FORMULAS FOR REGRESSION AND HYPOTHESIS TESTING

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$$

$$s_n^{*\,2} = \frac{n}{n-1} s_n^2$$

$$\hat{r} = \frac{\overline{xy} - \bar{x}\bar{y}}{s_X s_Y}$$

$y = ax + b$ linear regression: $\hat{a} = \hat{r}\frac{s_Y}{s_X} = \frac{\overline{xy} - \bar{x}\bar{y}}{s_X^2}, \qquad \hat{b} = \bar{y} - \hat{a}\bar{x}$

z-test:

1. 1-sample, two-sided: $z = \frac{\bar{x}-\mu}{\sigma}\sqrt{n}, \qquad z_{\varepsilon/2} = \Phi^{-1}(1-\varepsilon/2)$,
   confidence interval for μ: $\left[\bar{x} - z_{\varepsilon/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\varepsilon/2}\frac{\sigma}{\sqrt{n}}\right]$.
2. 1-sample, one-sided: $z = \frac{\bar{x}-\mu}{\sigma}\sqrt{n}, \qquad z_{\varepsilon} = \Phi^{-1}(1-\varepsilon)$.
3. 2-sample, two-sided: $z = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}, \qquad z_{\varepsilon/2} = \Phi^{-1}(1-\varepsilon/2)$.
4. 2-sample, one-sided: $z = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}, \qquad z_{\varepsilon} = \Phi^{-1}(1-\varepsilon)$.

t-test:

1. 1-sample, two-seded: $t = \frac{\bar{x}-\mu}{s_n^*}\sqrt{n}, \qquad t_{\varepsilon/2}$ is the $1-\varepsilon/2$ quantile value of the t_{n-1}-distribution.
   confidence interval for μ: $\left[\bar{x} - t_{\varepsilon/2}\frac{s_n^*}{\sqrt{n}}, \bar{x} + t_{\varepsilon/2}\frac{s_n^*}{\sqrt{n}}\right]$.
2. 1-sample, one-sided: $t = \frac{\bar{x}-\mu}{s_n^*}\sqrt{n}, \qquad t_{\varepsilon}$ is the $1-\varepsilon$ quantile value of the t_{n-1}-distribution
3. 2-sample, two-sided: $t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{(n_1-1)s_X^{*\,2}+(n_2-1)s_Y^{*\,2}}{n_1+n_2-2}}}\sqrt{\frac{n_1 n_2}{n_1+n_2}}, \qquad t_{\varepsilon/2}$ is the $1-\varepsilon/2$ quantile value of the $t_{n_1+n_2-2}$-distribution
4. 2-sample, one-sided: $t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{(n_1-1)s_X^{*\,2}+(n_2-1)s_Y^{*\,2}}{n_1+n_2-2}}}\sqrt{\frac{n_1 n_2}{n_1+n_2}}, \qquad t_{\varepsilon}$ is the $1-\varepsilon$ quantile value of the $t_{n_1+n_2-2}$ distribution

χ^2**-test**:

1. Test for **goodness of fit**: $\chi^2 = \sum_{i=1}^{r}\frac{(\nu_i - np_i)^2}{np_i}$ to be compared to the $1-\varepsilon$ quantile value of the χ^2_{r-1}-distribution
2. Test for **homogeneity**: $\chi^2 = nm\sum_{i=1}^{r}\frac{(\frac{\nu_i}{n} - \frac{\mu_i}{m})^2}{\nu_i+\mu_i}$ to be compared to the $1-\varepsilon$ quantile value of the χ^2_{r-1}-distribution
3. Test for **independence**: $\chi^2 = n\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(\nu_{ij} - \frac{\nu_{i.}\nu_{.j}}{n})^2}{\nu_{i.}\nu_{.j}}$ to be compared to the $1-\varepsilon$ quantile value of the $\chi^2_{(r-1)(s-1)}$-distribution

Welch-test:

$t'(\mathbf{X},\mathbf{Y}) = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_X^{*\,2}}{n_1}+\frac{S_Y^{*\,2}}{n_2}}}, \qquad c = \frac{S_X^{*\,2}/n_1}{S_X^{*\,2}/n_1 + S_Y^{*\,2}/n_2}, \qquad \frac{1}{f} = \frac{c^2}{n_1-1} + \frac{(1-c)^2}{n_2-1}$,

$\mathcal{X}_k = \{(\mathbf{x},\mathbf{y}) \big| \, |t'(\mathbf{x},\mathbf{y})| \geq t_{\varepsilon/2}(f)\}$ in the two-sided case,

$\mathcal{X}_k = \{(\mathbf{x},\mathbf{y}) \big| \, t'(\mathbf{x},\mathbf{y}) \geq t_{\varepsilon}(f)\}$ in the one-sided case, where f is to be rounded to the nearest integer.