Hypothesis Testing

Anirban Mondal

11/27/2020

Example: Sulfer level (from lecture slides)

A chemical plant is required to maintain ambient sulfur levels in the working environment atmosphere at an average level of no more than 12.5. The results of 15 randomly timed measurements of the sulfur level produced a sample mean of $\bar{x}$= 14.82. Assume that the population is approximately normal and standard deviation of these sulfur measurements is $\sigma = 3.35$.

a)

What is the evidence that the chemical plant is in violation of the working code?

Step 1: Parameter of interest

μ: The mean sulpher level in the environment

Step 2: Set up hypothesis $H_0 : \mu = 12.5 \; vs. H_1 : \mu > 12.5$ or $H_0 : \mu \leq 12.5 \; vs. H_1 : \mu > 12.5$

Step 3: Test Statistic and Critical Region:

Test Stistic: $Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

Critical Region: $z_{obs} > z_\alpha$

Step 4: Computation

```
x_bar=14.82
sig = 3.35
n=15
a=0.05
z_obs=(x_bar - 12.5)/(sig/sqrt(n))
z_a= qnorm(1-a)
z_obs
```

```
## [1] 2.682185
```

```
z_a
```

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

Ans: Since observed value of the test statistic is greater than $Z_{0.05}$ we reject H_0 at 5% level of significance and conclude that the chemical plant is in violation of the working code.

b)

Find the p-value of the test and use it to test the hypothesis in a)

P-value $= P(Z > z_{obs})$

```
pval = 1-pnorm(z_obs)
pval
```

```
## [1] 0.003657145
```

Ans: As p-value is less than 0.05, we reject H_0 at 5% level of significance and conclude that the chemical plant is in violation of the working code.

c)

Find the power of the test when $\mu = 13$.

$$\begin{aligned} Power = 1 - \beta &= P(Z_0 > z_{0.05} | \mu = 13) \\ &= P(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > z_{0.05} | \mu = 13) \\ &= P(\bar{X} > \mu_0 + z_{0.05} * \frac{\sigma}{\sqrt{n}} | \mu = 13) \\ &= P(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > z_{0.05} - (\mu - \mu_0)\sqrt{n}/\sigma) \\ &= P(Z > z_{0.05} - \delta\sqrt{n}/\sigma) \end{aligned}$$

```
delta=13-12.5 ## mu - mu_0
Power = 1-pnorm(qnorm(1-0.05) -delta*sqrt(n)/sig) ## z_\alpha
Power
```

```
## [1] 0.1430319
```

d)

What sample size do we need to detect a difference between the true and hypothesized mean of 2.0 with power at least 95%?

$n \approx \frac{(z_{\alpha/2}+z_\beta)^2\sigma^2}{\delta^2}$

```
delta = 2
b = 1-0.95
a=0.05
z_a= qnorm(1-a)
z_b= qnorm(1-b)
(z_a+z_b)^2*sig^2/(delta^2)
```

```
## [1] 30.36296
```

Ans: We need at least 31 samples.

d) Use the 95% one sided CI for μ to answer part a)

```
#LCL=x_bar - z_a*sig/sqrt(n)
LCL=x_bar - z_a*sig/sqrt(n)
LCL
```

```
## [1] 13.39726
```

Ans: The 95% CI for μ (13.397, ∞) does not include the value under H_0, (12.5), hence we reject H_0 at 5% level of significance and conclude that the chemical plant is in violation of the working code.

Example: Hot dogs calorie content from lecture slides

Data on calorie content in 20 different beef hot dogs from Consumer Reports (June 1986 issue): 186, 181, 176, 149, 184, 190, 158, 139, 175, 148, 152, 111, 141, 153, 190, 157, 131, 149, 135, 132. Assume that these numbers are observed values from a random sample of twenty independent $N(\mu, \sigma^2)$ random variables, where μ and σ^2 are unknown. Observed sample mean and standard deviations are $\bar{x} = 156.85$ and s = 22.64201.

a)

Do these data give strong evidence that the average calorie content in beef hot dogs is less than 180?

Step 1: Parameter of interest

μ: The mean calorie content of all such hotdogs.

Step 2: Set up hypothesis $H_0 : \mu = 180\ vs. H_1 : \mu < 180$ or $H_0 : \mu \geq 180\ vs. H_1 : \mu < 180$

Step 3: Test Statistic and Critical Region:

Test Stistic: $T_0 = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$

Critical Region: $t_{obs} < -t_{\alpha,n-1}$

```
data=c(186, 181, 176, 149, 184, 190, 158, 139, 175, 148, 152, 111, 141, 153, 190, 157, 131, 149, 135, 13

## using summary statistics and t interval formula
n=length(data)
x_bar=mean(data)
s = sd(data)
x_bar
```

```
## [1] 156.85
```

```
s
```

```
## [1] 22.64201
```

```
t_obs=(x_bar-180)/(s/sqrt(n))
t_a=qt(1-.05, n-1)
t_obs
```

```
## [1] -4.572472
```

```
-t_a
```

```
## [1] -1.729133
```

```
pt(t_obs,n-1) # P(T<t_obs)
```

```
## [1] 0.0001040172
```

```
# UCL=x_bar + t_a*s/sqrt(n)
# UCL

## using data directly
t.test(data, mu=180, alternative="less", conf.level=0.95)
```

```
##
##  One Sample t-test
##
## data:  data
## t = -4.5725, df = 19, p-value = 0.000104
## alternative hypothesis: true mean is less than 180
## 95 percent confidence interval:
##      -Inf 165.6044
## sample estimates:
```

```
## mean of x
##    156.85
```

Ans: Since test statistic is less than the critical value we reject H_0 and conclude that there is strong evidence that the average calorie content in beef hot dogs is less than 180.

b)

If the true mean calorie content is 166 is the sample size n = 20 adequate that the null hypothesis in a) would be rejected with probability at least 0.9.

Need to check if the power of the test at $\mu = 166$ is greater than 0.9 with n=20 and $\alpha = 0.05$

Power $= P(T_0 < -t_{0.05,n-1} | \mu = 166)$

```
a=0.05
delta=(166-180)/(s/sqrt(n)) ## (mu-mu_0)/(s/sqrt(n))
Power = pt(-qt(1-a, n-1),ncp=delta, df=n-1) #P(T<-t_a, ncp=d)
Power
```

```
## [1] 0.8460901
```

Example: Metal Rod from lectute slides

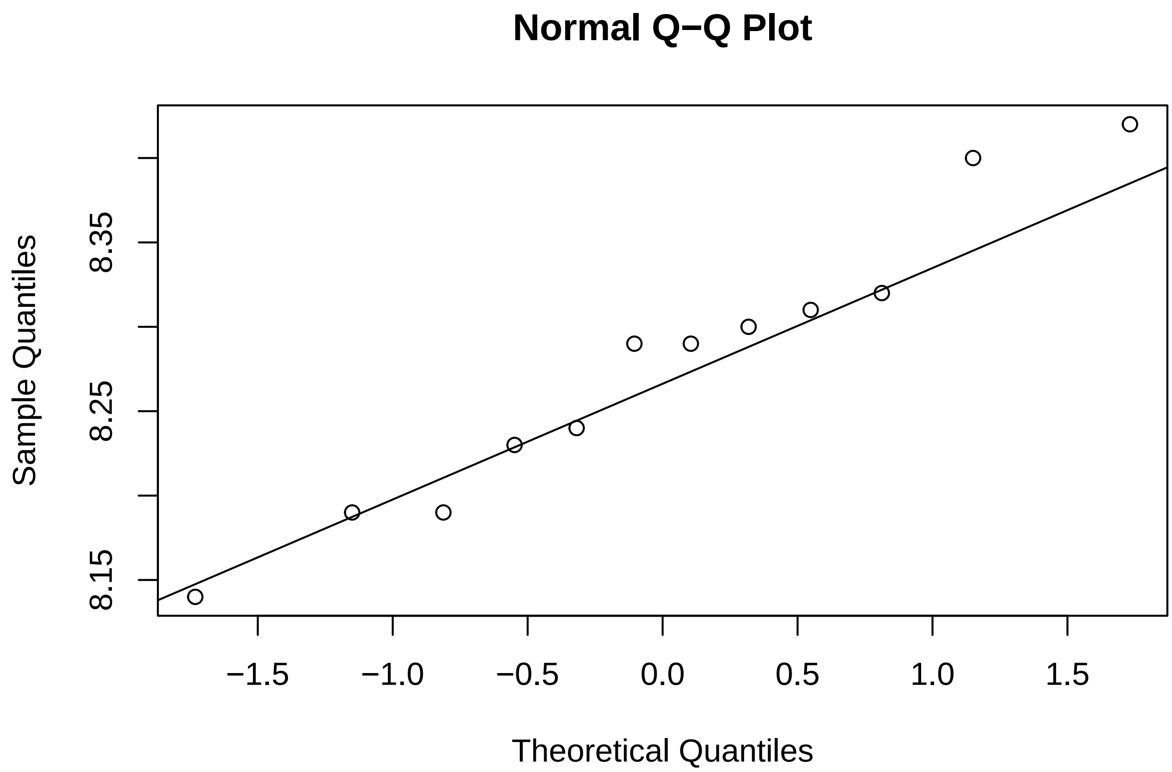
A machine produces metal rods used in an automobile suspension system. A random sample of 12 rods is selected and the diameter is measured. The resulting data is given below. 8.23, 8.29, 8.19, 8.14, 8.31, 8.19, 8.29, 8.32, 8.42, 8.24, 8.30, 8.40

a)

Check the assumption of normality for rod diameter.

```
data = c(8.23, 8.29, 8.19, 8.14, 8.31, 8.19, 8.29, 8.32, 8.42, 8.24, 8.30, 8.40)

### a) Normality check
qqnorm(data)
qqline(data)
```

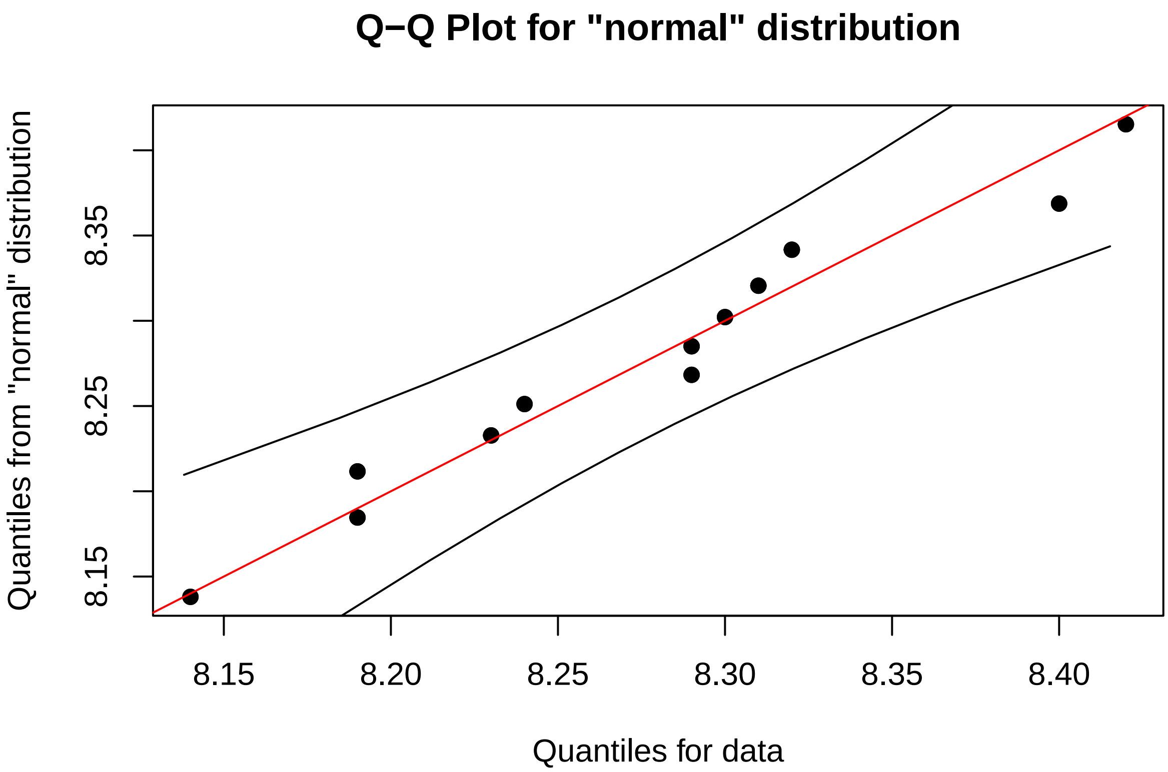



```
library(qualityTools)
```

```
##
## Attaching package: 'qualityTools'
## The following object is masked from 'package:stats':
##
##     sigma
```

```
qqPlot(data)
```

```
## Loading required package: MASS
```



The data appears to come from a Normal distribution as the q-q plot follows a straight line

b)

Is there a strong evidence that the mean rod diameter is not 8.20 using a fixed level 0.05 test.

H_0: $\mu = 8.2$ vs. H_1: $\mu \neq 8.2$

```
t.test(data, mu=8.2, alternative="two.sided", conf.level=0.95)
```

```
##
##  One Sample t-test
##
## data:  data
## t = 3.1771, df = 11, p-value = 0.008807
## alternative hypothesis: true mean is not equal to 8.2
## 95 percent confidence interval:
##  8.223554 8.329779
## sample estimates:
## mean of x
##  8.276667
```

Since the p-value is very close to 0 (<0.05) we reject H_0 and conclude the mean rod diameter is signifantly different from 8.20.

c)

Find the P-value for this test

d)

Find the 95% confidence interval on mean rod diameter.

Ans: the 95% CI for the mean rod diameter is given by (8.223554 8.329779)

Example: Defective calculator

A manufacturer of electronic calculators is interested in estimating the fraction of defective units produced. A random sample of 800 calculators contains 10 defectives.

a)

Formulate and test appropriate hypothesis to determine if the fraction of defectives exceeds 0.01 at 5% level of significance.

H_0: p=0.01 vs. H_1: p > 0.01

```
a=0.05
n=800
x=10
p_hat = x/n
p_hat
```

```
## [1] 0.0125
```

```
z_obs = (p_hat - 0.01)/sqrt(0.01*0.99/n)
z_a=qnorm(1-a)
z_obs
```

```
## [1] 0.7106691
```

```
z_a
```

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

```
1-pnorm(z_obs)
```

```
## [1] 0.2386447
```

Ans: Since the observed value of the test statistic 0.71 is less than the critical value 1.645 we fail to reject H_0 at level 0.05 and conclude that there is no signifant evidence in the data that fraction of defectives exceeds 0.01.

or

Since the p-value, 0.2386 is greater than 0.05 we fail to reject H_0 and conclude that there is no signifant evidence in the data that fraction of defectives exceeds 0.01.

b)

Suppose that the true p =0.02 and $\alpha = 0.05$. What is the power for this test?

```
p0=0.01
p1=0.02

## prob of fail to reject H0 when p=0.02 P(phat<z_a+sd(phat)|p=0.02)
beta = pnorm(p0+z_a*sqrt(p0*(1-p0)/n),p1, sqrt(p1*(1-p1)/n))
pow=1-beta
pow
```

```
## [1] 0.8026983
```

```r
## using the formula from book
1-pnorm((p0-p1+z_a*sqrt(p0*(1-p0)/n))/sqrt(p1*(1-p1)/n))
```

```
## [1] 0.8026983
```

c)

Suppose that the true p =0.02 and $\alpha = 0.05$. How large a sample would be required if we want the power to be at least 0.9.

```r
z_b=qnorm(0.9)
nb=(z_a*sqrt(p0*(1-p0)) + z_b*sqrt(p1*(1-p1)))^2/(p1-p0)^2
nb
```

```
## [1] 1177.026
```

```r
n=1178
1-pnorm((p0-p1+z_a*sqrt(p0*(1-p0)/n))/sqrt(p1*(1-p1)/n))
```

```
## [1] 0.9001778
```

Ans: The required sample size is 1178

Goodness of fit tests

SAT Exam score example

To study the SAT scores (verbal) of students admitted to all the universities in a state, 200 students are randomly sampled from the universities in that state and their scores are observed.

Question: Can you infer from this sample that the SAT scores of al students in that state can be described by a normal probability model?

Hypothesis: H_0: The population is normal vs H_1 The population is not normal

```r
## loadinf the data
data=read.csv("SAT_score.csv")
sat=data$Score

### Shapiro-Wilk normality test

## The test is based on the order statistics (quantiles)
shapiro.test(sat)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  sat
## W = 0.99475, p-value = 0.713
```

Ans: As the p value is greater than 0.05, we fail to reject the null hypothesis that the distribution of SAT scores is Normal.