

```

/*                                                    */
/* question 1                                         */
/*                                                    */
DATA q1_data;
    INPUT is_treated time status;
    DATALINES;
        0 2 1
        0 5 1
        0 7 1
        0 9 1
        0 11 1
        0 12 0
        0 13 1
        0 13 1
        0 17 1
        0 19 1
        0 19 1
        0 20 1
        0 22 1
        1 4 1
        1 9 1
        1 9 1
        1 9 1
        1 13 1
        1 14 1
        1 14 1
        1 18 1
        1 18 1
        1 21 0
        1 23 0
        1 26 0
        1 26 1
        1 28 1
        1 30 1
        1 34 0
        1 35 0
    ;
RUN;

/* kaplan-meier */
PROC LIFETEST DATA=q1_data PLOTS=survival(test atrisk) METHOD=km;
    STRATA is_treated / TEST = logrank;
    TIME time * status(0);
RUN;

```

```

/*                                                    */
/* question 2                                         */
/*                                                    */
DATA q2_data_raw;
    INPUT pair r_status t1 t2 did_relapse;
    DATALINES;
        1 1 1 10 1
        2 2 22 7 1
        3 2 3 32 0
        4 2 12 23 1
        5 2 8 22 1
        6 1 17 6 1
        7 2 2 16 1
        8 2 11 34 0
        9 2 8 32 0
        10 2 12 25 0
        11 2 2 11 0
        12 1 5 20 0
        13 2 4 19 0
        14 2 15 6 1
        15 2 8 17 0
        16 1 23 35 0
        17 1 5 6 1
        18 2 11 13 1
        19 2 4 9 0
        20 2 1 6 0
        21 2 8 10 0
    ;
RUN;

DATA temp (drop = t1 t2);
    SET q2_data_raw;
    ARRAY surv_time_arr[2] t1-t2;
    DO treat_num = 1 to 2;
        surv_time = surv_time_arr[treat_num];
        OUTPUT;
    END;
RUN;

DATA q2_data;
    SET temp;
    LENGTH treatment $ 10;
    IF treat_num = 1 THEN DO;
        did_relapse = 1;
        treatment = 'Placebo';
    END;
    ELSE treatment = '6-MP';
    DROP pair r_status;
RUN;

```

```

/* part a - proportional hazards */
PROC PHREG DATA = q2_data;
    CLASS treatment;
    MODEL surv_time*did_relapse(0) = treatment;
RUN;

/* part b - weibull proportional hazards */
PROC LIFEREG DATA = q2_data;
    CLASS treatment;
    MODEL surv_time*did_relapse(0) = treatment /
        COVB DIST = weibull;
RUN;

DATA temp;
    LENGTH statistic $ 10;
    INPUT statistic value;
    DATALINES;
        estimate 1.2673
    ;
RUN;

%LET scale = 0.7322;
%LET log_variance = 0.3106**2 + 0.1078**2 - 2*(0.011515);

DATA haz_vals;
    SET temp;
    beta = -value / &scale;
    haz_ratio = exp(-value / &scale);
    conf_low = haz_ratio - exp(1.96*sqrt(&log_variance));
    conf_high = haz_ratio + exp(1.96*sqrt(&log_variance));
RUN;

/* get medians for each group */
PROC SORT DATA = q2_data;
    BY treat_num;
RUN;

PROC LIFEREG DATA = q2_data;
    CLASS treat_num;
    MODEL surv_time*did_relapse(0) = treat_num /
        COVB DIST = weibull;
    BY treat_num;
RUN;

DATA temp;
    LENGTH treatment $ 10;
    INPUT treatment intercept shape scale;

```

```

        DATALINES;
            Placebo 2.2494 1.3705 0.7297
            6-MP 3.5194 1.3537 0.7387
        ;
RUN;

DATA median_vals;
    SET temp;
    lambda = exp(-intercept/scale);
    median = ((1/lambda)*log(100/(100-50)))*(1/shape);
RUN;

/* plots */
PROC LIFETEST DATA=q2_data PLOTS=(lls) METHOD=pl;
    TIME surv_time*did_relapse(0);
    STRATA treatment;
RUN;

/* part c - log logistic AFT */
PROC SORT DATA = q2_data;
    BY DESCENDING treat_num;
RUN;

PROC LIFEREG DATA=q2_data ORDER=DATA;
    CLASS treat_num;
    MODEL surv_time*did_relapse(0) = treat_num / DIST = llogistic;
RUN;

/* get acceleration factor */
DATA temp;
    INPUT value;
    DATALINES;
        1.2655
    ;
RUN;

DATA accel;
    SET temp;
    rel_accel = exp(-value);
    reduction = 1/rel_accel;
RUN;

/*
/* question 3
/*

```

```

/*                                                    */
DATA q3_data;
    INFILE "/home/u63563888/435/final/bducks.dat";
    INPUT duck time status age weight length;
    IF _N_ = 1 THEN delete;
RUN;

/* part a – proportional hazards */
PROC PHREG DATA = q3_data;
    CLASS age;
    MODEL time*status(0) = age weight length;
    OUTPUT OUT = res_data
            RESMART = val_mart
            RESSCH = val_sch_age val_sch_weight
val_sch_length;
RUN;

/* part b – martingale residuals */
PROC SGPLOT DATA = res_data;
    LOESS Y=val_mart X=weight;
    TITLE "Martingale Residual Plot for Weight";
RUN;

PROC SGPLOT DATA = res_data;
    LOESS Y=val_mart X=length;
    TITLE "Martingale Residual Plot for Length";
RUN;

/* part c – see if PH holds */
PROC LIFETEST DATA = q3_data METHOD = km OUTSURV = sch_age;
    TIME time*status(0);
    TEST age weight length;
RUN;

DATA sch_vals;
    SET sch_age(KEEP = time survival);
    logT = log(time);
    logLogSurv = log(-log(survival));
RUN;

PROC SGPLOT DATA = sch_vals;
    SERIES x=time y=logT;
    SERIES x=time y=logLogSurv;
    TITLE "Schoenfeld Residual Plot for Age";
    YAXIS LABEL="Value";
RUN;

```

```

PROC SGPLOT DATA = res_data;
    SCATTER Y=val_sch_weight X=time;
    REFLINE 0 / AXIS = y LINEATTRS = (COLOR = red PATTERN = dot);
    TITLE "Schoenfeld Residual Plot for Weight";
RUN;

PROC SGPLOT DATA = res_data;
    SCATTER Y=val_sch_length X=time;
    REFLINE 0 / AXIS = y LINEATTRS = (COLOR = red PATTERN = dot);
    TITLE "Schoenfeld Residual Plot for Length";
RUN;

/* part d - final model */
PROC PHREG DATA = q3_data;
    CLASS age;
    MODEL time*status(0) = age;
    OUTPUT OUT = phreg_out DFBETA = dfbAge LMAX = lmax;
RUN;

PROC RANK DATA = phreg_out
            OUT = ranked_data;
    VAR time;
    RANKS surv_rank;
RUN;

PROC GGPLOT DATA = ranked_data;
    TITLE "L-Max Plots";
    BUBBLE lmax*surv_rank=duck / BLABEL BSIZE=1 BCOLOR=bib;
    BUBBLE lmax*age=duck / BLABEL BSIZE=1 BCOLOR=bib;
RUN;

PROC GGPLOT DATA = ranked_data;
    TITLE "Delta-Beta Index Plots by Covariate";
    BUBBLE dfbAge*surv_rank=duck / BLABEL BSIZE=1 BCOLOR=bib;
RUN;

```