Hao-HW5

Bruce Hao

10/6/2016

```
setwd("~/Google Drive/CUNY/git/DATA606/Lab5")
```

5.6

```
n = 25
z = qnorm((1 + 0.9) / 2)  # using right tail to keep z positive
z
```

```
## [1] 1.644854
```

```
mu = (65 + 77) /2
mu
```

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

```
sd = (77 - mu) / z
sd
```

```
## [1] 3.647741
```

```
moe = z * sd
moe
```

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

5.14

a)

```
sd = 250
moe = 25
z = qnorm(0.95)
n = (z * sd / moe)^2
n
```

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

b) Luke's required sample size would be much larger because he needs a higher level of precision.

c)

```
sd = 250
moe = 25
z = qnorm(0.995)
n = (z * sd / moe)^2
n
```

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

5.20

a) Optically, it appears that reading scores are somewhat lower than writing scores.
b) Reading and writing scores of each student are unlikely to be completely independent of each other. One might think that on average strong readers are also stronger writers, and vice versa.
c) H_0: $\bar{x}_{read\ -\ write} = 0$ H_A: $\bar{x}_{read\ -\ write} <> 0$
d) The observations are based on a random sample from less than 10% of the population, so independence is established. There is no noticeable skew and the sample size is large, so the conditions for a paired t-test are met.
e) The two-sided p-value is about 0.19, so there is insufficient evidence to reject the null at the 5% significance level.

```
n = 200
mu = -0.545
sig = 8.887
se = sig / sqrt(n)
t = mu / se
df = n - 1
pt(t, df)
```

```
## [1] 0.1934182
```

f) Since we failed to reject the null, we might have made a Type II error - failing to reject the null when the alternative is actually true. In this context, a Type II error means failing to find a difference in reading and writing scores when a difference actually exists.
g) Yes, it is because the confidence interval includes zero that we reject the null.

5.32

H_0: $\bar{x}_{auto\ -\ manual} = 0$
H_A: $\bar{x}_{auto\ -\ manual} <> 0$

```
mu_diff = 16.12 - 19.85
sd1 = 3.58
sd2 = 4.51
n1 = 26
n2 = 26
se = sqrt(sd1^2 / n1 + sd2^2 / n2)
t = mu_diff / se
df = min(n1, n2) - 1
pt(t, df)
```

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

The p-value for t-test for the difference of means is less than 5%, so there is a statistically significant difference between the average city mpg for cars with automatic and manual transmissions at the 5% significance level.

5.48

a) H_0: $\bar{x}_{Less\ than\ HS} = \bar{x}_{HS} = \bar{x}_{JC} = \bar{x}_{Bachelor} = \bar{x}_{Graduate}$

b) H~A!: At least one mean is different

c)

```
k = 5
n = 1172
data.frame('stat' = c('degree', 'Residuals', 'Total'),
           'Df' = c(k - 1, n - k, n - 1),
           'SumSq' = c(501.54 * (k - 1), 267382, 501.54* (k - 1) + 267382),
           'MeanSq' = c(501.54, 267382 / (n - k), 501.54 + 267382 / (n - k)),
           'Fvalue' = c(round(501.54 / (267382 / (n - k)), 4),'',''),
           'Fcheck' = c(round(qf(0.0682, k - 1, n - k, lower.tail = F), 4),'',''),
           'Pr(>F)' = c(0.0682,'',''))
```

```
##        stat   Df     SumSq   MeanSq Fvalue Fcheck Pr..F.
## 1    degree    4   2006.16 501.5400  2.189 2.1889 0.0682
## 2 Residuals 1167 267382.00 229.1191
## 3     Total 1171 269388.16 730.6591
```

d) The ANOVA p-value of 0.0682 is greater than 5%, so we cannot reject the null that there are any differences between the means at a 5% significance level.