1. (a) categorical
   (b) quantitative
   (c) quantitative
   (d) quantitative
2. colleges
3. Some possibilities:
   - What is the mean graduation rate for colleges in the population under consideration?
   - What is the mean amount of time colleges in the population require students to live *on campus*?
   - What is the mean annual tuition for colleges in the population under consideration?
   - What proportion of colleges under consideration have a *zip code* that starts with 4?
4. Some possibilities:
   - Is there a positive correlation between tuition costs and graduation rate?
   - Do colleges in CA (requires identifying those zip codes that correspond to this state) have higher mean tuition than the rest of the country?
5. (a) scatterplot, since both variables are quantitative
   (b) stacked (or side-by-side) bar chart or two-way table, since both variables are categorical
6. (a) explanatory: *whether media is used in bedroom*; response: *level of tiredness*
   (b) The values are "yes" (if one watches TV or uses the internet in the bedroom), and "no" (otherwise).
   (c) This is an observational study. The researchers do not assign participants to one group (the "uses media" group) or the other (the "does not use media" group).
   (d) To automatically use the person who answers the door as the respondent may create a biased sample, tending to obtain responses from only the more "alert" members of households. (There are, of course, other ways this study might encourage the collection of a biased sample, and other steps one should take to avoid further sources of bias.)
7. (a) sample
   (b) population
   (c) sample
   (d) population
8. (a) median
   (b) IQR
9. (a) This study is observational. Presumably the explanatory variable would be "*which hand?*" (with values "dominant" and "off"), and this value cannot be assigned to participants.
   (b) The original design is preferable, as each participant contributes a value to both groups. Scars on her dominant hand can be compared with scars on her “off" hand, so if there are differences in those two numbers, it is suitable to attribute that difference to how she makes different use of dominant vs. off hand. When comparing her use of dominant hand, along with the other members of the dominant hand group, with scar counts on the off hand for a non-overlapping group of people, we will see variations in response values that arise for other “lifestyle" reasons.
10. (a) Write two phrases of equal difficulty (or perhaps one nonsense phrase which participants will tend not to "memorize" after one attempt) to be "texted." Randomly assign participants to two groups, one that will use the dominant hand first, the other that will use the "off" hand first. Have each participant text a phrase using both hands, and subtract the times (time with off hand minus time with dominant one) to find the *time difference*.

(b) Mine is matched pairs. The list of times with "off" hand are not independent from the list of times for dominant hand, as faster "texters" using the dominant hand are likely to tend toward being the faster texters with the off hand as well. Since each subject contributes a time in both list, we look at time differences on a subject-by-subject basis.

11. The mean is larger.

12. (a) $\dfrac{228+85}{1018} \doteq 0.307$

 (b) $\dfrac{63}{85} \doteq 0.741$

 (c) We have

$$\text{HS or less: } \frac{80}{380} \doteq 0.211, \quad \text{some college: } \frac{133}{325} \doteq 0.409, \quad \text{college grad: } \frac{121}{228} \doteq 0.531, \quad \text{postgrad: } \frac{63}{85} \doteq 0.741.$$

 As these proportions are not approximately equal, there is a strong suggestion that the variables have an association.

 (d) barchart, or frequency/relative frequency table

13. (a) The margin of error grows.

 (b) The margin of error grows.

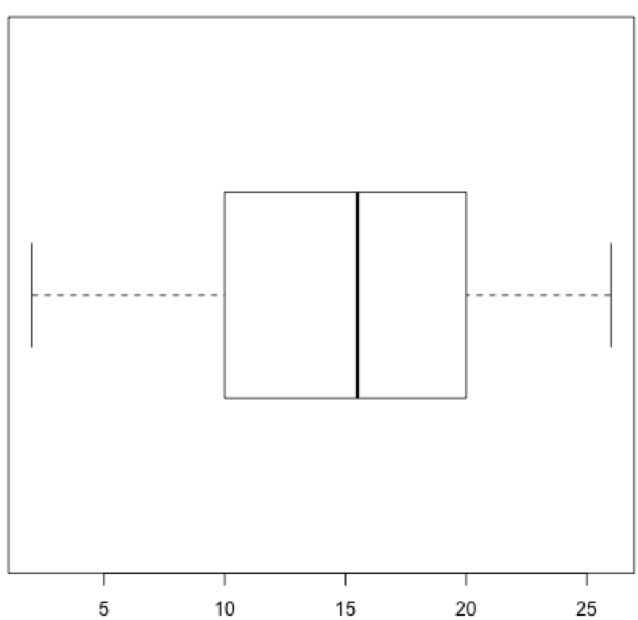
14. (a) 15.5

 (b) 14.786

 (c) Q1 = 10, Q3 = 20, and IQR = 10.

 (d) $26 - 2 = 24$

 (e)



15. (a) The mean is $\frac{1}{4}(13+22+31+34) = 25$.
 The variance is $\frac{1}{3}[(13-25)^2+(22-25)^2+(31-25)^2+(34-25)^2] = \frac{1}{3}(144+9+36+81) = \frac{270}{3} = 90$.
 The standard deviation is $s = \sqrt{90} \doteq 9.487$.

 (b) $z = \dfrac{34-25}{9.487} \doteq 0.949$

16. (a) A negative association here means that people with lower GPAs tend to have more facebook friends while people with higher GPAs tend to have fewer facebook friends.

 (b) The study is observational, not an experiment, and so even if a relationship is established, we cannot know (from this study) that it is a *causal* one. There are possible lurking variables linked both to *GPA* and *number of facebook friends* which remain unaltered by the action of "unfriending," and hence it is quite likely this action will have no effect on one's GPA.

17. (a) One dot represents the mean weight $\bar{x}$, in kg, computed from a random sample of 10$^{\text{th}}$ grade boys.

    (b) The population mean should be at the center of the distribution, making it appear to be about 65.

    (c) It is closest to 2. It appears that one would get about 95% of observed $\bar{x}$ values between 61 and 69, and the distance from 65 to 69 is twice the standard error.

    (d) The 95% CI from the last two answers would be

$$65 \pm 2(2), \qquad \text{or} \qquad [61, 69].$$

18. (a) The correlation is $r = 0.83$.

    (b) We have $\quad (\text{enrolled}) = 88.9 + (0.208)\,(\text{accepted}).$

    (c) For (roughly) every 5 additional students accepted, there is approximately one more that enrolls.

    (d) Using the regression line, we have

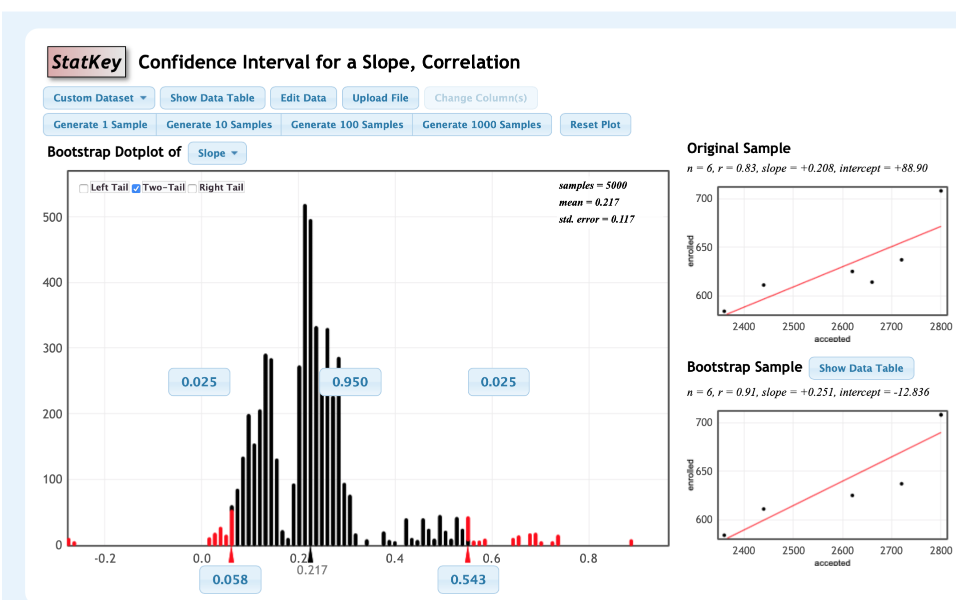
$$(\text{predicted enrollment}) = 88.9 + (0.208)(2575) = 624.5.$$

    (e) The residual is $\quad (\text{observed}) - (\text{predicted}) = 671 - 624.5 = 46.5.$

    (f) One could do this using percentiles or not. Either way, we must enter these data into the option "CI for Slope, Correlation" in StatKey and be sure to select the "Slope" option not "Correlation." Many bootstrap samples leads to the bootstrap distribution pictured below. From the two-tail info, we have the 95% bootstrap percentile CI $[0.058, 0.543]$. If, instead we do the construction as (point estimate) ± ME, then it is

$$0.208 \pm 2(0.117), \qquad \text{or} \qquad [-0.026, 0.442].$$

    Both methods are suspect, as the bootstrap distribution is not very *normal*-looking.



    (g) Under repetitions of the process we followed in constructing this 95% interval, there is a 95% success rate that the desired population parameter lies inside the result. Either our true slope lies inside this interval, or we got unlucky and began with one of those samples that, even under the best of circumstances (i.e., a random sample) occurs about 5% of the time, and results in a CI not containing the population parameter (true slope).