1. (i) and (iv)

2. (iii)

3. (a) Since 1470 is more than two standard deviations above the mean 1050, it is beyond the nearest 95% of observations in proximity to the mean. Said another way, it is past the 0.975 quantile, thus also beyond the 96$^{\text{th}}$ percentile.

   (b) The two standardized Z-scores:

   $$\text{You:} \quad \frac{1200-1050}{200} = 0.75$$

   $$\text{Your friend:} \quad \frac{27-20}{6} = 1.167$$

   Your friend's performance is the one with the higher Z-score, and should be ranked as higher.

4. (iii)

5. (a) The left tail consists of observations as extreme or more extreme than $\widehat{p}_1 - \widehat{p}_2 = -0.246$, and these occur with relative frequency 0.028. For a two-tailed alternative, we double this, arriving at a P-value of 0.056.

   (b) One approach is to write "Intelligence" on 29 slips of paper and "effort" on 30 slips, placing all 59 in one bag and mixing them up. In advance, you can have decided the first 15 draws (doing so without replacement) from the bag will be treated as the 15 liars, and the rest as the ones who do not lie. Once the bag is emptied, you can calculate the proportion of "Intelligence" slips considered honest (i.e., what proportion of the slips marked that way were among the final 44 slips drawn), calling that $\widehat{p}_1$, and calculating the analogous proportion of slips marked "Effort". Finally, you subtract the two: $\widehat{p}_1 - \widehat{p}_2$.

   (c) We first compute the pooled proportion:

   $$\tilde{p} = \frac{44}{59},$$

   and using that along with sample sizes n_1, n_2, we get an estimated standard error

   $$\text{SE}_{\widehat{p}_1-\widehat{p}_2} = \sqrt{\frac{44}{59}\cdot\frac{15}{59}\cdot\left(\frac{1}{29}+\frac{1}{30}\right)} \doteq 0.1134,$$

   not a huge amount different than the SE reported on the randomization distribution dot plot, but it is a bit different. To standardize:

   $$z = \frac{-0.246}{0.1134} = -2.1693.$$

   (d) In this case,

   $$n_1\widehat{p}_1 = 18, \quad n_1(1-\widehat{p}_1) = 11, \quad n_2\widehat{p}_2 = 26, \quad n_2(1-\widehat{p}_2) = 4,$$

   so one of the rules of thumb for justifying a normal model is not met. So, the randomization method is somewhat more reliable, as it just uses frequencies and makes no assumptions.

6. (a) (ii)

   (b)

   (c) We would use

```
qt(0.97, df=35)
```

(d) Since the null value 0 is outside the 94% confidence interval, out in the rejection region, our P-value must be less than 0.06.

7. (i)

8. (a) A bootstrap distribution for a sample mean $\overline{x}$ repeatedly draws a bootstrap sample, and computes $\overline{x}$ as the bootstrap statistic. Doing so makes the original sample mean, not the null value, the center of this distribution. The important extra step is to add to each $\overline{x}$ a number equal to the difference between the null value 5.7 and the test statistic, given in part (b) to be 5.552.

   (b) The standardized test statistic is

$$\frac{5.552 - 5.7}{0.552/\sqrt{50}} \doteq -1.8959.$$

   (c) *(3 pts)* Write an RStudio command that generates a P-value, using your test statistic from part (b). We could use

```
pt(-1.8959, df=49) * 2
```

9. (a) True A statistically significant result is not always of practical importance.

   (b) False When you reject a null hypothesis that is, in fact, true, you have committed a Type II error.

   (c) False When your P-value is 0.213, you will reject the null hypothesis at the significance level $\alpha = 0.05$.

   (d) True When you flip a coin 50 times, recording whether you got "heads" or "tails" on each flip, you can view this as an i.i.d. random sample taken from the population of all Hs and Ts that coin can possibly flip.

   (e) False A P-value represents the chance of obtaining a result at least as extreme as our test statistic when the null hypothesis is false.

   (f) True As you take steps to decrease the chance of Type I error, you correspondingly increase the chance of Type II error.