Exam 1 Practice Problems

Penguins are cute

package: `palmerpenguins`

Question Set 1

Let's MANUALLY calculate the correlation between two variables X and Y

```
##          X   Y
##  [1,]   6  45
##  [2,]  12  47
##  [3,]  13  39
##  [4,]  17  58
##  [5,]  22  68
##  [6,]  25  76
##  [7,]  27  75
##  [8,]  29  74
##  [9,]  30  78
## [10,]  32  81
```

Question Set 2

- Scientists have discovered that 14% of Adelie penguins have a genetic condition that prevents them from being able to reproduce. If 40 Adelie penguins are selected at random from any of the 3 islands, how many penguins could be expected to be able to reproduce?

```
## 
##             Biscoe Dream Torgersen
##   Adelie        44    56        52
##   Chinstrap      0    68         0
##   Gentoo       124     0         0
```

Question Set 3

- For an upcoming stats exam we have decided to be lazy and only make 5 multiple choice questions. The questions have 4 response options each. How likely are students to get *at least 4* of these questions correct simply by guessing? Calculate this "by hand" meaning do not use a R function

Question Set 4

- Get the z-scores of the `body_mass_g` variable "by hand" meaning do not use a R function
- Which value is furthest away from the mean? Which is closest?
- Interpret the 10th z-score
- Which of these is an outlier? How do you propose to deal with this outlier?

```
## # A tibble: 10 × 4
##    bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
##             <dbl>         <dbl>             <int>       <int>
##  1           39.1          18.7               181        3750
##  2           39.5          17.4               186        3800
##  3           40.3          18                 195        3250
##  4           36.7          19.3               193        3450
##  5           39.3          20.6               190        3650
##  6           38.9          17.8               181        3625
##  7           39.2          19.6               195        4675
##  8           34.1          18.1               193        3475
##  9           42            20.2               190        4250
## 10           37.8          17.1               186        3300
```

Question Set 5

- Which of these variables are qualitative? Quantitative?
- Which variable would be appropriate for a binomial question?
- Which variable do you assume is normally distributed? t distributed?
- Calculate the standard error of the mean for `bill_depth_mm` and describe what it means

```
## # A tibble: 10 × 8
##    species island    bill_length_mm bill_depth_mm flipper_length_mm
##    <fct>   <fct>              <dbl>         <dbl>             <int>
##  1 Adelie  Torgersen           39.1          18.7               181
##  2 Adelie  Torgersen           39.5          17.4               186
##  3 Adelie  Torgersen           40.3          18                 195
##  4 Adelie  Torgersen           36.7          19.3               193
##  5 Adelie  Torgersen           39.3          20.6               190
##  6 Adelie  Torgersen           38.9          17.8               181
##  7 Adelie  Torgersen           39.2          19.6               195
##  8 Adelie  Torgersen           34.1          18.1               193
##  9 Adelie  Torgersen           42            20.2               190
## 10 Adelie  Torgersen           37.8          17.1               186
## # i 3 more variables: body_mass_g <int>, sex <fct>, year <int>
```