

MA 331 Homework 4

Daniel Detore, Emma Erdtmann

I pledge my honor that I have abided in the Stevens Honor System.

Problem 1

```
> mpg = c(41.5, 50.7, 36.6, 37.3, 34.2, 45.0, 48.0, 43.2, 47.7, 42.2, 43.2,
44.6, 48.4, 46.4, 46.8, 39.2, 37.3, 43.5, 44.3, 43.3, 35.8, 33.9, 40.1, 41.3,
37.7, 39.6, 42.4, 41.7, 35.7)
```

(i)

Null hypothesis:

1. $\mu = 44$

Alternative hypotheses:

2. $\mu < 44$ (requires left-tail test)
3. $\mu > 44$ (requires right-tail test)
4. $\mu \neq 44$ (requires two-tail test)

(ii)

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = -3.189008$$

```
> mu = 44
> sigma=3.75
> n = length(mpg)
> xbar = mean(mpg)
> z <- (xbar - mu)/(sigma/sqrt(n))
> z
[1] -3.189008
```

Left-tail test:

$t_{left} = \Phi(-3.189008) = 0.0007138098 < \alpha = 0.05$; we reject H_0 .

```
> pnorm(z)
[1] 0.0007138098
> z.test(mpg, sigma.x = sigma, alternative=c("less"), mu = mu)
```

One-sample z-Test

```
data: mpg
z = -3.189, p-value = 0.0007138
alternative hypothesis: true mean is less than 44
95 percent confidence interval:
    NA 42.92472
sample estimates:
mean of x
41.77931
```

Right-tail test:

$t_{right} = 1 - \Phi(-3.189008) = 0.9992862 > \alpha = 0.05$; we cannot reject H_0 .

```
> 1-pnorm(z)
[1] 0.9992862
> z.test(mpg, sigma.x = sigma, alternative=c("greater"), mu = mu)
```

One-sample z-Test

```
data: mpg
z = -3.189, p-value = 0.9993
alternative hypothesis: true mean is greater than 44
95 percent confidence interval:
 40.6339      NA
sample estimates:
mean of x
 41.77931
```

Two-tail test:

$t_{two} = 2\Phi(-3.189008) = 0.00142762 > \alpha = 0.05$; we reject H_0 .

```
> 2*pnorm(-abs(z))
[1] 0.00142762
> z.test(mpg, sigma.x = sigma, alternative=c("two.sided"), mu = mu)
```

One-sample z-Test

```
data: mpg
z = -3.189, p-value = 0.001428
alternative hypothesis: true mean is not equal to 44
95 percent confidence interval:
 40.41447 43.14415
sample estimates:
mean of x
 41.77931
```

We can reject H_0 in the case of $\mu < 44$ or $\mu \neq 44$, but μ cannot be greater than 44. This means that $\mu > 44$ at the 0.05 significance level when $\sigma = 3.75$.

(iii)

```
> s = sd(mpg)
> s
[1] 4.506533
```

$$T = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = -3.189008$$

```
t <- (xbar - mu) / (s/sqrt(n))
```

```
> t
[1] -2.653654
> df = n-1      # = 28
```

Left-tail test:

$t_{left} = pt(-2.653654, 28) = 0.006487755 < \alpha = 0.05$; we reject H_0 .

```
> pt(t, df)
[1] 0.006487755
> t.test(mpg, alternative=c("less"), mu=mu, conf.level=(1-.05))
```

One Sample t-test

```
data: mpg
t = -2.6537, df = 28, p-value = 0.006488
alternative hypothesis: true mean is less than 44
95 percent confidence interval:
 -Inf 43.20289
sample estimates:
mean of x
 41.77931
```

Right-tail test:

$t_{right} = 1 - pt(-2.653654, 28) = 0.9935122 > \alpha = 0.05$; we cannot reject H_0 .

```
> 1-pt(t, df)
[1] 0.9935122
> t.test(mpg, alternative=c("greater"), mu=mu, conf.level=(1-.05))
```

One Sample t-test

```
data: mpg
t = -2.6537, df = 28, p-value = 0.9935
alternative hypothesis: true mean is greater than 44
95 percent confidence interval:
 40.35573      Inf
sample estimates:
mean of x
 41.77931
```

Two-tail test:

$t_{two} = 2pt(-2.653654, 28) = 0.01297551 < \alpha = 0.05$; we reject H_0 .

```
> 2*pt(-abs(t), df)
[1] 0.01297551
> t.test(mpg, alternative=c("two.sided"), mu=mu, conf.level=(1-.05))
```

One Sample t-test

```

data: mpg
t = -2.6537, df = 28, p-value = 0.01298
alternative hypothesis: true mean is not equal to 44
95 percent confidence interval:
 40.06512 43.49350
sample estimates:
mean of x
 41.77931

```

We can reject H_0 in the case of $\mu < 44$ or $\mu \neq 44$, but not $\mu > 44$. This means that $\mu < 44$ at the 0.05 significance level when we assume $\sigma = S = 4.506533$.

(iv)

If we do not know the population standard deviation, we are working with less accurate statistics extracted from our observations. In this case, we must calculate our data's standard deviation and use it in the population standard deviation's stead. Therefore we can come to a better conclusion when we know the population standard deviation.

Problem 2

```

> X = c(41.5, 50.7, 36.6, 37.3, 34.2, 45.0, 48.0, 43.2, 47.7, 42.2, 43.2, 44.6,
48.4, 46.4, 46.8, 39.2, 37.3, 43.5, 44.3, 43.3, 35.8, 33.9, 40.1, 41.3, 37.7,
39.6, 42.4, 41.7, 35.7)
> Z <- rnorm(n = length(X), mean = 0, sd = 4)
> Z
 [1] -5.59128318 -0.02533175  2.41195412 -2.96872360 -2.94679304  0.48833476
 [7]  0.28885157  3.45012260  0.72795278  2.29913207 -2.75541597 -2.50289990
[13] -2.53546465 -5.31092594 -1.73059182  8.94546869  2.72596957 -7.06020092
[19]  0.69744604 -1.90264708  2.55088541 -7.78399996 -0.17280706 -1.66184781
[25] -2.70987035 -6.08832618  4.83278493  7.90744134  3.64387596
> Y <- c()
> for (i in 1:length(X)) {
+   Y[i] = X[i] + Z[i]
+ }

```

(i)

```

> mean(Y) == mean(X)
[1] FALSE
> (1/length(Y)) * sum(Y)
[1] 41.33873
> (1/length(X)) * sum(X)
[1] 41.77931
> (1/length(Y)) * sum(Y) == (1/length(X)) * sum(X)
[1] FALSE

```

They are quite close, but they are not equal.

(ii)

Null hypothesis:

$$5. \mu_x = \mu_Y$$

Alternative hypotheses:

$$6. \mu_x < \mu_Y \text{ (requires left-tail test)}$$

$$7. \mu_x > \mu_Y \text{ (requires right-tail test)}$$

$$8. \mu_x \neq \mu_Y \text{ (requires two-tail test)}$$

(iii)

Based on code from Recitation5.R:

```
> ss1 = var(X)
> ss1
[1] 20.30884
> ss2 = var(Y)
> ss2
[1] 35.798
> n1 = length(X)    # = 29
> n2 = length(Y)    # = 29
> t = (mean(X)-mean(Y))/sqrt(ss1/n1+ss2/n2)
> t
[1] 0.3167517
> k = (ss1/n1+ss2/n2)^2/((ss1/n1)^2/(n1-1)+(ss2/n2)^2/(n2-1))
> k = ceiling(k)    # = 53
```

Recall $\alpha = 0.05$.

Left-tail test:

$t_{left} = pt(0.3167517, 53) = 0.6236618 > \alpha = 0.05$; we cannot reject H_0 .

```
> pt(t,k)          ## p-value for Ha:mu1 < mu2
[1] 0.6236618
> t.test(X, Y, alternative=c("less"), paired=FALSE, var.equal=FALSE)
```

Welch Two Sample t-test

```
data: X and Y
t = 0.31675, df = 52.034, p-value = 0.6237
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf 2.76995
sample estimates:
mean of x mean of y
41.77931 41.33873
```

Right-tail test:

$t_{right} = 1 - pt(0.3167517, 53) = 0.3763382 > \alpha$; we cannot reject H_0 .

```
> 1-pt(t,k)
[1] 0.3763382
> t.test(X, Y, alternative=c("greater"), paired=FALSE, var.equal=FALSE)
```

Welch Two Sample t-test

```
data: X and Y
t = 0.31675, df = 52.034, p-value = 0.3763
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -1.888784      Inf
sample estimates:
mean of x mean of y
 41.77931  41.33873
```

Two-tail test:

$t_{two} = 2pt(-0.3167517, 53) = 0.7526764 > \alpha$; we cannot reject H_0 .

```
> 2*pt(-abs(t),k)
[1] 0.7526764
> t.test(X, Y, alternative=c("two.sided"), paired=FALSE, var.equal=FALSE)
```

Welch Two Sample t-test

```
data: X and Y
t = 0.31675, df = 52.034, p-value = 0.7527
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.350502  3.231668
sample estimates:
mean of x mean of y
 41.77931  41.33873
```

In summary, we cannot reject H_0 , so we can be 95% confident that $\mu_x = \mu_y$.

Problem 3

```
> mpg = c(41.5, 50.7, 36.6, 37.3, 34.2, 45.0, 48.0, 43.2, 47.7, 42.2, 43.2,
44.6, 48.4, 46.4, 46.8, 39.2, 37.3, 43.5, 44.3, 43.3, 35.8, 33.9, 40.1, 41.3,
37.7, 39.6, 42.4, 41.7, 35.7)
> W <- ifelse(mpg > 42.5, 1, 0)
> W
[1] 0 1 0 0 0 1 1 1 1 0 1 1 1 1 1 0 0 1 1 1 0 0 0 0 0 0 0 0 0
```

(i)

Null hypothesis:

$$1. P_{H_0}(X > 42.5) = 0.4$$

Alternative hypotheses:

$$2. P_{H_0}(X > 42.5) < 0.4 \text{ (requires left-tail test)}$$

$$3. P_{H_0}(X > 42.5) > 0.4 \text{ (requires right-tail test)}$$

$$4. P_{H_0}(X > 42.5) \neq 0.4 \text{ (requires two-tail test)}$$

(ii)

```
> successes <- sum(W)    # = 13
```

```
> trials <- length(W)    # = 29
```

Recall $\alpha = 0.05$.

Z-value:

$$Z = \frac{\sqrt{n}(\hat{p} - p_0)}{\sqrt{p_0(1-p_0)}} \text{ where } \hat{p} = \frac{N}{n} \Rightarrow z = 0.53066$$

```
> z <- sqrt(trials)*((successes/trials)-0.4)/sqrt(0.4*(1-0.4))
```

```
[1] 0.5306686
```

Left-tail test:

$p_{left} = \Phi(z) = 0.7021758 > \alpha = 0.05$; we cannot reject H_0 .

```
> pnorm(z)
```

```
[1] 0.7021758
```

```
> prop.test(successes, trials, p = 0.4, alternative = c("less"), correct = FALSE)
```

1-sample proportions test without continuity correction

data: successes out of trials, null probability 0.4

X-squared = 0.28161, df = 1, p-value = 0.7022

alternative hypothesis: true p is less than 0.4

95 percent confidence interval:

0.0000000 0.5980325

sample estimates:

p

0.4482759

Right-tail test:

$p_{right} = 1 - \Phi(z) = 0.2978242 > \alpha$; we cannot reject H_0 .

```
> 1- pnorm(z)
```

```
[1] 0.2978242
```

```
> prop.test(successes, trials, p = 0.4, alternative = c("greater"), correct = FALSE)
```

1-sample proportions test without continuity correction

```
data: successes out of trials, null probability 0.4
X-squared = 0.28161, df = 1, p-value = 0.2978
alternative hypothesis: true p is greater than 0.4
95 percent confidence interval:
 0.3073468 1.0000000
sample estimates:
      p
0.4482759
```

Two-tail test:

$p_{left} = 2\Phi(-|z|) = 0.5956544 > \alpha$; we cannot reject H_0 .

```
> 2*pnorm(-abs(z))
[1] 0.5956484
> prop.test(successes , trials , p = 0.4, alternative = c("two.sided"), correct
= FALSE)
```

1-sample proportions test without continuity correction

```
data: successes out of trials, null probability 0.4
X-squared = 0.28161, df = 1, p-value = 0.5956
alternative hypothesis: true p is not equal to 0.4
95 percent confidence interval:
 0.2841317 0.6245204
sample estimates:
      p
0.4482759
```

(iii)

By converting the data into a binomial distribution, we are losing some important information that we could use to interpret the results, such as how far each value is from 42.5. Additionally, the z-test is approximate, according to the slides; while these z-values are all quite conclusive, if they were very near our α , we might need to worry about how that approximation is affecting our calculations.