BASICS:

Be able to compute the **mean** and **variance** of a distribution (both discrete - summation, and continuous - integration).

For continuous distributions, also the **median**.

For bivariate distributions, also the **covariance** (double integration).

And, of course, any **probability**.

Some important rules relating to two or more independent variables:

Expected value of a linear combination equals the linear combination of expected values.

Variance equals the sum of individual variances, each multiplied by the corresponding coefficient squared.

Expected value of an independent product is the product of the expected values.

Be familiar with these distributions:

Discrete: Binomial, Geometric, Poisson, Hypergoemetric.

Continuous: Uniform, exponential, Normal, Cauchy, gamma, χ^2, t and F

TRANSFORMING RVs

Given the distribution of X, find the distribution of $Y = \frac{1}{1+X}$.

Given the joint distribution of X and Y, find the distribution of $\frac{X}{X+Y}$.

SAMPLING

From any distribution:

Sample mean and **variance**. $\mathbb{E}(\bar{X}) = \mu$, $\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$. Distribution of $\bar{X}$ tends to Normal (CLT).

From $\mathcal{N}(\mu,\sigma)$: $\bar{X} \in \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$ exactly, $\frac{S_{xx}}{\sigma^2} \in \chi^2_{n-1}$, $\frac{\bar{X}-\mu}{s/\sqrt{n}} \in \mathrm{t}_{n-1}$

ORDER STATISTICS

$$\Pr[X_{(i)} < x] = \sum_{j=i}^{n} \binom{n}{j} p_x^j (1-p_x)^{n-j}$$

where $p_x = F(x)$.

The pdf of $X_{(i)}$ is

$$f_{(i)}(x) = \frac{n!}{(i-1)!(n-i)!} p_x^{i-1} (1-p_x)^{n-i} f(x)$$

Sample **median** is, for large n, approximately $\mathcal{N}(\tilde{\mu}, \frac{1}{2f(\tilde{\mu})\sqrt{n}})$.

Bivariate pdf of $X_{(i)}$ and $X_{(j)}$:

$$\frac{n!}{(i-1)!(j-i-1)!(n-j)!} F(x_i)^{i-1} [F(x_j) - F(x_i)]^{j-i-1}$$
$$[1 - F(x_j)]^{n-j} f(x_i) f(x_j)$$

where $x_i < x_j$.

PARAMETER ESTIMATION

Unbiased, asymptotically unbiased, consistent and MVUE estimator.

RCB:

$$\mathrm{Var}(\hat{\Theta}) \geq \frac{-1}{n\mathbb{E}\left[\frac{\partial^2 \ln f(x|\theta)}{\partial\theta^2}\right]}$$

Efficient and asymptotically efficient estimator.

Sufficient statistic and estimator.

Method-of-**M**oments and **M**aximum-**L**ikelihood estimators.

CONFIDENCE INTERVALS

For population mean μ (σ known or not, n large or not),
difference of two population means,
population proportion p, difference of two proportions (large n only),
population variance σ^2, ratio of two variances σ_1^2/σ_2^2.

HYPOTHESES TESTING

Parallel to CIs. **Standard error** (denominator of **test statistic**). One and two sided hypotheses.

Extra: H_0: $p_1 = p_2 = \ldots = p_k$ (same population **proportions**), **Contingency tables** (test of independence - often confused with the previous 'proportions' test) and **Goodness of Fit** (to a distribution). Significance level. P-value.

REGRESSION

Simple:

$$\beta_1 : \quad \frac{S_{xy}}{S_{xx}} \pm \sqrt{\frac{MS_e}{S_{xx}}} \cdot \mathsf{t}_{n-2}$$

$$\beta_0 : \quad \bar{y} - \bar{x}\hat{\beta}_0 \pm \sqrt{MS_e\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)} \cdot \mathsf{t}_{n-2}$$

Correlation (both x and y random and Normal):

$$\rho: \quad \tanh\left[\operatorname{arctanh}(r) \pm \frac{z_{\mathrm{crit}}}{\sqrt{n-3}}\right]$$

Multivariate:

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

The $\pm$ are computed by

$$\mathsf{t}_{\frac{\alpha}{2}, n-k-1} \cdot \sqrt{C_{jj} \cdot MS_e}$$

ANOVA

One-way.

Two-way with no interaction.

Two-way with interaction and replicates.