

Homework 3 - Key

Math 530/630

1. The longitudinal study *Bone mass is recovered from lactation to postweaning in adolescent mothers with low calcium intakes* examined total-body bone mineral content of young mothers during breast feeding and then in the postweaning period. We want to test the hypothesis that mothers gained more than 25 grams of bone mineral content in the postweaning period. The data for 10 mothers is provided below; use a significance level of 0.05. The column **bf** stands for the first measurement (during breastfeeding). The column **pw** stands for the second measurement (during postweaning period). All values are in grams of bone density.
 - a. State the null and alternative hypotheses.
 - b. First, do the test wrong: use an independent samples t-test, which ignores the paired nature of the dependent variables here, using the `t.test` function in R assuming equal variances. What would you conclude?
 - c. Now, do the test right: use a dependent samples t-test (also known as a paired t-test), using the `t.test` function in R. Do you change your conclusions? Explain why or why not. In your discussion, you must reference the degrees of freedom of each test.

```
bones <- data.frame(mother=1:10,
  bf=c(1928, 2549, 2825, 1924, 1628, 2175, 2114, 2621, 1843, 2541),
  pw=c(2126, 2885, 2895, 1942, 1750, 2184, 2164, 2626, 2006, 2627))
```

```
# Remember, the two-sample test is inappropriate.
```

```
t.test(bones$pw, bones$bf, mu=25, alternative="greater", var.equal = TRUE)
```

```
##
## Two Sample t-test
##
## data: bones$pw and bones$bf
## t = 0.44948, df = 18, p-value = 0.3292
## alternative hypothesis: true difference in means is greater than 25
## 95 percent confidence interval:
## -205.6358 Inf
## sample estimates:
## mean of x mean of y
## 2320.5 2214.8
```

```
# this is the right test
```

```
t.test(bones$pw, bones$bf, mu=25, paired = TRUE, alternative="greater")
```

```
##
## Paired t-test
##
## data: bones$pw and bones$bf
## t = 2.4575, df = 9, p-value = 0.01815
## alternative hypothesis: true difference in means is greater than 25
## 95 percent confidence interval:
## 45.50299 Inf
## sample estimates:
## mean of the differences
## 105.7
```

Answers

- a. * Null hypothesis: Mothers gained less than or equal to 25 grams of bone mineral content.
* Alternative hypothesis: Mothers gained more than 25 grams of bone mineral content.
 - b. As the p-value is greater than our alpha (e.i., $0.3292 > 0.05$), we fail to reject the null hypothesis.
 - c. Running the appropriate test, we now have a p-value of 0.01815, which is less than our alpha (0.05). Thus, we reject the null hypothesis. The degrees of freedom for the un-paired t.test is 18 ($n_1 + n_2 - 2$), whereas the degrees of freedom for the paired t.test is 9 ($n/2 - 1$). Although this higher degrees of freedom does mean a lower critical value for the un-paired t.test making it “easier” to reach statistical significance at the $\alpha=0.05$ level, it also means that the standard error will be inflated. As the standard error is in the denominator of the t-statistic calculation, this inflated standard error results in a smaller t-statistic, making it “harder” to reach statistical significance.
2. Your office mate ran an experiment with $N=50$ to test the hypothesis that her sample would have a mean different from the population mean, $\mu=0$, previously found by her advisor. She conducted a one-sample t test with $\alpha=.05$ (two-tailed), and reported the 95% confidence interval for μ of variable X is (8.979, 10.349). Note which of the following must also be true:

Answers

- * Null hypothesis: The sample mean is the same as the population mean of 0.
- * Alternate hypothesis: The sample mean is not equal to 0.

#First, let's look at some important numbers.

```
mu <- 0 #population mean
n <- 50
df <- n-1 #degrees of freedom
critical_value <- qt(.975, 49) #critical value.
mean <- sum(8.979 + 10.349)/2 # sample mean
sd <- (10.349 - mean) * (sqrt(n) / critical_value) # sample standard deviation
se <- sd/sqrt(n) # We could have computed the se directly, by leaving the sqrt(n) out of the above
t_statistic <- (mean-mu)/(se)
p_value <- 2*pt(t_statistic, df, lower=FALSE)
```

```
## [1] "critical value: 2.00957523712924"
## [1] "sample mean: 9.664"
## [1] "sample standard deviation: 2.41030112316036"
## [1] "standard error: 0.340868053777649"
## [1] "t statistic: 28.3511461191488"
## [1] "p_value: 4.68988749639606e-32"
```

- a. She rejected her null hypothesis. TRUE The population mean is well outside the critical value range.
- b. The t -statistic based on her sample was greater than 2.01. TRUE This is relevant given the critical value above.
- c. The p value for her t -statistic was less than her α -level. TRUE
- d. Her degrees of freedom were 51. FALSE
- e. Her sample mean of X was 9.664. TRUE
- f. Her sample mean of X was 5.664. FALSE

3. Suppose that the readings of a laboratory scale are normally distributed with unknown mean μ and standard deviation $\sigma = 0.01$ grams. To assess the accuracy of the laboratory scale, a standard weight that is known to weigh exactly 1 gram is repeatedly weighed a total of $N = 50$ times. Let $\bar{x} = 0.998$ be the average of the 50 readings. What is the 95% confidence interval for μ ?

Answer

Here we want to generate a 95% confidence interval estimated for an unknown population mean. This means that there is a 95% probability that the confidence interval will contain the true population mean. As this is a normal population, and we know the standard deviation, we can use `qnorm(0.975)` to find the critical value for a normal distribution. Using `qt(0.975,49)` to find the critical value for a t-distribution is a more conservative approach, but will result in a slightly wider interval.

```
n <- 50
x_bar <- 0.998
sd <- 0.01

mult_norm <- qnorm(0.975)
lower_norm <- x_bar - (mult_norm*(sd/sqrt(n)))
upper_norm <- x_bar + (mult_norm*(sd/sqrt(n)))
paste("Using the normal distribution: ", round(lower_norm,5), " < mu < " , round(upper_norm,5))

## [1] "Using the normal distribution:  0.99523  < mu <  1.00077"
# "0.99523  < mu <  1.00077"

mult_t <- qt(0.975,n-1)
lower_t <- x_bar - (mult_t*(sd/sqrt(n)))
upper_t <- x_bar + (mult_t*(sd/sqrt(n)))
paste("Using the t-distribution: ", round(lower_t,5), " < mu < " , round(upper_t,5))

## [1] "Using the t-distribution:  0.99516  < mu <  1.00084"
# "0.99516  < mu <  1.00084"
```