Project – Graduation tests

Tushar
407

❖ Original data

	AGE	ETR	DEATHS	CRUDE	GRADUATED	EXPECTED	ZX
1	25	78500	24	0	0	0	0
2	26	80425	24	0	0	0	0
3	27	81975	24	0	0	0	0
4	28	83725	24	0	0	0	0
5	29	84875	72	0	0	0	0
6	30	85075	48	0	0	0	0
7	31	85275	120	0	0	0	0
8	32	86250	24	0	0	0	0
9	33	87250	72	0	0	0	0
10	34	88300	72	0	0	0	0
11	35	90200	24	0	0	0	0
12	36	92500	48	0	0	0	0
13	37	95425	24	0	0	0	0
14	38	98550	168	0	0	0	0
15	39	99775	120	0	0	0	0
16	40	99125	240	0	0	0	0
17	41	99200	216	0	0	0	0
18	42	101525	144	0	0	0	0
19	43	104525	120	0	0	0	0
20	44	107075	24	0	0	0	0
21	45	109125	216	0	0	0	0
22	46	109425	360	0	0	0	0
23	47	109075	312	0	0	0	0
24	48	110175	120	0	0	0	0
25	49	111675	168	0	0	0	0
26	50	112725	432	0	0	0	0
27	51	115250	408	0	0	0	0
28	52	118225	600	0	0	0	0
29	53	120025	702	0	0	0	0
30	54	122150	891	0	0	0	0
31	55	124350	513	0	0	0	0
32	56	125750	675	0	0	0	0
33	57	126350	864	0	0	0	0
34	58	127100	837	0	0	0	0
35	59	129350	1242	0	0	0	0
36	60	132475	1593	0	0	0	0
37	61	134000	1539	0	0	0	0
38	62	133700	1998	0	0	0	0
39	63	134375	1728	0	0	0	0
40	64	136125	2403	0	0	0	0
41	65	136625	1971	0	0	0	0
42	66	136100	2835	0	0	0	0
43	67	135750	2889	0	0	0	0
44	68	134350	3348	0	0	0	0
45	69	131575	4212	0	0	0	0
46	70	129225	4428	0	0	0	0
47	71	128875	3915	0	0	0	0
48	72	130075	5103	0	0	0	0
49	73	130475	5454	0	0	0	0
50	74	129550	6453	0	0	0	0
51	75	129400	6453	0	0	0	0

❖ Data after calculating CRUDE

	AGE	ETR	DEATHS	CRUDE	GRADUATED	EXPECTED	ZX
1	25	78500	24	0.0003057325	0	0	0
2	26	80425	24	0.0002984147	0	0	0
3	27	81975	24	0.0002927722	0	0	0
4	28	83725	24	0.0002866527	0	0	0
5	29	84875	72	0.0008483063	0	0	0
6	30	85075	48	0.0005642081	0	0	0
7	31	85275	120	0.0014072120	0	0	0
8	32	86250	24	0.0002782609	0	0	0
9	33	87250	72	0.0008252149	0	0	0
10	34	88300	72	0.0008154020	0	0	0
11	35	90200	24	0.0002660754	0	0	0
12	36	92500	48	0.0005189189	0	0	0
13	37	95425	24	0.0002515064	0	0	0
14	38	98550	168	0.0017047184	0	0	0
15	39	99775	120	0.0012027061	0	0	0
16	40	99125	240	0.0024211854	0	0	0
17	41	99200	216	0.0021774194	0	0	0
18	42	101525	144	0.0014183699	0	0	0
19	43	104525	120	0.0011480507	0	0	0
20	44	107075	24	0.0002241420	0	0	0
21	45	109125	216	0.0019793814	0	0	0
22	46	109425	360	0.0032899246	0	0	0
23	47	109075	312	0.0028604171	0	0	0
24	48	110175	120	0.0010891763	0	0	0
25	49	111675	168	0.0015043653	0	0	0
26	50	112725	432	0.0038323353	0	0	0
27	51	115250	408	0.0035401302	0	0	0
28	52	118225	600	0.0050750687	0	0	0
29	53	120025	702	0.0058487815	0	0	0
30	54	122150	891	0.0072943103	0	0	0
31	55	124350	513	0.0041254524	0	0	0
32	56	125750	675	0.0053677932	0	0	0
33	57	126350	864	0.0068381480	0	0	0
34	58	127100	837	0.0065853659	0	0	0
35	59	129350	1242	0.0096018554	0	0	0
36	60	132475	1593	0.0120249104	0	0	0
37	61	134000	1539	0.0114850746	0	0	0
38	62	133700	1998	0.0149439043	0	0	0
39	63	134375	1728	0.0128595349	0	0	0
40	64	136125	2403	0.0176528926	0	0	0
41	65	136625	1971	0.0144263495	0	0	0
42	66	136100	2835	0.0208302719	0	0	0
43	67	135750	2889	0.0212817680	0	0	0
44	68	134350	3348	0.0249199851	0	0	0
45	69	131575	4212	0.0320121604	0	0	0
46	70	129225	4428	0.0342658154	0	0	0
47	71	128875	3915	0.0303782735	0	0	0
48	72	130075	5103	0.0392312128	0	0	0
49	73	130475	5454	0.0418011113	0	0	0
50	74	129550	6453	0.0498108838	0	0	0
51	75	129400	6453	0.0498686244	0	0	0

❖ Calculating Graduated rates using Gompertz law

```
> model= lm(log(CRUDE)~AGE, data= data2)
> summary(model)

Call:
lm(formula = log(CRUDE) ~ AGE, data = data2)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.07946 -0.09831  0.08351  0.21582  1.13949 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -11.000864   0.256884  -42.82   <2e-16 ***
AGE           0.106298   0.004929   21.57   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5181 on 49 degrees of freedom
Multiple R-squared:  0.9047,	Adjusted R-squared:  0.9028 
F-statistic: 465.2 on 1 and 49 DF,  p-value: < 2.2e-16

> coeff=as.numeric(coef(model))
> B=exp(coeff[1])
> C=coeff[2]
> B
[1] 1.668727e-05
> C
[1] 0.1062976
> data2$GRADUATED=B*exp(C*data2$AGE)
```

	AGE	ETR	DEATHS	CRUDE	GRADUATED	EXPECTED	ZX
1	25	78500	24	0.0003057325	0.0002379562	0	0
2	26	80425	24	0.0002984147	0.0002646437	0	0
3	27	81975	24	0.0002927722	0.0002943242	0	0
4	28	83725	24	0.0002866527	0.0003273335	0	0
5	29	84875	72	0.0008483063	0.0003640449	0	0
6	30	85075	48	0.0005642081	0.0004048735	0	0
7	31	85275	120	0.0014072120	0.0004502812	0	0
8	32	86250	24	0.0002782609	0.0005007816	0	0
9	33	87250	72	0.0008252149	0.0005569456	0	0
10	34	88300	72	0.0008154020	0.0006194087	0	0
11	35	90200	24	0.0002660754	0.0006888771	0	0
12	36	92500	48	0.0005189189	0.0007661366	0	0
13	37	95425	24	0.0002515064	0.0008520610	0	0
14	38	98550	168	0.0017047184	0.0009476220	0	0
15	39	99775	120	0.0012027061	0.0010539005	0	0
16	40	99125	240	0.0024211854	0.0011720985	0	0
17	41	99200	216	0.0021774194	0.0013035526	0	0
18	42	101525	144	0.0014183699	0.0014497497	0	0
19	43	104525	120	0.0011480507	0.0016123433	0	0
20	44	107075	24	0.0002241420	0.0017931721	0	0
21	45	109125	216	0.0019793814	0.0019942815	0	0
22	46	109425	360	0.0032899246	0.0022179458	0	0
23	47	109075	312	0.0028604171	0.0024666947	0	0
24	48	110175	120	0.0010891763	0.0027433414	0	0
25	49	111675	168	0.0015043653	0.0030510149	0	0
26	50	112725	432	0.0038323353	0.0033931949	0	0
27	51	115250	408	0.0035401302	0.0037737513	0	0
28	52	118225	600	0.0050750687	0.0041969882	0	0
29	53	120025	702	0.0058487815	0.0046676923	0	0
30	54	122150	891	0.0072943103	0.0051911872	0	0
31	55	124350	513	0.0041254524	0.0057733935	0	0
32	56	125750	675	0.0053677932	0.0064208959	0	0
33	57	126350	864	0.0068381480	0.0071410176	0	0
34	58	127100	837	0.0065853659	0.0079419029	0	0
35	59	129350	1242	0.0096018554	0.0088326098	0	0
36	60	132475	1593	0.0120249104	0.0098232120	0	0
37	61	134000	1539	0.0114850746	0.0109249130	0	0
38	62	133700	1998	0.0149439043	0.0121501729	0	0
39	63	134375	1728	0.0128595349	0.0135128491	0	0
40	64	136125	2403	0.0176528926	0.0150283534	0	0
41	65	136625	1971	0.0144263495	0.0167138257	0	0
42	66	136100	2835	0.0208302719	0.0185883286	0	0
43	67	135750	2889	0.0212817680	0.0206730622	0	0
44	68	134350	3348	0.0249199851	0.0229916046	0	0
45	69	131575	4212	0.0320121604	0.0255701781	0	0
46	70	129225	4428	0.0342658154	0.0284379460	0	0
47	71	128875	3915	0.0303782735	0.0316273420	0	0
48	72	130075	5103	0.0392312128	0.0351744379	0	0
49	73	130475	5454	0.0418011113	0.0391193507	0	0
50	74	129550	6453	0.0498108838	0.0435066965	0	0
51	75	129400	6453	0.0498686244	0.0483860956	0	0

❖ Checking for smoothness

```
> first_diff_G=diff(data2$GRADUATED)
> second_diff_G=diff(first_diff)
> third_diff_G=diff(second_diff)
> third_diff_G
 [1] 3.356818e-07 3.733294e-07 4.151994e-07 4.617651e-07 5.135534e-07 5.711499e-07
 [7] 6.352059e-07 7.064461e-07 7.856760e-07 8.737918e-07 9.717900e-07 1.080779e-06
[13] 1.201991e-06 1.336798e-06 1.486724e-06 1.653464e-06 1.838905e-06 2.045143e-06
[19] 2.274512e-06 2.529605e-06 2.813307e-06 3.128828e-06 3.479734e-06 3.869996e-06
[25] 4.304027e-06 4.786736e-06 5.323582e-06 5.920637e-06 6.584653e-06 7.323141e-06
[31] 8.144452e-06 9.057875e-06 1.007374e-05 1.120354e-05 1.246005e-05 1.385748e-05
[37] 1.541163e-05 1.714009e-05 1.906240e-05 2.120030e-05 2.357798e-05 2.622232e-05
[43] 2.916322e-05 3.243396e-05 3.607152e-05 4.011704e-05 4.461628e-05 4.962013e-05
```

```
> first_diff_C=diff(data2$CRUDE)
> second_diff_C=diff(first_diff)
> third_diff_C=diff(second_diff)
> third_diff_C
 [1] 3.356818e-07 3.733294e-07 4.151994e-07 4.617651e-07 5.135534e-07 5.711499e-07
 [7] 6.352059e-07 7.064461e-07 7.856760e-07 8.737918e-07 9.717900e-07 1.080779e-06
[13] 1.201991e-06 1.336798e-06 1.486724e-06 1.653464e-06 1.838905e-06 2.045143e-06
[19] 2.274512e-06 2.529605e-06 2.813307e-06 3.128828e-06 3.479734e-06 3.869996e-06
[25] 4.304027e-06 4.786736e-06 5.323582e-06 5.920637e-06 6.584653e-06 7.323141e-06
[31] 8.144452e-06 9.057875e-06 1.007374e-05 1.120354e-05 1.246005e-05 1.385748e-05
[37] 1.541163e-05 1.714009e-05 1.906240e-05 2.120030e-05 2.357798e-05 2.622232e-05
[43] 2.916322e-05 3.243396e-05 3.607152e-05 4.011704e-05 4.461628e-05 4.962013e-05
```

The values are very low compared to the original values and they progress regularly. The data is smooth.

❖ Chi-Square test to identify values of EXPECTED and ZX

```
> #4)Chi-Square test
> data2$EXPECTED=data2$GRADUATED*data2$ETR
> data2$ZX=(data2$DEATHS-data2$EXPECTED)/(sqrt(data2$EXPECTED))
> chisq_data=data.frame(data2$DEATHS,data2$EXPECTED)
> view(chisq_data)
> chisq.test(chisq_data)

        Pearson's Chi-squared test

data:  chisq_data
X-squared = 905.26, df = 50, p-value < 2.2e-16
```

	AGE	ETR	DEATHS	CRUDE	GRADUATED	EXPECTED	ZX
1	25	78500	24	0.0003057325	0.0002379562	18.67956	1.23101698
2	26	80425	24	0.0002984147	0.0002646437	21.28397	0.58872001
3	27	81975	24	0.0002927722	0.0002943242	24.12723	-0.02590141
4	28	83725	24	0.0002866527	0.0003273335	27.40600	-0.65061102
5	29	84875	72	0.0008483063	0.0003640449	30.89831	7.39421962
6	30	85075	48	0.0005642081	0.0004048735	34.44462	2.30967653
7	31	85275	120	0.0014072120	0.0004502812	38.39773	13.16889792
8	32	86250	24	0.0002782609	0.0005007816	43.19241	-2.92028868
9	33	87250	72	0.0008252149	0.0005569456	48.59351	3.35774143
10	34	88300	72	0.0008154020	0.0006194087	54.69378	2.34009304
11	35	90200	24	0.0002660754	0.0006888771	62.13671	-4.83803632
12	36	92500	48	0.0005189189	0.0007661366	70.86764	-2.71642235
13	37	95425	24	0.0002515064	0.0008520610	81.30792	-6.35547801
14	38	98550	168	0.0017047184	0.0009476220	93.38815	7.72079292
15	39	99775	120	0.0012027061	0.0010539005	105.15293	1.44787209
16	40	99125	240	0.0024211854	0.0011720985	116.18426	11.48688332
17	41	99200	216	0.0021774194	0.0013035526	129.31242	7.62319029
18	42	101525	144	0.0014183699	0.0014497497	147.18584	-0.26259786
19	43	104525	120	0.0011480507	0.0016123433	168.53018	-3.73829061
20	44	107075	24	0.0002241420	0.0017931721	192.00391	-12.12451420
21	45	109125	216	0.0019793814	0.0019942815	217.62596	-0.11021875
22	46	109425	360	0.0032899246	0.0022179458	242.69871	7.52955017
23	47	109075	312	0.0028604171	0.0024666947	269.05472	2.61815370
24	48	110175	120	0.0010891763	0.0027433414	302.24764	-10.48287624
25	49	111675	168	0.0015043653	0.0030510149	340.72209	-9.35723950
26	50	112725	432	0.0038323353	0.0033931949	382.49789	2.53109907
27	51	115250	408	0.0035401302	0.0037737513	434.92484	-1.29105871
28	52	118225	600	0.0050750687	0.0041969882	496.18893	4.66036720
29	53	120025	702	0.0058487815	0.0046676923	560.23977	5.98918118
30	54	122150	891	0.0072943103	0.0051911872	634.10351	10.20183000
31	55	124350	513	0.0041254524	0.0057733935	717.92148	-7.64802003
32	56	125750	675	0.0053677932	0.0064208959	807.42766	-4.66043982
33	57	126350	864	0.0068381480	0.0071410176	902.26757	-1.27398179
34	58	127100	837	0.0065853659	0.0079419029	1009.41586	-5.42677915
35	59	129350	1242	0.0096018554	0.0088326098	1142.49807	2.94376912
36	60	132475	1593	0.0120249104	0.0098232120	1301.33000	8.08533529
37	61	134000	1539	0.0114850746	0.0109249130	1463.93834	1.96180929
38	62	133700	1998	0.0149439043	0.0121501729	1624.47811	9.26742594
39	63	134375	1728	0.0128595349	0.0135128491	1815.78910	-2.06019297
40	64	136125	2403	0.0176528926	0.0150283534	2045.73460	7.89889439
41	65	136625	1971	0.0144263495	0.0167138257	2283.52644	-6.54009028
42	66	136100	2835	0.0208302719	0.0185883286	2529.87152	6.06643452
43	67	135750	2889	0.0212817680	0.0206730622	2806.36820	1.55982154
44	68	134350	3348	0.0249199851	0.0229916046	3088.92208	4.66151324
45	69	131575	4212	0.0320121604	0.0255701781	3364.39619	14.61299835
46	70	129225	4428	0.0342658154	0.0284379460	3674.89357	12.42321426
47	71	128875	3915	0.0303782735	0.0316273420	4075.97371	-2.52138553
48	72	130075	5103	0.0392312128	0.0351744379	4575.31501	7.80125075
49	73	130475	5454	0.0418011113	0.0391193507	5104.09728	4.89765118
50	74	129550	6453	0.0498108838	0.0435066965	5636.29253	10.87851872
51	75	129400	6453	0.0498686244	0.0483860956	6261.16077	2.42443191

- Standardized deviations test

```
> table(cut(data2$ZX, breaks = seq.int(from = -20,to = 20, by= 4)))

(-20,-16] (-16,-12]  (-12,-8]   (-8,-4]    (-4,0]     (0,4]     (4,8]    (8,12]
        0         1         2         6        11        12        11         5
  (12,16]   (16,20]
        3         0
```

5) i. The graph is very wide compared to a standard normal graph, with several values higher than abs(10)

ii. The Values of the absolute deviations are very high relative to the expected value

iii. The lower bound is approx. -2 and the upper bound is approx. 6. IQR is 8. There are some outliers on the lower side but none on the upper side

iv. The Graph is positively skewed. It looks like it is centered about 2

v. The graduated rates do not represent the underlying mortality rates with accuracy

- Sign Test

```
> signtest=sign(data2$ZX)
> table(signtest)
signtest
-1  1
20 31
> dbinom(31,51,0.5)
[1] 0.03443253
```

since it's a two tailed test, at 5% significance level, we accept the NULL hypothesis that the data is a true representation of the underlying mortality rates.

- Cumulative deviations test

```
> obs=sum(data2$DEATHS)
> exp=sum(data2$EXPECTED)
> z=(obs-exp)/exp
> pnorm(-z)
[1] 0.4697664
```

at 5% significance level, we accept the NULL hypothesis that the data is a true representation of the underlying mortality rates.

- Serial Correlations test

```
> serialCorrelationTest(data2$ZX)

        Rank von Neumann Test for Lag-1 Autocorrelation (Beta Approximation)

data:  data2$ZX
RVN = 1.8355, p-value = 0.5563
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.1477139
```

at 5% significance level, we accept the NULL hypothesis that the data is a true representation of the underlying mortality rates.