Statistical Analysis Cheatsheet

Compiled by Gejun Zhu (zhug3@miamioh.edu) in preparation for the analysis comprehensive exam by using William Chen's formula sheet template.

Last Updated August 18, 2015

Regression Analysis

Simple Linear Regression

Model : $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, $Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$.

$E_i = Y_i - \hat{Y}_i$, $\sum E_i = 0$, $\sum Y_i = \sum \hat{Y}_i$, $\sum X_i E_i = 0$, $\sum \hat{Y}_i E_i = 0$.
$B_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{S_{xy}}{S_{xx}}$, $S_{xy} = \sum XY - \frac{\sum X \sum Y}{n}$, $B_0 = \bar{Y} - B_1 \bar{X}$.
$B_1 = \sum_{i=1}^n Y_i \frac{(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \sum_{i=1}^n k_i Y_i$; $\sum k_i = 0$, $\sum k_i x_i = 1$, $\sum k_i^2 = \frac{1}{S_{xx}}$.
$B_1 \sim N(\beta_1, \frac{\sigma^2}{S_{xx}})$; $\frac{B_1 - \beta_1}{\sqrt{V(B_1)}} \sim N(0,1)$, $\frac{S_{B_1}^2}{V(B_1)} = \frac{MSE/S_{xx}}{\sigma^2/S_{xx}} = \frac{SSE}{(n-2)\sigma^2}$,
$\frac{SSE}{\sigma^2} \sim \chi^2_{n-2} \Rightarrow \frac{B_1 - \beta_1}{S_{B_1}} = \frac{B_1 - \beta_1}{\sqrt{MSE/S_{xx}}} \sim T_{n-2}$, CI: $b_1 \pm t_{1-\alpha/2, n-2} \cdot s_{b_1}$; $\frac{SSR}{\sigma^2} \sim \chi^2_{p-1}$.
Inference on $E(Y_h)$: $\hat{Y}_h = B_0 + B_1 X_h$, $\hat{Y}_h \sim N(\beta_0 + \beta_1 X_h, \sigma^2[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{S_{xx}}])$
Prediction on a new observation: $\hat{y} \pm t_{1-\alpha/2, n-2}\sqrt{mse[1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{X_{xx}}]}$.
$SST = \sum (Y_i - \bar{Y}_i)^2 = \sum (Y_i - \hat{Y}_i)^2 + \sum (\hat{Y}_i - \bar{Y}_i)^2 = SSE + SSR$
If $var(Y_i) = \sigma^2$, and $Y_i's$ are uncorrelated, then $Cov(\sum a_i Y_i, \sum b_i Y_i) = \sigma^2 \sum a_i b_i$.
B_1 and $\bar{Y}$ are uncorrelated, $Cov(B_1, \bar{Y}) = 0$ because
$Cov(B_1, \bar{Y}) = Cov(\sum k_i Y_i, \sum \frac{1}{n} Y_i) = \frac{\sigma^2}{n} \sum k_i = 0$.
Confidence intervals tell you about how well you have determined the mean. Prediction intervals tell you where you can expect to see the next data point sampled.

ANOVA Table - Analysis of variance for simple linear regression

Source	SS	DF	MS	Expected
Regression	SSR	1	MSR=SSR/1	$\sigma^2 + \beta_1^2 S_{xx}$
Error	SSE	n-2	MSE=SSE/(n-2)	σ^2
Total	SST	n-1		

Under $H_0: \beta_1 = 0$, $F^* = \frac{MSR}{MSE} \sim F_{1,n-2}$. $R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$. $B_1 = r\sqrt{\frac{S_{yy}}{S_{xx}}}$.
E(MSR) $= \sigma^2 + \beta_1^2 S_{xx}$, SSR $= B_1^2 S_{xx}$, $E(MSE) = E(\frac{SSE}{n-2}) = \frac{\sigma^2}{n-1} E(\frac{SSE}{\sigma^2}) = \sigma^2$
Studentized residuals: $E_i^* = \frac{E_i}{\sqrt{V(E_i)}}$

Assumptions - LINE

- Linearity: No curvature in the residual plot; (high-order, log/square)
- Independence: $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$;
- Normality of error: QQ plot; (GLM, poisson regression...)
- Equal Variance: standardized residual inside [-3, 3]. (weighting obs)

Matrix Approach - Matrix form
$\mathbf{Y_{n\times1}} = X_{n\times p}\boldsymbol{\beta}_{p\times1} + \boldsymbol{\epsilon}_{n\times1}$, $\boldsymbol{\epsilon} \sim MN(\mathbf{0}, \sigma^2\mathbf{I})$. $\boldsymbol{\beta} = (X'X)^{-1}X'\boldsymbol{Y}$.
$\boldsymbol{Y} \sim MN(\mu, \sum)$, then $A\boldsymbol{Y} + \boldsymbol{b} \sim MN(A\mu + b, A\sum A')$.
$\boldsymbol{B} \sim \boldsymbol{MN}(\boldsymbol{\beta}, \sigma^2(X'X)^{-1})$, $\hat{\boldsymbol{Y}} = X\mathbf{B} = H\boldsymbol{Y} \sim MN(X\boldsymbol{\beta}, \sigma^2 H)$, $H = X(X'X)^{-1}X'$.
$S^2(\boldsymbol{B}) = MSE(X'X)^{-1}$; $\boldsymbol{E} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = (I - H)\boldsymbol{Y} \sim MN(\mathbf{0}, \sigma^2(I - H))$.
$\sum Y_i^2 = \boldsymbol{Y}'J\boldsymbol{Y}$, $SSTO = \boldsymbol{Y}'(I - \frac{1}{n}J)\boldsymbol{Y}$, $SSE = \boldsymbol{Y}'(I - H)\boldsymbol{Y}$, $SSR = \boldsymbol{Y}'(H - \frac{1}{n}J)\boldsymbol{Y}$.
$\hat{Y} = HY$, $\hat{Y}' = Y'H$, $0 = \sum \hat{Y}_i E_i = \sum \hat{Y}_i Y_i - \sum \hat{Y}_i^2$.
Distribution of $\hat{Y}_h$: $\hat{Y}_h = \boldsymbol{X_h'B} \sim MN(\boldsymbol{X_h'\beta}, \sigma^2\boldsymbol{X_h'}(X'X)^{-1}\boldsymbol{X_h})$
$S^2(\hat{Y}_h) = MSE(\boldsymbol{X_h'}(X'X)^{-1}\boldsymbol{X_h})$
$pred = Y_{h(new)} - \hat{Y}_h$, $pred \sim N(0, \sigma^2(1 + \boldsymbol{X}'_{h(new)}(X'X)^{-1}\boldsymbol{X}_{h(new)}))$

Multiple Regression - Multiple regression
$\boldsymbol{Y} = X\beta + \boldsymbol{\epsilon}$, where $\epsilon \sim MN(\mathbf{0}, \sigma^2 I)$.
$df(SSE) = n - p$, where p is the number of parameters; $df(SSTo) = n - 1$,
$df(SSR) = p - 1$. Reject H_0 if $f^* = \frac{MSR}{MSE} > F_{1-\alpha; p-1, n-p}$. $R^2 = 1 - \frac{SSE}{SSTo}$,
adjusted $R^2 = 1 - (\frac{n-1}{n-p})\frac{SSE}{SSTo}$
CI: $b_1 \pm t_{1-\alpha/2, n-p} \cdot s_{b_1}$, Bonferroni CI: $b_1 \pm t_{1-\alpha/2g, n-p} \cdot s_{b_1}$
FWER - Bonferroni and Holm. Bonferroni: compare p-value with α/g, CL = $1 - \alpha/g$, where g is the number of tests; Holm: sort p-values and multiple g, g-1, ..., 1 in order and compare with α finally.
$SSTo = SSR(X_1) + SSE(X_1) = SSR(X_1, X_2) + SSE(X_1, X_2)$, where
$SSR(X_1, X_2) = SSR(X_1) + SSR(X_2|X_1)$, $SSE(X_1, X_2) = SSE(X_2) - SSR(X_1|X_2)$.
In general, $SSR(X_q, X_{q+1}, ..., X_{p-1}|X_1, X_2, ..., X_{q-1}) =$
$SSE(X_1, X_2, ..., X_{q-1}) - SSE(X_1, X_2, ..., X_{p-1}) = SSE_R - SSE_F$.
Partial F-test (Reduced vs. Full): $H_0: \beta_q = \beta_{q+1} = ... = \beta_{p-1} = 0$, H_a: At least one $\beta_k \neq 0$, $k = q, q+1, .., p-1$. Test statistic
$F^* = \frac{\frac{SSE_R - SSE_F}{df_R - df_F}}{MSE_F} = \frac{\frac{SSR(X_q, X_{q+1}, ..., X_{p-1}|X_1, X_2, ..., X_{q-1})}{p-q}}{\frac{SSE(X_1, X_2, ..., X_{p-1})}{n-p}}$. If H_0 is true, then
$F^* \sim F_{p-q, n-p}$.
coefficient of partial determination is the proportion of the variation in Y "explained" by an indep. variable when other indep. variables are in the model.
$R^2_{Y1|2} = \frac{SSE(X_2) - SSE(X_1, X_2)}{SSE(X_2)} = \frac{SSR(X_1|X_2)}{SSE(X_2)} = 1 - \frac{SSE(X_1, X_2)}{SSE(X_2)}$
$R_{Y1|2,3,4} = 1 - \frac{SSE(X_1, X_2, X_3, X_4)}{SSE(X_2, X_3, X_4)}$
multicollinearity diagnostic: Variance Inflation Factor (VIF) = $(1 - R_k^2)^{-1}$, where $R_k^2 =$ coefficient of determination when X_k is regressed upon other predictors. If $VIF > 1$, variance of B_k is inflated due to correlations b/w X_k and other predictors. If X_k is uncorrelated with other predictors, then $R_k^2 = 0$ and $VIF_k = 1$.

Model Diagnostics - More about model diagnostics
Added-variable Plots (1) regress Y on predictors except X_k and obtain the residuals; (2) regress X_k on other predictors and obtain residuals; (3) plot (1) vs (2);
Leverage: A measure of how unusual an X is. (diagonal values of Hat matrix, $\sum h_{ii} = tr(H) = tr[X(X'X)^{-1}X'] = tr[X'X(X'X)^{-1}] = tr[I_{p\times p}] = p$)
Influence: An influence point is its exclusive causes substantial changes to the fitted data. Just because a point has high leverage doesn't mean it has high influence.
Measures of influence include:
$DFFITS_i = \frac{\hat{Y}_i - \hat{Y}_{i(i)}}{\sqrt{MSE_{(i)} h_{ii}}}$ - a measure of an observation on its own fitted value.
Cook's Distance = $\frac{e_i^2}{pMSE}[\frac{h_{ii}}{(1 - h_{ii})^2}]$ - a measure of influence of observation i on all the fitted value.
$DFBETAS_{k(i)} = \frac{b_k - b_{k(i)}}{\sqrt{MSE_{(i)} c_{kk}}}$ where c_{kk} is $(k+1)^{th}$ diagonal in $(X'X)^{-1}$ - a measure of the influence of observation i on the parameter estimate b_k. It measures the difference b/w the parameter with / without observation i.

Design of experiments

CRD with one factor

Model one factor with $a \geq 2$ levels. $H_0: \mu_1 = \mu_2, ..., = \mu_a$ or $\hat{\tau}_i = 0$.

- $y_{ij} = \mu + \tau_i + \epsilon_{ij}$, $i = 1, 2, ..., a$, $j = 1, 2, ..., n_j$, $\epsilon_{ij} \sim N(0, \sigma^2)$;
- $y_{i.} = \sum_j y_{ij}$, $E(y_{ij}) = \mu_i$, $var(y_{ij}) = \sigma^2$;
- LSE estimator $\hat{\mu} + \hat{\tau}_i = \bar{y}_{i.}$, if $\sum n_i \hat{\tau}_i = 0$ or $\hat{\mu} = 0$ or $\hat{\tau}_a = 0$;
- $MS_{trt} = \frac{SS_{trt}}{a-1} = \frac{\sum_{i=1}^a n_i(\bar{y}_i - \bar{y}_{..})^2}{a-1}$, $MS_E = \frac{SSE}{N-a} = \frac{\sum_{i=1}^a \sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})}{N-a}$;
- $E(MS_E) = \sigma^2$, $E(MS_{trt}) = \sigma^2 + \frac{\sum_{i=1}^a n_i \tau_i^2}{a-1}$, $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$;
- $SS_T = \sum_{i=1}^a \sum_{j=1}^{n_i} y_{ij}^2 - \frac{(y_{..})^2}{N}$, $SS_{trt} = \sum_{i=1}^a \frac{y_i^2}{n_i} - \frac{(y_{..})^2}{N}$;
- Fact: Under H_0, $SSE/\sigma^2 \sim \chi^2_{N-a}$, $SS_{trt}/\sigma^2 \sim \chi^2_{a-1}$, independent;
- $\frac{SS_{trt}/(a-1)\sigma^2}{SS_E/(N-a)\sigma^2} \sim F_{a-1,N-a}$; rej $F_0 > F(\alpha, a-1, N-a)$, $p = P(F_{a-1,N-a} > F_0)$;
- $E(\bar{y}_{i.}) = \mu_i$, $V(\bar{y}_{i.}) = \sigma^2/n_i$, $\frac{\bar{y}_{i.} - \mu_i}{\sqrt{MSE/n_i}} \sim T_{N-a}$;
- CI: $\bar{y}_{i.} \pm t_{\alpha/2,N-a}\sqrt{MSE/n_i}$, $\bar{y}_{s.} - \bar{y}_{t.} \pm t_{\alpha/2,N-a}\sqrt{MSE/n_s + MSE/n_t}$.
- Linear contrasts: $\Gamma = \sum c_i \mu_i$, $C = \sum c_i \bar{y}_{i.}$ with $\sum c_i = 0$. $E(C) = \Gamma$, $V(C) = \sigma^2 \sum \frac{c_i^2}{n_i}$. CI: $\sum c_i \bar{y}_{i.} \pm t_{\alpha/2,N-a}\sqrt{MSE \sum \frac{c_i^2}{n_i}}$, $t = \frac{\sum c_i \bar{y}_{i.} - c}{\sqrt{MSE \sum \frac{c_i^2}{n_i}}}$.

ANOVA Table Analysis of variance for three factor fixed effects model.

Source	DF	Expected Mean Square
A	$a-1$	$\sigma^2 + \frac{bcn \sum \tau_i^2}{a-1}$
AB	$(a-1)(b-1)$	$\sigma^2 + \frac{cn \sum\sum (\tau\beta)_{ij}^2}{(a-1)(b-1)}$
ABC	$(a-1)(b-1)(c-1)$	$\sigma^2 + \frac{n \sum\sum\sum (\tau\beta\gamma)_{ijk}^2}{(a-1)(b-1)(c-1)}$
Error	$abc(n-1)$	σ^2

Basic Blocking Designs

Model two factors - the treatment factor τ_i and the block factor β_j.
$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$, $i = 1, 2, ..., a$, $j = 12, ..., b$. ($\sum \tau_i = 0$ and $\sum \beta_j = 0$)
$H_0: \tau_0 = \tau_1 = ... = \tau_a = 0$ or $\mu_1 = \mu_2 = ... = \mu_a$, $H_a: Not\ H_0$ (at least two means differs) . $E(\bar{Y}_i) = \mu + \tau_i$
A balanced incomplete block design (BIBD) includes a treatment factor with a levels, a blocking factor with b levels, each block includes k experimental units, which implies a total of bk runs. This means that each treatment occurs $r = bk/a$ times. Each treatment occurs either 0 or 1 times, and each pair of treatments occurs together in a block exactly λ times. $N = bk$. (1) $ar = bk$; (2) $r(k-1) = \lambda(a-1)$; (3) $b \geq a$.

Source	DF	Sum of Squares
Treatments	$a-1$	$\sum_i \frac{y_{i.}^2}{b} - \frac{y_{..}^2}{N}$
Blocks	$b-1$	$\sum_j \frac{y_{.j}^2}{a} - \frac{y_{..}^2}{N}$
Error	$N-a-b+1$	$SS_{total} - SS_{trts} - SS_{blocks}$
Total	$N-1$	$\sum\sum y_{ij}^2 - \frac{y_{..}^2}{N}$

$E(MS_{trt}) = \sigma^2 + \frac{b\sum\tau_i^2}{a-1}$, $E(MS_{blk}) = \sigma^2 + \frac{a\sum\beta_j^2}{b-1}$, $E(MSE) = \sigma^2$
$F_0 = MS_{trt}/MSE$, p-value = $P(F_{a-1,(a-1)(b-1)} > F_0)$.
$Q_i = y_{i.} - \frac{1}{k}\sum_j n_{ij} y_{.j}$, $\hat{\tau}_i = \frac{kQ_i}{\lambda a}$, $\hat{\mu} = \frac{y_{..}}{N} = \frac{y_{..}}{bk}$, $LSMean(\mu_i) = \hat{\mu} + \hat{\tau}_i$

Factorial Designs

Model $Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}$, $i = 1, 2, ..., a$, $j = 1, 2, ..., b$, $k = 1, 2, ..., n$
$\sum \tau = 0$, $\sum \beta = 0$, $\sum_i (\tau\beta)_{ij} = 0$, $\sum_j (\tau\beta)_{ij} = 0$
$\hat{\mu} = \bar{y}_{...}$, $\hat{\tau}_i = \bar{y}_{i..} - \bar{y}_{...}$, $\hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...}$, $\hat{\tau\beta}_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$
$\mu_{i.} = \mu + \tau_i$ =mean of ith level of A; $\mu_{.j} = \mu + \beta_j$ =mean of jth level of B;
$\mu_{ij} = \mu + \tau_i + \beta_j + (\mu\beta)_{ij}$ =mean of $ijth$ treatment.
$E(\hat{\mu}) = E(\bar{Y}_{...}) = \mu$, $var(\hat{\mu}) = \frac{\sigma^2}{abn}$;

Source	DF	Sum of squares
Treatments	$ab-1$	$\frac{1}{n}\sum_i\sum_j y_{ij.}^2 - \frac{(y_{...})^2}{abn}$
A	$a-1$	$\frac{1}{bn}\sum_i y_{i..}^2 - \frac{(y_{...})^2}{abn}$
B	$b-1$	$\frac{1}{an}\sum_j y_{.j.}^2 - \frac{(y_{...})^2}{abn}$
AB	$(a-1)(b-1)$	$SS_T - SS_A - SS_B$
Error	$ab(n-1)$	$SS_T - SS_{trts} = \sum_i\sum_j\sum_k (y_{ijk} - \bar{y}_{ij.})^2$
Total	$abn-1$	$\sum_i\sum_j\sum_k y_{ijk}^2 - \frac{(y_{...})^2}{abn}$

Overall test: $\mu_{11}=\mu_{12}=...=\mu_{ab}$, test statistic $F_0=\frac{MS_{trt}}{MSE}$;
Interaction test: $(\alpha\beta)_{ij}=0$ for all ij, test statistic $F_0=\frac{MS_{AB}}{MSE}$.
sample size: $\delta^2=\frac{nb\Delta^2}{2\sigma^2}$, assuming 2 levels of A differ by Δ; $\delta^2=\frac{n\Delta^2}{2\sigma^2}$ from the interaction terms.

2^k Factorial Designs

Two-level Fractional Factorial Designs

Design resolution - A fractional factorial design's resolution is the length of the shortest word and its defining relation.
2^{k-p} terms, 2^p alias.

Random Effects and Mixed Models

Model $Y_{ij}=\mu+\tau_i+\epsilon_{ij}$, $i=1,2,...,a; j=1,2,...,n_i$, where τ_i are assumed to be independent $N(0,\sigma_\tau^2)$ random variables.
$H_0:\sigma_\tau^2=0$ vs. $H_a:\sigma_\tau^2>0$, test stat: $F_0=\frac{MS_{trt}}{MSE}$, $F_0\sim F_{a-1,N-a}$ under H_0.
Some facts: $Y_{ij}\sim N(\mu,\sigma_\tau^2+\sigma^2)$ (1) if $i\neq k$ - different treatment levels, $Cov(Y_{ij},Y_{kj})=0$ since τ_i and τ_k are independent and $E(\tau_i\tau_k)=E(\tau_i)E(\tau_k)=0$; (2) if $k\neq l$ - same treatment different obs, $Cov(Y_{ij},Y_{kj})=\sigma_\tau^2$.
two-way random model: $Cov(Y_{ijk},Y_{ijk'})=\sigma_\tau^2+\sigma_\beta^2+\sigma_{\tau\beta}^2$ if $k\neq k'$;
$Cov(Y_{ijk},Y_{ij'k})=\sigma_\tau^2$ if $j\neq j'$.
$E(MSE)=\sigma^2$, $E(MS_{trt})=\sigma^2+n_0\sigma_\tau^2$ where $n_0=n$ if all $n_i=n$ and $n_0=\frac{1}{a-1}[N-\frac{\sum n_i^2}{N}]$.
Estimates: $\hat\sigma^2=MSE$ and $\hat\sigma_\tau^2=\frac{MS_{trt}-MSE}{n_0}$.
Confidence interval for $\frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2}$: $\frac{MS_{trts}/(n\sigma_\tau^2+\sigma^2)}{MS_E/\sigma^2}\sim F_{a-1,N-a}$,
$(F_{1-\alpha/2,a-1,N-a}\leq\frac{MS_{trts}}{MS_E}\frac{\sigma^2}{n\sigma_\tau^2+\sigma^2}\leq F_{\alpha/2,a-1,N-a})=1-\alpha$, $P(L\leq\frac{\sigma_\tau^2}{\sigma^2}\leq U)=1-\alpha$,
$L=\frac{1}{n}(\frac{MS_{trts}}{MS_E}\frac{1}{F_{\alpha/2,a-1,N-a}}-1)$, $U=\frac{1}{n}(\frac{MS_{trts}}{MS_E}\frac{1}{F_{1-\alpha/2,a-1,N-a}}-1)$,
$\frac{L}{L+1}\leq\frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2}\leq\frac{U}{1+U}$
Two-factor factorial with random factors: $Y_{ijk}=\mu+\tau_i+\beta_j+(\tau\beta)_{ij}+\epsilon_{ijk}$, $i=1,2,...,a$, $j=1,2,...,b$, $k=1,2,...,n$, $V(\tau_i)=\sigma_\tau^2$, $V(\beta_j)=\sigma_\beta^2$, $V[(\tau\beta)_{ij}]=\sigma_{\tau\beta}^2$, and $V(\epsilon)=\sigma^2$.
Expected mean squares: $E(MS_A)=\sigma^2+n\sigma_{\tau\beta}^2+bn\sigma_\tau^2$;
$E(MS_B)=\sigma^2+n\sigma_{\tau\beta}^2+an\sigma_\beta^2$; $E(MS_{AB})=\sigma^2+n\sigma_{\tau\beta}^2$; $E(MSE)=\sigma^2$.
Two-factor mixed model: Factor A is fixed; factor B is random.
$Y_{ijk}=\mu+\tau_i+\beta_j+(\tau\beta)_{ij}+\epsilon_{ijk}$, where

- $i=1,2,...,a$, $j=1,2,...,b$, $k=1,2,...,n$;
- τ_i is a fixed effect with $\sum\tau_i=0$;
- $\beta_j\sim N(0,\sigma_\beta^2)$, $(\tau\beta)_{ij}\sim N(0,\sigma_{\tau\beta}^2)$, and $\epsilon_{ijk}\sim N(0,\sigma^2)$.

$Y_{ijk}\sim N(\mu+\tau_i,\sigma^2+\sigma_\beta^2+\sigma_{\tau\beta}^2)$.
Expected mean squares: $E(MSE)=\sigma^2$, $E(MS_A)=\sigma^2+n\sigma_{\tau\beta}^2+bn\frac{\sum\tau_i}{a-1}$,
$E(MS_B)=\sigma^2+n\sigma_{\tau\beta}^2+an\sigma_\beta^2$, $E(MS_{AB})=\sigma^2+n\sigma_{\tau\beta}^2$
Variance components estimates: $\hat\sigma^2=MSE$, $\hat\sigma_{\tau\beta}^2=\frac{MS_{AB}-MSE}{n}$, $\hat\sigma_\beta^2=\frac{MS_B-MS_{AB}}{an}$
Hypothesis Tests: (1) $H_0:\sigma_{\tau\beta}^2=0$ vs. $H_a:\sigma_{\tau\beta}^2>0$ using $F=\frac{MS_{AB}}{MSE}$; (2) $H_0:\sigma_\beta^2=0$ vs. $H_a:\sigma_\beta^2>0$ using $F=\frac{MS_B}{MS_{AB}}$; (3) $H_0:\tau_i=0$ vs. $H_a:not\ H_0$ using $F=\frac{MS_A}{MS_{AB}}$.

Approximate F-test: degree of freedom $\nu=\frac{(\sum c_iMS_i)^2}{\sum\frac{c_i^2MS_i^2}{\nu_i^2}}$

Nested Designs

Model $Y_{ijk}=\mu+\tau_i+\beta_{j(i)}+\epsilon_{ijk}$, $i=1,2,...,a; j=1,2,...,b, k=1,2,...,n$.
Crossed vs. nested factors: Two factors are considered crossed if every level of one factor occurs with every level of the other factor. In two factor design, one factor is nested with another when the levels of one factor are different within each level of the other factor.
If fixed: $\sum\tau_i=0$, $\sum\beta_{j(i)}=0$ for all i.
If random: $\beta_{j(i)}\sim N(0,\sigma_\beta^2)$, $\tau_i\sim N(0,\sigma_\tau^2)$, independent;
If mixed (τ is fixed, and β is random): $\sum\tau_i=0$, $\beta_{j(i)}\sim N(0,\sigma_\beta^2)$, independent.
$SSE=\sum_i\sum_j\sum_k(Y_{ijk}-\bar Y_{ij.})^2$ with $df=N-ab$
$SS_{B(A)}=n\sum_i\sum_j(\bar Y_{ij.}-\bar Y_{i..})^2$ with $df=a(b-1)$
A random, B(A) random: $Cov(Y_{ijk},Y_{mno})=\sigma_\beta^2+\sigma_\tau^2$ if $i=m, j=n, k\neq o$;
$Cov(Y_{ijk},Y_{mno})=\sigma_\tau^2$ if $i=m, j\neq n$; $Cov(Y_{ijk},Y_{mno})=0$ if $i\neq m$;
A fixed, B(A) random: $Cov(Y_{ijk},Y_{mno})=\sigma_\beta^2$ if $i=m, j=n$; 0 otherwise.

Generalized Linear Models

Textbook 1

An othogonal matrix $C_{k\times k}$ has the property $C'C=CC'=I$, i.e. $C'=C^{-1}$. The eigenvalues of $A_{k\times k}$ are the same as $C'AC$.
P and Q are nonsingular, then $rank(AQ)=rank(PA)=rank(A)$.
$A_{n\times n}$, symmetric, then $\mathbf{x}_i'\mathbf{x}_j=0$ for $i\neq j$. $P_{n\times n}$ nonsingular, then $Tr(P^{-1}AP)=Tr(A)$.
$A_{n\times n}$, symmetric, then A can be factorized as $A=P\Lambda P^{-1}$, where $\Lambda_{ii}=\lambda_i$, P is an orthogonal matrix, i.e. $PP'=I$.
$A_{n\times n}$, symmetric, idempotent, then $r(A)=tr(A)=r(P'AP)=tr(P'AP)$.
$z=a'Y$, $\frac{\partial z}{\partial Y}=a$; $z=Y'Y$, $\frac{\partial z}{\partial Y}=2Y$; $z=Y'AY$, $\frac{\partial z}{\partial Y}=AY+A'Y$.
$E(Y)=\mu$, $E(a'Y)=a'E(Y)=a'\mu$; $V(Y)=V$, $V(a'Y)=a'V(Y)a$,
$V(AY)=AV(Y)A'$.
$E(Y'AY)=tr(AV)+\mu'A\mu$.
If $Y_{k\times 1}\sim N(\mu,I)$, then $Y'Y\sim\chi^2_{k,\lambda=\frac{1}{2}(\mu'\mu)}$.
$Y_{n\times 1}\sim N(\mu,I)$, $A=A'$, then $Y'AY\sim\chi^2_{k,\lambda}$ with $k=r(A)$, $\lambda=\frac{1}{2}(\mu'A\mu)$ iff $A=A^2$.
$Y_{n\times 1}\sim N(\mu,\sigma^2I)$, $A=A'$, then $Y'AY\sim\chi^2_{k,\lambda}$ with $k=r(A)$, $\lambda=\frac{1}{2\sigma^2}(\mu'A\mu)$ iff $A=A^2$.
$Y_{n\times 1}\sim N(\mu,V)$, $A=A'$, then $Y'AY\sim\chi^2_{k,\lambda}$ with $k=r(AV)=r(A)$, $\lambda=\frac{1}{2}(\mu'A\mu)$ iff $AV=(AV)^2$.
$Y_{n\times 1}\sim N(\mu,V)$, then $Y'V^{-1}Y\sim\chi^2_{k,\lambda}$, with $k=n$, $\lambda=\frac{1}{2}(\mu'V^{-1}\mu)$.
$Y_{n\times 1}\sim N(\mu,V)$, then AY and BY are independent iff $AVB'=0$.
$Y_{n\times 1}\sim N(\mu,V)$, $A_{n\times n}=A'$, $B_{m\times n}$, then $Y'AY$ and BY are independent iff $BVA=0$.
$Y_{n\times 1}\sim N(\mu,V)$, $A_{n\times n}=A'$, $B_{n\times n}=B'$, then $Y'AY$ and $Y'BY$ are independent iff $AVB=0$.
$B=(X'X)^{-1}XY$, $\hat Y=XB=X(X'X)^{-1}XY=HY$, $E(B)=\beta$,
$var(B)=\sigma^2(X'X)^{-1}$, $E(\hat Y)=X\beta$, $var(\hat Y)=\sigma^2H$.
$SSE=Y'(I-H)Y$ with $df=n-p$, $SSR=Y'\left(H-\frac{1}{n}J\right)Y$ with $df=p-1$,
$SST=Y'\left(I-\frac{1}{n}\right)Y$ with $df=n-1$.
If $Y=X\beta+\epsilon$, $\epsilon\sim N(0,\sigma^2I)$, then $B=(X'X)^{-1}XY\sim N\left(\beta,\sigma^2(X'X)^{-1}\right)$.

$\frac{(n-p)s^2}{\sigma^2}=\frac{(n-p)MSE}{\sigma^2}=\frac{SSE}{\sigma^2}=\frac{1}{\sigma^2}Y'(I-H)Y\sim\chi^2_{n-p}$.
B and $\frac{SSE}{\sigma^2}$ are independent.
$\frac{b_j-\beta_j}{\sqrt{var(b_j)}}=\frac{b_j-\beta_j}{\sigma\sqrt{c_{jj}}}\sim N(0,1)$, c_{jj} is jth diag entry of $(X'X)^{-1}$.
$\frac{(b_j-\beta_j)/(\sigma\sqrt{c_{jj}})}{\sqrt{\frac{SSE}{\sigma^2}/(n-p)}}\sim t_{n-p}\Rightarrow b_j\pm t_{n-p}\sqrt{MSEc_{xx}}$
$LB\sim N\left(L\beta,\sigma^2L(X'X)^{-1}L'\right)$.
Let $M=(LB)'\left(\sigma^2\left(L(X'X)^{-1}L'\right)^{-1}\right)^{-1}(LB)\sim\chi^2_{r,\lambda}$, where
$\lambda=\frac{1}{2\sigma^2}(LB)'\left(L(X'X)^{-1}L'\right)^{-1}(LB)$.
$E(M)=r\sigma^2+(L\beta)'\left(L(X'X)^{-1}L'\right)^{-1}(L\beta)$
$F^*=\frac{(Lb)'\left(L(X'X)^{-1}L'\right)^{-1}(Lb)/r}{\frac{SSE}{\sigma^2}/(n-p)}=\frac{MSQ}{MSE}\sim F_{r,n-p}$ under $H_0:L\beta=0$.
$\frac{SSR}{\sigma^2}\sim\chi^2_{p,\lambda}$, where $\lambda=\frac{1}{2\sigma^2}\beta'(X'X)\beta$.
$MSQ(L\beta)=(LB)'\left(\sigma^2\left(L(X'X)^{-1}L'\right)^{-1}\right)^{-1}(LB)=\frac{SSR}{\sigma^2}$.
$A=X(X'X)^{-1}X'-X_2(X_2'X_2)^{-1}X_2'$ is idempotent; $r(A)=r$.

Textbook 2

standardized residual $r_i=\frac{y_i-\hat\mu_i}{\hat\sigma}$.
$f(y;\theta)=exp\{a(y)b(\theta)+c(\theta)+d(y)\}$
glm = exp family + link func (mono + diff)
$E[a(y)]=-c'(\theta)/b(\theta)$
$var[a(y)]=\frac{b''(\theta)c'(\theta)-c''(\theta)b'(\theta)}{[b'(\theta)]^3}$
score info: $U=\frac{\partial ln[f(\theta;y)]}{\partial\theta}=a(y)b'(\theta)+c'(\theta)$
sampling dist of score statistics: $E(U)=0$;
$J=var(U)=-E(U')=-E(\frac{\partial^2 lnf(x;\theta)}{\partial\theta^2})=\frac{b''(\theta)c'(\theta)}{b'(\theta)}-c''(\theta)$
One parameter: $\frac{U-0}{\sqrt{J}}\sim N(0,1)$, $U'J^{-1}U=\frac{U^2}{J}\sim\chi^2_{(1)}$; for a vector of parameters: $U\sim MVN(0,J)$, then $U'J^{-1}U\sim\chi^2_p$.
Approximate l and U with Taylor series:
$l(\beta)=l(b)+(\beta-b)'U(b)-1/2(\beta-b)'J(b)(\beta-b)$; $U(\beta)=U(b)-J(b)(\beta-b)$.
sampling dist of MLE's: $U(b)=0$, $J^{-1}U\sim N(0,J^{-1})$, $(b-\beta)'J(b)(b-\beta)\sim\chi^2_p$.
wald statistic $(b-\beta)'J(b)(b-\beta)\sim\chi^2(p)$
GLiM: (1) response follows independent exponential family dist; (2) predictors $X's$ available; (3) g(): monotone differentiable link function.
saturated model is a model with the maximum number of parameters that can be estimated. β_{max} is the parameter for saturated model, b_{max} is MLE for saturated model.
$\lambda=\frac{L(b_{max};y)}{L(b;y)}$, b_{max}/b, - MLE for saturated/reduced model
Deviance: $D=2[l(b_{max};y)-l(b;y)]$
$AIC=-2l(\hat\pi;y)+2p$; $BIC=-2l(\hat\pi;y)+2p\times$(#of obs)