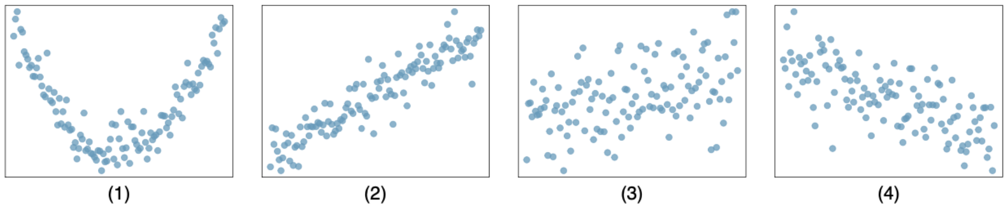
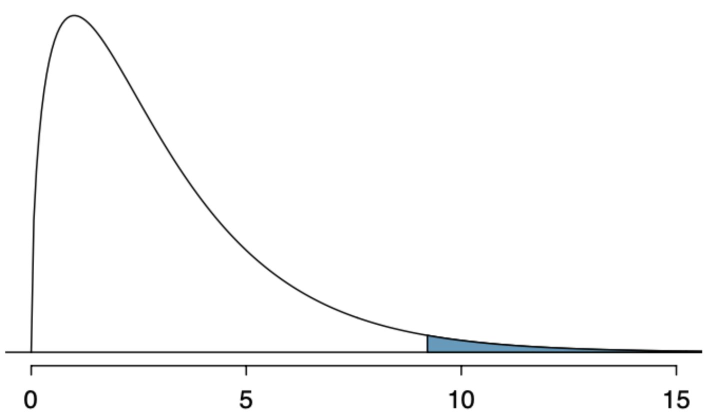
1. For each of the following scatter plots, match them to the Pearson correlation coefficient that best describes the correlation displayed.



   (a) $R = -0.7$
   (b) $R = 0.45$
   (c) $R = 0.06$
   (d) $R = 0.92$

2. A test of homogeneity is performed to determine if the distributions of two populations are the same and finds a χ^2 statistic of 10. Pictured below is the appropriate chi-square distribution with a shaded area of .05.



   Based on this information, what is the appropriate conclusion of the hypothesis test at a 10% significance level?

3. The table below summarizes a data set examining the responses of a random sample of 850 college graduates and non-graduates on the topic of oil drilling. We want to know if graduating college affects people's positions on oil drilling.

	College Grad	Non-College Grad
Support	154	155
Oppose	175	131
No Opinion	104	131

(a) Write appropriate hypotheses for your test.

(b) Is it expected for there to be more college grads in support or opposition of oil drilling? (show work to justify your answer)

(c) How many degrees of freedom define the appropriate chi-square distribution to use for your test. (show work to justify your answer)

4. The following table lists the data observed from 60 die rolls of a 4 sided die.

Outcome	1	2	3	4
Frequency	14	12	19	15

We will use a chi-square goodness-of-fit test to detect if the die is fair (each outcome is equally likely) at a 5% significance level.

(a) Compute χ^2. (show work to justify your answer)

(b) To find the p-value using `R` you will need the value of χ^2 found in the previous part and which of the following `R` commands? (circle the correct option)

i. `pchisq(`χ^2`,15)`
ii. `1-pchisq(`χ^2`,15)`
iii. `pchisq(`χ^2`,3)`
iv. `1-pchisq(`χ^2`,3)`
v. `pchisq(`χ^2`,8)`
vi. `1-pchisq(`χ^2`,8)`

(c) The output of the correct command from the previous part is `0.629`. Write the appropriate conclusion of the test within the context of the problem.

5. Determine if the following are true or false. Explain your reasoning.
   (a) The chi-square distribution, just like the normal distribution, has two defining parameters, mean and standard deviation.

   (b) The chi-square statistic is always positive.

   (c) As you increase the defining parameters, the shape of the chi-square distribution becomes more skewed.

   (d) A chi-square test is never two-tailed.

6. How do you distinguish among the infinitely many different chi-square distributions and their corresponding χ^2-curves?

7. Explain why a chi-square goodness-of-fit test, a chi-square independence test, or a chi-square homogeneity test is always right tailed.

8. If the observed and expected frequencies for a chi-square goodness-of-fit test, a chi-square independence test, or a chi-square homogeneity test matched perfectly, what would be the value of the test statistic?

9. Researchers studying anthropometry collected body girth measurements and skeletal diameter measurements, as well as age, weight, height and gender for 507 physically active individuals. The mean shoulder girth is 107.20 cm with a standard deviation of 10.37 cm. The mean height is 171.14 cm with a standard deviation of 9.41 cm. The correlation between height and shoulder girth is 0.67.

   (a.) Write the equation of the regression line for predicting height.

   (b.) Interpret the slope and the intercept in this context.

   (c.) Calculate R^2 of the regression line for predicting height from shoulder girth, and interpret it in the context of the application.

   (d.) A randomly selected student from your class has a shoulder girth of 100 cm. Predict the height of this student using the model.

   (e.) The student from part (d) is 160 cm tall. Calculate the residual, and explain what this residual means.

   (f.) A one year old has a shoulder girth of 56 cm. Would it be appropriate to use this linear model to predict the height of this child?

10. Below is a least squares linear model for predicting wife's age based on husband's age in a sample of 170 married British couples.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.5740	1.1501	1.37	0.1730
age_husband	0.9112	0.0259	35.25	0.0000
				$df = 168$

(a) Write the equation of the regression line for predicting wife's age from husband's age.

(b) Interpret the slop and intercept in the context of the problem

(c) Given that $R^2 = 0.88$, what is the correlation of ages in this data set?

(d) You meet a married man from Britain who is 55 years old. What would you predict his wife's age to be?

(e) Provide a 95% confidence interval for the slope.

Throughout, X is a random variable and x_i is a particular value of X.

$$E(X) = \sum x_i P(X = x_i)$$

$$\mu = \frac{\sum x_i}{N}$$

$$\sigma(E(X)) = \sqrt{\sum (x_i - E(X))^2 P(X = x_i)}$$

$$\overline{X} = \frac{\sum x_i}{n}$$

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$\sigma = \sqrt{\frac{\sum(\mu - x_i)^2}{N}}$$

$$\left(\frac{\bar{x} - \mu}{s/\sqrt{n}}\right) \sim t(n-1)$$

$$s = \sqrt{\frac{\sum(\overline{X} - x_i)^2}{n-1}}$$

$$\hat{P} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

$$IQR = Q_3 - Q_1$$

$$\overline{X_1} - \overline{X_2} \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

$$\text{whiskers} = \begin{cases} Q_1 - 1.5 \cdot IQR \\ Q_3 + 1.5 \cdot IQR \end{cases}$$

$$z\text{-score} = \frac{\text{observation} - \text{expected}}{\text{standard deviation}}$$

$$\hat{P}_1 - \hat{P}_2 \sim N\left(p_1 - p_2, \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right)$$

$$t\text{-score} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

$$R = \frac{1}{N-1} \sum \frac{x_i - \bar{X}}{s_X} \frac{y_i - \bar{Y}}{s_Y}$$

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

$$b_1 = \frac{s_Y}{s_X} R$$