		○ $E(x+y) = E(x) + E(y)$ ○ $var(x+y) = var(x) + var(y) + 2cov(x,y)$ ○ $E(Ax+b) = AE(x) + b$ ○ $var(Ax+b) = A\ var(x)\ A'$ (remember the transpose!)	○ $var(\hat{\beta}_{OLS}) = \sigma^2(X^TX)^{-1}$ ○ Gauss-Markov Theorem, i.e., among all unbiased linear estimators, $\hat{\beta}_{OLS}$ has the maximum variance, and hence $\hat{\beta}_{OLS}$ is the Best Linear Unbiased Estimator. • Linearity: $\hat{\beta}_{OLS} = Cy$, where C denotes a matrix ○ Unbiasedness: $E(\hat{\beta}_{OLS}) = \beta = E(Cy)$
	Least Squares as a Statistical Problem	• Assumption of a statistical relationship between X and y. ○ X is a fixed, non-stochastic matrix with rank k. ○ X^TX also has rank k and is invertible. • U is a random vector with $E(u) = 0$ and $var(u) = E(uu') = \sigma^2 I$ • $var(u)$ is the error covariance matrix, which gives the covariance between any 2 elements in u. ○ Assume that each observation has the same variance (homoskedasticity). ○ Assume that all observations have zero covariance (no autocorrelation). • Key properties of the OLS estimator! Need to prove these too, -_-) ○ $E(\hat{\beta}_{OLS}) = \beta$ ○ $var(\hat{\beta}_{OLS}) = \sigma^2(X^TX)^{-1}$ ○ Gauss-Markov Theorem, i.e., among all unbiased linear estimators, $\hat{\beta}_{OLS}$ has the maximum variance, and hence $\hat{\beta}_{OLS}$ is the Best Linear Unbiased Estimator. ○ Linearity: $\hat{\beta}_{OLS} = Cy$, where C denotes a matrix ○ Unbiasedness: $E(\hat{\beta}_{OLS}) = \beta = E(Cy)$	
	OLS Estimate of σ^2	• Residuals: $\hat{u} \equiv y - X\hat{\beta} = (I - X(X^TX)^{-1}X^T)u$ • Expected Sum of Squared Residuals: $E(\hat{u}^T\hat{u}) = \sigma^2(n-k)$ $$\Rightarrow \sigma^2_{OLS} = \frac{\hat{u}^T\hat{u}}{n-k}$$ • Since σ^2_{OLS} is an unbiased estimator, $E(\sigma^2_{OLS}) = E\left(\frac{\hat{u}^T\hat{u}}{n-k}\right) = \sigma^2$	
Hypothesis Testing	**Common Probability Distributions**	• **Normal Distribution, N(μ, σ^2)** ○ μ is the mean, controls center point. ○ σ^2 is the variance, controls dispersion (spread) ○ Formula of PDF: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$	• Key relationships between distributions: ○ If $z_i \sim N(0,1), i = 1, \dots, n$ and z_i are all independent, then $\sum_{i=1}^{n} z_i^2 \sim \chi^2(n)$

Marginal Effects

- Helps to interpret $\widehat{\beta_{ML}}$, as a derivative indicating how much a one unit increase in each X variable will increase the expected value of y by.

$$\frac{\partial E(y_i)}{\partial x_i'} = \frac{\partial P(y_i = 1)}{\partial x_i'}$$

- Expressed as a column vector.
- In probit, marginal effects: $\frac{\partial E(y_i)}{\partial x_i'} = \frac{\partial}{\partial x_i'} \Phi\left(\frac{x_i\beta}{\sigma}\right) = \phi\left(\frac{x_i\beta}{\sigma}\right)\left(\frac{\beta}{\sigma}\right)$
- Can be evaluated at any value of x_i, however, it is more common to use the mean values of each X variable, denoted $\bar{x}$, a row vector.

$$\frac{\partial E(y_i)}{\partial x_i'}|_{x_i=\bar{x}} = \phi\left(\frac{\bar{x}\beta}{\sigma}\right)\left(\frac{\beta}{\sigma}\right)$$

- If you know $\widehat{\beta_{ML}}$ and $\widehat{\sigma_{ML}}$. Then the estimated marginal effect just uses the probit coefficients in place of the true model parameters.

$$\frac{\partial E(y_i)}{\partial x_i'}|_{x_i=\bar{x}} = \phi\left(\frac{\bar{x}\widehat{\beta_{ML}}}{\sqrt{\hat{\sigma}_{ML}^2}}\right)\left(\frac{\widehat{\beta_{ML}}}{\sqrt{\hat{\sigma}_{ML}^2}}\right)$$

- **Average ME:** The average of predicted changes in fitted values for one unit change in X (if it's continuous) for each observation.