Name:
Date:

Quiz #2

In 2021, California passed Senate Bill 129, which appropriated funds for wildfire prevention, forest resiliency, and related matters. The California Energy Commission received $20M under this bill to “assist local governments in California with establishing online, automated solar permitting”, creating the CalAPP program. CalAPP’s mission is “to support a grant program for California cities, counties, or cities and counties to establish online solar permitting.” CalAPP funded proposals from 350 jurisdictions.

Imagine it’s now 2035 and you want to know if CalAPP succeeded. Consider how you would design a study to test the impact of the policy and answer the questions below.

1. What variable(s) would you want to collect for your study?

2. What’s the population you’re interested in?

3. What’s a sample you could draw?

4. What statistic would you use to numerically summarize the sample?

5. On the back of this page, sketch a figure or table that depicts what the data in your sample could look like.

Name:
Date:

Quiz #3

There are three code chunks below and two outputs. Match the code chunks to the outputs and identify the distractor.

```
penguins %>%
  count(species, island) %>%
  group_by(island) %>%
  mutate(n = n / sum(n)) %>%
  ungroup() %>%
  pivot_wider(names_from = island,
              values_from = n,
              values_fill = 0)
```

```
# A tibble: 3 × 4
  species   Biscoe Dream Torgersen
  <fct>      <dbl> <dbl>     <dbl>
1 Adelie     0.262 0.452         1
2 Chinstrap  0     0.548         0
3 Gentoo     0.738 0             0
```

```
penguins %>%
  count(species, island) %>%
  group_by(species) %>%
  mutate(n = n / sum(n)) %>%
  ungroup() %>%
  pivot_wider(names_from = species,
              values_from = n,
              values_fill = 0)
```

```
# A tibble: 3 × 4
  island    Adelie Chinstrap Gentoo
  <fct>      <dbl>     <dbl>  <dbl>
1 Biscoe     0.289         0      1
2 Dream      0.368         1      0
3 Torgersen  0.342         0      0
```

```
penguins %>%
  count(species, island) %>%
  group_by(species, island) %>%
  mutate(n = n / sum(n)) %>%
  ungroup() %>%
  pivot_wider(names_from = island,
              values_from = n,
              values_fill = 0)
```

Name:
Date:

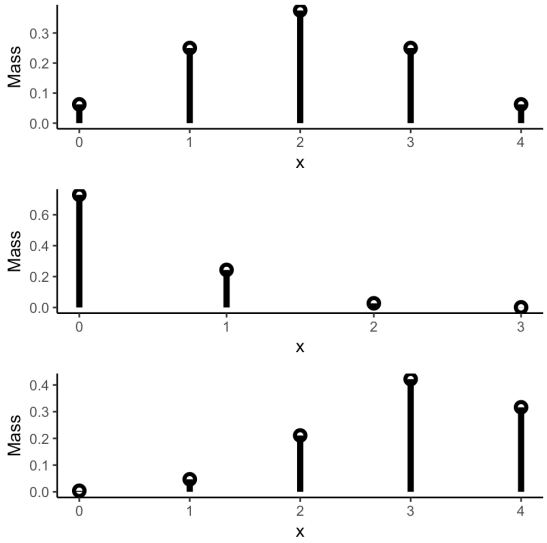
Quiz #4

The PMFs on the right correspond to the following random variables. Which is which?

Binomial(3, 0.1)

Binomial(4, 0.5)

Binomial(4, 0.75)

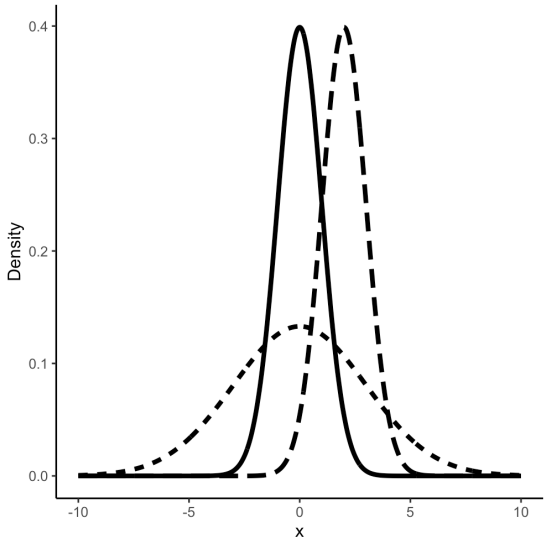


The figure on the right has PDFs for the following random variables. Which is which?

Normal(0, 1)

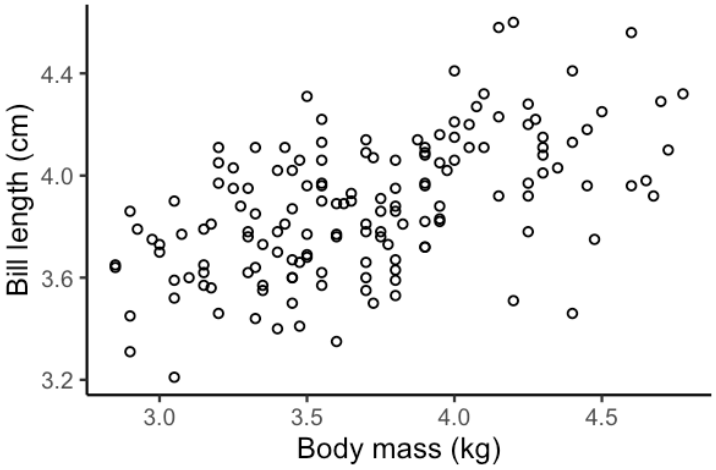
Normal(0, 3)

Normal(2, 1)



Name:
Date:

Quiz #5



The figure above displays body size data for Adélie penguins. Use it to answer the following questions.

1. Write the statistical notation for a linear regression model using these data.

2. Draw a line through the data that you think approximates the best fit line
3. Using the line you drew, estimate values for the coefficients from question 1

4. How would you estimate σ?

Name:
Date:

Quiz #6

```
> penguin_mod <- lm(bill_length_mm ~ bill_depth_mm, penguins)
> summary(penguin_mod)

Call:
lm(formula = bill_length_mm ~ bill_depth_mm, data = penguins)

Residuals:
     Min       1Q   Median       3Q      Max
-12.8949  -3.9042  -0.3772   3.6800  15.5798

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    55.0674     2.5160  21.887  < 2e-16 ***
bill_depth_mm  -0.6498     0.1457  -4.459 1.12e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.314 on 340 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.05525,   Adjusted R-squared:  0.05247
F-statistic: 19.88 on 1 and 340 DF,  p-value: 1.12e-05

> summary(penguin_mod)$sigma
[1] 5.314418
```

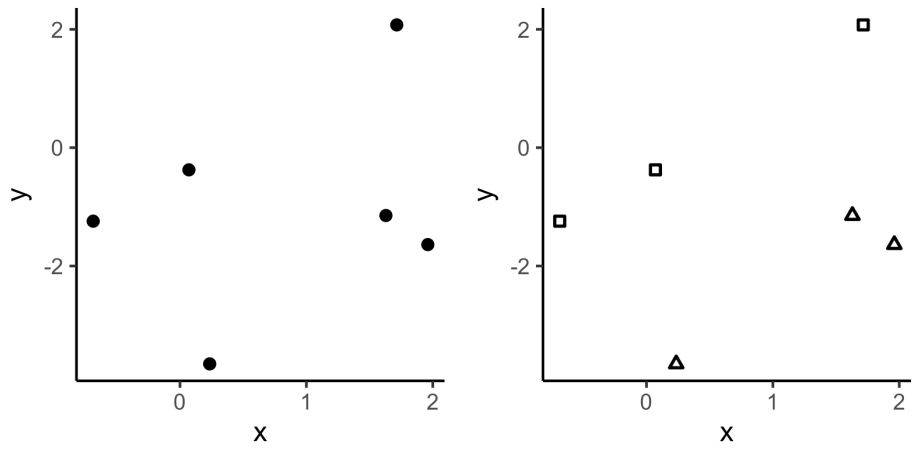
Using the model shown above, answer the following questions.

1. Approximately, what's the expected bill length of a penguin with a 20mm deep bill?

2. Approximately, in what range would you expect to find 67% of the observed bill lengths for penguins with 20mm deep bills?

Name:
Date:

Quiz #7



The two figures above show the same x-y data. The one on the right encodes a second predictor variable as shapes.

On the left figure, sketch the line for $\hat{y}$ for the following model.

$$y \sim Normal(\mu, \sigma)$$
$$\mu = \beta_0 + \beta_1 x$$

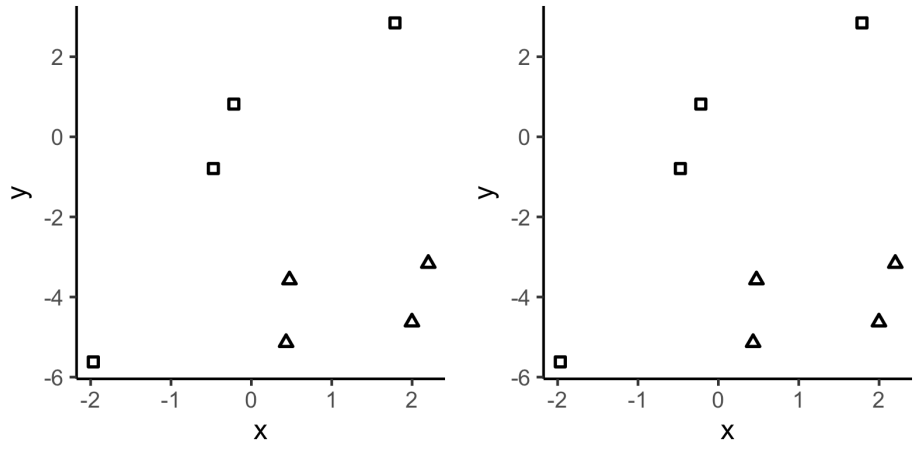
On the right figure, sketch the line for $\hat{y}$ for the following model (where z is represented by the shapes).

$$y \sim Normal(\mu, \sigma)$$
$$\mu = \beta_0 + \beta_1 x + \beta_2 z$$

In both figures, graphically indicate what each coefficient represents.

Name:
Date:

Quiz #8



The two figures above show the same data: a continuous response with two predictors (one continuous, one categorical).

On the left figure, sketch the line for $\hat{y}$ for the following model, where z is represented by the shapes.

$$y \sim Normal(\mu, \sigma)$$
$$\mu = \beta_0 + \beta_1 x + \beta_2 z$$

On the right figure, sketch the line for $\hat{y}$ for the following model.

$$y \sim Normal(\mu, \sigma)$$
$$\mu = \beta_0 + \beta_1 x + \beta_2 z + \beta_3 xz$$

Roughly estimate β_2 and β_3 in the second model.