Some Useful Formulas From LM's, GLM's, and MLM's

Do not bring this to the exam. A copy will be provided for you.

Basic Statistics

	Traditional	Simulation
Confidence Intervals	If parameter estimate (mean, MLE, etc.) follows normal distribution, use "estimate ± 2×SE"	Estimate parameter(s) from real data, simulate 1000 data sets from estimated parameters, compute 1000 estimates of what you want, compute 2.5%-ile and 97.5%-ile.
Hypothesis Tests	Mathematically derive (or look up) distribution of T(data) under H_0. If T(real data) is in tail of distribution, reject H_0.	Estimate parameters from real data, simulate 1000 data sets to get distribution of T(data). If T(real data) is in tail of distribution, reject H_0.

Linear Regression

- *Lazy Way:* $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \epsilon$
- *Long Way:* $y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_K X_{iK} + \epsilon_i = X_i\beta + \epsilon_i,\ i = 1, \ldots n$
  - y_i are responses, X_{ik} are intercept (1's) and predictors, β_k are coef's, ϵ_i are iid $N(0, \sigma^2)$ "errors".
- *Matrix Way:* $Y = X\beta + \epsilon$
  - $Y_{n\times 1}$ are responses, $X_{n\times K}$ are intercept (1's) and predictors, $\beta_{K\times 1}$ are coef's, $\epsilon_{n\times 1}$ are iid $N(0, \sigma^2)$ "errors".

Some other formulae

- $\hat\sigma^2 = \frac{1}{n-k}\sum_{i=1}^n (y_i - X_i\hat\beta)^2 = \frac{1}{n-k}(y - X\hat\beta)^T(y - X\hat\beta)$
- $Y \sim N(X\beta, \sigma^2 I) \Rightarrow \hat\beta \sim N(\beta, (X^TX)^{-1}\sigma^2)$, $\hat y \sim N(Hy, H\sigma^2)$, where $H = X(X^TX)^{-1}X^T$
- $R^2 = \frac{\text{(raw variance)–(residual variance)}}{\text{(raw variance)}} = 1 - \frac{\hat\sigma^2}{\text{Var}(Y)} = [Cor(Y, \hat Y)]^2$

Generalized Linear Models

GLM's

- $y_i \sim f(y|\mu_i, \ldots),\ \mu_i = E[y_i], i = 1 \ldots n$
- $g(\mu_i) = \theta_i = X_i\beta$

Normal Linear Model

- $y_i \in \mathbb{R};\ y_i \sim N(\mu_i, \sigma^2),\ \mu_i = E[y_i]$
- $g(\mu) = \mu$

Logistic Regression

- $y_i \in \{0, 1\};\ y_i \sim \text{Bern}(p_i),\ p_i = P[y_i = 1] = E[y_i]$
- $g(p) = \log p/(1-p);\ g^{-1}(\theta) = \exp(\theta)/(1 + \exp(\theta))$

Poisson Regression

- $y_i \in \{0, 1, 2, 3, \ldots\};\ y_i \sim \text{Poisson}(\mu_i),\ \mu_i = E[y_i]$
- $g(\mu) = \log\mu;\ g^{-1}(\theta) = \exp(\theta)$

Causal Inference

- $\text{ACE} = \frac{1}{N}\sum_{i=1}^N (y_i^1 - y_i^0) = \frac{1}{N}\sum_{i=1}^N y_i^1 - \frac{1}{N}\sum_{i=1}^N y_i^0 = E[y^1] - E[y^0]$; but can't observtve both y_i^1 and y_i^0.
- $\widehat{\text{ACE}} = \hat\beta_1$ in $y_i = \beta_0 + \beta_1 T_i + \beta_2 x_{2i} + \cdots + \beta_K x_{Ki} + \epsilon_i$; x_{ik}'s are confounders and/or pre-treatment covariates.
- *Obs. Studies:* What can a "cause" be?
- *Obs. Studies:* Instrumental Variables, Propensity Scores, Regression Discontinuity, Bounding Confounders

Multilevel Models – Example: Random Intercept

Hierarchical Form

$$\text{Level 2: } \alpha_j \overset{iid}{\sim} N(\beta_0, \tau^2)$$
$$\text{Level 1: } y_i \overset{indep}{\sim} N(\alpha_{j[i]}, \sigma^2)$$

Multi-Level Form

$$y_i = \alpha_{j[i]} + \epsilon_i,\ \epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$
$$\alpha_j = \beta_0 + \eta_j,\ \eta_j \overset{iid}{\sim} N(0, \tau^2)$$

Variance-Components Form

$$y_i = \beta_0 + \eta_{j[i]} + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$
$$\eta_j \overset{iid}{\sim} N(0, \tau^2)$$

"lmer" Form

```
lmer(y ~ 1 + (1 | group))
```

Laird-Ware Form

$$y = X\beta + Z\eta + \epsilon$$

Model selection / model criticism

- AIC, BIC, DIC: 2, 3, 10
- Marginal Residuals, Conditional Residuals, Rand. Eff Residuals

MLE Facts

- Likelihood $L(\theta) \propto f(\text{data} \mid \theta)$
- $LL(\theta) = \log L(\theta)$
- MLE $\hat{\theta}$ solves $0 = L'(\theta)$
- Fisher information $I(\theta) = E[-LL''(\theta)]$
- $SE(\hat{\theta}) = 1/\sqrt{I(\theta)}$

Model Selection Criteria

Let $\mathcal{M}$ be a model $f_{\mathcal{M}}(\text{data} \mid \text{parameters})$ with k parameters for data with n observations.

- Deviance = $D(\mathcal{M}) = -2\log f_{\mathcal{M}}(\text{data} \mid \text{MLE's of parameters})$.
- AIC = $D(\mathcal{M}) + 2k$
- BIC = $D(\mathcal{M}) + k\log(n)$
- DIC = $D(\mathcal{M}) + 2k_{est}$

Bayesian Inference

- The slogan
- Posterior mode [or median, or mean] is like MLE
- Posterior SD is like SE
- Shrinkage

Conjugate Priors – Two Examples

- *Gamma/Poisson:*
  - Prior: $f(\lambda) = \text{Gamma}(\lambda|\alpha, \beta)$
  - Likelihood: $x_1, \ldots, x_n \overset{iid}{\sim} \text{Poisson}(x|\lambda)$
  - Posterior: $f(\lambda|x_1, \ldots, x_n) = \text{Gamma}(\lambda|\alpha^*, \beta^*)$, where $\alpha^* = \alpha + n\overline{x}$, $\beta^* = \beta + n$
- *Normal/Normal:*
  - Prior: $f(\mu) = N(\mu|\mu_0, \tau_0^2)$
  - Likelihood: $x_1, \ldots, x_n \overset{iid}{\sim} N(x|\mu, \sigma^2)$
  - Posterior: $f(\mu|x_1, \ldots, x_n) = N(\mu|\mu_n, \tau_n^2)$ where $\tau_n^2 = 1/(n/\sigma^2 + 1/\tau_0^2)$ and

$$\mu_n = \left(\frac{\tau_0^2}{\tau_0^2 + \sigma^2/n}\right)\overline{x} + \left(\frac{\sigma^2/n}{\tau_0^2 + \sigma^2/n}\right)\mu_0$$

Some Common Distributions

Name	**Density**	**Range**	$E[X]$	**Var(X)**
$N(x|\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$	$-\infty < x < \infty$	μ	σ^2
Binomial$(x|n, p)$	$\binom{n}{x} p^x (1-p)^{n-x}$	$0 \le x \le 1$	np	$np(1-p)$
Poisson$(x|\lambda)$	$\lambda^x e^{-\lambda}/x!$	$x = 0, 1, 2, \ldots$	λ	λ
Exponential$(x|\lambda)$	$\lambda e^{-\lambda x}$	$0 \le x < \infty$	λ^{-1}	λ^{-2}
Unif$(x|a, b)$	$1/(b-a)$	$a \le x \le b$	$(a+b)/2$	$(b-a)^2/12$
Beta$(x|\alpha, \beta)$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$	$0 \le x \le 1$	$\alpha/(\alpha+\beta)$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Gamma$(x|\alpha, \beta)$	$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$	$0 \le x < \infty$	α/β	α/β^2

JAGS – Example: Random Intercept

```
model {
  for (i in 1:n) {
    y[i] ~ dnorm(mu[i],sig2inv)
    mu[i] <- b0 + a0[group[i]]
  }
  for (j in 1:n.groups) {
    a0[j] ~ dnorm(0,tau2inv)
  }
  b0 ~ dnorm(0,0.000001)
  tau2inv <- pow(tau,-2);  tau ~ dunif(0,1000)
  sig2inv <- pow(sig,-2);  sig ~ dunif(0,1000)
}
```