^T(\mathbb{X}^T\mathbb{X})^{-1}\mathbf{x}_0 \cdot MS_E}$

PI for a new y taken at $\mathbf{x}_0$:

$\widehat{\boldsymbol{\beta}}^T\mathbf{x}_0 \pm \mathtt{t}_{\frac{\alpha}{2}, n-k-1} \cdot \sqrt{\left(1 + \mathbf{x}_0^T(\mathbb{X}^T\mathbb{X})^{-1}\mathbf{x}_0\right) \cdot MS_E}$

Backward Elimination (testing H_0: $\beta_i = 0$ against H_A: $\beta_i \neq 0$):

Test statistic: $\dfrac{\beta_i}{\sqrt{MS_E \cdot [(\mathbb{X}^T\mathbb{X})^{-1}]_{i,i}}}$ (has $\mathtt{t}_{n-k-1}$ distribution).

– –

Redundancy Test:

H_0: A few, specifically selected β's are all equal to zero

H_A: Not so, at least one of them is non-zero

SS_E is now called SS_E^{full}

Remove the corresponding $\mathbf{x}$'s from $\mathbb{X}$, recompute SS_E, call it SS_E^{red}

Test statistic: $\dfrac{SS_E^{rest} - SS_E^{full}}{SS_E^{full}} \cdot \dfrac{n-(k+1)}{k-\ell}$ has, under H_0,

$F_{k-\ell, n-(k+1)}$ distribution (always a right-tail test)!

– –

Weighted Case:

Now, $\varepsilon_i \in \mathcal{N}(0, \frac{\sigma^2}{w_i})$

Define $\mathbb{W}$ as a matrix with w_i on the main diagonal (0 otherwise).

$\boldsymbol{\beta}$ estimators: $\widehat{\boldsymbol{\beta}} = (\mathbb{X}^T\,\mathbb{W}\mathbb{X})^{-1}\mathbb{X}^T\,\mathbb{W}\mathbf{y}$

Their variance-covariance matrix: $\sigma^2(\mathbb{X}^T\,\mathbb{W}\mathbb{X})^{-1} \simeq MS_E \cdot (\mathbb{X}^T\,\mathbb{W}\mathbb{X})^{-1}$

$SS_E = \mathbf{y}^T\mathbb{W}\mathbf{y} - \widehat{\boldsymbol{\beta}}^T\mathbb{X}^T\mathbb{W}\mathbf{y}$, $MS_E = \dfrac{SS_E}{n-(k+1)}$

Everything else remains the same.

3 Nonlinear Regression

Model: $y = f(\mathbf{x}, \mathbf{b}) + \varepsilon$

$\mathbb{X}_{\ell i} = \dfrac{\partial f(\mathbf{x}_\ell, \mathbf{b})}{\partial b_i}$, where $\ell = 1,\, 2, \ldots n$ and $i = 1,\, 2, \ldots k$

(n is the number of observations, k the number of b-parameters).

Residuals: $\mathbf{e}_\ell = y_\ell - f(\mathbf{x}_\ell, \mathbf{b})$, $\ell = 1,\, 2, \ldots n$

Levenberg-Marquardt:

$\mathbf{b}_{(j+1)} = \mathbf{b}_{(j)} + \left[\mathbb{X}_{(j)}^T \mathbb{X}_{(j)} + \lambda \text{ diag } (\mathbb{X}_{(j)}^T \mathbb{X}_{(j)})\right]^{-1} \mathbb{X}_{(j)}^T \mathbf{e}_{(j)}$

where (j) denotes the *iteration*, and λ is increased if SS_e gets bigger (is decreased otherwise), until $\mathbf{b}$ no longer changes (this yields $\widehat{\mathbf{b}}$).

The remaining formulas are identical with the linear case, but the corresponding distributions (Normal, $\mathbf{t}$, χ^2, etc.) are only *approximate*.

In particular: $\widehat{\sigma}^2 = MS_e = \dfrac{\sum_{\ell=1}^n e_\ell^2}{n-k}$, where $\dfrac{\sum_{\ell=1}^n e_\ell^2}{\sigma^2}$ is, approximately, χ_{n-k}^2 $\Rightarrow$ Standard error of $\widehat{\sigma}$ is $\sqrt{\frac{2}{n-k}} \cdot \widehat{\sigma}$.

and $MS_e(\mathbb{X}^T\mathbb{X})^{-1}$ is the V-C matrix of the $\widehat{\mathbf{b}}$ estimators, where the *final* values of e_ℓ and $\mathbb{X}$ are to be used.

4 Robust Regression

Laplace Distribution:

Pdf. of error terms: $\dfrac{\exp\left(\dfrac{|\varepsilon|}{\gamma}\right)}{2\gamma}$, note that $\sigma = \sqrt{2}\gamma$.

$\widehat{\beta}_0$ and $\widehat{\beta}_1$ found by minimizing $\sum\limits_{i=1}^n |e_i|$ (graphically), $\widehat{\gamma} = \dfrac{\sum\limits_{i=1}^n |e_i|}{n}$.

Standard error of $\widehat{\gamma}$: $\dfrac{\gamma}{\sqrt{n}}$, V-C matrix of $\widehat{\beta}_0$ and $\widehat{\beta}_1$: $\gamma^2(\mathbb{X}^T\mathbb{X})^{-1}$.

– –

Cauchy distribution:

Pdf. of error terms: $\dfrac{1}{\pi} \cdot \dfrac{\sigma}{\sigma^2 + \varepsilon^2}$.

$\widehat{\beta}_0$ and $\widehat{\beta}_1$ found by solving $\sum\limits_{i=1}^n \dfrac{e_i}{\sigma^2 + e_i^2} = 0$ and $\sum\limits_{i=1}^n \dfrac{e_i x_i}{\sigma^2 + e_i^2} = 0$,

$\widehat{\sigma}$ by solving $\sum\limits_{i=1}^n \dfrac{\sigma^2}{\sigma^2 + e_i^2} = \dfrac{n}{2}$ (iteratively).

Standard error of $\widehat{\sigma}$: $\sigma\sqrt{\dfrac{2}{n}}$, V-C matrix of $\widehat{\beta}_0$ and $\widehat{\beta}_1$: $2\sigma^2(\mathbb{X}^T\mathbb{X})^{-1}$.

5 Autoregressive Model (for the ε's)

Model: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, where $\varepsilon_i = \rho \cdot \varepsilon_{i-1} + \delta_i$
with $\delta_i \in \mathcal{N}(0, \sigma)$ and independent.
Implies: $\text{Var}(\varepsilon_i) = \frac{\sigma^2}{1-\rho^2}$, and $\text{Corr}(\varepsilon_i, \varepsilon_{i-k}) \equiv \rho_k = \rho^k$.
$\widehat{\beta}_0$, $\widehat{\beta}_1$ and $\widehat{\sigma}$ are computed using weighted-regression formulas with

$$\mathbb{W} = \begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

the ρ estimator is find by solving (iteratively)

$$\rho = \frac{\sum_{i=1}^{n-1} e_i e_{i+1}}{\sum_{i=2}^{n-1} e_i^2 + \frac{\sigma^2}{1-\rho^2}}$$