Workshop Statistics: Discovery with Data, Second Edition

Topic 15: Normal Distributions

Activity 15-4: Normal Curves

(a) mu = 50, sigma = 5
(b) mu = 1100, sigma = 300
(c) mu = -15, sigma = 45
(d) mu = 225, sigma = 75

Activity 15-5: Pregnancy Durations

(a) The z-score is (244-270)/15 = -1.73 . Looking up this value in Table II reveals that the proportion of pregnancies lasting less than 244 days is .0418.
(b) The z-score is (275-270)/15 = 0.33 . Looking up this value in Table II reveals that the proportion of pregnancies lasting more than 275 days is 1-.6293 = .3707.
(c) The z-score is (300-270)/15 = 2.00 . Looking up this value in Table II reveals that the proportion of pregnancies lasting over 300 days is 1-.9772 = .0228.
(d) The z-scores are (260-270)/15 = -0.67 and (280-270)/15 = 0.67. Looking up these two values in Table II reveals that the proportion of pregnancies lasting between 260 and 280 days is .7486-.2514 = .4972.
(e) 36 or fewer weeks corresponds to 36*7 = 252 or fewer days. The z-score is (252-270)/15 = -1.20, so Table II reveals that the normal model predicts that .1151 of all pregnancies last for 36 or fewer weeks. The observed proportion of such pregnancies is 436600/3880894 = .1125. The approximation of the normal model is very accurate in this case.

Activity 15-6: Professors' Grades

(a)

(b) Fisher: z-score = (90-74)/7 = 2.29, so the proportion scoring above 90 is 1-.9890 = .0110
Savage: z-score = (90-78)/18 = 0.67, so the proportion scoring above 90 is 1-.7486 = .2514
Savage gives a much higher proportion of A's.
(c) Fisher: z-score = (60-74)/7 = -2.00, so the proportion scoring below 60 is .0228
Savage: z-score = (60-78)/18 = -1.00, so the proportion scoring below 60 is .1587
Savage gives a much higher proportion of F's, as well.
(d) To be in the top 10%, the proportion below that score must be .9000. The z-score corresponding to this area is 1.28, so the score required is mu+1.28*sigma = 69+1.28*9 = 80.52.

Activity 15-7: Professors' Grades *(cont.)*

(a) A score of 90 is more impressive from Zeddes. Because his standard deviation is lower, he gives fewer scores farther from the mean of 75. With Zeddes the z-score is (90-75)/5 = 3.00, so only 1-.9987 = .0013 of his students score above 90. With Wells the z-score is (90-75)/10 = 1.50, so 1-.9332 = .0668 of his students score above 90.
(b) Likewise, a score of 60 is more discouraging from Zeddes. Again, because his standard deviation is lower, he gives fewer scores farther from the mean of 75. With Zeddes the z-score is (60-75)/5 = -3.00, so only .0013 of his students score below 60. With Wells the z-score is (60-75)/10 = -1.50, so .0668 of his students score below 60.

Activity 15-8: IQ Scores

(c) The z-score is (100-15)/12 = -1.25, so Table II reveals that .1056 of the students have IQ's below 100.
(d) The z-score is (130-15)/12 = 1.25, so Table II reveals that 1-.8944 = .1056 of the students have IQ's above 130.

(e) The z-scores are (130-15)/12 = 1.25 and (110-115)/12 = -0.42, so Table II reveals that .8944-.3372 = .5572 of the students have IQ's between 110 and 130.
(f) The z-score is (75-115)/12 = -3.33, so Table II reveals that only .0004 of the students would have an IQ lower than Forrest's.
(g) To be in the top 1% means that .9900 of the IQ's are lower. The z-score corresponding to this probability is 2.33. Thus, the IQ needs to be 115+2.33*12 = 142.96, or about 143.

Activity 15-9: Candy Bar Weights

(a) The z-score is (2.13-2.2)/.04 = -1.75, so Table II reveals that .0401 of the candy bars weigh less than the advertised weight of 2.13 ounces.
(b) The z-score is (2.25-2.2)/.04 = 1.25, so Table II reveals that 1-.8944 = .1056 of the candy bars weigh more than 2.25 ounces.
(c) The z-scores are (2.2-2.2)/.04 = 0 and (2.3-2.2)/.04 = 2.50, so Table II reveals that .9938-.5000 = .4938 of the candy bars weigh between 2.2 and 2.3 ounces.
(d) In order for only 1/1000 = .0010 of the candy bars to weigh less than 2.13 ounces, Table II reveals the necessary z-score to be -3.09. If the mean is changed from 2.2 to a new value of mu, the z-score would be (2.13-mu)/.04. Setting this equal to -3.09 and solving for mu gives mu = 2.13+3.09*.04 = 2.2536.
(e) The required z-score is still -3.09, but now that must be set equal to (2.13-2.2)/sigma. Solving gives that sigma = (2.13-2.2)/(-3.09) = .02265.
(f) The required z-score is still -3.09, but now that must be set equal to (2.13-2.15)/sigma. Solving gives that sigma = (2.13-2.15)/(-3.09) = .00647.

Activity 15-10: SATs and ACTs *(cont.)*

(a) Bobby's z-score is (1080-896)/174 = 1.06. Table II reveals that 1-.8554 = .1446 of the applicants scored higher than Bobby.
(b) Kathy's z-score is (28-20.6)/5.2 = 1.42. Table II reveals that 1-.9222 = .0778 of the applicants scored higher than Kathy.
(c) The z-scores and the normal probability calculations indicate that Kathy scored better relative to her peers than Bobby did.

Activity 15-11: Heights

(a) Since 5'6" is a total of 66 inches, the z-score is (66-70)/3 = -1.33. Table II reveals that .0918, or 9.18%, of American men in this age group are shorter than 5'6" tall.
(b) Since 6' is 72 inches, the z-score is (72-70)/3 = 0.67. Table II reveals that 1-.7486, or 25.14%, of American men in this age group are taller than 6' tall.
(c) To be in the tallest 10% requires a z-score of 1.28. Thus, the height needed to be in the top 10% is 70+1.28*3 = 73.84 inches.
(d) The z-score for 66 inches is (66-65)/3 = 0.33, so Table II says that .6293, or 62.93%, of American women in this age group are shorter than 5'6" tall. The z-score for 72 inches is (72-65)/3 = 2.33, so Table II says that 1-.9901, or 0.99%, of American women in this age group are taller than 6' tall. A z-score of 1.28 corresponds to 65+1.28*3 = 68.84 inches, which is how tall a woman has to be in order to reach the tallest 10%.

(e) The normal model predicts 25.14% of men to be 6' or taller, which is only somewhat close to the reported 29.9%. The normal model predicts 0.99% of men to be 6' or taller, which again is only somewhat close to the reported 0.5%.

Activity 15-12: Weights

(a) The respective z-scores for these weights are (150-175)/35 = -0.71, (200-175)/35 = 0.71, and (250-175)/35 = 2.14. Table II reveals the associated percentages weighing less than these weights to be .2389, .7611, and .9838, respectively.
(b) The respective z-scores for these weights are (150-140)/30 = 0.33, (200-140)/30 = 2.00, and (250-140)/30 = 3.67. Table II reveals the associated percentages weighing less than these weights to be .6293, .9772, and .9999, respectively. (Actually, the z-score of 3.67 is off the table, so the table only reveals that the proportion below is at least .9998. Technology gives a more accurate value of .9999.)
(c) The normal model gives predictions somewhat close to the observed proportions.

Activity 15-13: Coin Ages

The distribution of coin ages must be skewed to the right, because no coin can have an age less than zero years. The large standard deviation (compared to the mean) indicates that there must be some very large values of age, whereas there can not be small ones below zero. Thus, the normal distribution does not provide a reasonable model for these data.

Activity 15-14: Empirical Rule

(a) Table II reports the area to the left of 1.00 to be .8413 and the area to the left of -1.00 to be .1587, so the probability of falling within one standard deviation of the mean with a normal curve is .8413-.1587 = .6826.
(b) Table II reports the area to the left of 2.00 to be .9772 and the area to the left of -2.00 to be .0228, so the probability of falling within two standard deviations of the mean with a normal curve is .9772-.0228 = .9544.
(c) Table II reports the area to the left of 3.00 to be .9987 and the area to the left of -3.00 to be .0013, so the probability of falling within three standard deviations of the mean with a normal curve is .9987-.0013 = .9974.
(d) Due to the symmetry of the standard normal curve, we want the z-score for which the area to the left is .7500 and then also .2500. From Table II, these z-scores are (roughly) 0.67 and -0.67. Thus, the IQR of the standard normal distribution is 0.67-(-0.67) = 1.34. The IQR of any normal distribution is therefore 1.34*sigma.
(e) 1.5*IQR = 1.5*1.34 = 2.01. Thus, Q3+1.5IQR = 2.68 and Q1-1.5IQR = -2.68, so outliers are those values lying above 2.68 or below -2.68. Table II reveals each of these probabilities to be .0037, so the probability of an outlier in either direction is 2*.0037 = .0074.

Activity 15-15: Critical Values

(a) The probability of being less than z* must be .9000, and Table II reveals that this z-score is 1.28.
(b) The probability of being less than z* must be .9500, and Table II reveals that this z-score is 1.645 (the midpoint between 1.64 and 1.65)..
(c) The probability of being less than z* must be .9750, and Table II reveals that this z-score is 1.96.
(d) The probability of being less than z* must be .9900, and Table II reveals that this z-score is 2.33.
(e) The probability of being less than z* must be .9950, and Table II reveals that this z-score is 2.58.

Activity 15-16: Random Normal Data

The samples that come from non-normal populations are (b) because of the skew to the right, (e) because of the skew to the left, and (c) because of the dual peaks.