Exam 2 Practice
Stat 630, Fall 2021

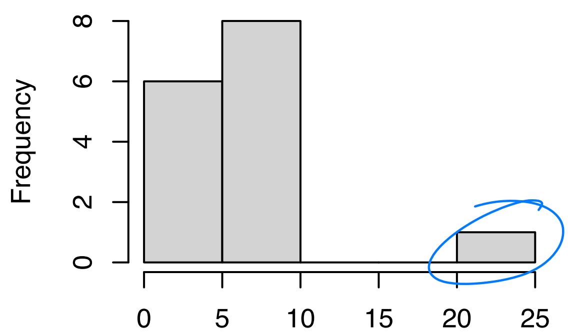
Exercise 1. In each of the following scenarios determine if the data are paired.

(a) Paired We would like to know if Intel's stock and Southwest Airlines' stock have similar rates of return. To find out, we take a random sample of 50 days, and record Intel's and Southwest's stock on those same days.

(b) Paired We randomly sample 50 items from Target stores and note the price for each. Then we visit Walmart and collect the price for each of those same 50 items.

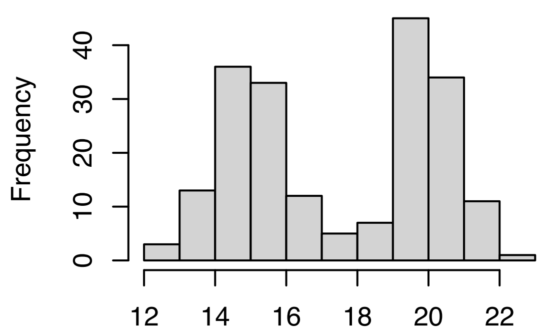
(c) Not Paired A school board would like to determine whether there is a difference in average SAT scores for students at one high school versus another high school in the district. To check, they take a simple random sample of 100 students from each high school.

Exercise 2. The following is a histogram of a random sample of $n = 15$ observations from a population. Would it be reasonable to compute a 95% confidence interval for μ using this data set? That is, are the conditions for inference satisfied?



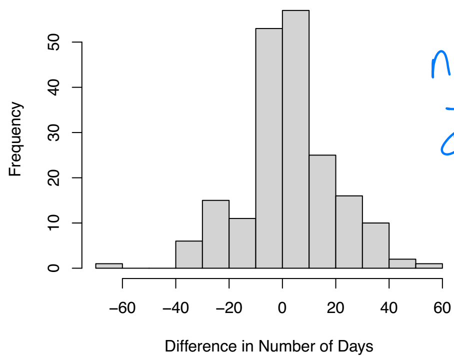
No. The sample size $n = 15 < 30$ is small, and there is an outlier.

Exercise 3. The following is a histogram of a random sample of $n = 200$ observations from a population. Would it be reasonable to compute a 95% confidence interval for μ using this data set? That is, are the conditions for inference satisfied?



Yes. The sample size $n = 200 > 30$ is large. So the sampling distribution for $\bar{X}$ is approx normal according to the CLT.

Exercise 4. [From *OpenIntro* Ch. 7.2] Let's consider a limited set of climate data, examining temperature differences in 1948 vs 2018. We sampled 197 locations from the National Oceanic and Atmospheric Administration's (NOAA) historical data, where the data was available for both years of interest. We want to know: were there more days with temperatures exceeding 90°F in 2018 or in 1948? The difference in number of days exceeding 90°F (number of days in 2018 - number of days in 1948) was calculated for each of the 197 locations. The average of these differences was 2.9 days with a standard deviation of 17.2 days. We are interested in determining whether these data provide strong evidence that there were more days in 2018 that exceeded 90°F from NOAA's weather stations.



$n = 197$

$\bar{d} = 2.9$

$s_d = 17.2$

location	1948	2018	Diff
1			
2			
⋮			
197			

(a) Are these data from two independent samples, or are these data paired?

Paired

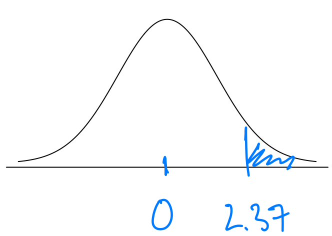
(b) Write the null and alternative hypothesis for a one-sided test.

$H_0: \mu_d = 0$ $\quad H_A: \mu_d > 0$

(c) Are the conditions for the hypothesis test satisfied? Yes

- Locations randomly sampled (independence)
- Large sample size: $n = 197 > 30$ locations

(d) Calculate the test statistic and p-value, and make a decision using $\alpha = 0.05$ significance level.



$$z = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{2.9}{17.2/\sqrt{197}} = 2.37$$

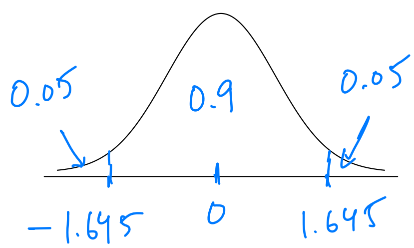
p-value $= 1 - \text{pnorm}(2.37) = 0.009$

Since p-value < 0.05, we reject H_0

(e) What is the conclusion of the test in the context of the data?

The data provide strong evidence that there were more days in 2018 than 1948 that exceeded 90°F at locations monitored by NOAA's weather stations.

(f) Calculate and interpret a 90% confidence interval for the population mean difference between the number of days exceeding 90°F in 1948 and 2018.



critical value

$$\text{qnorm}(0.95) = 1.645$$

$$\bar{d} \pm z_{\alpha/2} \cdot \frac{s_d}{\sqrt{n}} \Rightarrow 2.9 \pm 1.645 \cdot \frac{17.2}{\sqrt{197}}$$

$$\Rightarrow \boxed{(0.88, 4.92)}$$

We are 90% confident that, on average, between 0.88 and 4.92 more days exceeded 90°F in 2018 than 1948, at locations monitored by NOAA weather stations.