



Homework 4

Statistical Inference, Spring 2021



- 1- Determine if the following statements are true or false, and explain your reasoning for the statements you identify as false.
- The standard deviation of the bootstrap distribution will be approximately the same as the standard deviation of the original sample.
 - When generating the resamples, it is best to use a sample size smaller than the size of the original sample.
 - The bootstrap distribution is created by resampling with replacement from the population.
 - As the total sample size increases, the degrees of freedom for the residuals increase as well.
 - The constant variance condition can be somewhat relaxed when the sample sizes are relatively consistent across groups.
 - In a paired analysis, we first take the difference of each pair of observations, and then we do inference on these differences.
 - Each observation in one data set has a natural correspondence with exactly one observation from the other data set.
 - The test that 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 is a paired t test.
 - If you care more about the TypeII error, you must assign a large amount of β .
- 2- A marine biologist wants to determine if the body temperature of crabs exposed to ambient air temperature is different from the ambient air temperature. The biologist exposed a sample of 25 crabs to an air temperature of 24.3 °C for several minutes and then measured the body temperature of each crab (shown below).
- | | | | | | | | | |
|------|------|------|------|------|------|------|------|------|
| 22.9 | 22.9 | 23.3 | 23.5 | 23.9 | 23.9 | 24.0 | 24.3 | 24.5 |
| 24.6 | 24.6 | 24.8 | 24.8 | 25.1 | 25.4 | 25.4 | 25.5 | 25.5 |
| 25.8 | 26.1 | 26.2 | 26.3 | 27.0 | 27.3 | 28.1 | | |
- Test the biologist's question at the 5% level and compute the p-value for the test statistic.
 - Place a 95% confidence interval on μ .
- 3- (R) The dataset "Chick" contains weight 23 chickens that have been fed in two different types. We want to determine whether or not there is a difference in the mean of weight between these two groups.
- Bootstrap the difference in means with 1000 sample and describe the bootstrap distribution.
 - Compute p-value with both original and bootstrap samples and compare the results.
 - Obtain a 95% bootstrap percentile confidence interval with both original and bootstrap samples and compare the results.



Homework 4

Statistical Inference, Spring 2021



- 4- One way to measure a person's fitness is to measure their body fat percentage. The normal range of average body fat percentage for men is 15-20% body fat, and for women is 20-25% body fat. Our sample data is from a group of men and women who did workouts at a gym three times a week for a year. Then, their trainer measured the body fat. The table below shows the data. Determine if there is any major difference between men and women body fat percentage at a 5% significance level.

Table 1: Body fat percentage data grouped by gender

Group	Body Fat Percentages				
Men	13.3	6.0	20.0	8.0	14.0
	19.0	18.0	25.0	16.0	24.0
	15.0	1.0	15.0		
Women	22.0	16.0	21.7	21.0	30.0
	26.0	12.0	23.2	28.0	23.0

- 5- Melli owns a Spa salon and is trying to decide whether or not to purchase a vending machine so customers can buy coffee while they wait. She'll get the machine if she is convinced that more than 30% of her customers would buy coffee. She plans on taking a random sample of n customers and asking them whether or not they would buy coffee from the machine, and she'll then do a significance test using $\alpha = 0.05$ to see if the sample proportion who say "yes" is significantly greater than 30%. Which situation below would result in the highest power for her test? What is the effect of sample size and actual mean on power?
- She uses a sample size of $n=200$ and 32% of all customers would actually buy coffee.
 - She uses a sample size of $n=50$ and 50% of all customers would actually buy coffee.
 - She uses a sample size of $n=100$ and 32% of all customers would actually buy coffee.
 - She uses a sample size of $n=200$ and 50% of all customers would actually buy coffee.
- 6- Employees at multiple levels of a large company are stating that they are receiving, on average, fewer annual paid vacation days than the national average. Suppose it is known that the national average of annual paid vacation days in 2014 is known to be 7.5 days with a standard deviation of 1.6 days. To assess the validity of the employee's claims, the company randomly samples 55 employees and finds that the average number of annual vacation days for this sample is 7.22 days. The boss wanted to evaluate whether the average of annual paid vacation days has less than the national average or not.
- Using $\alpha=0.05$, what conclusions can you draw about the hypothesis?
 - What is the probability of making a Type II error if the actual value of μ is 7.4?
- 7- Calcium is a mineral that is necessary for life. In addition to building bones and keeping them healthy, calcium enables our blood to clot, our muscles to contract, and our heart to beat. The Institute of Medicine recommends a daily calcium intake of 1,000 milligrams per day, and the upper level (safe) limit is 2,500 milligrams per day. While calcium is contained in some foods, most adults do not get enough calcium in their diets and take supplements. Unfortunately, some of the supplements have side effects such as gastric distress, making them difficult for some patients to take on a regular basis.



Homework 4

Statistical Inference, Spring 2021



A study is designed to test whether there is a difference in mean daily calcium intake in adults with normal bone density, adults with osteopenia and adults with osteoporosis. Adults 60 years of age with normal bone density, osteopenia and osteoporosis are selected at random from hospital records and invited to participate in the study. Each participant's daily calcium intake is measured based on reported food intake and supplements. The data are shown below.

Normal Bone Density	Osteopenia	Osteoporosis
1200	1000	890
1000	1100	650
980	700	1100
900	800	900
750	500	400
800	700	350

- Write the hypotheses for testing for a difference between the average score of three different states.
- Do calculation and conduct analysis using one-way ANOVA and complete the table.

		DF	Sum SQ	Mean SQ	F value
Group	Class				
Error	Residuals				
	Total				

- What is the conclusion of the test?
- 8- (R) The dataset “Diet” contains information about 79 people with different health backgrounds began to undertake 3 different types of diets (referred to as diet A, B and C). This exercise aims to see which diet is best for losing weight. First and foremost look at the data, defining a new column weight loss, corresponding to the difference between the initial and final weights (respectively the corresponding to the columns initial weight and final weight of the dataset).
- Plot the three group's data using side-by-side boxplots.
 - Test either there is significant evidence of a difference in the mean weight loss between groups using ANOVA in R.
 - Display and analyze the results.
 - Compare the mean of weight loss for each two groups in R and:
 - Write the hypothesis.
 - Report the level of significance of the test and the decision about hypothesis.
 - Estimate the size of the difference in the mean drop, Use $\alpha=0.05$.