CHAPTER 10

1. **When would the mean grade in a class on a final exam be considered a statistic? When would it be considered a parameter?**
The mean grade in a class on a final exam will be considered a statistic if this was a sample population used to estimate a population parameter. However, in this case, this is a parameter as it is an estimate of the entire population being evaluated.

5. **When you construct a 95% confidence interval, what are you 95% confident about?**
A 95% confidence interval means that we are 95% confident that the true population parameter will be within the lower and upper bounds of the confidence interval. If repeated samples were taken and the 95% confidence interval was computed for each sample, 95% of the intervals would contain the population parameter. 5% of the intervals would not contain the population parameter.

10. **The effectiveness of a blood-pressure drug is being investigated. How might an experimenter demonstrate that, on average, the reduction in systolic blood pressure is 20 or more?**
The experimenter would calculate the confidence intervals using one of the below ways to demonstrate that on average the reduction in systolic blood pressure is 20 or more:
    a. If the population mean and standard deviation are known, then the sample mean (μ) is same as the population mean and the standard error of the mean is
    $$\sigma_M = \frac{\sigma}{\sqrt{N}}\ where\ \sigma\ is\ population\ standard\ deviation, N\ is\ sample\ size$$
    For a confidence level of 95% to state that the experimenter is 95% confident about the true population mean existing in this interval, it is found that in a standard normal distribution, this area is within 1.96 standard deviations of the mean and is computed as follows:
    $$Upper\ bound = \mu + (1.96)(\sigma_M)$$
    $$Lower\ bound = \mu - (1.96)(\sigma_M)$$
    This will then show an interval as 20 ≤ μ ≤ Upper bound.
    Based on the confidence level, the value Z can be found using the normal distribution calculator and the interval can be calculated as follows:
    $$Upper\ bound = \mu + (Z)(\sigma_M)$$
    $$Lower\ bound = \mu - (Z)(\sigma_M)$$
    b. If the population standard deviation is not known, then t-distribution can be used. The values to be used in the above formula can be found using a t-table with the help of degrees of freedom = N-1.
    c. Calculate the confidence interval using the difference of means where the following assumptions are made:
        i. The two populations have the same variance. This assumption is called the assumption of homogeneity of variance.
        ii. The populations are normally distributed.
        iii. Each value is sampled independently from each other value.
    The confidence interval is then calculated using the below formula:
    $$Upper\ bound = M_1 - M_2 + (t_{CL})(S_{M_1-M_2})$$
    $$Lower\ bound = M_1 - M_2 - (t_{CL})(S_{M_1-M_2})$$

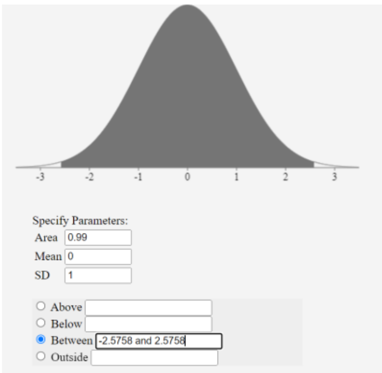
Where $M_1 - M_2$ is difference of means, t_{CL} is the desired level of confidence and $S_{M_1-M_2}$ is the estimated standard error of the difference between sample means.

15. **You take a sample of 22 from a population of test scores, and the mean of your sample is 60. (a) You know the standard deviation of the population is 10. What is the 99% confidence interval on the population mean.**
Mean =60
Standard Deviation =10

$$\sigma_M = \frac{\sigma}{\sqrt{N}} \; where\ \sigma\ is\ population\ standard\ deviation, N\ is\ sample\ size$$

$$\sigma_M = \frac{10}{\sqrt{22}} = 2.13200716356$$

The $Z_{.99}$ is calculated using the normal distribution calculator for a standard normal deviaton:



Therefore, $Z_{.99}$ is ± 2.5758.

$$Upper\ Limit = 60 + (2.5758)(2.13200716356)$$
$$Upper\ Limit = 65.4916240519$$
$$Lower\ Limit = 60 - (2.5758)(2.13200716356)$$
$$Lower\ Limit = \ 54.5083759481$$

The confidence interval is (54.5083759481, 65.4916240519)

(b) Now assume that you do not know the population standard deviation, but the standard deviation in your sample is 10. What is the 99% confidence interval on the mean now?

As the standard deviation is unknown, we compute an estimate of the standard error (s_M) using the formula:

$$s_M = \frac{s}{\sqrt{N}} = 2.13200716356$$

The value of t for 99% confidence interval for degrees of freedom = N-1 = 22-1 = 21 is 2.831

$$Lower\ Limit = 60 - (2.518)(2.13200716356)$$
$$Lower\ Limit = \ 53.96428772$$
$$Upper\ Limit = 60 + (2.518)(2.13200716356)$$
$$Upper\ Limit = \ 66.03571228$$

The confidence interval is (53.96428772, 66.03571228).

20. **True/false: You have a sample of 9 men and a sample of 8 women. The degrees of freedom for the t value in your confidence interval on the difference between means is 16.** *False*.

24. **(AM#6c) Is there a difference in how much males and females use aggressive behavior to improve an angry mood? For the "Anger-Out" scores, compute a 99% confidence interval on the difference between gender means.**

Gender	Sample size (n)	Mean	Variance
Male	30	16.56667	19.01264

Female	48	15.77083	17.15913

As can be seen from above, there is a difference in the sample size, mean and variance for Anger-Out scores.

$$Lower\ Limit = M_1 - M_2 - (t_{CL})(S_{M_1} - S_{M_2})$$
$$Upper\ Limit = M_1 - M_2 + (t_{CL})(S_{M_1} - S_{M_2})$$

Therefore, $M_1 - M_2$ = 16.56667-15.77083 = 0.79584

t_{CL} = t value for confidence level 99% and degrees of freedom 76 = 2.642

Mean squared Error (MSE) = $\frac{(s_1^2+s_2^2)}{2} = \frac{19.01264+17.15913}{2}$ =18.08588744

$$S_{M_1-M_2} = \sqrt{\left(\frac{18.08588744}{30}\right) + \left(\frac{18.08588744}{48}\right)} = 0.98977383089$$

Therefore,

$$Lower\ Limit = 0.79584 - (2.642)(0.98977383089) = -1.81914246121$$
$$Upper\ Limit = 0.79584 + (2.642)(0.98977383089) = 3.41082246121$$

The 99% confidence level of difference of means is *-1.81914246121 ≤ μ ≥ 3.41082246121*

25. Calculate the 95% confidence interval for the difference between the mean Anger-In score for the athletes and non-athletes. What can you conclude?

Condition	Sample size (n)	Mean	Variance
Athlete	25	16.68	13.56
Non-Athlete	53	19.4717	23.86938

$$Lower\ Limit = M_1 - M_2 - (t_{CL})(S_{M_1} - S_{M_2})$$
$$Upper\ Limit = M_1 - M_2 + (t_{CL})(S_{M_1} - S_{M_2})$$

Therefore, $M_1 - M_2$ = 19.4717 – 16.68 = 2.791698113

t_{CL} = t value for confidence level 95% and degrees of freedom 76 = 1.992

Mean squared Error (MSE) = $\frac{(s_1^2+s_2^2)}{2} = \frac{13.56+23.86938}{2}$ =18.71468795

$$S_{M_1-M_2} = \sqrt{\left(\frac{18.71468795}{25}\right) + \left(\frac{18.71468795}{53}\right)} = 1.04961651941$$

Therefore,

$$Lower\ Limit = 2.791698113 - (1.992)(1.04961651941) = 0.70086200633$$
$$Upper\ Limit = 2.791698113 + (1.992)(1.04961651941) = 4.88253421966$$

The 95% confidence level of difference of means is *0.70086200633 ≤ μ ≥ 4.88253421966*

26. Find the 95% confidence interval on the population correlation between the Anger-Out and Control-Out scores.

The correlation between the Anger-Out and Control-Out scores = -0.582683

```
In [45]:  import pandas as pd
          import numpy as np
          from scipy.stats.stats import pearsonr
          df = pd.read_excel('C:/Users/Administrator/Downloads/Stats Class/angry_moods.xls')

          df[['Anger-Out', 'Control-Out']].corr(method = 'pearson')
```

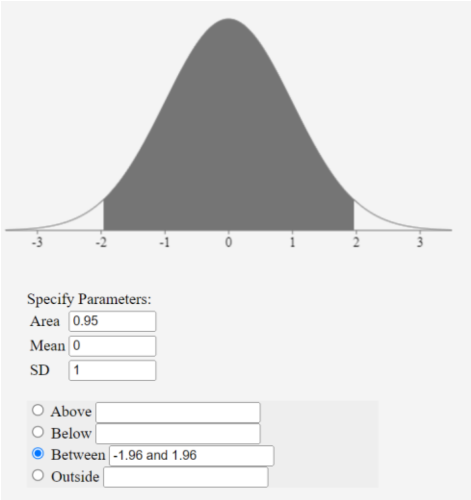
Out[45]:

	Anger-Out	Control-Out
Anger-Out	1.000000	-0.582683
Control-Out	-0.582683	1.000000

The z' associated with r of -0.582683 is -0.667.

The sampling distribution of z' is approximately normally distributed and has a standard error of $\frac{1}{\sqrt{N-3}} = \frac{1}{\sqrt{78-3}} = 0.11547005383$

The Z for a 95% confidence interval (Z.95) is 1.96 as seen below:



Therefore,

$$Lower\ Limit = -0.667 - (1.96)(0.11547005383) = -0.8933213055$$
$$Upper\ Limit = -0.667 + (1.96)(0.11547005383) = -0.44067869449$$

The r associated with a z' of -0.8933213055 is -0.713 and the r associated with a z' of -0.44067869449 is -0.414. Therefore, the population correlation (ρ) is likely to be between –0.713 and -0.414. The 95% confidence interval is:

$$-0.713 \leq \rho \geq -0.414$$

CHAPTER 11

4. **State the null hypothesis for:**
   a. **An experiment testing whether echinacea decreases the length of colds.**
   Echinacea increases or has no effect on the length of colds
   b. **A correlational study on the relationship between brain size and intelligence.**
   The correlation between brain size and intelligence is 0.
   c. **An investigation of whether a self-proclaimed psychic can predict the outcome of a coin flip.**
   The self-proclaimed psychic cannot predict the outcome of a coin flip
   d. **A study comparing a drug with a placebo on the amount of pain relief. (A one-tailed test was used.)**
   The drug provides lesser or equal pain relief than placebo.

8. **A significance test is performed and p = .20. Why can't the experimenter claim that the probability that the null hypothesis is true is .20?**

The probability value is not stating that the probability of the null hypothesis is true or false. It is instead the probability of the data given the null hypothesis meaning it is the probability of obtaining a statistic from the sample data under the assumption that the null hypothesis is true. Thus, the experimenter cannot make the above claim.

14. Why is "Ho: "M1 = M2" not a proper null hypothesis?
The null hypothesis should be stated in terms of the population parameters and not the statistics and should show no significant difference between them.

18. You choose an alpha level of .01 and then analyze your data. (a) What is the probability that you will make a Type I error given that the null hypothesis is true?
The probability that you will make a Type I error given that the null hypothesis is true is same as the alpha level which is 0.01
(b) What is the probability that you will make a Type I error given that the null hypothesis is false?
If the null hypothesis is false, then the probability of making a Type I error is 0 as it is impossible to make a Type I error.

20. True/false: It is easier to reject the null hypothesis if the researcher uses a smaller alpha (α) level. *False*
21. True/false: You are more likely to make a Type I error when using a small sample than when using a large sample. *True*
22. True/false: You accept the alternative hypothesis when you reject the null hypothesis. *True*
23. True/false: You do not accept the null hypothesis when you fail to reject it. *True*
24. True/false: A researcher risks making a Type I error any time the null hypothesis is rejected. *True*

CHAPTER 12

8. Participants threw darts at a target. In one condition, they used their preferred hand; in the other condition, they used their other hand. All subjects performed in both conditions (the order of conditions was counterbalanced). Their scores are shown below.

  a. **Which kind of t-test should be used?** A paired t test
  b. **Calculate the two-tailed t and p values using this t test.**

$$t = \frac{\text{statistic - hypothesized value}}{\text{estimated standard error of the statistic}}$$

The special case of this formula applicable to testing a single mean is

$$t = \frac{M - \mu}{s_M}$$

where t is the value we compute for the significance test, M is the sample mean, μ is the hypothesized value of the population mean, and s_M is the estimated

Mean of preferred hand = 10.6
Mean of other condition = 8.6
Difference of means = 10.6 – 8.6 = 2
The Hypothesized value = 0
Standard deviation (s) = 2.645751311

$$t = \frac{2-0}{(2.645751311/\sqrt{5})} = 1.690308509$$

P value for degrees of freedom N-1 (4) = 0.1662

c. **Calculate the one-tailed t and p values using this t test.**
T value = 1.690308509
P value = 0.1662/2 = 0.0831

9. **Assume the data in the previous problem were collected using two different groups of subjects: One group used their preferred hand and the other group used their non-preferred hand. Analyze the data and compare the results to those for the previous problem**

$$t = \frac{\text{statistic - hypothesized value}}{\text{estimated standard error of the statistic}}$$

Mean of preferred hand = 10.6
Mean of other condition = 8.6
Difference of means = 10.6 – 8.6 = 2
The Hypothesized value = 0

$$\text{MSE} = \frac{s_1^2 + s_2^2}{2}$$

Variance of preferred hand = 5.3
Variance of non-preferred hand = 1.3
MSE = (5.3+1.3)/2 = 3.3

$$S_{M_1 - M_2} = \sqrt{\frac{2(MSE)}{n}} = \sqrt{\frac{2(3.3)}{5}} = 1.148912529$$

$$t = \frac{2}{1.148912529} = 1.74077656$$

P value for degrees of freedom (8) = *0.1199*

11. **In an experiment, participants were divided into 4 groups. There were 20 participants in each group, so the degrees of freedom (error) for this study was 80 - 4 = 76. Tukey's HSD test was performed on the data.**
(a) Calculate the p value for each pair based on the Q value given below. You will want to use the Studentized Range Calculator.

Comparison of Groups	Q	p-value
A - B	3.4	0.0849
A - C	3.8	0.043
A - D	4.3	0.0167
B - C	1.7	0.6274
B - D	3.9	0.0359
C - D	3.7	0.0513

(b) Which differences are significant at the .05 level?
The pairs A-C, A-D and B-D are significant at the 0.05 level as they are ≤ 0.05.

21. **Do athletes or non-athletes calm down more when angry? Conduct a t test to see if the difference between groups in Control-In scores is statistically significant.**
Athletes calm down more when angry.

```
[1] 0
> mean_g0 <- mean(x_g0)
> mean_g1 <- mean(x_g1)
> n_g0 <- length(x_g0)
> n_g1 <- length(x_g1)
> var_pooled <- (sum((x_g0 - mean_g0)**2) + sum((x_g1 - mean_g1)**2)) / (n_g0 + n_g1 - 2)
> d <- mean_g1 - mean_g0
> se_d <- var_pooled * (1/n_g0 + 1/n_g1)
> t_stat <- d / sqrt(se_d)
> t_df <- (n_g0 + n_g1 - 2)
> p_value <- pt(-abs(t_stat), t_df, lower.tail = TRUE) + pt(abs(t_stat), t_df, lower.tail = FAL
SE)
> # Call the r function
> t.test(x_g0, x_g1, var.equal = TRUE)

        Two Sample t-test

data:  x_g0 and x_g1
t = 3.04431023461, df = 76, p-value = 0.0032030099814
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.2001554806912 5.7417313117616
sample estimates:
      mean of x       mean of y
24.320000000000 20.849056603774
```

The p-value = 0.0032030099814 which is less than 0.05 and hence is statistically significant.

22. Do people in general have a higher Anger-Out or Anger-In score? Conduct a t test on the difference between means of these two scores. Are these two means independent or dependent?

```
[1] 7
> # Manually calculate equal variance Student's t test
> mean_g0 <- mean(x_g0)
> mean_g1 <- mean(x_g1)
> n_g0 <- length(x_g0)
> n_g1 <- length(x_g1)
> var_pooled <- (sum((x_g0 - mean_g0)**2) + sum((x_g1 - mean_g1)**2)) / (n_g0 + n_g1 - 2)
>
> d <- mean_g1 - mean_g0
> se_d <- var_pooled * (1/n_g0 + 1/n_g1)
> t_stat <- d / sqrt(se_d)
> t_df <- (n_g0 + n_g1 - 2)
> p_value <- pt(-abs(t_stat), t_df, lower.tail = TRUE) + pt(abs(t_stat), t_df, lower.tail = FAL
SE)
>
> # Call the r function
> t.test(x_g0, x_g1, var.equal = TRUE)

        Two Sample t-test

data:  x_g0 and x_g1
t = -3.49755188061, df = 154, p-value = 0.00061388680242
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.9120505754208 -1.0879494245792
sample estimates:
      mean of x       mean of y
```

The p-value = 0.00061388680242

These two means are independent.

CHAPTER 13

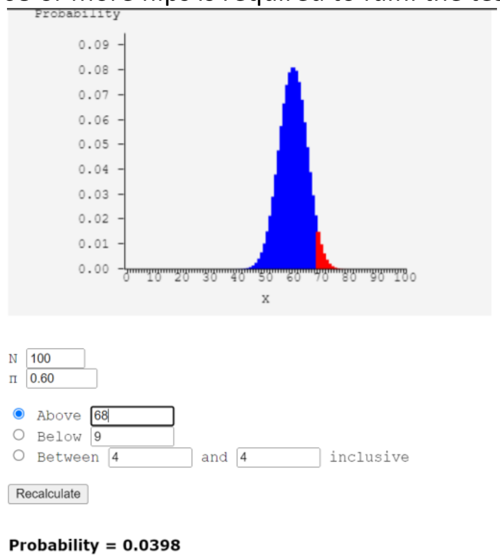
5. Alan, while snooping around his grandmother's basement stumbled upon a shiny object protruding from under a stack of boxes . When he reached for the object a genie miraculously materialized and stated: "You have found my magic coin. If you flip this coin an infinite number of times you will notice that heads will show 60% of the time." Soon after the genie's declaration he vanished, never to be seen again. Alan, excited about his new magical discovery, approached his friend Ken and told him about what he had found.

Ken was skeptical of his friend's story, however, he told Alan to flip the coin 100 times and to record how many flips resulted with heads.

(a) What is the probability that Alan will be able convince Ken that his coin has special powers by finding a p value below 0.05 (one tailed).

Use the Binomial Calculator (and some trial and error)

The probability = *0.0398,* 68 or more flips is required to fulfil the test of ≤ 0.05



(b) If Ken told Alan to flip the coin only 20 times, what is the probability that Alan will not be able to convince Ken (by failing to reject the null hypothesis at the 0.05 level)?

Probability = 1 - 0.051 = *0.949,* 16 or more flips is required to fulfil the test of ≤ 0.05