Lecture 7

Monday, October 30, 2023 10:03

BIOS522_Slides7

EMORY ROLLINS SCHOOL OF PUBLIC HEALTH | Department of Biostatistics and Bioinformatics

BIOS 522: Survival Analysis Methods

Lecture 7:

Cox model diagnostics

Why do we need model diagnostics?

- **Identify outliers**
  - Check data quality
  - Test robustness of the model
- **Check functional form of the covariates**
  - Properly characterize the covariate effect(s) (e.g. linear log hazard ratio)
  - Reveal relationships that were otherwise missed
  - Optimize model fit
- **Check proportional hazards assumption**
  - Ensure statistical inference is valid
  - Properly characterize the covariate effect(s)

2

Residuals and model diagnostics

- **Residuals**
  - Cox-Snell
  - Martingale
  - Deviance
  - Schoenfeld, scaled Schoenfeld
  - Others, like DFBeta – *not discussed*
- **Plots**
  - Kaplan-Meier
  - Log-log survival
- **Hypothesis tests**
  - Grambsch-Therneau
  - Covariate-by-time interaction

3

Example: Multiple myeloma

- Consider data on the survival of **48 patients with multiple myeloma** – a malignant disease characterized by the accumulation of abnormal white blood cells in the bone marrow.
- The primary endpoint was the time in months from **diagnosis until death** from multiple myeloma.

4

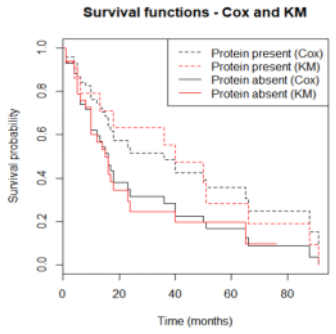
Example: Multiple myeloma

- At the time of diagnosis, **data were collected** on:
  - Patient age
  - Sex (1=male, 2=female)
  - Levels of blood urea nitrogen (Bun)
  - Levels of serum calcium (Ca)
  - Levels of hemoglobin (Hb)
  - The percentage of plasma cells in the bone marrow (Pcells)
  - An indicator variable (Protein) that denotes whether or not the Bence-Jones protein was present in the urine (0=absent, 1=present).
- The main purpose of the analysis was to investigate of **the effect of the risk factors** (Bun, Ca, Hb, Pcells, and protein) **on survival time**. The effects of these risk factors may be modified by age and sex.

5

Kaplan-Meier plot

- **X-axis**: time
- **Y-axis**: Kaplan-Meier curve for each group, plotted on top of Cox predicted survival for each group
- **Expected shape**: The KM and Cox survival curves are similar within group
- **Purpose**: Assess overall model fit

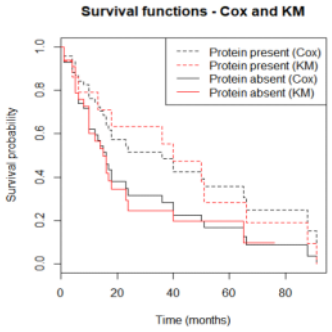


6

Limited in its applications

Example: Multiple myeloma

- We see relatively good agreement between the Cox and Kaplan-Meier curves, although there is some divergence in the protein present group between 20 and 40 months.



- Continuous variables must be broken into groups → which means we need a more refined approach

7

Residuals for survival data

- Consider residuals for linear regression models:

observed - predicted

$$r_i = Y_i - \hat{Y}_i$$

- Instead of Y_i, we have T_i^*. But what if T_i^* is a censoring time?

Time
↳ right-censored

8

Cox-Snell (generalized) residuals

- If survival time T_i has cumulative hazard function $H_i(t)$, then $H_i(T_i)$ is exponentially distributed with rate $\lambda = 1$

- The generalized or Cox-Snell residual is:

$$r_{Ci} = \tilde{H}_i(T_i^*)$$

9

Example: Multiple myeloma

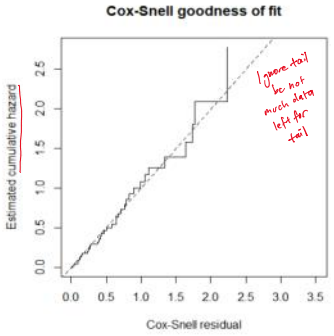
- Consider a model with a single covariate for blood urea nitrogen with fitted coefficient $\hat{\beta} = 0.02$.
- Patient 1:
  - Blood urea nitrogen of 25 mg/dL
  - This patient failed at 13 months ($T_1^* = 13, \delta_1 = 1$)
  - This patient's estimated cumulative hazard at 13 months is 0.434
- Patient 2:
  - Blood urea nitrogen of 13 mg/dL
  - This patient was censored at 52 months ($T_2^* = 52, \delta_2 = 0$)
  - This patient's estimate cumulative hazard at 52 months is 1.120

Patient (i)	Time (T_i^*)	Status (δ_i)	BUN ($X_{i,Bun}$)	$r_{Ci} = \hat{H}_i(T_i^*)$
1	13	1	25	0.434
2	52	0	13	1.120

Cox-Snell residual plot

→ Becomes basis for other residuals

- **X-axis**: Cox-Snell residual
- **Y-axis**: Cumulative hazard of Cox-Snell residuals
- **Expected shape**: Straight line with slope 1 that passes through the origin
- **Purpose**: Assess overall model fit



Martingale residuals

- r_{Ci} is the expected number of failures based on the fitted model

observed # failures

$$r_{Mi} = \delta_i - r_{Ci}$$

expected # failures

- The difference between the observed number of failures (δ_i =0 or 1) for subject i between time 0 and T_i^* and the expected number based on the fitted model
- Can take values between $-\infty$ and 1, and have mean 0

Example: Multiple myeloma

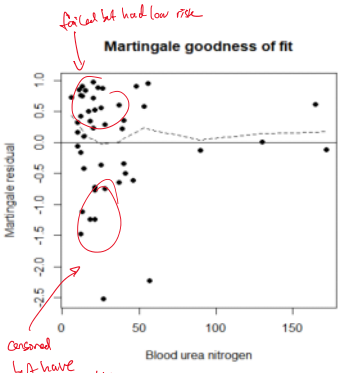
- Patient 1:
  - Blood urea nitrogen of 25 mg/dL
  - This patient failed at 13 months ($T_1^* = 13, \delta_1 = 1$)
  - This patient's estimated cumulative hazard at 13 months is 0.434
- Patient 2:
  - Blood urea nitrogen of 13 mg/dL
  - This patient was censored at 52 months ($T_2^* = 52, \delta_2 = 0$)
  - This patient's estimate cumulative hazard at 52 months is 1.120

Patient (i)	Time (T_i^*)	Status (δ_i)	BUN ($X_{i,Bun}$)	$r_{Ci} = \hat{H}_i(T_i^*)$	$r_{Mi} = \delta_i - r_{Ci}$
1	13	1	25	0.434	0.566
2	52	0	13	1.120	-1.120

13

Martingale residual plot – version 1

- **X-axis**: Covariate X_i or linear predictor $(\beta_1 X_{i1} + \cdots + \beta_k X_{ik})$
- **Y-axis**: Martingale residual for model with all covariates
- **Expected shape**: Random scatter around mean 0, may be asymmetric around 0
- **Purpose**: Assess functional form of the covariates, assess overall model fit when using linear predictor



14

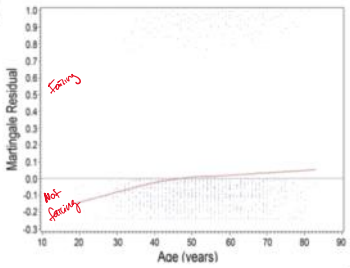
Martingale residual plot – version 2

- **X-axis**: Covariate X_i
- **Y-axis**: Martingale residuals for a model with all covariates *except* X_i
- **Expected shape**: Functional form of X_i
- **Purpose**: Reveal the appropriate functional form (a straight line indicates no transformation needed)

Cai M, Wei J, Zhang Z, Zhao H, Qiu Y, Fang Y, et al. (2012) Impact of Age on the Cancer-Specific Survival of Patients with Localized Renal Cell Carcinoma: Martingale Residual and Competing Risks Analysis. PLoS ONE 7(10): e48489. https://doi.org/10.1371/journal.pone.0048489

Goal: To select the optimal age cutpoint that would that would maximize the predictive value of age on the CSS of localized RCC

A



15

Deviance residuals

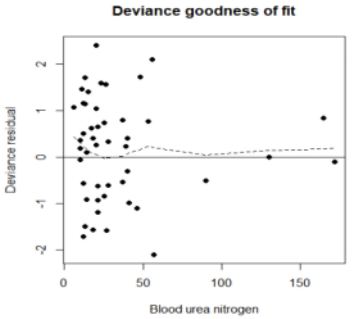
$$r_{Di} = sign(r_{Mi})\sqrt{-2[r_{Mi} + \delta_i \log(\delta_i - r_{Mi})]}$$

- Designed to be symmetric around 0
- Easier to visually examine than Martingale residuals

16

Deviance residual plot

- **X-axis**: Covariate X_i or linear predictor $(\beta_1 X_{i1} + \cdots + \beta_k X_{ik})$
- **Y-axis**: Deviance residual
- **Expected shape**: Random scatter around mean 0, symmetric around 0
- **Purpose**: Assess functional form of the covariates, assess overall model fit when using linear predictor; check for outliers



17

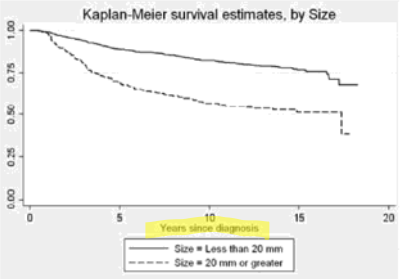
Model fit is poor – now what?

- May be missing an important covariate
- May need to change the functional form of a covariate (transform the variable, add a spline, etc.)
- May need interaction term(s) between covariates

18

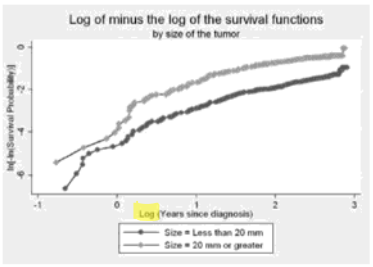
Kaplan-Meier plot – PH assumption

- **X-axis**: time
- **Y-axis**: Kaplan-Meier curve $\hat{S}(t)$ for each group
- **Expected shape**: Diverging curves that do not cross
- **Purpose**: Examine proportional hazards assumption



Complementary log-log plot → additive

- **X-axis**: time or log(time)
- **Y-axis**: Log-log Kaplan Meier, $\log\left(-\log\left(\hat{S}(t)\right)\right)$, which is equal to $\log\left(\hat{H}(t)\right)$, plotted for each group
- **Expected shape**: Parallel lines
- **Purpose**: Examine proportional hazards assumption



Schoenfeld residuals

only calculate residuals for those who FAILED, and there's a diff. residual for every person & every covariate

- Example shown for a model with only a single covariate
- For person i who fails at time T_i:

$$r_{Si} = X_i - \frac{\sum_{i'=1}^{n_t} X_{i'} \exp(\hat{\beta} X_{i'})}{\sum_{i'=1}^{n_t} \exp(\hat{\beta} X_{i'})}$$

- Compare his/her covariate X_i with the n_t individuals in the risk set
- When there is more than one covariate, each person has one residual per covariate
- Residuals only defined for people who fail
- Can be scaled by its standard error

e.g. $n = 20$ $p = 2$ covariates
$d = 10$ failures
Martingale = 20
Schoenfeld: One set of 10 for first covariate
One set of 10 for 2nd covariate

Example: Multiple myeloma

↑ BUN = ↑ risk

- Consider our model with a single covariate for BUN
- Individual $i = 7$ fails at time $T_i = 66$ months. At that time, there are $n_t = 4$ individuals remaining at risk.

hazard ratio?

Patient (i)	Time (T_i^*)	Status (δ_i)	BUN ($X_{i,Bun}$)	$exp(\hat{\beta} X_{i,Bun})$
7	66	1	21	1.527
19	76	0	12	1.274
21	88	1	21	1.527
36	91	1	27	1.723

22

Example: Multiple myeloma

- The Schoenfeld residual for person $i = 7$ is:

$$r_{Si} = 21 - \frac{\left(21(1.527) + 12(1.274) + 21(1.527) + 27(1.723)\right)}{(1.527 + 1.274 + 1.527 + 1.723)}$$

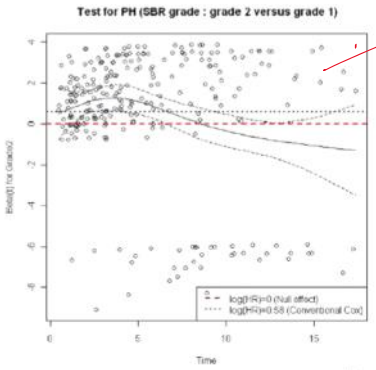
$$= 21 - 20.8 = 0.2$$

If all of the highest covariates are significantly sharing...

23

Scaled Schoenfeld residual plot

- **X-axis**: time
- **Y-axis**: Scaled Schoenfeld residuals for a given covariate
- **Expected shape**: Random scatter around a horizontal line at $\hat{\beta}$ (the log HR for that covariate)
- **Purpose**: Directly visualize the log hazard ratio over time; assess time-varying effects and proportional hazards assumption



Schoenfeld consistently high means highest value is what used for Schoenfeld

At all times, the difference between the groups should be the same

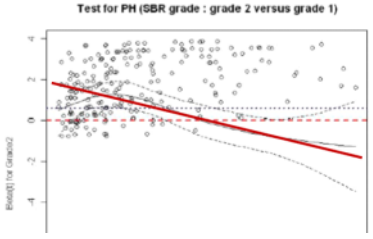
Anything that isn't a flat line is a violation

We want scatter around any flat line

24

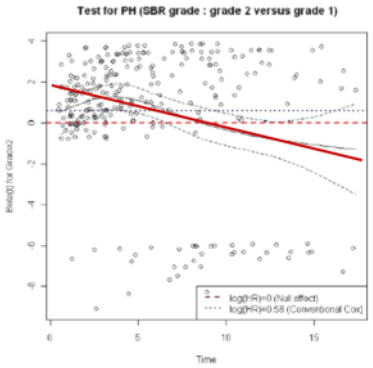
Grambsch-Therneau test

- We can run a simple linear regression to test that the slope of the Schoenfeld residuals is equal to 0



Grambsch-Therneau test

- We can run a simple linear regression to test that the slope of the Schoenfeld residuals is equal to 0



25

Test of time-by-covariate interactions

- Add a new covariate $X_i Q(t)$ to your model
  - $Q(t) = t$
  - $Q(t) = \log(t)$
  - $Q(t) = \log(t) - \overline{log}(t)$
- If the log hazard ratio for this new covariate is significantly different from 0, then there is evidence of non-proportionality

26

Proportional hazards is violated – now what?

- Relax the proportional hazards assumption (next week!)

27

Tutorial Paper

Survival Analysis Part III: Multivariate data analysis – choosing a model and assessing its adequacy and fit

MJ Bradburn[*,1], **TG Clark**[1], **SB Love**[1] **and DG Altman**[1]

[1]*Cancer Research UK/NHS Centre for Statistics in Medicine, Institute of Health Sciences, Old Road, Oxford OX3 7LF, UK*

Goal: Investigators sought to develop a **prognostic index for overall survival** among ovarian cancer patients.

Statistical analysis: The original analysis included 10 variables, but the tutorial focuses on five for simplicity. All are **measured at diagnosis (baseline)**.

- FIGO stage (an ordinal covariate taking values of 1, 2, 3 or 4)
- Histology (one of seven subtypes)
- Grade (1, 2, or 3)
- Ascites (yes/no)
- Patient age.

Researchers fit a Cox proportional hazards model with all covariates.

28

Results: Advanced FIGO stage, higher grade, presence of ascites, and increased age all impaired survival to varying degrees. The mucinous and serous histology types had a better prognosis, and undifferentiated and mixed mesodermal a lesser one.

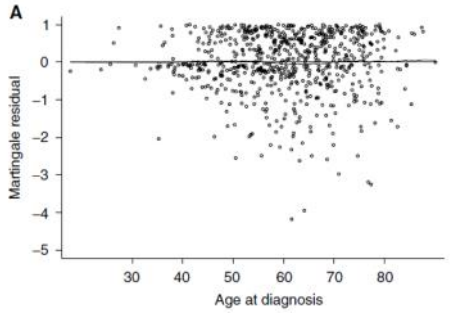
Table 2 Cox model applied to the ovarian data

Covariate	Coefficient (b_i)	HR [exp(b_i)]	95% CI	*P*-value
FIGO stage	0.731	2.08	(1.82–2.37)	<0.001
Histology				<0.001
Serous	(0.000)	(1.00)		
Mucinous	−0.422	0.66	(0.50–0.85)	
Endometroid	0.198	1.22	(0.80–1.85)	
Clear cell	0.342	1.41	(0.99–2.00)	
Adenocarcinoma	0.501	1.65	(0.91–2.99)	
Undifferentiated	0.746	2.11	(1.03–4.29)	
Mixed mesodermal	0.789	2.20	(1.45–3.35)	
Grade				<0.001
1	(0.000)	(1.00)		
2	0.885	2.42	(1.40–4.19)	
3	0.885	2.42	(1.40–4.18)	
Absence of ascites	−0.396	0.67	(0.54–0.84)	<0.001
Age (per 5-year increase)	0.133	1.14	(1.09–1.19)	<0.001

HR = hazard ratio; CI = confidence interval.

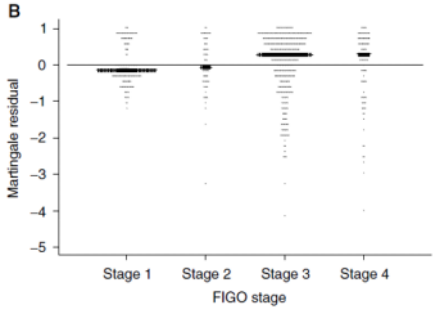
29

Results: The figure below plots the **Martingale residuals** against the **patient's age**, with a smoother marked as the dashed line. Age was modeled as a linear effect. If age had been modeled incorrectly (i.e. nonlinear), the figures should display a trend rather than a strictly horizontal line. The age residual plot shows to no evidence of a trend. Thus, the model fit appears adequate.



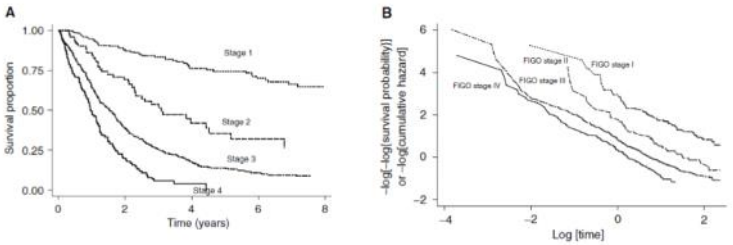
30

Results: The figure below plots the **Martingale residuals** against **FIGO stage** (1, 2, 3 or 4), with the median for each stage represented by the solid bar. FIGO stage was modeled as a linear effect. Although there appears to be evidence of a trend in the FIGO plot, the inclusion of this covariate as a categorical covariate fails to improve the fit to a significant degree. Thus, the model fit appears adequate.



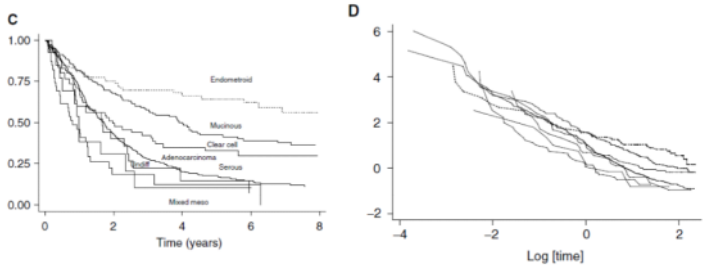
31

Results: Figure A below plots **Kaplan-Meier survival according to FIGO stage**. The lines appear to diverge over time and do not cross. Figure B plots **log(-log(survival)) against log(time) for each FIGO stage**. The lines appear to be parallel. These results suggest that the assumption of proportionality is reasonable for FIGO stage.



32

Results: Figure C below plots **Kaplan-Meier survival according to histology**. The endometroid group is shown with a dotted line. The prognosis for the endometroid group sits in the middle of all other groups in the first year but improves thereafter. Figure B plots **log(-log(survival)) against log(time) for each histology**. The endometroid group's curve is not parallel to the other groups. These results suggest that the assumption of proportionality is violated for this group.



33

Results: **Scaled Schoenfeld residuals** showed significant non-proportionality for the **endometroid group** and the **presence/absence of ascites**. This was also assessed by adding **interactions with log(time)** to the model. The time-dependent covariates are statistically significant for both variables.

Table 4 The Cox model applied to the ovarian data, with a time dependency added to ascites and endometroid terms

Covariate	Coefficient (b_i)	HR [exp(b_i)]	95% CI	P-value
FIGO	0.734	2.09	(1.83–2.38)	<0.001
Histology				<0.001
Serous	(0.000)	(1.00)		
Mucinous	−0.432	0.65	(0.50–0.85)	
Clear cell	0.344	1.41	(0.99–2.01)	
Adenocarcinoma	0.494	1.64	(0.91–2.96)	
Undifferentiated	0.769	2.16	(1.06–4.40)	
Mixed mesodermal	0.825	2.28	(1.50–3.47)	
Endometroid	0.312	1.37	(0.90–2.07)	
Endometroid × log(time)	−0.500	0.61	(0.45–0.82)	0.001
Grade				<0.001
1	(0.000)	(1.00)		
2	0.826	2.28	(1.32–3.95)	
3	0.843	2.32	(1.35–4.00)	
Absence of ascites	−0.466	0.63	(0.50–0.80)	<0.001
Ascites × log(time)	0.233	1.26	(1.01–1.58)	0.04
Age (per 5-year increase)	0.134	1.14	(1.09–1.20)	<0.001

HR = hazard ratio; CI = confidence interval.

34

Results: The suggested final model is a Cox PH model that relaxes the proportional hazards assumption for endometroid histology and the absence of ascites.

Table 4 The Cox model applied to the ovarian data, with a time dependency added to ascites and endometroid terms

Covariate	Coefficient (b_i)	HR [exp(b_i)]	95% CI	P-value
FIGO	0.734	2.09	(1.83–2.38)	<0.001
Histology				<0.001
Serous	(0.000)	(1.00)		
Mucinous	−0.432	0.65	(0.50–0.85)	
Clear cell	0.344	1.41	(0.99–2.01)	
Adenocarcinoma	0.494	1.64	(0.91–2.96)	
Undifferentiated	0.769	2.16	(1.06–4.40)	
Mixed mesodermal	0.825	2.28	(1.50–3.47)	
Endometroid	0.312	1.37	(0.90–2.07)	
Endometroid × log(time)	−0.500	0.61	(0.45–0.82)	0.001
Grade				<0.001
1	(0.000)	(1.00)		
2	0.826	2.28	(1.32–3.95)	
3	0.843	2.32	(1.35–4.00)	
Absence of ascites	−0.466	0.63	(0.50–0.80)	<0.001
Ascites × log(time)	0.233	1.26	(1.01–1.58)	0.04
Age (per 5-year increase)	0.134	1.14	(1.09–1.20)	<0.001

HR = hazard ratio; CI = confidence interval.

35

Model building

- Given a data set with lots of potential covariates, how we do construct our **final model**?
- This depends on the *scientific goal* of our study
  - Test a hypothesis of primary interest
  - Identify a set of variables to aid in modeling survival
- A **parsimonious model** is a model that accomplishes a desired level of explanation or prediction with as few predictor variables as possible

36

Purposeful selection (Hosmer, Lemeshow, May)

- **Step 1:** Screen covariates using univariable models (p<0.20 or 0.25)
- **Step 2:** Fit multivariable model and use p-values to identify covariates to delete
- **Step 3:** Assess whether removal of each covariate produces an "important" change in other coefficients
- **Step 4:** Confirm eliminated variables are not significant or confounders
- **Step 5:** Examine the scale of continuous covariates
- **Step 6:** Determine whether interactions are needed
- **Step 7:** Check for influential observations and test overall goodness-of-fit

37

Information criteria

- Akaike information criterion:

$$AIC = 2p - 2\ell(\widehat{\boldsymbol{\beta}})$$

where p is the number of parameters and $\ell(\widehat{\boldsymbol{\beta}})$ is the log partial likelihood at the MPLE

- Bayesian information criterion:

$$BIC = p\log(d) - 2\ell(\widehat{\boldsymbol{\beta}})$$

where d is the number of events (failures) in the data set

- Information criteria, e.g. AIC, BIC, etc
  - Discourages models with too many covariates that don't actually improve the fit of the model
  - (Similar to adjusted R^2 used in linear regression)

38

Today's activity

- R practice session

39