Preston Robertson
Summer 2022: Data Analysis
Homework IV
Due: 6/20
Submit through Canvas

Instructions: Provided solutions to these questions using this template. Include graphics with your solutions. Put all code an appendix to this homework. Use the *verbatim* command to leave code unchanged.

1. We investigate the t-distribution and how it relates to the standard normal distribution.

a) Find a number z_α such that $P(Z > z_\alpha) = \alpha$ where $\alpha = 0.01$.

Solution:
We observe from the standard normal table that $P(Z > 2.32) = 0.01$. Therefore, $z_{0.01} = 2.32$.

b) Find t-score t^α_{df} with degrees of freedom $df = 10$ where again $\alpha = 0.025$. This t-score will represent $P(T_{df=10} > t^\alpha_{df}) = \alpha$.

Solution:
Note that since $P(T^{df=10} - 2.228$ we see that $t^\alpha_{df=10} = 2.228$.

c) Repeat part b) but with $df = 50$.
Solution:
Using R, we find that $P(TT^{df=50} > 2.008559) = 0.025$, thus $t^{df=50}_{0.025} = 2.008559$
d) Repeat part b) but with $df = 100$. What distribution does the t distribution come to resemble as $df \to \infty$?

Solution:
When df $= 100$, we find that $P(T_{df} > 1.984) = 0.025$. It appears that as $df \to \infty$ the t distribution tends to the Gaussian distribution.

2) Generate a sample of size $n = 10$ from a Normal, mean $\mu = 0$, $\sigma^2 = 1$. Make sure and use *set.seed(123)*.

a) Compute $\bar{x}$ and s^2. Program in long format by writing a loop. (I know, but you got to do it once). Make sure an put this code in the Appendix.

Solution:

After running the code in the appendix, R computes $\bar{x} = 0.07462564$ and $\hat{sigma}^2 = s^2 = 0.909704$

b) Use the formula from class to compute a 95% confidence interval for μ when $\sigma = 1$ is assumed known.

Solution:

First note that $\alpha = 0.05$ and $z_{\frac{\alpha}{2}} = 1.96$ on the standard normal table. As derived in class, the 95% confidence interval is $\bar{x} \pm z_{\alpha/2}\sigma/\sqrt{n} = 0.07462564 \pm 0.6198064$ which is the interval $[-0.5451808, 0.6944321]$.

c) Repeat part b), but incorporate $\hat{\sigma}$ instead of σ. Don't forget to switch to the t distribution with the appropriate degrees of freedom.

Solution:

Since $n = 10$, we find that $df = 9$, thus $t_{0.975}^{df=9} = 2.262$. We now find that the confidence interval is $\bar{x} \pm t_{0.975}^{df=9}/\sqrt{9} = 0.07462564 \pm 2.262(0.953784)/3$ which is the interval [-0.6445275, 0.7937788]

3) Take a random sample $\{X_i\}_{i=1}^{n}$ where $X_i \sim N(0,1)$. Don't for get to set the seed.

a) Construct an two-tailed test for the hypothesis $H_0 : \mu = 0$ at the 95% level of confidence. What are the critical values for this test? Do you accept or reject the hypothesis? Provide a clear explanation of your results.

Solution:

First, the critical values when $\alpha = 0.05$ for a two-sided hypothesis test of $H_0 : \mu = 0$ v.s. $H_a : \mu \neq 0$ are ± 1.96. The rejection region will be $(-\infty, -1.96) \cup (1.96, \infty)$. Any test statistic falling in the rejection region will result in a rejection of the null hypothesis (H_0). After sampling the random normal data, we obtain a test statistic

$$z = \frac{\bar{x} - 0}{1/\sqrt{100}} = 0.9040591,$$

therefore we failed to reject the null hypothesis (H_0).

b) Re-sample using $set.seed(124)$ such that $X_i \sim N(0.50, 1)$. Construct a two tailed test for the hypothesis $H_0 : \mu = 0$ at the 95% level of confidence. What are the critical values? Provide a clear explanation of your results.

Solution:

First, the critical values when $\alpha = 0.05$ for a two-sided hypothesis test of $H_0 : \mu = 0$ v.s. $H_a : \mu \neq 0$ are ± 1.96. The rejection region will be $(-\infty, -1.96) \cup (1.96, \infty)$. Any test statistic falling in the rejection region will result in a rejection of the null hypothesis (H_0). After sampling the random normal data, we obtain a test statistic

$$z = \frac{\bar{x} - 0}{1/\sqrt{100}} = 5.096206,$$

therefore we reject the null hypothesis.

Appendix: Code

Below is the code for Q.2A.

```
# HOMEWORK 4
set.seed(123)
dat <- rnorm(10, mean = 0, sd = 1)

# MEAN
datasum = 0
t <- c(1:10)
for(i in t){
datasum = datasum + dat[i]
}
xbar = datasum / 10

# SAMPLE VARIANCE
datvar = 0
for(i in t){
datvar = datvar + (dat[i] - xbar)^2
}
sqr = datvar / (10-1)
```

Below is the code for Q3.

```
set.seed(123)
dat = rnorm(100, 0,1)
hist(dat)

tsa = sqrt(100)*mean(dat)
tsa

set.seed(124)
datb = rnorm(100, .5, 1)
hist(datb)

tsb = sqrt(100)*mean(datb)
tsb
```