Isaac Ng Yi Ming 1002174 Statistics Homework 2

1)

Let $X\ Uniform(1, N)$

$\mathbb{E}(\bar{x}) = \frac{\mathbb{E}(X_1)+\mathbb{E}(X_2)+...+\mathbb{E}(X_n)}{n}$

as $\mathbb{E}(X_n) = \mu$,

$\mathbb{E}(\bar{x}) = \frac{n\mu}{n}$

$\mathbb{E}(\bar{x}) = \mu$

$\mathbb{E}(\bar{X}) = \mu = \frac{N+1}{2}$ (since X is a uniform distribution)

$N = 2\,\mathbb{E}(\bar{X}) - 1$ (since X is a uniform distribution)

2)

The probaility is $P(S^2 > 2\sigma^2)$

$P\left(\frac{S^2}{2\sigma^2} \geq 1\right)$

$P\left(\frac{(2(n-1))S^2}{2\sigma^2} \geq 2(n-1)\right)$

$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

$\frac{2(8-1)S^2}{2\sigma^2} \sim \chi^2_{8-1}$

$\frac{(14)S^2}{2\sigma^2} \sim \chi^2_{8-1}$

$P\left(\chi^2_{8-1} \geq 14\right)$

$1 - P\left(\chi^2_7 \leq 14\right)$

In [1]:

```
lapply(list(8,17,21), function(x){1-pchisq(2*(x-1),x-1)})
```

1. 0.0511813534130654
2. 0.00999978095310483
3. 0.00499541230830758

$n = 8$ does not pass at the $\alpha = 0.05$ confidence level, while $n = 17$ and $n = 21$ do.

3)

In [2]:

```
#3a: we use the normal test as \sigma is known
print("the lower bound of the CI is:")
qnorm(0.025, mean = 16.3, sd = 6/sqrt(100))
print("the upper bound of the CI is:")
qnorm(0.975, mean = 16.3, sd = 6/sqrt(100))
```

```
[1] "the lower bound of the CI is:"
```

15.124021609276

```
[1] "the upper bound of the CI is:"
```

17.475978390724

In [3]:

```
#3b: we use the t test because \sigma is unknown and because we can
print("the lower bound of the CI is:")
qt(0.025, df = 99)*6/sqrt(100)+16.3
print("the upper bound of the CI is:")
qt(0.975, df = 99)*6/sqrt(100)+16.3
```

```
[1] "the lower bound of the CI is:"
```

15.1094698290481

```
[1] "the upper bound of the CI is:"
```

17.4905301709518

In [4]:

```
#3c
print("the lower bound of the CI is:")
qt(0.05, df = 99)*6/sqrt(100)+16.3
print("the upper bound of the CI is:")
qt(0.95, df = 99)*6/sqrt(100)+16.3
```

```
[1] "the lower bound of the CI is:"
```

15.3037653063898

```
[1] "the upper bound of the CI is:"
```

17.2962346936102

3d)

We used the t-distribution, which has heavier tails. As such the confidence interval is wider.

4)

$$Var(X) = \mathbb{E}\left(\bar{X^2}\right) - \mathbb{E}\left(\bar{X}\right)^2$$

$$Var(X) = \mathbb{E}\left(\bar{X^2}\right) - \mu^2$$

when $Var(X) \neq 0$:

$$\mathbb{E}\left(\bar{X^2}\right) \neq \mu^2$$

$$\therefore \mathbb{E}\left(\bar{X^2}\right) \neq \mu^2$$

in general unless

$$Var(X) = 0$$

Hence, $\mathbb{E}\left(\bar{X^2}\right)$ is a biased estimator of μ^2

5)

5a)

H_0 : The Yoghurt is at least 98% fat free.

H_a : The Yoghurt is not at least 98% fat free.

5b)

H_0 : Cloud seeding is an effective technique to increase precipitation

H_a : Cloud seeding is not an effective technique to increase precipitation

6)

$H_0 : \mu = 0$

$H_a : \mu \neq 0$

if $\mu = 0$, the probability of getting a value of $\bar{x}$ at least as extreme as 0.1 at $n = 100$ is:

$$P\left(\left|\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| \leq Z\right) = 1 - P\left(\left|\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| > Z\right)$$

$$1 - P\left(\left|\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| > Z\right) = 1 - P\left(-\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < Z < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$1 - P\left(\left|\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| > Z\right) = 1 - P\left(-\frac{0.1}{\frac{1}{10}} < Z < \frac{0.1}{\frac{1}{10}}\right)$$

$$1 - P\left(\left|\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| > Z\right) = 1 - P\left(-1 < Z < 1\right)$$

$$1 - P\left(\left|\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| > Z\right) = 1 - \left(P\left(Z < 1\right) - P\left(-1 < Z\right)\right)$$

$n = 400$:

$$P\left(\left|\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| > Z\right) = P\left(-\frac{0.1}{\frac{1}{20}} < Z < \frac{0.1}{\frac{1}{20}}\right)$$

$$P\left(\left|\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| > Z\right) = P\left(-2 < Z < 2\right)$$

$n = 900$:

$$P\left(\left|\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| > Z\right) = P\left(-3 < Z < 3\right)$$

In [5]:

```
#6a
twonormtest <- function(x){1-(pnorm(x)-pnorm(-x))}

pval <- lapply(list(1,2,3), twonormtest)

cat("the values for n = 100, 400, 900 are ")
cat(sprintf("%.3f, ", pval))
cat("respectively")

lapply(pval, function(x) {x>0.01})
```

```
the values for n = 100, 400, 900 are 0.317,  0.046,  0.003, respectively
```

1. TRUE
2. TRUE
3. FALSE

6b)

0.317310508 > 0.01, hence we fail to reject H_0 for $\alpha = 0.01$ when n = 100

0.045500264 > 0.01, hence we fail to reject H_0 for $\alpha = 0.01$ when n = 400

0.002699796 < 0.01, hence we reject H_0 for $\alpha = 0.01$ when n = 900

7)

$H_a : \mu \geq 60,000$

Assuming that the null hypothesis is true,

The test statistic is $\frac{60758-60000}{1500/\sqrt{16-1}}$

$= 1.9571$

The P-value is $P\left(1.9571 > Z\right)$

In [6]:

```
zscore <- (60758-60000)/(1500/(sqrt(16-1)))
cat(sprintf("The test statistic is %.3f\n", zscore))
pvalue <- pnorm(zscore)
cat(sprintf("The P-value is %.3f\n", pvalue))
```

```
The test statistic is 1.957
The P-value is 0.975
```

7b)

Assuming that $\mu = 61000$, the power of the test is $1 - \beta$, where $\beta = P\left(H_0|\mu = 61000\right)$

Find the critical value to reject H_a

$P\left(\frac{c-60000}{1500/\sqrt{16-1}} < Z\right) = 0.01$

Where c is the critical value to reject at the $\alpha = 0.01$ level

In [7]:

```
critval <- qnorm(0.99)*(1500/sqrt(16-1))+60000
cat(sprintf("The critical value is %.3f\n", critval))


beta <- pnorm((critval-61000)/(1500*sqrt(16-1)))
power <- 1-beta
cat(sprintf("The power of the test is %.3f\n", power))
```

```
The critical value is 60900.991
The power of the test is 0.507
```

7c

$0.90 = 1 - \beta$

$\beta = 0.1$

Assuming $\alpha = 0.01$;

$n = \left\lceil \frac{\sigma(z_\alpha + z_\beta)}{\mu - \mu_0} \right\rceil^2$

In [8]:

```
sigma <- 1500
alpha <- 0.01
beta <- 0.1
mu <- 61000
mu_0 <- 60000
n <- (sigma*(qnorm(1-alpha)+qnorm(1-beta))/(mu-mu_0))**2
cat(sprintf("At least %i tires should be tested", ceiling(n)))
```

```
At least 30 tires should be tested
```

8a)

Assuming $\sigma_1^2 = \sigma_2^2$,

In [9]:

```
x1 <- c(12.0129, 12.0072, 12.0064, 12.0054, 12.0016, 11.9853, 11.9949, 11.9985, 12.0077
, 12.0061)
x2 <- c(12.0318, 12.0246, 12.0069, 12.0006, 12.0075)
ssq <- ((length(x1)-1)*sd(x1)**2+(length(x2)-1)*sd(x2)**2)/(length(x1)+length(x2)-2)
t <- (mean(x1)-mean(x2))/(sqrt(ssq*((1/length(x1))+(1/length(x2)))))
pval <- pt(t, length(x1)+length(x2)-2)
print(pval)
cat(sprintf("As %.3f <  0.05, we reject the null hypothesis and say that they are not e
qual", pval))
```

```
[1] 0.02494432
As 0.025 <  0.05, we reject the null hypothesis and say that they are not
equal
```

8b)

Assuming $\sigma_1^2 \neq \sigma_2^2$,

In [10]:

```
x1 <- c(12.0129, 12.0072, 12.0064, 12.0054, 12.0016, 11.9853, 11.9949, 11.9985, 12.
0077, 12.0061)
x2 <- c(12.0318, 12.0246, 12.0069, 12.0006, 12.0075)
omega_1 <- sd(x1)**2
omega_2 <- sd(x2)**2
nu <- ((omega_1+omega_2)**2)/(((omega_1**2)/(length(x1)-1))+(omega_2/(length(x2)-1
)))
t <- (mean(x1) - mean(x2))/sqrt((sd(x1)**2)/length(x1)+(sd(x2)**2)/length(x2))
pval <- pt(t,nu)

print(pval)
cat(sprintf("As %.3f > 0.05, we fail to reject the null hypothesis and say that the
y are equal", pval))
```

```
[1] 0.4970267
As 0.497 > 0.05, we fail to reject the null hypothesis and say that they a
re equal
```

9

H_0: There is no difference, $\sigma_1 = \sigma_2$

H_a: There is a difference, $\sigma_1 \neq \sigma_2$

Taking H_0 to be true, $\sigma_1 = \sigma_2$,

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2}{S_2^2} \sim F_{9-1,9-1}$$

In [11]:

```
s_1 <- 2.3
s_2 <- 1.1

Fstat <- s_1/s_2

lowerbound <- qf(0.025, 8, 8)
upperbound <- qf(0.975, 8, 8)
cat(sprintf("The 95 percent confidence interval is [%.3f, %.3f]. \n", lowerbound, upper
bound))
print(Fstat)
print(lowerbound<Fstat & Fstat < upperbound)
```

```
The 95 percent confidence interval is [0.226, 4.433].
[1] 2.090909
[1] TRUE
```

As the F-statistic is within the confidence interval, we fail to reject H_0 and say that there is no differene.

10

H_0: She can tell how the cups are prepared, i.e. $n_{correct} = n_{cups}$

H_a: She cannot tell how the cups are prepared, i.e. $n_{correct} \neq n_{cups}$

We can look at the 12 cups as 6 correct cups and 6 incorrect cups. As she knows how the test works, and under the null hypothesis she can tell the way the cups are prepared, she should guess 6 correct and 6 incorrect cups.

Therefore we can reduce the problem to seeing what she guesses for the 6 cups.

The probability that we can do at least as well as her given random guesses is
$P(X \geq 5), X \sim Binomial(6, 0.5)$

In [12]:

```
1-pbinom(4, 6, 0.5)
```

0.109375

The P-value is lower than $\alpha = 0.05$. Thus we reject the null hypothesis and conclude that she cannot tell how the cups were prepared.