Lin-HW5

Bin Lin

2016-10-28

5.6 Working backwards, Part II.

Mean equals 71, margin of error equals 30, and standard deviation equals 3.658637.

```
Mean <- (77 + 65) / 2
Mean
```

```
## [1] 71
```

```
margin_of_error <- (77 - 65) / 2
margin_of_error
```

```
## [1] 6
```

```
z <- 1.64
sd <- (77 - Mean) / z
sd
```

```
## [1] 3.658537
```

5.14 SAT scores. (a) Raina wants to use a 90% confidence interval. How large a sample should she collect? At least 271

b. Luke wants to use a 99% confidence interval. Without calculating the actual sample size, determine whether his sample should be larger or smaller than Raina's, and explain your reasoning. The sample size will be larger than Raina's, because for Luke, his z-score will be larger. That means sd/ sqrt(n) will need to be smaller. So that n will haver to be larger.
c. Calculate the minimum required sample size for Luke. At least 664

```
z <- qnorm(0.95)
sd <- 250
margin_of_error <- 25

n <- (sd / (margin_of_error / z))^2
n
```

```
## [1] 270.5543
```

```
z <- qnorm(0.995)
sd <- 250
margin_of_error <- 25

n <- (sd / (margin_of_error / z))^2
n
```

```
## [1] 663.4897
```

5.20 High School and Beyond, Part I. (a) Is there a clear difference in the average reading and writing scores? No, the second graph show the difference between reading and writing score. The distribution is unimodel and centered at 0.

b. Are the reading and writing scores of each student independent of each other? I do not think they are independent to each other. Because if someone who is good at reading, he or she will be more likely to be good at writing also.

c. Create hypotheses appropriate for the following research question: is there an evident difference in the average scores of students in the reading and writing exam?

H0 : ??diff == 0 HA : ??diff != 0

d. Check the conditions required to complete this test.

This is a simple random sample. The sample size is greater than 30 and less than 10% of population. the distribution is unimodel and bell-shaped. It does not appear to be skewed. Therefore we can assume it is normally distributed.

e. The average observed difference in scores is ¯xread???write = ???0.545, and the standard deviation of the differences is 8.887 points. Do these data provide convincing evidence of a difference between the average scores on the two exams?

These data does not provide convincing evidence of a difference between the average scores on the two exams. Because the p-value equals 0.192896 which is significantly larger than alpha (0.005 for one-sided or 0.01 for two sided). Therefore, we fail to reject the null hypothesis.

```
z <- (-0.545 / (8.887 / sqrt(200)))
z
```

```
## [1] -0.867274
```

```
pnorm(z)
```

```
## [1] 0.192896
```

f. What type of error might we have made? Explain what the error means in the context of the application.

Type II errors, because we fail to reject the null hypothesis, however, the null hypothesis could still be false

g. Based on the results of this hypothesis test, would you expect a confidence interval for the average difference between the reading and writing scores to include 0? Explain your reasoning.

Yes, including 0 means there is no difference between reading and writing score. Which is the same idea as the null hypothesis. Since we fail to reject null hypothesis, so that it could still be true.

5.32 Fuel efficiency of manual and automatic cars, Part I. H0 : ??diff = 0 HA : ??diff != 0

```
t <- (16.12 - 19.85) / sqrt((3.58^2) / 26 + (4.51^2) / 26)
t
```

```
## [1] -3.30302
```

```
pt(t, df = min(26 - 1, 26 -1))
```

```
## [1] 0.001441807
```

The p-value equals 0.00144, which is significantly lower than 0.05. Therefore, these data provide strong evidence of a difference between the average fuel efficiency of cars with manual and autonomic transmissions in terms of their average city mileage.

5.48 Work hours and education.

a. Write hypotheses for evaluating whether the average number of hours worked varies across the five groups. H0 : Average number of hours worked across five groups are same HA : least one pair of means is different.

b. Check conditions and describe any assumptions you must make to proceed with the test.

The data is simple random sample from less than 10% of the population. The obeservations are independent within and between groups. The distribution of each group are assume to be nearly normal

d. What is the conclusion of the test?

Because Pr(>F) equals 0.0682, which is greater than 0.05, therefore we can not reject the null hypothesis. That means there is no enough data to prove the null hypothesis is wrong at significance level alpha = 0.05.

```
df_degree <- 5 - 1
df_total <- 1172 - 1
df_residual <- df_total - df_degree
df
```

```
## function (x, df1, df2, ncp, log = FALSE)
## {
##     if (missing(ncp))
##         .Call(C_df, x, df1, df2, log)
##     else .Call(C_dnf, x, df1, df2, ncp, log)
## }
## <bytecode: 0x000000000b0a4628>
## <environment: namespace:stats>
```

```
sumsquare_degree <- 501.54 * df_degree
sumsquare_degree
```

```
## [1] 2006.16
```

```
sumsquare_total <- sumsquare_degree + 267382
sumsquare_total
```

```
## [1] 269388.2
```

```
MSG <- sumsquare_degree / df_degree
MSE <- 267382 / df_residual
F_value <- MSG / MSE
F_value
```

```
## [1] 2.188992
```