Experimental Statistics for Engineers I

Kamala Dadashova

Problem 1: Consider the sample observations on stabilized viscosity of asphalt specimens: 2781, 2900, 3013, 2856, 2888. Assume they come from a normal distribution. Suppose that for a particular application it is required that the true mean viscosity be less than 3000. Does this requirement appear to have been satisfied? State and test the appropriate hypotheses. Use a level of significance $\alpha = 0.05$, and all the three testing approaches. State your decision and conclusion.

SOLUTION:

Hypotheses: $H_0 : \mu = 3000 \quad H_a : \mu < 3000$

Significance level: $\alpha = 0.05$

Now, let us find average mean of the data,

```
data=c(2781, 2900, 3013, 2856, 2888)
n=length(data)
x_bar=mean(data)
x_bar
```

```
## [1] 2887.6
```

```
s=sd(data)
s
```

```
## [1] 84.02559
```

Test statistic will be

```
t <- (x_bar-3000)*sqrt(n)/s
t
```

```
## [1] -2.991161
```

```
#1) P-value: P(T<t)
pt(t,n-1)
```

```
## [1] 0.02014595
```

```
##0.02014595<0.05. So, we reject  null hypotheses.

# 2)Rejection region:
#Find z_alpha such that P(T<t_alpha)=0.05.
qt(0.05,n-1)
```

```
## [1] -2.131847
```

```
#So, rejection region is  t<-2.131847 which does contain t=-2.991161. Thus, we reject null hypotheses.

#3)CI:95% CI upper bound
x_bar+2.131847*s/sqrt(n)
```

```
## [1] 2967.709
```

```r
#95 %CI=(-\infinity,2967.709) which does not contain 3000. Thus, we reject null hypotheses.

#Verification:
t.test(data,alternative="less", mu=3000)
```

```
##
##  One Sample t-test
##
## data:  data
## t = -2.9912, df = 4, p-value = 0.02015
## alternative hypothesis: true mean is less than 3000
## 95 percent confidence interval:
##      -Inf 2967.709
## sample estimates:
## mean of x
##    2887.6
```

According to all three hypotheses tests, this requirement appear to be not satisfied.

Problem 2: Judy wants to test whether a coin is fair or not. Suppose she observes 477 heads in 900 tosses. Perform an appropriate p-value test using a 10% level of significance, and state your decision and conclusion.

SOLUTION:

We know it is binomial distribution because there are success and fail. A coin is fair if it has probability 0.5. Since n is large, and $n \cdot p = n \cdot (1-p) = 450 > 10$.

Hypotheses: $H_0 : p = 0.5 \quad H_a : p \neq 0.5$.

Significance level: $\alpha = 0.1$

$\hat{p} = \frac{477}{900}$

```r
s=sqrt(0.5*(1-0.5))
s
```

```
## [1] 0.5
```

Test statistic:

```r
n = 900;
p_bar=477/n;
p=0.5;
z <- (p_bar-p)*sqrt(n)/s
z
```

```
## [1] 1.8
```

```r
#1) P-value:
pnorm(-z)+1-pnorm(z)
```

```
## [1] 0.07186064
```

```r
#0.07186064<0.1. We  reject null hypotheses.

#2) Rejection region. P(|Z|>z_alpha/2)=0.1. So, z_alpha/2
qnorm(0.95)
```

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

```r
#So, rejection area is z<-1.644854 or z>1.644854. It does cover z, which is 1.8. We  reject null hypoth

#3) CI: 95%.
ci=c(ci_lower=p_bar-1.644854*s/sqrt(n), ci_upper=p_bar+1.644854*s/sqrt(n))
ci
```

```
##  ci_lower  ci_upper
## 0.5025858 0.5574142
```

```r
#p=0.5 is not in this interval. We reject null hypotheses.

#Verification:

prop.test(477,n,p=0.5,alternative="two.sided",conf.level=0.90,correct=F)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  477 out of n, null probability 0.5
## X-squared = 3.24, df = 1, p-value = 0.07186
## alternative hypothesis: true p is not equal to 0.5
## 90 percent confidence interval:
##  0.5025861 0.5572340
## sample estimates:
##    p
## 0.53
```

Conclusion we reject that the coin is fair.

Problem 3: A manufacturer of sprinkler systems used for fire protection in office buildings claims the true average system activation temperature is greater than 130F. A sample of n = 9 systems yields a sample average activation temperature of 131.08F with standard deviation 1.5F. Does the data contradict the manufacturer's claim at the significance level $\alpha = 0.01$? Use all the three methods we discussed, and assume the activation temperature is normally distributed. State your decision and conclusion.

SOLUTION: Since sample is samll and standard deviation of sample is given, we use t distribution.

Hypotheses:

$H_0 : \mu = 130 \quad H_a : \mu > 130$

Significance level: $\alpha = 0.01$

```r
t <- (131.08-130)*sqrt(9)/1.5
t
```

```
## [1] 2.16
```

```r
# 1) P-value. P(T>t)
1-pt(t,9-1)
```

```
## [1] 0.0313947
```

```r
#0.0313947>0.01. We fail to reject null hypothesis.

#2) Rejection region. P(T>t_alpha)=0.1. So, t_alpha
qt(0.99,9-1)
```

```
## [1] 2.896459
```

```
#Rejection region is t>2.896459. which does not contain t=2.16. We fail to reject null hypothesis.

#3) CI: 99%.
ci=131.08-2.896459*1.5/sqrt(9)
ci
```

```
## [1] 129.6318
```

```
#Confidence interval is t>129.6318, which covers mu=130. We fail to reject null hypothesis.
```

Conclusion: According to all three tests above, we accept that a manufacturer of sprinkler systems used for fire protection in office buildings claims the true average system activation temperature is greater than 130F.

Problem4: Simulation can be used to investigate the performance of the large sample Z-test statistic and gain better understanding of the meaning of p-value. Let μ be the true mean of a chisquare distribution with degrees of freedom $\nu = 5$. Assume we want to test the performance of the large sample Z-test for the hypotheses $H_0 : \mu = 5$ versus $H_a : \mu > 5$ at the significance level $\alpha = 0.05$ for a sample size $n = 50$. For this simulation, you should:

a) Generate a random sample of 50 observations drawn from a χ^2_5 distribution.

```
set.seed(987985)
chi_sample=rchisq(n=50, df=5, ncp = 0)
chi_sample
```

```
##  [1]  2.369276  4.014316  3.307747  2.598740  4.904609  4.931678  9.067833
##  [8]  3.500958  5.330246  9.895637  6.686952  2.729916  5.465436  4.782427
## [15]  1.517509  3.823234  3.731172  1.467169  3.247867  5.083666  2.036608
## [22]  4.176472  3.640991  6.612324  6.728202  8.993820  2.543248  3.327020
## [29] 11.195243  4.047645 10.725311  3.504616  1.729198  6.428600  2.676569
## [36] 17.312705  1.223412  1.529453  9.132694  2.705878 10.586320  3.564754
## [43]  2.760987  1.136014  7.928279  8.953137 11.679166  5.771199  1.772298
## [50]  7.777884
```

b) Calculate the appropriate test statistic and p-value.

```
n=50;
nu=5;
s=sd(chi_sample)
s
```

```
## [1] 3.414265
```

```
x_bar = mean(chi_sample)

#z
z <- (x_bar-5)*sqrt(n)/s
z
```

```
## [1] 0.4413972
```

```
#P-value.
p.value <- 1 - pnorm(z)
p.value
```

```
## [1] 0.3294627
```

```
z.alpha <- qnorm(0.95)
z.alpha
```

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

```
ci <- x_bar-abs(z.alpha)*s/sqrt(n)
ci
```

```
## [1] 4.418911
```

c) Reject or fail to reject the null hypothesis; record the result as 1 if you reject and 0 if you fail to reject.

Since p values test result is $0.3294627 > 0.05$, we fail to reject null hypothesis. Record is 0.

d) Repeat this process for 5000 random samples.

```
set.seed(5)
chi_sample=rchisq(n=5000, df=5, ncp = 0)
n=5000;
nu=5;
s=sd(chi_sample)
s
```

```
## [1] 3.205982
```

```
x_bar = mean(chi_sample)

#z
z <- (x_bar-5)*sqrt(n)/s
z
```

```
## [1] 0.5387529
```

```
#P-value.
p.value <- 1 - pnorm(z)
p.value
```

```
## [1] 0.2950287
```

```
z.alpha <- qnorm(0.95)
z.alpha
```

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

```
ci <- x_bar-abs(z.alpha)*s/sqrt(n)
ci
```

```
## [1] 4.94985
```

Since p value , $0.2950287 > 0.05$, we fail to reject null hypothesis. Record is zero.

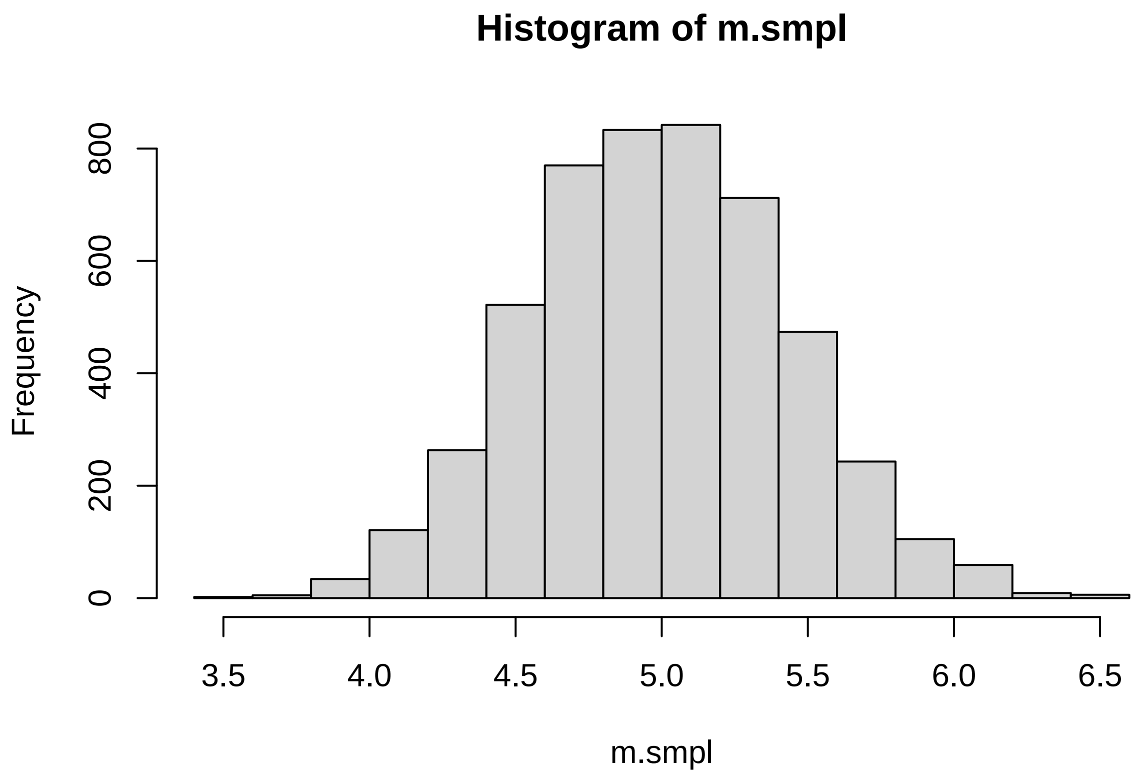
e) Report the proportion of times that you reject the null hypothesis in the 5000 simulations. Comment on how the test performed relative to the significance level.

```
set.seed(1)
B <- 5000  # set number of simulated tests
n <- 50    # sample size
m0 <- 5     # null hypothesis for true mean; alt mu > m0
m.smpl <- rep(0,B)   # create dummy vectors to store simulation statistics and p-values
stat <- rep(0,B)
p.app <- rep(0,B)  # large sample p-value
rej.app <- rep(0,B) # "1" if reject for large sample test, "0" if fail to reject
for (i in 1:B){
  smpl <- rchisq(n=n,df=5)
  m.smpl[i] <- mean(smpl)  # calculate sample mean
```

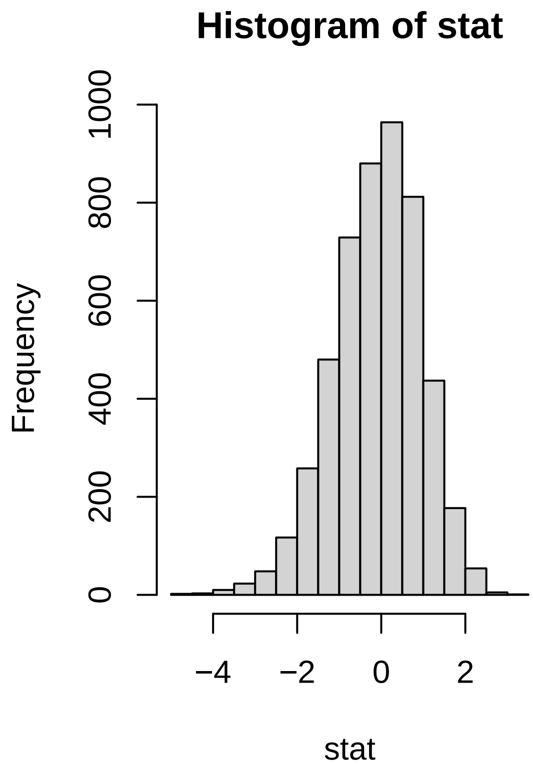
```
  stat[i] <- (mean(smpl)-m0)/(sd(smpl)/sqrt(n))  # test statistic
  p.app[i] <- pnorm(stat[i],lower.tail=F)  # large sample p-value; Ha: mu>m0
  if(p.app[i]<0.05) rej.app[i] <- 1  # if p-value < 0.05, set rej array value to "1"
}
mean(rej.app)  # % rejected for large sample test
```

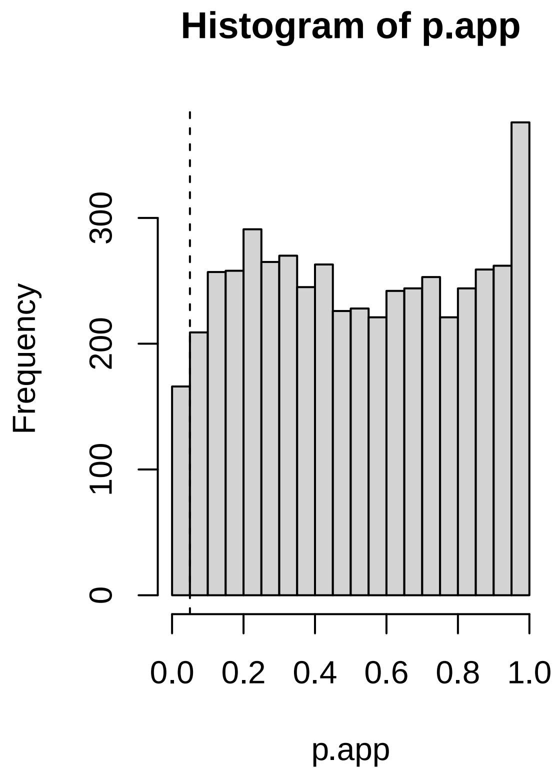
```
## [1] 0.0332
```

```
par(mfrow=c(1,1))   # plot distribution of sample means
hist(m.smpl,breaks=20)
```




```
par(mfrow=c(1,2))  # plot distribution of sample statistics and large sample p=values
hist(stat,breaks=20)
hist(p.app,breaks=20)
abline(v=0.05,lty=2)
```





The proportion of times that we reject the null hypothesis in the 5000 simulations is 0.0332. It is less than significance level, 0.05.