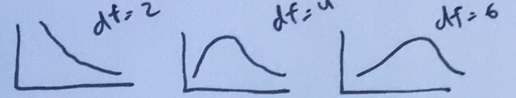
CHI SQUARE DISTRIBUTION

- Distribution of SS of k ind. normal variables
- use w/ Chi 2 test

ex:



df = (~~categories~~ levels in 1)(levels in 2)

- only need 1 parameter, m, to define it

$$\chi^2 = \chi^2(m) \qquad \chi^2 = \sum_{i=1}^{m} z_i^2$$

errs for 1 obs.

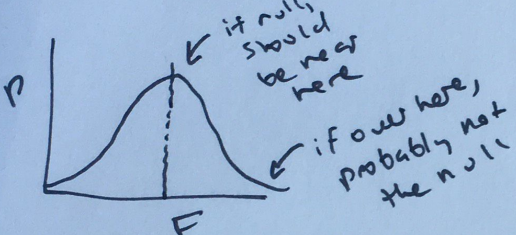
$$z_i = \frac{x_i - \bar{x}}{\sigma}$$

← sample mean

← sample s.d.

F-DISTRIBUTION

· test statistic in F-tests follow this distribution under the null.



WRITTEN MIDTERM NOTES PSYCH 205

LIKELIHOOD

$L = P(\text{DATA} \mid \text{MODEL})$ Probability of data given the model

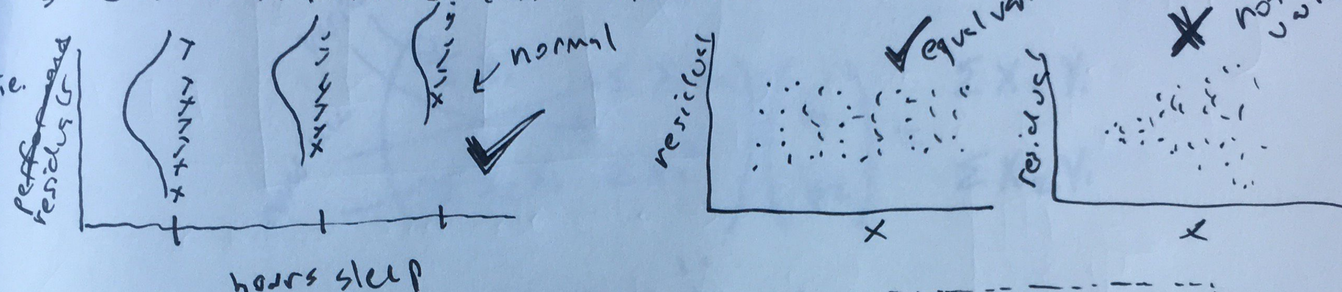
$$L = \prod_{i=1}^{n} P(y \mid \hat{y})$$

LM ASSUMPTIONS

1- OBSERVATIONS ARE INDEPENDENT

2- DISTRIBUTION IS NORMAL / GAUSSIAN (OF RESIDUALS

3- EQUAL VARIANCE (AKA HOMOSKEDASTIC)



MODEL COMPARISON

$$\boxed{SSE = \sum_{i=1}^{n} (\hat{y} - y)^2}$$

$$\boxed{F = \frac{\frac{(SS_1 - SS_2)}{(K_2 - K_1)}}{\frac{SS_2}{n - K_2}}}$$

SS = sum of squares
K = # parameters
n/m = sample size

$$\boxed{R^2 = \frac{\frac{SS_1}{n} - \frac{SS_2}{n}}{\frac{SS_1}{n}} = 1 - \frac{SS_2}{SS_0}}$$ (for null)

$$\boxed{R^2_{adj} = 1 - \frac{\frac{SS_2}{(n-k_2)}}{\frac{SS_1}{(n-K_1)}} = 1 - \frac{\frac{SS_2}{n-k_2}}{\frac{SS_0}{(n-1)}}}$$ (for null)

$$\boxed{R^2_{cv} = \frac{SS_1 - SS_2}{SS_1}}$$ } calculated on hold-out set of test data the model was ∅ trained on

SSE = way to quantify errors absolutely

R^2 = % of variance model accts for vs. another baseline

R^2_{adj} = like R^2, but punishes extra parameters

R^2_{cv} = like R^2, but controls for overfit using test data set

F-stat = usually for testing ~~into~~ if ∅ null (independent distributions)
test stat has an F distribution under the null

NORMAL EQUATION FOR BI-/MULTIVARIATE

BIVARIATE

$$\hat{y} = b_0 + b_1 X_1$$

(if means = 0)

$$b_1 = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\Sigma (x-\bar{x})^2} = \frac{\sum_{i=1}^{n} X_{1i} Y_{1i}}{\sum_{i=1}^{n} X_{1i} X_{1i}}$$

MULTI-VARIATE

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 (+ \dots + b_j x_j)$$

$$\frac{dL}{db_1} = 0 \quad \frac{dC}{db_2} = 0 \Rightarrow \begin{cases} b_1 \Sigma X_{1i}^2 + b_2 \Sigma X_{1i} X_{2i} = \Sigma X_i Y_i \\ b_2 \Sigma X_{1i} X_{2i} + b_2 X_{2i}^2 = \Sigma X_{2i} Y_i \end{cases}$$

$$\parallel$$

variance

covariance

$$\begin{pmatrix} \Sigma X_1^2 & \Sigma X_1 X_2 \\ \Sigma X_2 X_1 & \Sigma X_2^2 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{matrix} \Sigma X_1 Y_i \\ \Sigma X_2 Y_i \end{matrix}$$

$$\parallel$$

$$C_{xx} \vec{b} = C_{xy}$$

PEARSON'S CORRELATION COEFF.

(if means = 0)

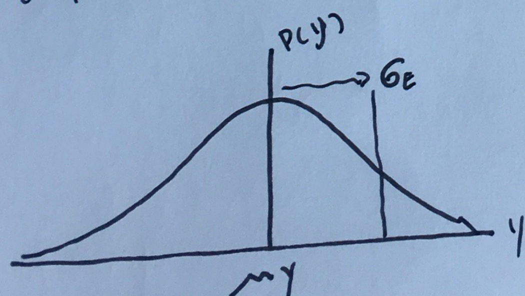
$$\boxed{cor(x,y) = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma (x-\bar{x})^2 \Sigma (y-\bar{y})^2}}} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

NORMAL DISTRIBUTION (AKA GAUSSIAN)

$$PDF = \frac{1}{\sqrt{2\pi}\,\sigma_E} e^{\left(\frac{-\frac{1}{2}(y-\mu y)^2}{\sigma_E^2}\right)}$$

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma_E} e^{\frac{-\frac{1}{2}(y_i - \hat{y}_i)^2}{\sigma_E^2}}$$

} when you assume = variance



- defined by μ + σ (mean and sd.)

eg $N(\mu, \sigma)$