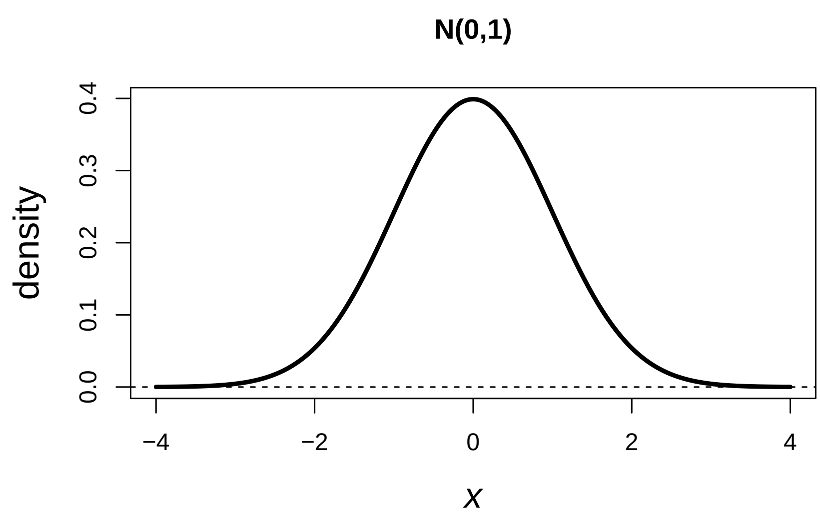
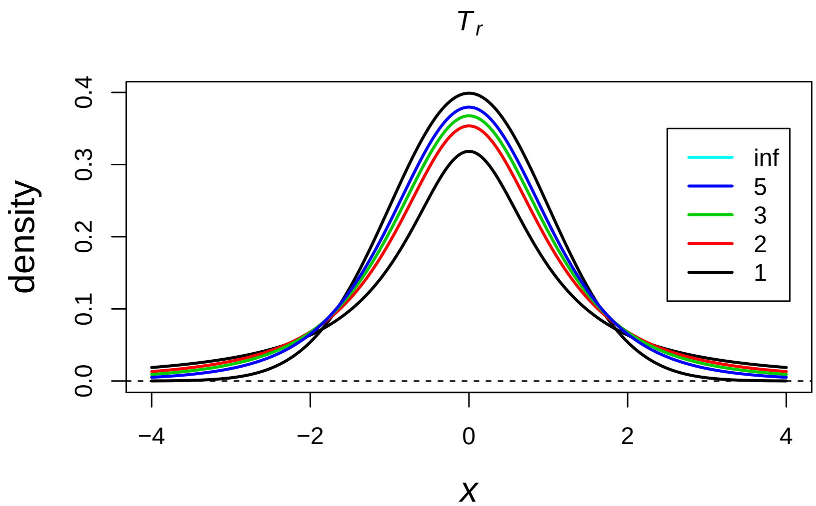
1 Statistical Tests

- In this course, we will develop a theoretical framework for statistical inference and tests.
- Although our favorite high-school level AP Statistics course is not a prerequisite, it is useful to know the formulae presented there.
- In fact, we will not have time to derive each of these!
- We begin by listing the 4 most important sampling distributions.
- All statistical tests basically attempt to characterize when the data are too far away from the assumed hypothesis.

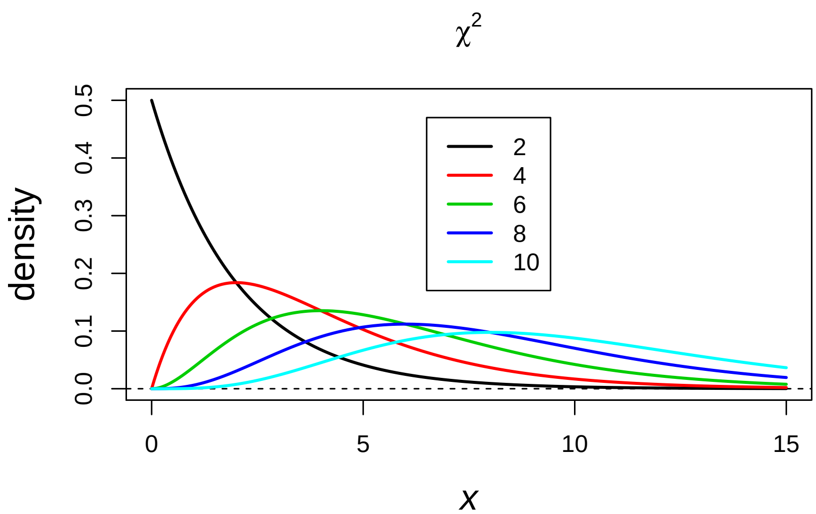
2 Sampling Distributions: Normal



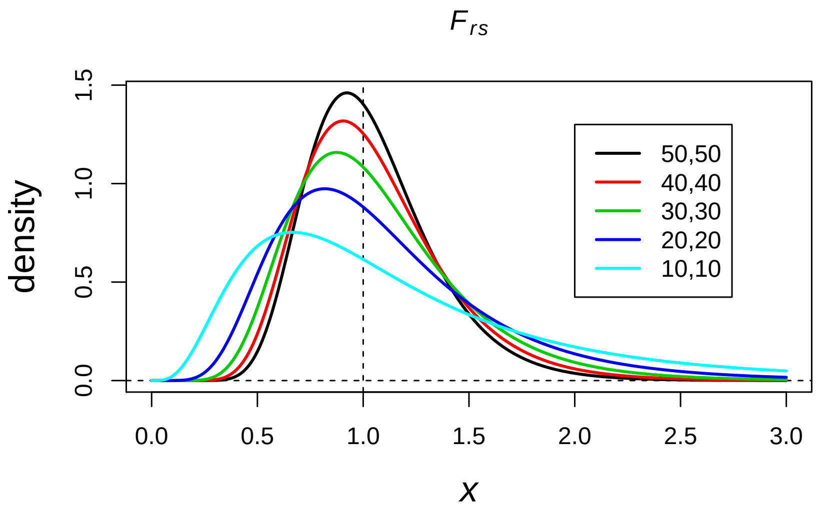
3 Sampling Distributions: Student's T_r



4 Sampling Distributions: Chi-Squared χ_r^2



5 Sampling Distributions: Snedecor's $F_{r,s}$



6 Most Important Tests: One-Sample T-test

To test the null hypothesis

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0 \, ,$$

the test statistic is

$$T_{n-1} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

where

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$
$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \, .$$

7 Two-Sample T-test

To test the null hypothesis

$$H_0 : \mu_x = \mu_y$$
$$H_1 : \mu_x \neq \mu_y \,,$$

the test statistic is

$$T_{n-2} = \frac{\bar{X} - \bar{Y}}{S_P \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$$

where

$$S_P^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}, .$$

8 F-test for Equality of Variances

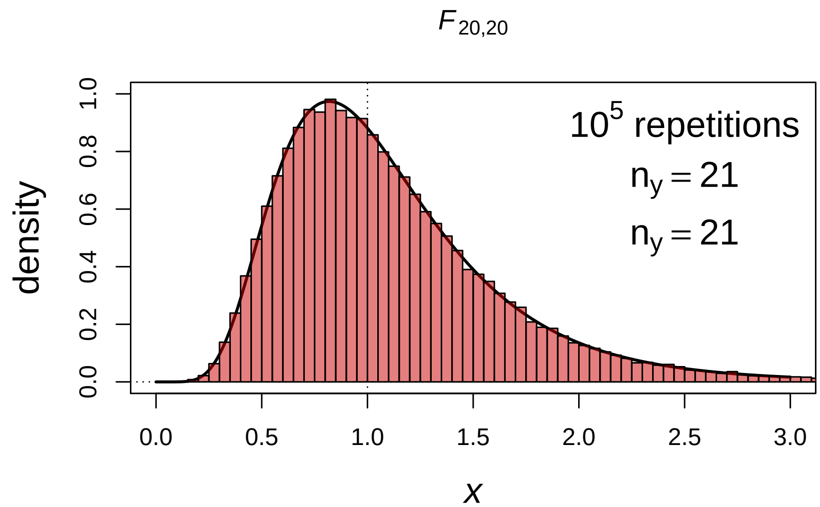
To test the null hypothesis

$$H_0 : \sigma_x^2 = \sigma_y^2$$
$$H_1 : \sigma_x^2 \neq \sigma_y^2 \,,$$

the test statistic is

$$F_{n_x - 1, n_y - 1} = \frac{S_x^2}{S_y^2} \,.$$

9 F-test Simulation (Check)



10 χ^2-tests for Goodness-of-Fit

To test the null hypothesis

H_0 : model gives predictions $e_1, e_2, \ldots, e_k$

H_1 : predictions not close to observed counts $o_1, o_2, \ldots, o_k$,

the test statistic is either

$$\sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} \qquad \text{or} \qquad \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{(o_{ij} - e_{ij})^2}{e_{ij}},$$

which is approximately χ^2_{df} . The number of degrees of freedom (df) depends on which of the 3 types of models are under consideration: goodness-of-fit, contingency table, or multinomial.

11 One-Way ANOVA F-Test

To test the null hypothesis

$$\begin{aligned} H_0 &: \ \mu_1 = \mu_2 = \cdots = \mu_k \\ H_1 &: \ \text{the } k \text{ means are not all equal}\,, \end{aligned}$$

the test statistic is

$$F_{k,n-k} = \frac{MS_{treatment}}{MS_{error}}\,.$$

12 T-Test for Correlation Coefficient

To test the null hypothesis

$$\begin{aligned} H_0 &: \ \rho = 0 \\ H_1 &: \ \rho \neq 0\,, \end{aligned}$$

the test statistic is

$$T_{n-2} = R\sqrt{\frac{n-2}{1-R^2}}\,,$$

where

$$R = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2} \cdot \sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}\,.$$

13 T-Test for Linear Regression Coefficient

To test the null hypothesis

$$\begin{aligned} H_0 &: \ \beta = \beta_0 \\ H_1 &: \ \beta \neq \beta_0\,, \end{aligned}$$

the test statistic is

$$T_{n-2} = \frac{\hat{\beta} - \beta_0}{SE_{\hat{\beta}}}\,.$$