P8110 Applied Regression II — Final Exam Reference

I. Survival Analysis

A. Core Functions & Properties

$S(t) = P(T > t) = \exp[-H(t)]$; $h(t) = \frac{f(t)}{S(t)} = -\frac{d \log S(t)}{dt}$; $H(t) = \int_0^t h(u)du = -\log S(t)$

Properties: $S(0) = 1$, $S(\infty) = 0$; $h(t) \geq 0$ always; $H(0) = 0$, $H(t)$ monotone non-decreasing

Censoring: Right-censoring (most common); observe (Y, δ) where $Y = \min(T, C)$, $\delta = I(T \leq C)$

Assumption: Non-informative censoring (censoring independent of failure time)

B. Kaplan-Meier Estimator

$\hat{S}(t) = \prod_{t_i \leq t}(1 - \frac{d_i}{n_i}) = \prod_{t_i \leq t} \frac{n_i - d_i}{n_i}$ where d_i=events, n_i=at risk just before t_i

KM Table (7 columns): j, t_j, interval, n_j, d_j, $p_j = 1 - d_j/n_j$, $\hat{S}(t_j)$

Key: Censored at t_j removed from risk set at t_j^+; censoring $\downarrow n_{j+1}$ but NOT $\uparrow d_j$; if max time censored, $\hat{S}(t)$ doesn't reach 0

Greenwood Variance: $\widehat{\text{Var}}(\hat{S}) = \hat{S}^2 \sum_{t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}$ (undefined when $n_i = d_i$)

95% CI: Log-log: $\exp(-e^{C_u}), \exp(-e^{C_l})$ where $\sigma^2 = \frac{1}{[\log \hat{S}]^2} \sum \frac{d_i}{n_i(n_i - d_i)}$; $C_u = \log[-\log \hat{S}] + 1.96\sigma$

Nelson-Aalen: $\hat{H}(t) = \sum_{t_i \leq t} \frac{d_i}{n_i}$; $\tilde{S}(t) = \exp[-\hat{H}(t)]$ (asymptotically $\equiv$ KM)

Quantiles: $\hat{t}_p = \min\{t_j : \hat{S}(t_j) < 1 - p\}$; if $\hat{S}(t_j) = 1 - p$ exactly, $\hat{t}_p = (t_j + t_{j+1})/2$

Median: $\hat{t}_{0.5} = \min\{t_j : \hat{S}(t_j) < 0.5\}$; Q1: $\hat{t}_{0.25} = \min\{t_j : \hat{S}(t_j) < 0.75\}$

Mean: $\hat{\mu} = \int_0^\infty \hat{S}(t)dt$ (underestimated if max censored)

Discrete: $\hat{\mu} = \sum_{j=1}^{J-1} \hat{S}(t_{j-1})(t_j - t_{j-1})$ where $t_0 = 0$

Variance (HW2): $Var(\hat{\mu}) = \frac{n_d}{n_d - 1} \sum_{j=1}^{J-1} \frac{A_j^2 d_j}{n_j(n_j - d_j)}$ where $A_j = \sum_{k=j+1}^{J} \hat{S}(t_{k-1})(t_k - t_{k-1})$, $n_d = \sum d_j$

RMST: $\hat{\mu}(\tau) = \int_0^\tau \hat{S}(t)dt$ (more robust; choose τ=max follow-up)

C. Survival Curve Comparison

Both Log-rank and Wilcoxon can be extended to K groups

$H_0 : S_1(t) = S_0(t)$ for all $t \leq \tau$

Weighted Test: $Q = \frac{[\sum w_j(d_{1j} - e_{1j})]^2}{\sum w_j^2 \nu_{1j}} \sim \chi_1^2$

$e_{1j} = \frac{n_{1j} d_j}{n_j}$; $\nu_{1j} = \frac{n_{1j} n_{0j} d_j (n_j - d_j)}{n_j^2 (n_j - 1)}$

- **Log-rank:** $w_j = 1$ (equal weights, optimal under PH)
- **Wilcoxon:** $w_j = n_j$ (early-time sensitive, since n_j larger early)
- **Curves cross $\Rightarrow$ both tests lose power**; use RMST/landmark
- K groups: df=$K - 1$

D. Cox Proportional Hazards

$h(t, \mathbf{x}) = h_0(t) \exp(\boldsymbol{\beta}^T \mathbf{x})$ **(no β_0!)**

$\log \frac{h(t,\mathbf{x})}{h_0(t)} = \boldsymbol{\beta}^T \mathbf{x}$; $S(t, \mathbf{x}) = [S_0(t)]^{\exp(\boldsymbol{\beta}^T \mathbf{x})}$

HR Interpretation:

- Binary (x=0 vs 1): HR $= e^\beta$
- Continuous (k-unit $\uparrow$): HR $= e^{k\beta}$
- Multi-class: HR$_j = e^{\beta_j}$ (vs ref); HR$(j : k) = e^{\beta_j - \beta_k}$
- **HR = instantaneous relative risk $\neq$ cumulative risk ratio**

95% CI: $\exp(\hat{\beta} \pm 1.96 \cdot SE)$

Partial Likelihood: $L_p(\boldsymbol{\beta}) = \prod_{j=1}^{m} \frac{\exp(\mathbf{x}_j^T \boldsymbol{\beta})}{\sum_{k \in R(t_j)} \exp(\mathbf{x}_k^T \boldsymbol{\beta})}$

m=distinct failure times; $R(t_j)$=risk set; $\log L_p = \sum[\mathbf{x}_j^T \boldsymbol{\beta} - \log \sum_{k \in R} \exp(\mathbf{x}_k^T \boldsymbol{\beta})]$

Ties: EFRON (recommended) approximates exact; BRESLOW faster but less accurate

PH Assumption Checks:

1. Log-log plot: $\log[-\log \hat{S}(t)]$ vs $\log t$ (parallel $\Rightarrow$ PH)
2. Schoenfeld residuals: `assess ph / resample;`
3. Time interaction: add $X \times \log(t)$; if sig $\Rightarrow$ violates PH

Non-PH Solutions:

- Stratify: `strata Z;` (cannot estimate HR for Z)
- Time-varying coef: $X \times t$ or $X \times \log(t)$
- TDC: `model (tstart,tstop)*status(0)=...;`

Residuals:

- **Martingale** $\in (-\infty, 1]$: $M_i = \delta_i - \hat{H}_i(t_i)$; check functional form via plot vs X+lowess; **horizontal lowess $\Rightarrow$ correct form**, non-horizontal $\Rightarrow$ need transformation
- **Deviance**: $D_i = \text{sign}(M_i)\sqrt{-2[M_i + \delta_i \log(\delta_i - M_i)]}$; $|D_i| > 2$ noteworthy, > 3 outlier
- **Dfbeta**: $dfbeta_i = \hat{\beta} - \hat{\beta}_{(-i)}$; $|dfbeta| > 2/\sqrt{n}$ influential

Confounding: $\Delta\hat{\beta}\% = 100 \times \frac{|\hat{\theta} - \hat{\beta}|}{|\hat{\beta}|}$; $> 20\% \Rightarrow$ confounder

Interaction: $h(t) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2)$; HR$_{X_1}(X_2 = a) = e^{\beta_1 + \beta_3 a}$

Variance (group comparison): $Var(a \times \hat{\beta}_j - b \times \hat{\beta}_k) = a^2 \times Var(\hat{\beta}_j) + b^2 \times Var(\hat{\beta}_k) - 2abCov(\hat{\beta}_j, \hat{\beta}_k)$

Hypothesis Tests:

- **Wald:** $(\hat{\beta}/SE)^2 \sim \chi_1^2$ (single param); Type 3 (multiple)
- **LRT:** $G = -2[\log L_{reduced} - \log L_{full}] \sim \chi_{df}^2$ (preferred for nested models)
- **Score:** Based on $U(\beta)$; used in PH test

E. Key Points

Semi-parametric: Cox has nonparametric $h_0(t)$ + parametric $\boldsymbol{\beta}$

Model comparison: LRT (Likelihood Ratio Test, preferred when nested), Wald, Score test; df=# parameters difference

Stratified analysis: Control confounders; assumes common treatment effect across strata

Baseline hazard: Cannot estimate $h_0(t)$ directly, but can get $S_0(t)$ via `baseline`

II. Categorical Data

A. GLM Framework

1. Distribution: $Y \sim$ Normal/Binomial/Poisson/NB
2. Linear predictor: $\eta = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$
3. Link: $g\{E(Y|X)\} = \eta$

Model	Dist	Link	$g(\mu)$
Linear	Normal	Identity	μ
Logistic	Binomial	Logit	$\log \frac{\mu}{1-\mu}$
Poisson	Poisson	Log	$\log \mu$
NB	NB	Log	$\log \mu$

Variance: Normal=σ^2; Binomial=$\mu(1 - \mu)$; Poisson=μ; NB=$\mu + \alpha\mu^2$

Canonical link: Logit for Binomial, Log for Poisson

Fitted values: Logistic: $\hat{p} = \frac{e^{\hat{\eta}}}{1+e^{\hat{\eta}}}$ where $\hat{\eta} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots$

Interpretation:

- Linear: "β_j" = unit change in mean
- Logistic: "e^{β_j}" = odds ratio (OR)
- Poisson/NB: "e^{β_j}" = rate ratio (RR)

B. Multinomial Logit (unordered)

$\log \frac{P(Y=j)}{P(Y=K)} = \alpha_j + \boldsymbol{\beta}_j^T \mathbf{X}$, $j = 1, \ldots, K - 1$

Expanded: $\log \frac{P(Y=j)}{P(Y=K)} = \alpha_j + \beta_{j1} X_1 + \beta_{j2} X_2 + \cdots + \beta_{jp} X_p$ (each class j has own $\boldsymbol{\beta}_j$)

Example (Y=3 levels): $\log \frac{P(Y=1)}{P(Y=3)} = \alpha_1 + \beta_{11} Age + \beta_{12} Sex$; $\log \frac{P(Y=2)}{P(Y=3)} = \alpha_2 + \beta_{21} Age + \beta_{22} Sex$

df: Continuous X: $K - 1$; q-level categorical X: $(K - 1)(q - 1)$

Interpretation:

- β_{jk}: change in log-odds of class j vs ref K per unit increase in X_k
- $e^{\beta_{jk}}$=RRR (relative risk ratio): multiplicative change in odds(Y=j)/odds(Y=K)
- Example: $\beta_{11} = 0.5 \Rightarrow$ 1-yr age increase multiplies odds(Y=1 vs Y=3) by $e^{0.5} = 1.65$
- Compare classes: $e^{\beta_{1k} - \beta_{2k}}$=RRR for class 1 vs class 2

Probability: $P(Y = j) = \frac{\exp(\alpha_j + \boldsymbol{\beta}_j^T \mathbf{X})}{1 + \sum_{k=1}^{K-1} \exp(\alpha_k + \boldsymbol{\beta}_k^T \mathbf{X})}$; $P(Y = K) = \frac{1}{1 + \sum_{k=1}^{K-1} \exp(\alpha_k + \boldsymbol{\beta}_k^T \mathbf{X})}$

C. Ordinal (Cumulative)Logit (ordered)

$\text{logit}[P(Y \leq j)] = \log \left[\frac{P(Y \leq j)}{P(Y > j)}\right] = \alpha_j + \boldsymbol{\beta}^T \mathbf{X}$, $j = 1, \ldots, K - 1$

Expanded: $\text{logit}[P(Y \leq j)] = \alpha_j + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$ (same $\boldsymbol{\beta}$ for all j)

Example (Y=4 levels):

- $\log \frac{P(Y \leq 1)}{P(Y > 1)} = \alpha_1 + \beta_1 Age + \beta_2 Sex$
- $\log \frac{P(Y \leq 2)}{P(Y > 2)} = \alpha_2 + \beta_1 Age + \beta_2 Sex$ (note: same β_1, β_2!)
- $\log \frac{P(Y \leq 3)}{P(Y > 3)} = \alpha_3 + \beta_1 Age + \beta_2 Sex$

Proportional Odds: same $\boldsymbol{\beta}$ for all cutpoints; only α_j differs

df: Continuous X: 1; q-level X: $q - 1$ (only ONE $\boldsymbol{\beta}$ across all equations!)

Proportional Odds Assumption Test (HW6):

1. **Hypotheses:** H_0: slopes $(\beta_1, \beta_2, \ldots)$ equal across logit equations; H_a: different slopes needed
2. **Test:** Score chi-square test; df=$(K-2)(q-1)$ where K=response levels, q=predictor levels
3. Example: Y=4 levels, X 4-group $\Rightarrow$ df=$(4 - 2) \times (4 - 1) = 6$
4. **P-value:** $P(\chi_{df}^2 \geq S)$ where S=test statistic
5. **Decision:** $p > 0.05 \Rightarrow$ fail to reject $H_0 \Rightarrow$ use Ordinal; $p < 0.05 \Rightarrow$ reject $H_0 \Rightarrow$ use Multinomial

Interpretation:

- β_k: change in log-cumulative-odds per unit increase in X_k (**same effect for all cutpoints**)
- e^{β_k}=cumulative OR: multiplicative change in odds(Y≤j) for all j

- Default $P(Y \le j)$: $\beta > 0 \Rightarrow$ higher odds of $Y \le j \Rightarrow$ tend to lower levels
- Example: $\beta_1 = 0.3 \Rightarrow$ 1-yr age increase multiplies odds(Y≤any cutpoint) by $e^{0.3} = 1.35$
- Key: **ONE** β per predictor applies to **ALL** $(K-1)$ equations; only intercepts α_j vary

Hypothesis Test for β_k:

- $H_0 : \beta_k = 0$ vs $H_a : \beta_k \ne 0$ (no effect of X_k on cumulative odds)
- Wald test: $z = \frac{\hat\beta_k}{SE(\hat\beta_k)} \sim N(0,1)$; $p = 2P(|Z| > |z|)$
- $p < 0.05 \Rightarrow$ reject $H_0 \Rightarrow X_k$ significantly affects ordered outcome

Probability Calculation (HW6):

Ordinal: $P(Y \le j|X) = \frac{e^{\alpha_j + \beta X}}{1+e^{\alpha_j+\beta X}}$; then $P(Y=j) = P(Y \le j) - P(Y \le j-1)$; $P(Y=K) = 1 - P(Y \le K-1)$

Multinomial: $P(Y=j|X) = \frac{e^{\beta_{0j}+\beta_j X}}{1+\sum_{k=1}^{K-1} e^{\beta_{0k}+\beta_k X}}$; $P(Y=K) = \frac{1}{1+\sum_{k=1}^{K-1} e^{\beta_{0k}+\beta_k X}}$

D. Poisson Regression

$\log\{E(Y|X)\} = \beta_0 + \boldsymbol\beta^T\mathbf{X}$

Rate model: $\log\{E(Y|X)\} = \log(n) + \beta_0 + \boldsymbol\beta^T\mathbf{X}$

$\log(n)$=**offset** (known, no estimate); n=exposure (person-years, months, etc.)

Rate Ratio Derivation (HW7): Binary X: $RR = \frac{E(Y|X=1)}{E(Y|X=0)} = \frac{\exp(\beta_0+\beta_1\cdot 1+\cdots)}{\exp(\beta_0+\beta_1\cdot 0+\cdots)} = e^{\beta_1}$

Interpretation: Binary: $RR = e^{\beta}$; Continuous k-unit: $RR = e^{k\beta}$

Goodness Of Fit: Deviance$=-2(\log L_{current} - \log L_{saturated})$; Pearson $\chi^2 = \sum \frac{(y_i - \hat\mu_i)^2}{\hat\mu_i}$

Overdispersion Check: Scale=Deviance/df or Pearson χ^2/df; $\approx 1 \Rightarrow$ ok; $> 1 \Rightarrow$ mild; $> 1.5 \Rightarrow$ serious

Pearson residual: $r_i = \frac{y_i - \hat y_i}{\sqrt{\hat y_i}}$; $|r_i| > 2 \Rightarrow$ poor fit

Overdispersion Solutions:

- **Quasi-Poisson:** `scale=pearson` or `scale=deviance`; $\hat\beta$ unchanged, $SE_{new} = SE_{old} \times \sqrt{\phi}$ where ϕ=Scale; CI_{new} wider
- **NB:** `dist=negbin`; $Var(Y) = \mu + \alpha\mu^2$; when $\alpha \to 0 \Rightarrow$ Poisson
- **ZIP/ZINB:** Excess zeros; two-part: Logistic (0 vs > 0) + Poisson/NB (counts)

NB vs Poisson Test (HW7):

1. **Hypotheses:** $H_0 : \alpha = 0$ vs $H_a : \alpha \ne 0$ (no overdispersion vs overdispersion exists)
2. **Test:** t-test for α parameter (in SAS output: "Dispersion" parameter)
3. **From SAS:** Find "Analysis Of Maximum Likelihood Parameter Estimates" table; look for "Dispersion" row
4. Value: estimate of α (e.g., $\hat\alpha = 10.10$); Std Error; Wald 95% CI
5. **P-value:** from "Pr > ChiSq" column (e.g., $p < 0.0001$)
6. **Decision:** $p < 0.05 \Rightarrow$ reject $H_0 \Rightarrow$ NB preferred; $p > 0.05 \Rightarrow$ Poisson adequate
7. Example: $\hat\alpha = 10.10$, $p < 0.0001 \Rightarrow$ reject H_0 at 95% level $\Rightarrow$ NB more appropriate

Rate Ratio Estimation (HW7):

- **Point estimate:** $\widehat{RR} = \exp(\hat\beta)$ (same formula for both Poisson and NB)
- **95% CI:** $\exp(\hat\beta \pm 1.96 \cdot SE)$ where SE from respective model
- **Compare models:** NB typically gives wider CI (larger SE) than Poisson due to accounting for overdispersion
- Example: Male vs Female absence rate ratio
  - Poisson: $\widehat{RR} = e^{0.321} = 1.38$, 95% CI: $(1.25, 1.52)$
  - NB: $\widehat{RR} = e^{0.321} = 1.38$, 95% CI: $(1.15, 1.65)$ (wider!)
- **Conclusion change?** If Poisson CI excludes 1 but NB CI includes 1 $\Rightarrow$ significance changes (Poisson overconfident)

E. Key Concepts & Formulas

Multinomial vs Ordinal:

- Multinomial: unordered, $(K-1)$ equations, each with own $\boldsymbol\beta_j$
- Ordinal: ordered, PO $\Rightarrow$ same $\boldsymbol\beta$ for all cutpoints
- PO test $p < 0.05 \Rightarrow$ reject PO $\Rightarrow$ use Multinomial

Parameter & df count:

- Multinomial: cont X: $K-1$ param; q-level X: $(K-1)(q-1)$ param
- Ordinal: cont X: 1 param; q-level X: $q-1$ param
- PO test df: $(K-2)(q-1)$

CI for linear combination: $\hat\beta_1 - \hat\beta_2$: $SE = \sqrt{Var(\hat\beta_1) + Var(\hat\beta_2) - 2Cov(\hat\beta_1, \hat\beta_2)}$

Model selection: LRT for nested models (df=# param diff); AIC/BIC for non-nested; QIC for GEE

III. Longitudinal Data

A. GEE (population-averaged)

$g\{E[Y_{ij}|X_{ij}]\} = \mathbf{X}_{ij}^T\boldsymbol\beta$

Working Correlation:

Type	Matrix	Param
Independence	$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$	0
Compound Symmetry	$\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}$	1
AR(1)	$\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}$ $\rho^{\lvert j-k\rvert}$	1
Unstructured	$\begin{pmatrix}1 & \rho_{12}\\ \rho_{21} & 1\end{pmatrix}$	$k(k-1)/2$

SE: Robust/Sandwich (default, preferred); Model-based (if corr correct)

Selection: QIC (smaller better); cannot use AIC/BIC

Interpretation: Marginal — "at population level, treatment increases odds by e^{β}"

Key: Even if working correlation misspecified, $\hat\beta$ consistent with robust SE; efficiency loss only

B. Model Selection & Strategy

Correlation selection: Try multiple structures (IND, EXCH, AR(1), UN); compare using QIC (smaller is better)

- **EXCH:** All correlations equal; good for repeated measures with no time order
- **AR(1):** Correlation decays with time distance; good for equally spaced time series
- **UN:** Most flexible but many parameters; may not converge with many timepoints
- **IND:** No correlation; baseline comparison

Convergence: UN may fail with many time points; use simpler structures (EXCH, AR)

Time as Categorical (HW8): $E[Y_{ij}] = \beta_0 + \beta_1 X_{t_2} + \beta_2 X_{t_4} + \beta_3 X_{t_8} + \beta_4 trt + \beta_5(X_{t_2} \times trt) + \beta_6(X_{t_4} \times trt) + \beta_7(X_{t_8} \times trt)$ where X_{t_j}=dummy for time j

Interaction Test (HW8): $H_0 : \beta_5 = \beta_6 = \beta_7 = 0$ vs H_a: not H_0; Type 3 Wald test; df=# interaction terms

Interaction in GEE (time continuous): $E[Y_{ij}] = \beta_0 + \beta_1 time + \beta_2 trt + \beta_3(time \times trt)$

- β_3: difference in time trend between treatment groups
- Treatment A change: β_1; Treatment B change: $\beta_1 + \beta_3$; Difference: β_3

Interpretation guidelines:

- Marginal interpretation: "at population level, treatment increases odds by e^{β}"
- Always report with 95% CI
- Robust SE preferred (default in GEE)

C. Step-by-Step Calculation Guide

1. KM Table Construction (HW1):

1. List all unique event times (exclude censored-only times)
2. For each t_j: count n_j (at risk before t_j), d_j (events at t_j)
3. Censored at t: include in n_j for $t_j \le t$, exclude from n_{j+1} if $t_j = t$
4. Compute $p_j = 1 - d_j/n_j$, then $\hat S(t_j) = \prod_{i \le j} p_i$

2. Interaction HR with CI (HW5 - Complete Steps):

1. Model: $h = h_0 \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2)$; Compare: $X_1 = 0$ vs $X_1 = 1$ at $X_2 = a$
2. **Point estimate:** $HR = \frac{h(X_1=0, X_2=a)}{h(X_1=1, X_2=a)} = \exp(-\beta_1 - \beta_3 a)$; plug in $\hat\beta$'s
3. **Variance:** $Var(-\hat\beta_1 - \hat\beta_3 a) = Var(\hat\beta_1) + a^2 Var(\hat\beta_3) + 2a \cdot Cov(\hat\beta_1, \hat\beta_3)$
4. Get $Var(\hat\beta_1)$, $Var(\hat\beta_3)$, $Cov(\hat\beta_1, \hat\beta_3)$ from CovB matrix; compute variance
5. $SE = \sqrt{Var}$
6. 95% CI: $\exp[(-\hat\beta_1 - \hat\beta_3 a) \pm 1.96 \cdot SE]$

Example (HW5): $HR = \exp(-1.09033 - 0.40491) = 0.224$; $Var = 0.2302 + 0.4539 + 2(-0.227) = 0.2301$; $SE = 0.4797$; CI=$[0.088, 0.574]$

3. Ordinal Probability (HW6):

1. Get $\hat\alpha_j$ and $\hat\beta$ from output
2. Compute $P(Y \le j) = \frac{\exp(\hat\alpha_j + \hat\beta X)}{1+\exp(\hat\alpha_j + \hat\beta X)}$ for $j = 1, \ldots, K-1$
3. $P(Y=1) = P(Y \le 1)$; $P(Y=j) = P(Y \le j) - P(Y \le j-1)$ for $j \ge 2$
4. $P(Y=K) = 1 - P(Y \le K-1)$

4. Quasi-Poisson CI Adjustment (HW7):

1. Get $\hat\beta$ and SE_{old} from Poisson output
2. Get Scale=ϕ (Deviance/df or Pearson χ^2/df)
3. $SE_{new} = SE_{old} \times \sqrt{\phi}$
4. New 95% CI: $\exp(\hat\beta \pm 1.96 \cdot SE_{new})$

5. Mean Variance Calculation (HW2):

1. Compute $\hat\mu$ using discrete formula
2. For each j, calculate $A_j = \sum_{k=j+1}^{J} \hat S(t_{k-1})(t_k - t_{k-1})$ (remaining area)
3. Compute $Var(\hat\mu) = \frac{n_d}{n_d - 1}\sum_{j=1}^{J-1} \frac{A_j^2 d_j}{n_j(n_j - d_j)}$ where $n_d = \sum d_j$
4. $SE = \sqrt{Var}$; 95% CI: $\hat\mu \pm 1.96 \cdot SE$

6. GEE Mean Change (HW8 - Categorical Time):

1. Model: $E[Y] = \beta_0 + \beta_1 X_{t_2} + \beta_2 X_{t_4} + \beta_3 X_{t_8} + \beta_4 trt + \beta_5(X_{t_2} \times trt) + \beta_6(X_{t_4} \times trt) + \beta_7(X_{t_8} \times trt)$
2. **Trt A:** baseline=β_0; at time 2=$\beta_0 + \beta_1$; change=β_1
3. **Trt B:** baseline=$\beta_0 + \beta_4$; at time 2=$\beta_0 + \beta_1 + \beta_4 + \beta_5$; change=$\beta_1 + \beta_5$
4. **Difference in change:** $(\beta_1 + \beta_5) - \beta_1 = \beta_5$ (this is the interaction coefficient!)
5. Interpretation: β_5=mean difference of temperature change from baseline to 2hrs between two treatments

7. Log-log CI for S(t) (HW1):

1. Compute $Var[\log\{-\log(\hat S)\}] = \frac{1}{[\log(\hat S)]^2}\sum_{t_j \le t} \frac{d_j}{n_j(n_j - d_j)}$
2. $SE = \sqrt{Var}$; compute limits: $L = \log[-\log(\hat S)] - 1.96 \cdot SE$, $U = \log[-\log(\hat S)] + 1.96 \cdot SE$
3. 95% CI for $S(t)$: $[\exp(-e^U), \exp(-e^L)]$

Typical Exam Question Types (Based on HW1-8)

I. Survival Analysis

- **KM table:** Construct 7-column table; handle censoring ($\downarrow n_{j+1}$, NOT $\uparrow d_j$)
- **Median & mean:** Find $\min\{t_j : \hat{S}(t_j) < 0.5\}$; mean area under curve
- **Log-rank vs Wilcoxon:** Compare test stats; explain why differ (early weight)
- **Interaction HR:** $\log(HR) = \beta_1 + \beta_3 x_2$; $SE^2 = Var(\beta_1) + x_2^2 Var(\beta_3) + 2x_2 Cov(\beta_1, \beta_3)$
- **Survival prediction:** $S(t|X) = [S_0(t)]^{\exp(\beta^T X)}$ using baseline
- **Group comparison:** $Var(\beta_j - \beta_k) = Var(\beta_j) + Var(\beta_k) - 2Cov(\beta_j, \beta_k)$

II. Categorical Data

- **PO test:** Check p-value; $p < 0.05 \Rightarrow$ use Multinomial; $p > 0.05 \Rightarrow$ use Ordinal
- **Probability calc:** Ordinal: cumulative then subtract; Multinomial: softmax formula
- **OR interpretation:** Ordinal: cumulative OR (same for all cutpoints); Multinomial: RRR vs ref
- **df calculation:** Ordinal cont X: df=1; Ordinal q-level X: df=$q-1$; Multinomial: df=$(K-1)(q-1)$
- **Overdispersion check:** Look at Scale (Dev/df or χ^2/df); if $> 1.5 \Rightarrow$ serious problem
- **Quasi-Poisson:** $\hat{\beta}$ stays same, $SE_{new} = SE_{old} \times \sqrt{\phi}$; recalculate CI
- **NB vs Poisson:** Test $H_0 : \alpha = 0$; compare point estimates and CI width
- **RR interpretation:** e^{β} with 95% CI; state "adjusting for other variables"
- **Conclusion change:** NB gives wider CI $\Rightarrow$ may lose significance

III. Longitudinal Data

- **Correlation structure:** Try EXCH, AR(1), UN; compare QIC (smaller better)
- **Time as categorical:** Multiple dummy variables; different β for each time
- **Interaction** $time \times trt$**:** β_3=difference in time trend between groups
- **Mean change calc:** Trt A: β_1 (time main); Trt B: $\beta_1 + \beta_3$; Diff: β_3
- **Hypothesis test:** Test $H_0 : \beta_3 = 0$ (or all interaction terms=0); use Type 3 Wald; df=# interaction terms
- **Robust SE:** Always use (default); model-based SE assumes correct correlation