Solutions to Worktext Exercises

Chapter 13

Visualizing Goodness-of-Fit Tests

Basic Learning Exercises

1. The variable is the winning time of the Kentucky Derby winner from 1950 through 1999. The units are seconds (measured to the nearest 1/5 second). The data are continuous, although measured in a slightly discrete way because the rules of horse racing evolved before electronic timing was possible (mechanical stopwatches were graduated in units of 1/5 second). Yes, a sample of 50 races should reveal the shape.

2. Secretariat (in 1973) had the fastest winning time of 119.4 seconds (z-value is –2.640). Tiny Tam (in 1958) had the slowest winning time of 125.0 seconds (z-value is 2.675). Neither is an outlier since both are within 3 standard deviations of the mean.

3. Mean 122.2 Median 122.2 1st Quartile 121.75 3rd Quartile 122.85
The mean and median are equal, suggesting symmetry. However, the distance from the first quartile to the mean (0.45) is slightly less than the distance from the mean to the third quartile (0.65), which could indicate a slightly longer right tail.

4. The data are symmetric (skewness coefficient only slightly less than zero) but the sample is more peaked than a normal distribution. The percent within 1 SD suggests that the data may have somewhat more concentration than a normal distribution.

	Skewness	Kurtosis	% within 1 SD	% within 2 SD	% within 2 SD
Sample	–0.13	3.71	72	94	100
If normal	0.00	3.00	68.27	95.44	99.73

5. Each histogram is somewhat bell-shaped, but there is a persistent spike near the middle of the distribution that exceeds the normal probability density function. This spike is most noticeable when k=5, k=7, or k=9.

6. The p-values range from 0.009 to 0.338. The hypothesis of normality is rejected twice (using 7 and 9 classes). It seems that this data differs somewhat from a normal distribution, but not enough to get a rejection most of the time at $\alpha = 0.05$.

	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9
Chi-square	1.405	3.416	1.999	9.406	3.372	10.452
D.F.	1	2	1	2	3	4
P-value	0.236	0.181	0.157	0.009	0.338	0.033
Decision	Accept	Accept	Accept	Reject	Accept	Reject

7. For k classes, the optimal expected frequency option defines class limits so that each expected frequency is n/k. The p-values range from 0.001 to 0.157. At $\alpha = 0.05$ the hypothesis of normality is rejected in three tests (using 5, 6, and 8 classes). This data differs somewhat from a normal distribution, but we still only get a rejection about half the time at $\alpha = 0.05$. The p-values are consistently smaller using optimal frequencies than using equal class intervals. This result suggests that the power of the test is enhanced when class limits are set to optimize the expected frequencies.

	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9
Chi-square	2.000	13.600	8.320	7.680	14.000	11.920
D.F.	1	2	3	4	5	6
P-value	0.157	0.001	0.040	0.104	0.016	0.064
Decision	Accept	Reject	Reject	Accept	Reject	Accept

Intermediate Learning Exercises

8. The respective d.f. are 1, 2, 1, 2, 3, 4. With equal width class intervals, the end class frequencies often are small. To enlarge them, end classes are combined. The number of groups for the chi-square test therefore varies, causing erratic d.f.

9. With optimal expected frequencies, each class is defined so as to force its expected frequency to be n/k. This creates unequal classes, but produce a more orderly d.f. progression (adding 1 d.f. each time we add a class). Even for k=9, we have n/k of at least 5 in every class, so there is no need to combine end classes.

10. Not always, since eventually n/k becomes small. Cochran's Rule suggests that there is cause for concern when expected frequencies fall below 5 (shown by yellow highlighting of cells in the chi-square table). The problem is that the chi-square statistic becomes inflated when the denominator in the sum approaches zero.

11. Two classes (121.1 to 121.6 and 122.0 to 122.4) account for well over half the chi-square test statistic (the sum at the bottom of the chi-square column). These are the categories in which there is an especially large difference between observed and expected frequencies. This shows up clearly on the graphs.

12. No. With equal expected frequencies the denominator in the chi-square calculation is the same for all classes, so a given difference always has the same weight. But for equal class intervals, each class generally has a different expected frequency, so the relative contribution of a given difference the chi-square calculation will vary.

13. The near-linearity of the point suggests that the normal model gives a good fit. Secretatiat's actual time (119.4) is below his expected time (about 119.8) if the data were normal. Conversely, Tiny Tam's actual time (125.0) is above what would be expected (about 119.8) if the data were normal. However, the distances from the 45^o line are not great, and the majority of the data points are near the 45^o line.

14. The huge chi-square test statistic (51.360) and small p-value ($p < 0.000$) indicate strong rejection of the hypothesis of a uniform distribution. The non-linear S-shape of the probability plot shows that the uniform distribution is a poor fit. The former test is mathematically precise, while the latter is visually vivid.

15. The frequency histogram's overlaid normal distribution is shifted right. The huge chi-square test statistic (31.686) and small p-value ($p < 0.000$) indicate rejection of the specified normal distribution. All the points on the probability plot lie above the 45^o line. The data are nearly normal, but the parameters are misspecified. The specified standard deviation is close to the sample standard deviation, but the specified mean is considerably above the sample mean.

16. The ECDF corresponds to the probability plot. You can see that the lowest data value is 119.4 and the highest is 125.0. The ECDF reveals that the data corresponds moderately well to a fitted normal distribution. The estimated median (0.50 proportion) is slightly above 122 which is consistent with the sample median (Q_2) of 122.2.

Advanced Learning Exercises

17. The K-S graph is identical to the ECDF graph except that the largest vertical difference is shown in magenta. This distance corresponds to the test statistic shown in the table of calculations as D-, the maximum difference of 0.138 for observation 36. The observed cumulative frequency is $F(x) = 0.720$ compared with a hypothesized cumulative frequency of 0.582, a difference of 0.138. At the bottom of the table of calculations, we learn that $p > 0.20$, so we know that this difference is not significant at $\alpha = 0.20$. In other words, the sample evidence is insufficient to cause us to reject H_0, the hypothesis that Kentucky Derby winning times are normally distributed with the fitted parameters.

18. The help file indicates that, using Harter's small-sample correction, for $\alpha = 0.20$ the Kolmogorov-Smirnov critical value is $1.073/(n + \Delta n)^{.5}$ where $\Delta n = (n+5)^{.5}/3.5$. Using $n = 50$ we get $D_{critical} = 0.1486$. Since $D_{calculated} = 0.138$ the calculated value does not quite exceed the critical value, so at $\alpha = 0.20$ we cannot reject the null hypothesis.

19. The help file says the K-S test may be preferred to the chi-square test for small samples, and is thought to be more powerful then the chi-square test in under most circumstances. This means that the K-S test might improve the chances of correctly rejecting a false hypothesis (e.g., detecting a departure from normality) compared with a chi-square test. It is applicable when the distribution is completely specified and the data are continuous.

20. The uniform distribution does not correspond to the data. The K-S test statistic is 0.213 ($p < 0.02$). In other words, we would reject the hypothesis of a uniform distribution at $\alpha = 0.02$. The exact p-value is not given.

21. All three tests strongly indicate that the hypothesis is to be rejected. However, the histogram suggests that the binomial distribution may be *a propos* but is shifted to the right. In other words, the students' scores may follow a binomial with a higher mean. The histogram is clearly discrete, and classes correspond to integers.

22. A success probability of 0.60 gives a good fit to the sample. This suggests that the number of correct answers may be regarded as a random variable with a binomial distribution, and that the students had about a 60 percent chance of getting each question right.

23. The normal approximation to this discrete data seems to give a good fit. The K-S hypothesized distribution is a step function rather than a smooth curve, and the chi-square classes are simply integers except for the open-ended classes at the ends.

24. The histogram resembles the fitted exponential, the ECDF follows the stated model, and the chi-square test gives a large p-value (varies with the number of classes). We cannot prove a null hypothesis, but using any of the common tests there is insufficient evidence to reject the stated hypothesis at $\alpha = 0.20$ except in the K-S test.