Stat 3202 Lab 10

Nathan Johnson.9254

2023-04-11

Guided Example 1:

```
library(readr)
vanillin <- c(38,46,41,36,42,40,34,41,33,
46,41,39,44,45,41,37,39,36,
40,43,37,41,39,36,47,36,41,
38,38,38,33,39,41,36,38,44,39)
mean(vanillin)
```

```
## [1] 39.54054
```

1. Confidence interval Test $H_0 : \mu = 42$ vs. $H_1 : \mu \neq 42$ using a 99% confidence interval for μ.

```
alpha = 0.01
n = length(vanillin)
xbar = mean(vanillin)
sigma = 3.52

z = qnorm(1-alpha/2)

CIlower = xbar - z * sigma/sqrt(n)
CIupper = xbar + z * sigma/sqrt(n)

c(CIlower, CIupper)
```

```
## [1] 38.04995 41.03113
```

We are 99% confident that the true mean vanillin level of this batch is between 38.05 and 41.03 ppt. Since 42 is not contained within this confidence interval, we reject the null hypothesis, H_0.

We conclude that the true mean vanillin level is statistically significantly different from 42.

2. Rejection region We are testing the hypotheses $H_0 : \mu = 42$ vs. $H_1 : \mu \neq 42$.

$$Z = \frac{\bar{x} - 42}{\sigma/\sqrt{n}}$$

Under the null hypotheses, the test statistic Z follows the distribution $Z \sim N(0, 1)$.

We reject the null hypothesis H_0 when we observe our test statistic Z_{obs} to be either too high or too low.

```
alpha = 0.01

rrlower = qnorm(alpha/2)
rrupper = qnorm(1-alpha/2)
c(rrlower, rrupper)
```

```
## [1] -2.575829  2.575829
```

The rejection region is $(-\infty, -2.58]$ and $[2.58, \infty)$, and we would reject the null hypothesis when our observed test statistic falls within either of these regions.

```
Z.obs = (xbar - 42) / (sigma / sqrt(n))
Z.obs
```

```
## [1] -4.250087
```

We observe $Z_{obs} = -4.25$. Since this falls within the rejection region, we reject the null hypothesis H_0. And we conclude that the $\mu \neq 42$. That is the true mean vanillin is statistically significantly different from 42 ppt.

3. Verify with Z Test

```
library(DescTools)
```

```
ZTest(vanillin, alternative = "two.sided", mu = 42, sd_pop = 3.52, conf.level = 1-alpha)
```

```
##
##  One Sample z-test
##
## data:  vanillin
## z = -4.2501, Std. Dev. Population = 3.52, p-value = 2.137e-05
## alternative hypothesis: true mean is not equal to 42
## 99 percent confidence interval:
##  38.04995 41.03113
## sample estimates:
## mean of x
##  39.54054
```

Problem 2:

```
library(readr)
vanillin <- c(38,46,41,36,42,40,34,41,33,
46,41,39,44,45,41,37,39,36,
40,43,37,41,39,36,47,36,41,
38,38,38,33,39,41,36,38,44,39)
mean(vanillin)
```

```
## [1] 39.54054
```

1. Confidence interval Test $H_0 : \mu = 42$ vs. $H_1 : \mu \neq 42$ using a 99% confidence interval for μ.

```r
alpha = 0.01
n = length(vanillin)
xbar = mean(vanillin)
sigma = sd(vanillin)

t = qt(1-alpha/2, n - 1)

CIlower = xbar - t * sigma/sqrt(n)
CIupper = xbar + t * sigma/sqrt(n)

c(CIlower, CIupper)
```

```
## [1] 37.96482 41.11626
```

We are 99% confident that the true mean vanillin level of this batch is between 37.96 and 41.12 ppt. Since 42 is not contained within this confidence interval, we reject the null hypothesis, H_0.

We conclude that the true mean vanillin level is statistically significantly different from 42.

2. Rejection region We are testing the hypotheses $H_0 : \mu = 42$ vs. $H_1 : \mu \neq 42$.

$$t = \frac{\bar{x} - 42}{s/\sqrt{n}}$$

Under the null hypotheses, the test statistic t follows the distribution $t \sim N(0, 1)$.

We reject the null hypothesis H_0 when we observe our test statistic t_{obs} to be either too high or too low.

```r
alpha = 0.01

rrlower = qt(alpha/2, n - 1)
rrupper = qt(1-alpha/2, n - 1)
c(rrlower, rrupper)
```

```
## [1] -2.719485  2.719485
```

The rejection region is $(-\infty, -2.72]$ and $[2.72, \infty)$, and we would reject the null hypothesis when our observed test statistic falls within either of these regions.

```r
t.obs = (xbar - 42) / (sigma / sqrt(n))
t.obs
```

```
## [1] -4.244692
```

We observe $t_{obs} = -4.24$. Since this falls within the rejection region, we reject the null hypothesis H_0. And we conclude that the $\mu \neq 42$. That is the true mean vanillin is statistically significantly different from 42 ppt.

3. Verify with t Test

```
library(DescTools)

t.test(vanillin, alternative = "two.sided", mu = 42, sd_pop = sigma, conf.level = 1-alpha)
```

```
##
##  One Sample t-test
##
## data:  vanillin
## t = -4.2447, df = 36, p-value = 0.0001467
## alternative hypothesis: true mean is not equal to 42
## 99 percent confidence interval:
##  37.96482 41.11626
## sample estimates:
## mean of x
##  39.54054
```

Problem 3:

    a. We would use a z test statistic for this hypothesis.
    b. Z distribution.

$$Z_{obs} = \frac{\bar{x} - 8}{1/\sqrt{30}} \sim^{H_0} N(0, 1)$$

    .

```
alpha = 0.05
zlower = qnorm(alpha/2)
zupper = qnorm(1-alpha/2)
c(zlower, zupper)
```

```
## [1] -1.959964  1.959964
```

The rejection region is $(-\infty, -1.96], [1.96, \infty)$.

c & d & e.

```
set.seed(1)
n = 30
mu = 8
sigma = 1

zstat = c()
rrcoverage = c()

for(i in 1:20000) {
  samp = rnorm(n, mu, sigma)

  zstat[i] = (mean(samp) - mu) / (sigma / sqrt(n))

  rrcoverage[i] = zstat[i] <= zlower || zstat[i] >= zupper
}
mean(rrcoverage)
```
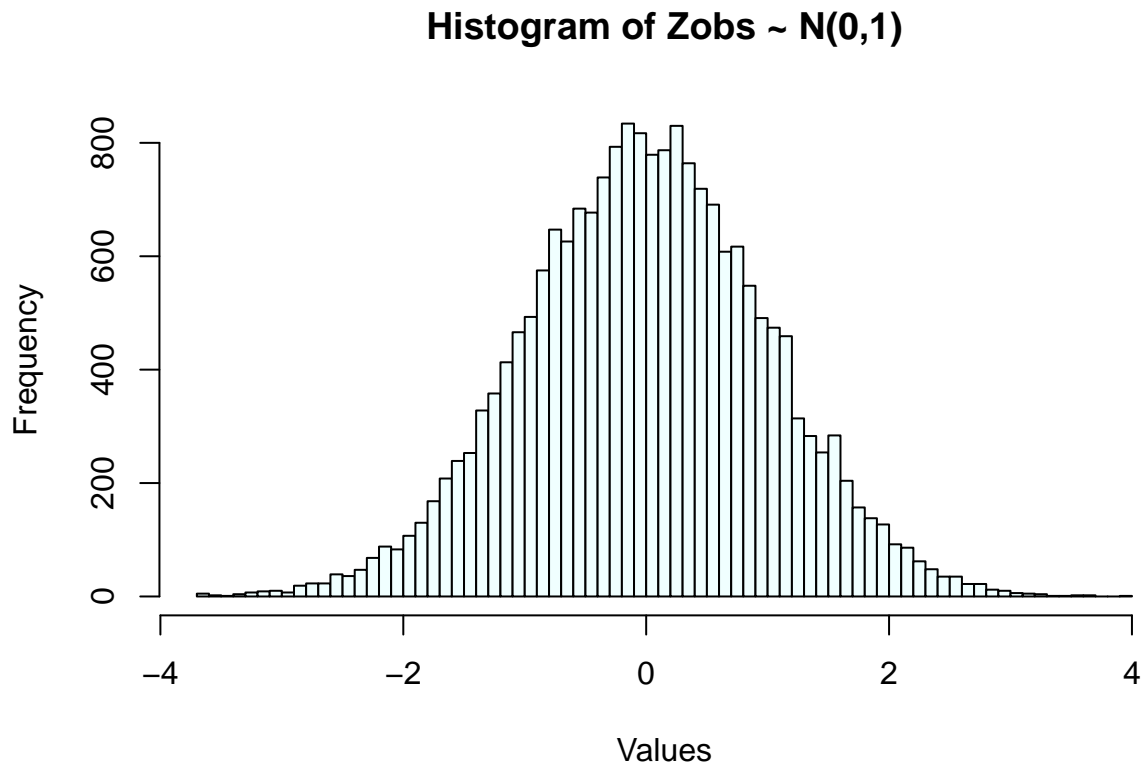
```
## [1] 0.05025
```

```
hist(zstat, col = "azure", xlab = "Values", main = "Histogram of Zobs ~ N(0,1)", breaks = 100)
```

Histogram of Zobs ~ N(0,1)

d. My answer is 0.05025 which means 5% of the samples fall within the rejection region, thus is a 95% confidence interval. We are 95% confident that Z_{obs} is captured in the confidence interval.

e. As can be seen, the histogram of Z_obs does simulate a normal distribution.

Problem 4:

a. We would use a χ^2 test statistic for this hypothesis.
b. Chi-square distribution.

$$\chi^2_{obs} = (19-1)s^2/1 \sim^{H_0} \chi^2_{19-1}$$

.

```
alpha = 0.01
n = 19

chiupper = qchisq(1-alpha, n - 1)
chiupper
```

```
## [1] 34.80531
```

The rejection region is $[34.81, \infty)$.

c & d & e.

```
set.seed(1)
mu = 8
sigma = 1

chistat = c()
rrcoverage = c()

for(i in 1:20000) {
  samp = rnorm(n, mu, sigma)
  var = var(samp)
  chistat[i] = (n - 1)*var

  rrcoverage[i] = chiupper <= chistat[i]
}
mean(rrcoverage)
```
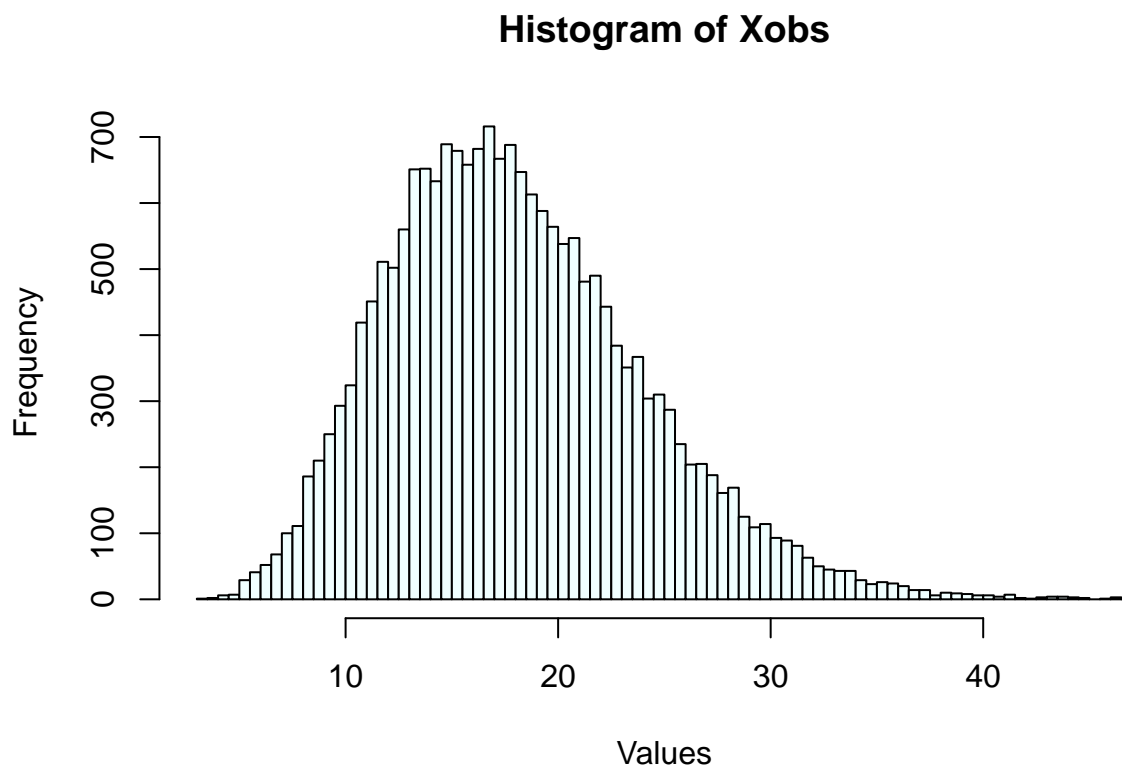
```
## [1] 0.0092
```

```
hist(chistat, col = "azure", xlab = "Values", main = "Histogram of Xobs", breaks = 100)
```

Histogram of Xobs

d. My answer is 0.0092 which means roughly 1% of the samples fall within the rejection region, thus is a 99% confidence interval. We are 99% confident that χ^2_{obs} is captured in the confidence interval.

e. As can be seen, the histogram of χ^2_{obs} does simulate a chi-squared distribution.