Minitab Summary characteristics Evaluating fit Parameter estimation Summary

> Parametric models continued: summary characteristics, fit, and estimation

> > Shannon Pileggi

STAT 417

STAT 417: Set 3

Parameter estimation

1 / 39

Summary

Minitab Summary characteristics Evaluating fit Parameter estimation Summary

#### **OUTLINE**

Summary characteristics

Evaluating fit

Parameter estimation

Minitab

Summary

STAT 417: Set 3

00000000

2 / 39

Summary characteristics Evaluating fit Parameter estimation Summary

## Summary characteristics of T

Evaluating fit

Summary characteristics

•00000000

With a specified probability distribution for T, we can answer:

- ▶ What is the average time it takes for a motorist to react aggressively?
- ▶ What is the median age that an individual has his/her first drink?
- ▶ For patients who have survived 100 days after being diagnosed with lung cancer, what is the average time they have left to live?

# Mean survival time

The mean (or expected value) of the time-to-event random variable T ( $T \ge 0$ ), denoted by E(T) or  $\mu$ , is given by:

◆□▶◆御▶◆恵▶◆恵▶○恵 3 / 39

Minitab Summary characteristics Evaluating fit Parameter estimation Summary 00000000

# Motorist reaction time: compute E(T)

Suppose that the time until a motorist reacts aggressively is exponentially distributed with parameter  $\lambda = 5.77$ . Find the expected time it takes for a motorist to react aggressively.

5 / 39 STAT 417: Set 3

000000000

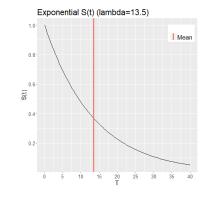
Evaluating fit

#### **Group Exercise**

Summary characteristics

STAT 417: Set 3

Another measure of the "center" of the distribution for T is the median survival time.



The median survival time is the mean survival time of

Minitab

Summary

13.5.

1. greater than

Parameter estimation

- 2. less than
- 3. equal to
- 4. not enough information to determine

6 / 39

Summary characteristics Minitab Evaluating fit Parameter estimation Summary 000000000

#### Percentiles of T

STAT 417: Set 3

- ▶ The **median survival time** is a special case of a *percentile* of the distribution of T (the  $50^{th}$  percentile). This is the time at which 50% of the individuals have experienced the event.
- ▶ The  $p^{th}$  percentile of the distribution of the time-to-event random variable T is the value of T, denoted  $t_p$  such that:
- ▶ **Interpretation**: The  $p^{th}$  percentile,  $t_p$ , can be interpreted as:

Summary characteristics Evaluating fit Parameter estimation Summary 00000 0000

## Graphical representation of percentiles

《中》《圖》《意》《意》。意 STAT 417: Set 3 8 / 39

ベロト (個) (を見) (見) (見)

7 / 39

Minitab Summary characteristics Evaluating fit Parameter estimation Summary 000000

## Age at first drink: median survival time

Suppose that the age at first drink of alcohol follows an exponential distribution with parameter  $\lambda = 13.5$ . Find the median age at first drink. Recall that  $S(t) = \exp(-t/\lambda)$ .

9 / 39

Summary characteristics

00000000

Evaluating fit

Parameter estimation

Minitab

Minitab

Summary

Summary

#### Mean residual lifetime

- **Example: Lung Cancer**. Suppose an individual with lung cancer has survived 100 months since being diagnosed. Then how much time (on average) does this individual have left to live? We can set this problem up as a conditional expectation:
- ▶ This is an example of quantity known as the mean residual **lifetime (mrl)**, denoted  $mrl(t^*)$ , and is defined to be the average remaining lifetime of all individuals given that the individuals have survived to time  $t^*$ . i.e.

STAT 417: Set 3 10 / 39

Parameter estimation

Summary characteristics Evaluating fit Parameter estimation Summary 00000000

# Lung cancer example: mean residual lifetime

STAT 417: Set 3

If the time until death from cancer follows an exponential distribution with parameter  $\lambda = 130.18$ , then find the average remaining lifetime for those patients who have survived at least 100 days.

Evaluating fit

•000000

**Evaluating fit** 

Summary characteristics

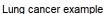
《中》《圖》《意》《意》。意 STAT 417: Set 3 12 / 39 11 / 39

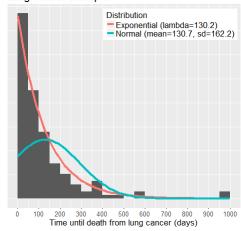
《□》《圖》《意》《意》。 意

Minitab Summary characteristics Evaluating fit Parameter estimation Summary 000000 Considerations for parametric models 1. 2. 3. < □ > → □ > → □ > → □ ≥ → □ ≥ → STAT 417: Set 3 13 / 39

Summary characteristics Evaluating fit Parameter estimation Minitab Summary 000000

# Veteran's Administration Lung Cancer Study (VALCS)





Lifetimes (since treatment) of 137 lung cancer patients taken from another lung cancer study with two different probability density curves superimposed (exponential and normal).

14 / 39

STAT 417: Set 3

Evaluating fit Minitab Summary characteristics Parameter estimation Summary 0000000

# Strategies for fitting parametric models

1.

2.

3.

Summary characteristics Evaluating fit Parameter estimation Summary 000000

# Probability plots for any distribution

- ► Can construct a probability plot for any distribution
- ▶ With incomplete data, special techniques are used in the probability plot to account for censoring
- ► Assess if an exponential distribution fits the data well  $H_0$ :

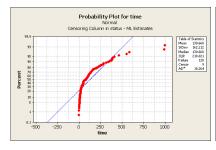
H<sub>2</sub>:

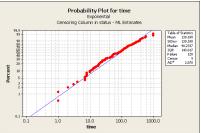
1.

2.

《中》《圖》《意》《意》。意 STAT 417: Set 3 16 / 39 
 Summary characteristics
 Evaluating fit 00000000
 Parameter estimation 0000000
 Minitab 000000
 Summary 0000000

# VALCS study: probability plots





#### Which distribution appears to fit the data better, and why?

- 1. normal
- 2. exponential

 ★□トイラトイミト ミ ぐへで

 STAT 417: Set 3

17 / 39

Median 4.35127 4.18097 Median TOR 3.46761 IOR 3.47690 Failure 42 Failure 42 Censor 15 Censor 15 An\* 8.509 AD\* 6.163 Probability Plot for Motorist Reaction Times Probability Plot for Motorist Reaction Times 5.76595 Shape 1.60295 StDev 5.76595 5.74178 Scale Median 3.99665 5.14706 Mean 6.33455 StDev 3.28809 IQR Failure 42 Median 4.56820 Censor IQR 4.40022 AD\* 8.111 Failure 42 Censor 15 AD\* 6.959 <ロト (部) (注) (注) STAT 417: Set 3 18 / 39

Parameter estimation

Probability Plot for Motorist Reaction Times

Minitab

Loc

Scale

Mean

StDev

∢ロト∢御ト∢差ト∢差ト 差

1.43054

5.00540

3.29467

0.599958

Summary

Summary characteristics

**Evaluating** fit

Parameter estimation

Minitab

Summary

STAT 417: Set 3

 Summary characteristics
 Evaluating fit 0000000
 Parameter estimation 000000
 Minitab 000000
 Summary 000000

#### Parameter estimation

Summary characteristics

Evaluating fit

Motorist reaction times: probability plots

Loc

Scale

Mean

4.35127

1.57818

4.35127

StDev 2.86250

000000

Probability Plot for Motorist Reaction Times

- We have worked through many examples of parametric models for T, and in each example, the value(s) of the parameter(s) were given. For example, the time until death for the lung cancer data was assumed to follow an exponential distribution with parameter  $\lambda=130.18$ .
- ▶ Also, in the previous probability plots (observe the box on the right side), values of the parameters were provided. How are these values computed?

STAT 417: Set 3 20 / 39

Summary characteristics

Evaluating fit

Parameter estimation

Minitab

Summary Summary characteristics

Evaluating fit

Parameter estimation

Minitab

Summary

#### Maximum likelihood estimation

Suppose we have a random sample of observations  $T_1, T_2, \ldots, T_n$  on n individuals, each with the same pdf  $f(t_i)$ ,  $i = 1, \ldots, n$  and parameter  $\theta$  (or possibly more than one parameter).

► Then **likelihood function** is given by:

- ► The likelihood function tells how likely the observed sample is as a function of the parameter values.
- ▶ By maximizing the likelihood function, we get the parameter value(s) that most likely would generate our given observed sample.



# Maximum likelihood estimation, cont.

Goal: Find an estimate of  $\theta$  that maximizes the likelihood function,  $L(\theta)$ .

Process:

2.

3.

STAT 417: Set 3

> ◆ □ > ◆ ≣ > ◆ ≣ > ○ ○ ○ 22 / 39

Summary characteristics

STAT 417: Set 3

Evaluating fit

Parameter estimation

Minitab

Summary

STAT 417: Set 3

Summary characteristics

Evaluating fit

Parameter estimation

Minitab

Summary

# MLE example: exponential distribution

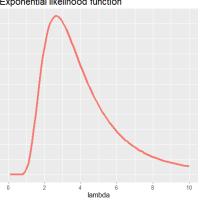
Suppose we obtain a random sample of 5 observations,  $T_1, T_2, \ldots, T_5$ , from an exponential distribution with parameter  $\lambda$  and pdf:

$$f(t) = \frac{1}{\lambda}e^{-t/\lambda}$$

1. Construct the likelihood function  $L(\lambda)$ .

## MLE example: exponential distribution, cont.

2. Suppose the observed values of  $T_1, T_2, \ldots, T_5$  are 5.3, 4.8, 0.4, 2.3, and 0.4. Plug these into  $L(\lambda)$  and examine the graph of  $L(\lambda)$  as a function of  $\lambda$ . At which value of  $\lambda$  does the likelihood function appear to be maximized?



 lambda
 ◄ □ ▶ ◀ শ ▶ ◀ 분 ▶ ◀ 분 ▶ 월
 ♥ ♀ ♀ ○

 STAT 417: Set 3
 24 / 39

MLE example: exponential distribution, cont.

3. Find the value of  $\lambda$  that maximizes  $L(\lambda)$ .

25 / 39

STAT 417: Set 3

Summary characteristics

Evaluating fit

26 / 39

Minitab

Summary

STAT 417: Set 3

Summary characteristics

Evaluating fit

Parameter estimation 000000000000

Summary

Likelihood function for right censored data

If there are right censored event times, then can't simply compute the likelihood function L as the product of the pdf's. Contributions to the likelihood function differ:

- ▶ If the event time for observation *i* is complete, then that observation contributes:
- ▶ If the event time for observation *i* is (right) censored, then that observation contributes:
- ▶ Putting the contributions of all observation together, the likelihood function is given by:

Parameter estimation Summary characteristics Evaluating fit Summary 000000000000

Parameter estimation

0000000000000

## MLE for right censored data

The following data consist of the times (in months) until relapse of 10 bone marrow transplant patients. Patients 7-10 were free of relapse at the end of the study. Suppose time to relapse follows an exponential distribution with parameter  $\lambda$ .

Patient			-		-	-	7	8	9	10
Relapse Time	5	8	12	24	32	17	16 <sup>+</sup>	17 <sup>+</sup>	19 <sup>+</sup>	30 <sup>+</sup>

- 1. Construct the likelihood function for the parameter  $\lambda$  and find the MLE for  $\lambda$ .
- 2. Assuming that time until relapse follows an exponential probability distribution with  $\lambda$  found in Part (1), what is the mean time until relapse?

28 / 39 STAT 417: Set 3 STAT 417: Set 3 27 / 39

Summary characteristics Minitab Evaluating fit Parameter estimation Summary 00000000000 MLE for right censored data, cont.

Summary characteristics

Evaluating fit

Parameter estimation 00000000000

Minitab

Summary

### MLE for right censored data, cont.

2. Assuming that time until relapse follows an exponential probability distribution with  $\lambda$  found in Part (1), what is the mean time until relapse?

29 / 39

30 / 39

STAT 417: Set 3

Summary characteristics

000000

Summary

Minitab Summary characteristics Evaluating fit Parameter estimation Summary •00000

Minitab

STAT 417: Set 3

Parametric distribution analysis in Minitab

Evaluating fit

► Along with the probability plots, the MLE's, and summary characteristics about the distribution of T are also computed in Minitab when Parametric Distribution Analysis is performed.

Parameter estimation

- ▶ The MLE's are provided in the box to the right of the plot, as well as in the output in the Minitab Session Window.
- ▶ The mean, median, quartiles, and several other percentiles are displayed in the Session Window.

《□》《圖》《意》《意》。 意 31 / 39

STAT 417: Set 3

《中》《圖》《意》《意》、意 200 32 / 39

Summary characteristics

Evaluating fit

Parameter estimation

Minitab

Summary characteristics

Evaluating fit

Parameter estimation

Minitab

Summary

## Motorist reaction times in Minitab (lognormal model)

Variable: Seconds until Horn Honk

Censoring Information Count Uncensored value 42 Right censored value 15

Censoring value: Censor Horn = 0

 ✓□ → ✓□ → ✓□ → ✓□ → ✓□ → ✓□ →
 № ○○

 STAT 417: Set 3
 33 / 39

Parametric distribution analysis in Minitab, cont.

Estimation Method: Maximum Likelihood

Distribution: Lognormal

Parameter Estimates

Standard 95.0% Normal CI
Parameter Estimate Error Lower Upper
Location 1.43054 0.0851775 1.26360 1.59749
Scale 0.599958 0.0665560 0.482718 0.745673

Log-Likelihood = -100.729

Goodness-of-Fit
Anderson-Darling (adjusted) = 6.163

. . .

STAT 417: Set 3

Summary characteristics Evaluating fit P

Parameter estimation

Minitab 0000•0 Summary

Parametric distribution analysis in Minitab, cont.

Characteristics of Distribution

		Standard	95.0% N	rmal CI	
	Estimate	Error	Lower	Upper	
Mean(MTTF)	5.00540	0.500060	4.11529	6.08804	
Standard Deviation	3.29467	0.672961	2.20774	4.91673	
Median	4.18097	0.356124	3.53813	4.94060	
First Quartile(Q1)	2.78954	0.249485	2.34102	3.32400	
Third Quartile(Q3)	6.26644	0.643405	5.12418	7.66334	
Interquartile Range(IQR)	3.47690	0.541312	2.56254	4.71752	

 Summary characteristics
 Evaluating fit occoods
 Parameter estimation occoods
 Minitab occood
 Summary occoods

 000000000
 00000000000
 00000€
 000

# Parametric distribution analysis in Minitab, cont.

Table of Percentiles

		Standard	95.0% No	rmal CI
Percent	Percentile	Error	Lower	Upper
1	1.03545	0.169643	0.751052	1.42753
2	1.21943	0.181146	0.911401	1.63155
3	1.35276	0.188314	1.02974	1.77711

. . .

## What do you think the first line means?

- 1. Pr(T < 1.03545) = 0.01
- 2. Pr(T < 1) = 0.0103545
- 3. Pr(T > 1.03545) = 0.01
- 4. Pr(T > 1) = 0.0103545

STAT 417: Set 3

4 D > 4 D > 4 E > 4 E > 9 Q Q

36 / 39

Summary characteristics

**Evaluating fit** 

Parameter estimation

Minitab

Summary

STAT 417: Set 3 37 / 39

Summary characteristics Evaluating fit Parameter estimation Minitab Summary

#### **Future**

STAT 417: Set 3

**Nonparametric** methods for summarizing and describing samples of time-to-event data:

- ▶ Estimators for S(t), h(t), and H(t).
- $\triangleright$  Estimates of E(T) and  $t_p$
- ▶ Inference methods for the estimator of S(t)
- ► Inference procedures for comparing two or more survival experiences.
- ► Survival analysis in R

## Recap

STAT 417: Set 3

- ► Features of time-to-event data and survival studies (e.g. censoring, truncation, etc.)
- Parametric models for the time-to-event random variable T: f(t), F(t), S(t), h(t), H(t)
- ▶ Descriptive measures for T: E(T),  $t_p$ , and,  $mrl(t^*)$
- ► Strategies for selecting a probability model: probability plots, AD test statistic
- ▶ Parameter estimation: maximum likelihood estimation

< □ ▷ < ∰ ▷ < 볼 ▷ < 볼 ▷ ○ 볼 · ◇ ○ ○ 38 / 39

4□▶ 4∰▶ 4½▶ 4½▶ ½
 30 / 30