Distribution Selection

CIVL 7012/8012

Possible Outcomes

		Truth Situation	
		Two Distributions Identical	Two Distributions Different
Predicted Situation	No evidence of difference (Fail to Reject Hypothesis)	√	Type-II error
	Evidence of difference (Reject Hypothesis)	Type-I Error	√

Hypothesis

- Fail to Reject Hypothesis
  - "There is no evidence of a statistical difference between the two distributions and the measured distribution could be identical to the mathematical distribution"
- Reject Hypothesis
  - "There is evidence of a statistical difference between the two distributions and it is unlikely that the measured distribution is identical to the mathematical distribution"

Error

- Type – I error
  - When prediction shows that the two distributions are different, but in truth situation two distributions are identical
- Type-II error
  - When prediction shows that two distributions are identical but in truth situation two distributions are different

Chi-square Test

$$\chi^2_{calc} = \sum_{i=1}^{I} \frac{(f_0 - f_t)^2}{f_t}$$

where.,

χ^2_{calc} -> Calculated chi-square value

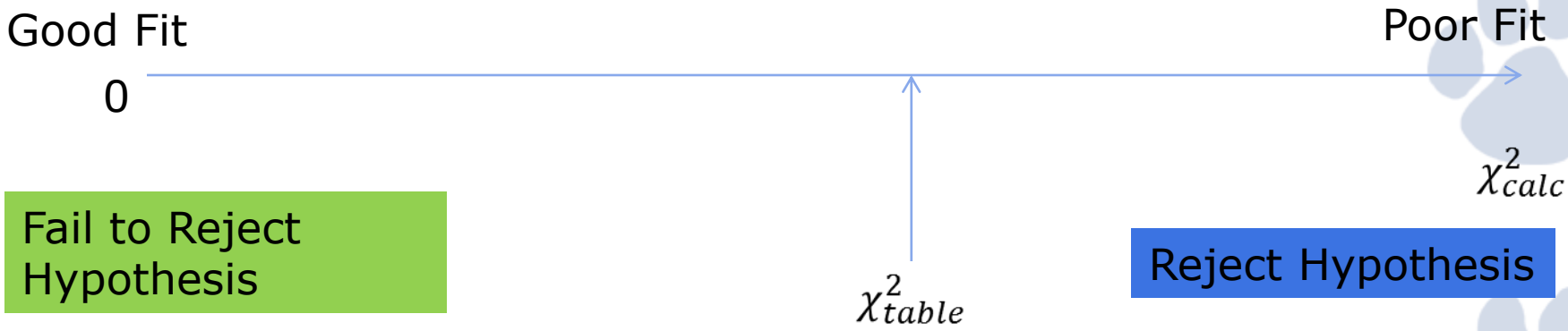
f_0-> Observed number or frequency of observations in time headway interval *i*

f_t-> Theoretical number or frequency of observations in time headway interval I

i-> Any time headway interval

I-> Number of time headway intervals

Accept or Reject Hypothesis



$\chi^2_{calc} \geq \chi^2_{table}$ -> Reject Null hypothesis

How to find Chi-square (Table)

$$n=(I-1)-p$$

Where,

n -> Number of degrees of freedom

I -> Number of time headway intervals being compared

1 = constant

p= Number of parameters estimated in defining the

Why 1 is subtracted

- A constant “1” is subtracted from the number of time headway groups since the total frequency of the two distributions are set equal
- Therefore, the theoretical frequency of the last group is not dependent on I-1 frequencies

Number of parameters needed

Distribution	Parameters (p)
Measured	0
Negative Exponential	1 $(\overline{t})$
Shifted Negative Exponential	2 $(\bar{t}, \alpha)$
Normal	2 $(\bar{t}, s)$
Pearson Type-III	2 (K, λ)
Composite	4

Example-1 Speed Data

- Consider speed data is collected and the following data is obtained.

ui	fi
30	-
31	-
32	0
33	1
34	2
35	1
36	1
37	-
38	1
39	1
40	2
41	1
42	5
43	4
44	1
45	7
46	4
47	8
48	8
49	15
50	8
51	8
52	10
53	23
54	15
55	16
56	9
57	14
58	6
59	3
60	9
61	3
62	6
63	3
64	3
65	2
66	-
67	-
68	-
69	-
70	-
Sum	200

Example-1 Speed Data

- Let us calculate descriptive statistics of the data

$$\overline{\mu} = \frac{\sum_{i=1}^{g} f_i \mu_i}{N}$$

$$s^2 = \frac{\sum_{i=1}^{g} f_i (\mu_i)^2 - \frac{1}{N}\left(\sum_{i=1}^{g} f_i \mu_i\right)^2}{N-1}$$

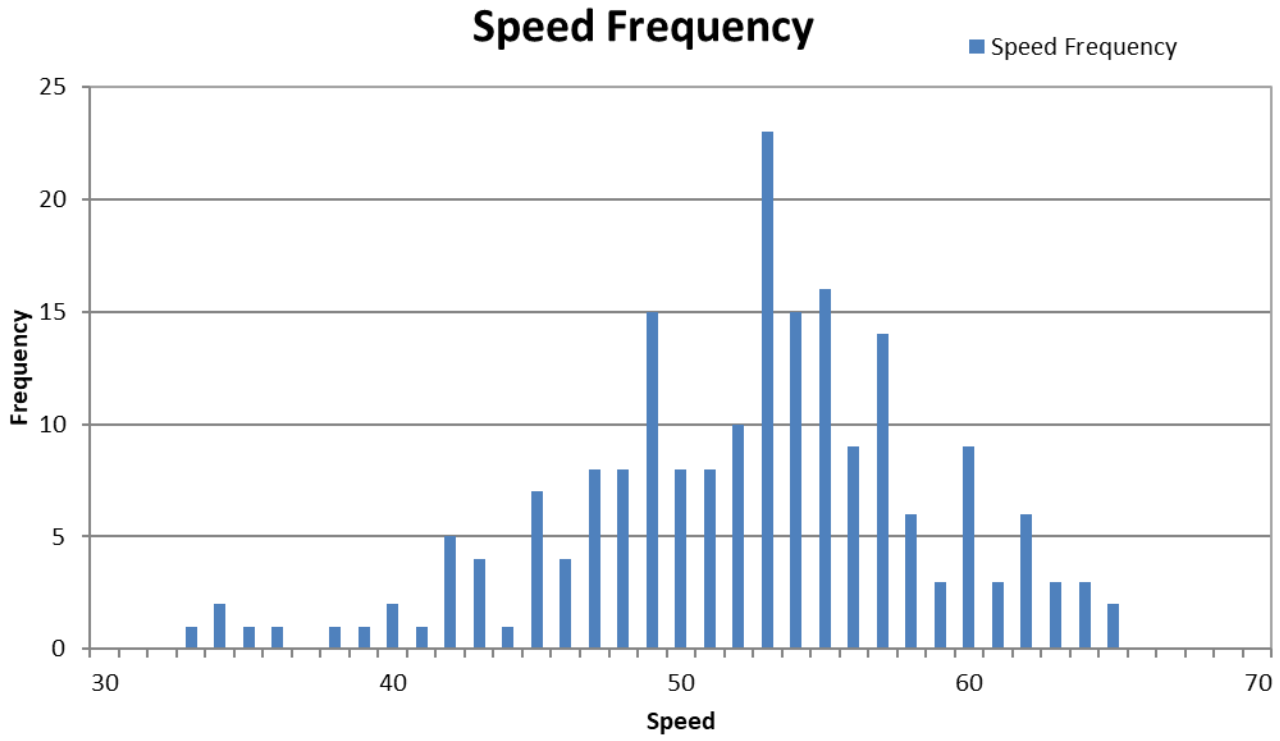
		Cumulative		$f_i\mu_i$	$f_i(\mu_i)^2$
ui	fi	fi	%		
30	-	-		-	-
31	-	0	0	-	-
32	0	0	0	0	0
33	1	1	0.5	33	1,089
34	2	3	1.5	68	2,312
35	1	4	2	35	1,225
36	1	5	2.5	36	1,296
37	-	5	2.5	-	-
38	1	6	3	38	1,444
39	1	7	3.5	39	1,521
40	2	9	4.5	80	3,200
41	1	10	5	41	1,681
42	5	15	7.5	210	8,820
43	4	19	9.5	172	7,396
44	1	20	10	44	1,936
45	7	27	13.5	315	14,175
46	4	31	15.5	184	8,464
47	8	39	19.5	376	17,672
48	8	47	23.5	384	18,432
49	15	62	31	735	36,015
50	8	70	35	400	20,000
51	8	78	39	408	20,808
52	10	88	44	520	27,040
53	23	111	55.5	1,219	64,607
54	15	126	63	810	43,740
55	16	142	71	880	48,400
56	9	151	75.5	504	28,224
57	14	165	82.5	798	45,486
58	6	171	85.5	348	20,184
59	3	174	87	177	10,443
60	9	183	91.5	540	32,400
61	3	186	93	183	11,163
62	6	192	96	372	23,064
63	3	195	97.5	189	11,907
64	3	198	99	192	12,288
65	2	200	100	130	8,450
66	-	200	100	-	-
67	-	200	100	-	-
68	-	200	100	-	-
69	-	200	100	-	-
70	-	200	100	-	-
Sum	200			10,460	554,882

Example-1 Speed Data

- Mean = 52.3
- Median = 49
- Mode = 53
- Variance = 39.13
- Standard deviation = 6.27
- Range = 32 (33-65)

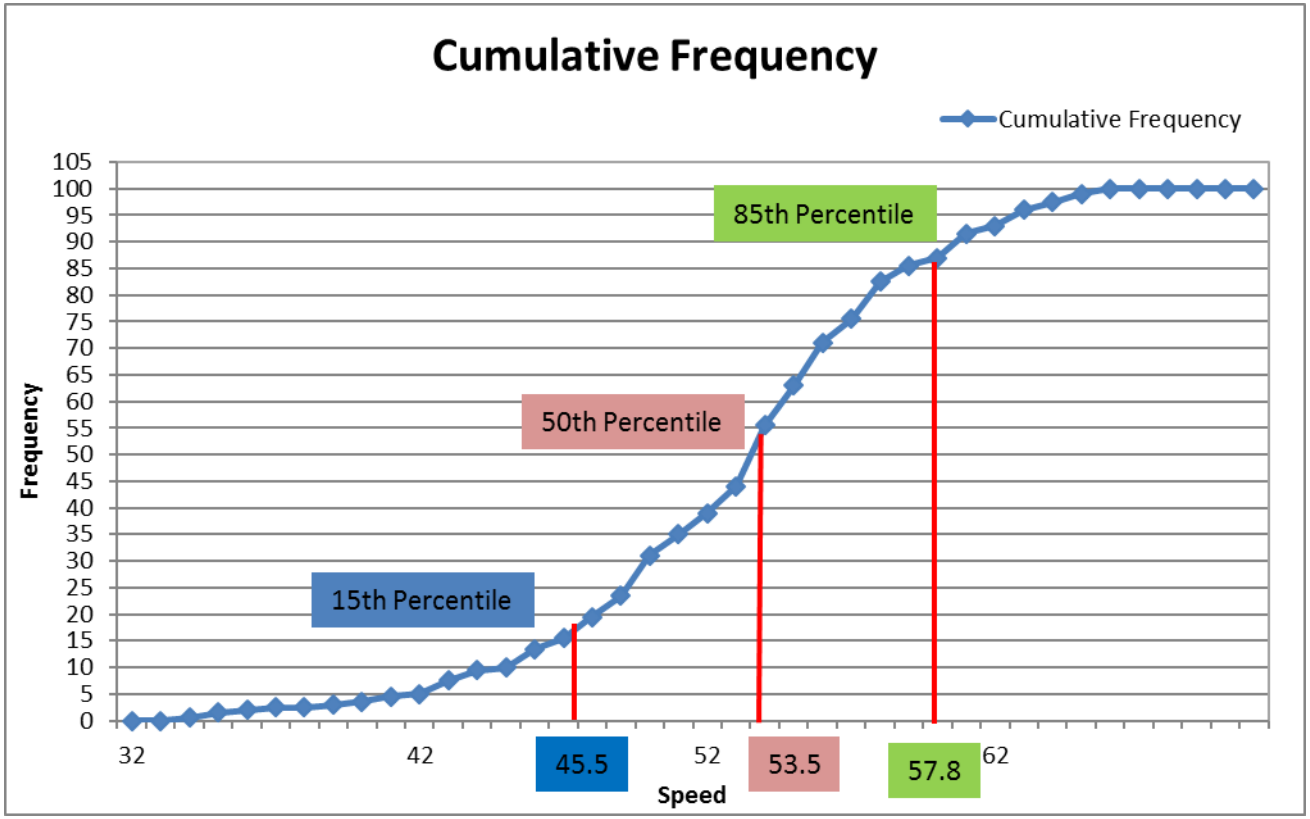
Example-1 Speed Data

- Frequency plot



Example-1 Speed Data

- Cumulative Frequency Plot



How to find interval width

$$I = \frac{\text{Range}}{1 + (3.322) \log N}$$

where I = size of the class interval
Range = total range (largest observed value minus smallest observed value)
N = number of observations

Example-1 Speed Data

- Normal Distribution

Class Interval Limit	z	z/s	P	Pt	Ft
				0.0038	0.77
35.5	16.8	2.666667	0.4962	0.0104	2.08
38.5	13.8	2.190476	0.4858	0.0290	5.80
41.5	10.8	1.714286	0.4568	0.0646	12.92
44.5	7.8	1.238095	0.3922	0.1152	23.04
47.5	4.8	0.761905	0.2769	0.1645	32.90
50.5	1.8	0.285714	0.1125	0.1880	37.60
53.5	1.2	0.190476	0.0755	0.1720	34.40
56.5	4.2	0.666667	0.2475	0.1259	25.19
59.5	7.2	1.142857	0.3735	0.0738	14.77
62.5	10.2	1.619048	0.4473	0.0527	10.54

Example-1 Speed Data

Hypothesis Test

Null Hypothesis: There is no statistical difference between the measured distribution and normal distribution
Alternate Hypothesis: There existis statistical difference between the measured distribution and normal distribution

Class Interval Limit	f0	ft	f0-ft	(f0-ft)^2	[(f0-ft)^2]/ft
	4	0.766076			
35.5	2	2.082896			
38.5	4	5.798655	1.352373	1.82891395	0.315403
41.5	10	12.92045	-2.92045	8.52901878	0.660118
44.5	19	23.04361	-4.04361	16.35078051	0.709558
47.5	31	32.89801	-1.89801	3.602447455	0.109504
50.5	41	37.5967	3.403296	11.5824244	0.30807
53.5	40	34.39509	5.604908	31.41498944	0.913357
56.5	23	25.18872	-2.18872	4.790480419	0.190184
59.5	18	14.76609	3.233911	10.4581796	0.708257
62.5	8	10.5437	-2.5437	6.470418735	0.613676
				Sum_Calculated	4.528126
				Table	12.59159
				Conclusion	Fail to reject Null Hypothesis