5.4 Comparing Two Proportions

We look at the *2-proportion Z-test* for the difference in two proportions, \( p_1 - p_2 = c \),

\[
z = \frac{\hat{p}_1 - \hat{p}_2 - c}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}},
\]

where we assume the samples random and there are at least 5 successes and 5 failures in each sample and

\[
\hat{p}_1 = \frac{x_1}{n_1}, \quad \hat{p}_2 = \frac{x_2}{n_2}, \quad \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}
\]

and where binomial conditions are satisfied. Two-sided, right-sided and left-sided tests are considered.

**Exercise 5.4 (Comparing Two Proportions)**

1. Inference \( p_1 - p_2 \), large independent samples: doctors.
   Consider number of male doctors in military and civilian hospitals. Test the claim there is a smaller proportion of male doctors in military than in civilian life at \( \alpha = 0.05 \).

<table>
<thead>
<tr>
<th></th>
<th>military (1)</th>
<th>civilian (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>male doctors</td>
<td>358</td>
<td>6786</td>
</tr>
<tr>
<td>total doctors</td>
<td>407</td>
<td>7363</td>
</tr>
</tbody>
</table>

   (a) Hypothesis test.

   i. Check assumptions.
   Since there are at least 5 male and 5 female doctors in both military and civilian life, the assumptions (i) are (ii) are not satisfied, so continue.

   ii. Statement. Choose one.
   A. \( H_0 : p_1 - p_2 = 0 \) versus \( H_1 : p_1 - p_2 < 0 \)
   B. \( H_0 : p_1 - p_2 = 0 \) versus \( H_1 : p_1 - p_2 > 0 \)
   C. \( H_0 : p_1 - p_2 = 0 \) versus \( H_1 : p_1 - p_2 \neq 0 \)

   iii. Test.
   Test statistic of \( \hat{p}_1 - \hat{p}_2 \), with pooled proportion,

   \[
   \hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{358 + 6786}{407 + 7363} = \]

   (i) 0.9194 (ii) 0.9934 (iii) 0.9993, is

   \[
z = \frac{(\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}}) - c}{\sqrt{\frac{358}{407} - \frac{6786}{7363} - 0}} = \]
(i) \(-1.23\) (ii) \(-3.03\) (iii) \(-4.56\), so chance \(\hat{p}_1 - \hat{p}_2 = \frac{358}{407} - \frac{6786}{7363} = -0.042\) or less, if \(p_1 - p_2 = 0\), is

\[
P(\hat{p}_1 - \hat{p}_2 \leq -0.042) = P \left( \frac{(\hat{p}_1 - \hat{p}_2) - c}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}} \leq -3.03 \right) \approx P(Z \leq -3.03) \approx
\]

(i) 0.001 (ii) 0.025 (ii) 0.109.

```
prop2.test <- function(x, n, p.null, signif.level, type) {
  x1 <- x[1]; x2 <- x[2]; n1 <- n[1]; n2 <- n[2]
  p.hat1 <- x1/n1; p.hat2 <- x2/n2
  p.hat <- (x1+x2)/(n1+n2)
  z.test.statistic <- ((p.hat1-p.hat2)-p.null)/(sqrt(p.hat*(1-p.hat)*(1/n1+1/n2)))
  if(type=="right") {
    z.crit <- -1*qnorm(signif.level)
    p.value <- 1-pnorm(z.test.statistic)
  }
  if(type=="left") {
    z.crit <- qnorm(signif.level)
    p.value <- pnorm(z.test.statistic)
  }
  if(type=="two.sided") {
    z.crit <- c(qnorm(signif.level/2),-1*qnorm(signif.level/2))
    p.value <- 2*min(1-pnorm(z.test.statistic),pnorm(z.test.statistic))
  }
  dat <- c(p.null, p.hat, z.crit, z.test.statistic, p.value)
  if(type != "two.sided") names(dat) <- c("p.null", "p.hat", "z crit value", "z test stat", "p value")
  return(dat)
}
```

```
prop2.test(c(358,6786), c(407,7363), 0, 0.05, "left") # approx 2-proportion z-test for p, left-side

p.null  p.hat  z crit value  z test stat  p value
0.000000000 0.919433719 -1.644853627 -3.032629048 0.001212167
```

Level of significance \(\alpha =\) (i) 0.01 (ii) 0.05 (iii) 0.10.

iv. Conclusion.

Since \(p\)-value = 0.001 < \(\alpha = 0.050\),

(i) do not reject (ii) reject null guess: \(H_0: p_1 - p_2 = 0\).

Sample \(\hat{p}_1 - \hat{p}_2\) indicates population \(p_1 - p_2\)

(i) is less than (ii) equals (iii) is greater than 0: \(H_1: p_1 - p_2 < 0\).

In other words, the population proportion of male military doctors is

(i) less than (ii) equal to (iii) greater than (iv) different from

the population proportion of male civilian doctors.


<table>
<thead>
<tr>
<th>terms</th>
<th>doctor example</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) population</td>
<td>(a) male doctor or not, 7363 civilian, 407 military doctors</td>
</tr>
<tr>
<td>(b) sample</td>
<td>(b) (\hat{p}_1 - \hat{p}_2)</td>
</tr>
<tr>
<td>(c) statistic</td>
<td>(c) male doctor or not, all civilian/military doctors</td>
</tr>
<tr>
<td>(d) parameter</td>
<td>(d) (p_1 - p_2)</td>
</tr>
</tbody>
</table>
2. *Inference* $p_1 - p_2$, *large independent samples: smokers*. Consider number of male smokers and female smokers. Test the claim there is 2% more male smokers (1) than female smokers (2) at $\alpha = 0.05$, where $x_1 = 358$, $n_1 = 958$, $x_2 = 267$ and $n_2 = 869$.

(a) *Hypothesis test.*

i. *Check assumptions.*
Since there are at least 5 male and female smokers and nonsmokers, the assumptions (i) are (ii) are not satisfied, so continue.

ii. *Statement.* Choose one.
A. $H_0 : p_1 - p_2 = 0.02$ versus $H_1 : p_1 - p_2 < 0.02$
B. $H_0 : p_1 - p_2 = 0$ versus $H_1 : p_1 - p_2 > 0$
C. $H_0 : p_1 - p_2 = 0.02$ versus $H_1 : p_1 - p_2 > 0.02$

iii. *Test.*
Test statistic of $\hat{p}_1 - \hat{p}_2$, with *pooled* proportion,

\[
\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{358 + 267}{958 + 869} = \frac{625}{1827} = 0.34234177
\]

\[
\begin{align*}
(i) & \quad 0.2344 \quad (ii) \quad 0.3421 \quad (iii) \quad 0.5993, \\
& \quad \text{is}
\end{align*}
\]

\[
\begin{align*}
z = \frac{(\hat{p}_1 - \hat{p}_2) - c}{\sqrt{\hat{p}(1-\hat{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} & = \frac{(\frac{358}{958} - \frac{267}{869}) - 0.02}{\sqrt{\frac{0.3421(1-0.3421)}{958} + \frac{0.3421(1-0.3421)}{869}}} = \frac{1.31}{0.17} \\
& \quad \text{so}
\end{align*}
\]

\[
P(\hat{p}_1 - \hat{p}_2 \geq 0.02) = P \left( \frac{(\hat{p}_1 - \hat{p}_2) - 0.02}{\sqrt{\hat{p}(1-\hat{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \geq 2.09 \right) \approx P(Z \geq 2.09) \approx 0.018.
\]

\[
\begin{align*}
& \quad (i) \quad 0.001 \quad (ii) \quad 0.003 \quad (iii) \quad 0.018. \\
& \quad \text{prop2.test(c(358,267), c(958,869), 0.02, 0.05, "right") \# approx 2-proportion z-test for p, right-sided}
\end{align*}
\]

\[
\begin{array}{cccc}
p.null & p.hat & z & \text{crit value} \quad z \text{ test stat} \quad p \text{ value} \\
0.02000000 & 0.34209086 & 1.64485363 & 2.08984071 & 0.01831606
\end{array}
\]

Level of significance $\alpha = (i) \quad 0.01 \quad (ii) \quad 0.05 \quad (iii) \quad 0.10.$

iv. *Conclusion.*
Since $p$-value = 0.018 < $\alpha = 0.050$,

(i) *do not reject*  (ii) *reject* null guess: $H_0 : p_1 - p_2 = 0.02$. 

Sample $\hat{p}_1 - \hat{p}_2$ indicates population $p_1 - p_2$ is
(i) less than (ii) greater than 0.02: $H_1 : p_1 - p_2 > 0.02$.
In other words, population proportion of male smokers is
(i) less than (ii) equal to (iii) greater than (iv) different from
population proportion of female smokers by 2%.

### 5.5 Comparing Two Variances

We use the $F$ distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom to perform a $F$-test of the ratio of two variances, $\frac{\sigma_1^2}{\sigma_2^2}$, with test statistic

$$f = \frac{s_1^2}{s_2^2}$$

and also determine a $100(1 - \alpha)\%$ confidence interval with numbers $f_L$ and $f_R$ such that $P(f_L \leq F \leq f_R) = 1 - \alpha$,

$$\left( \frac{1}{f_R} \cdot \frac{s_1^2}{s_2^2}, \frac{1}{f_L} \cdot \frac{s_1^2}{s_2^2} \right),$$

where both can be used when both underlying distributions are normal with no outliers and random samples have been collected and where it must be $s_1^2 > s_2^2$.

**Exercise 5.5 (Comparing Two Variances)**

1. Test $\frac{\sigma_1^2}{\sigma_2^2}$: plasma levels.
   Consider statistics on plasma levels for males and females given in table. Test if $\sigma_1^2 > \sigma_2^2$ at $\alpha = 0.05$. Assume both simple random samples normal with no outliers and collected independently of one another.

<table>
<thead>
<tr>
<th></th>
<th>males (1)</th>
<th>females (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{x}$</td>
<td>3.259</td>
<td>1.413</td>
</tr>
<tr>
<td>$s$</td>
<td>0.16</td>
<td>0.09</td>
</tr>
<tr>
<td>$n$</td>
<td>9</td>
<td>6</td>
</tr>
</tbody>
</table>

   (a) Statement. Choose one.
   i. $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_1 : \sigma_1^2 > \sigma_2^2$
   ii. $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_1 : \sigma_1^2 < \sigma_2^2$
   iii. $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_1 : \sigma_1^2 \neq \sigma_2^2$

   Since $\sigma_1^2 = \sigma_2^2$, $\frac{\sigma_1^2}{\sigma_2^2} = 1$,

   so $H_0 : \sigma_1^2 = \sigma_2^2$ is the same as $H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$. 
Chapter 5. Hypothesis Testing (LECTURE NOTES 10)

(b) Test.
With \( n_1 - 1 = 9 - 1 = 8 \) df and \( n_2 - 1 = 6 - 1 = 5 \) df,
chance observed \( \frac{s_1^2}{s_2^2} = \frac{0.16^2}{0.09^2} \approx 3.16 \) or more, if \( \sigma_1^2 = \sigma_2^2 \), is

\[
p \text{-value} = P \left( F \geq \frac{s_1^2}{s_2^2} \right) = P \left( F \geq \frac{0.16^2}{0.09^2} \right) \approx P \left( F \geq 3.16 \right) \approx \]

(i) \( 0.02 \)  (ii) \( 0.04 \) (iii) \( 0.11 \).

\[
\text{var2.F.test} <- \text{function}(s1, n1, s2, n2, s.null, signif.level, type) \{ \\
  \text{F.test.statistic} <- s1^2/s2^2 \\
  \text{if(type=="right")} \{ \\
    \text{F.crit} <- \text{qf(signif.level,n1-1,n2-1,lower.tail=FALSE)} \# F, n1,n2 df, alpha critical value \\
    \text{p.value} <- \text{1-pf(F.test.statistic,n1-1,n2-1)} \\
  \} \\
  \text{if(type=="two.sided")} \{ \\
    \text{F.crit} <- \text{qf(signif.level/2,n1-1,n2-1,lower.tail=FALSE)} \# F, n1,n2 df, alpha/2 critical value \\
    \text{p.value} <- 2*(\text{1-pf(F.test.statistic,n1-1,n2-1)}) \# twice area to right \\
  \} \\
  \text{dat} <- \text{c(s.null, F.crit, F.test.statistic, p.value)} \\
  \text{names(dat)} <- \text{c("s.null", "F crit value", "F test stat", "p value")} \\
  \text{return(dat)} \\
\}
\]

\[
\text{var2.F.test}(0.16, 9, 0.09, 6, 1, 0.05, "right") \# approx 2-variance F-test, right-sided \\
\text{s.null} \text{ F crit value} \text{ F test stat} \text{ p value} \\
1.0000000 4.8183195 3.1604938 0.1102753 \\
\]

Level of significance \( \alpha = (i) 0.01 \)  (ii) 0.05 (iii) 0.11.

(c) Conclusion.
Since \( p \text{-value} = 0.11 > \alpha = 0.05 \),
(i) do not reject  (ii) reject null guess: \( H_0 : \sigma_1^2 = \sigma_2^2 \).
Sample \( \frac{s_1^2}{s_2^2} \) indicates population \( \frac{\sigma_1^2}{\sigma_2^2} \)
(i) is greater than  (ii) equals  (iii) does not equal 1: \( H_0 : \sigma_1^2 = \sigma_2^2 \).
In other words, population SD in male plasma is
(i) less than  (ii) equal to  (iii) greater than  (iv) different from
population proportion SD in female plasma.

2. Confidence interval \( \frac{s_1^2}{s_2^2} : \text{plasma levels.} \) Calculate at 95% confidence interval of \( \frac{s_1^2}{s_2^2} \) for the following data. Assume both simple random samples normal with no outliers and collected independently of one another.

<table>
<thead>
<tr>
<th></th>
<th>males (1)</th>
<th>females (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{x} )</td>
<td>3.259</td>
<td>1.413</td>
</tr>
<tr>
<td>( s )</td>
<td>0.16</td>
<td>0.09</td>
</tr>
<tr>
<td>( n )</td>
<td>9</td>
<td>6</td>
</tr>
</tbody>
</table>

(a) The 95% confidence interval for \( \frac{s_1^2}{s_2^2} \) is

\[
\left( \frac{1}{f_R} \cdot \frac{s_1^2}{s_2^2} \cdot \frac{1}{f_L} \cdot \frac{s_1^2}{s_2^2} \right) \approx \left( \frac{1}{6.76} \cdot \frac{0.16^2}{0.09^2} \cdot \frac{1}{0.21} \cdot \frac{0.16^2}{0.09^2} \right) \approx \]

\[
\]
(i) $(0.69, 3.90)$  (ii) $(0.21, 6.76)$  (iii) $(0.47, 15.22)$.

(b) and so the 95% confidence interval for $\frac{\sigma_1^2}{\sigma_2^2}$ is $(\sqrt{0.47}, \sqrt{15.22}) \approx (0.69, 3.90)$  (ii) $(0.21, 6.76)$  (iii) $(0.47, 15.22)$.

\[
\text{var2.F.interval <- function(s1, n1, s2, n2, conf.level) \{}
\text{F.test.statistic <- s1^2/s2^2}
\text{F.crit.lower <- qf((1-conf.level)/2,n1-1,n2-1) \# F, n1,n2 df, alpha critical value}
\text{F.crit.upper <- qf(conf.level+(1-conf.level)/2,n1-1,n2-1) \# F, n1,n2 df, alpha critical value}
\text{ci.lower <- (1/F.crit.upper)*F.test.statistic}
\text{ci.upper <- (1/F.crit.lower)*F.test.statistic}
\text{dat <- c(F.test.statistic, F.crit.lower, F.crit.upper, ci.lower, ci.upper)}
\text{names(dat) <- c("F test stat", "lower crit value", "upper crit value", "CI lower", "CI upper")}
\text{return(dat)}
\text{\}}}
\[
\text{var2.F.interval(0.16, 9, 0.09, 6, 0.95) \# approx 2-variance F-interval}
\]
\[
\text{F test stat lower crit value upper crit value CI lower CI upper}
\text{3.1604938 0.2075862 6.7571720 0.4677243 15.2249697}
\]

3. Test of $\sigma_1^2 / \sigma_2^2$, $s_1 \neq s_2$, raw data: company returns.

Consider simple random sample of returns for small (1) and large (2) companies, collected independently of one another. Test $\sigma_1^2 \neq \sigma_2^2$ at $\alpha = 0.05$.

<table>
<thead>
<tr>
<th>company</th>
<th>return</th>
<th>company</th>
<th>return</th>
</tr>
</thead>
<tbody>
<tr>
<td>small</td>
<td>1.3</td>
<td>large</td>
<td>9.1</td>
</tr>
<tr>
<td>small</td>
<td>22.3</td>
<td>small</td>
<td>0.3</td>
</tr>
<tr>
<td>small</td>
<td>-30.3</td>
<td>large</td>
<td>9.4</td>
</tr>
<tr>
<td>large</td>
<td>15.4</td>
<td>small</td>
<td>-13.3</td>
</tr>
<tr>
<td>small</td>
<td>23.1</td>
<td>small</td>
<td>-3.5</td>
</tr>
<tr>
<td>small</td>
<td>-56.7</td>
<td>small</td>
<td>45.4</td>
</tr>
<tr>
<td>large</td>
<td>3.4</td>
<td>small</td>
<td>10.3</td>
</tr>
<tr>
<td>large</td>
<td>-7.5</td>
<td>large</td>
<td>11.9</td>
</tr>
<tr>
<td>large</td>
<td>-8.8</td>
<td>large</td>
<td>-9.2</td>
</tr>
</tbody>
</table>

small <- c(1.3,22.3,-30.3,23.1,-56.7,0.3,-13.3,-3.5,45.4,10.3)
large <- c(15.4,3.4,-7.5,-8.8,9.1,9.4,11.9,-9.2,9)

(a) Check assumptions.

i. Data normal?

Normal probability plots for both samples indicates

(i) **normal**  (ii) **not normal** because data within dotted bounds.
(Notice dotted lines flare out at lower end, so includes points.)

ii. Outliers?

Both boxplots indicates (i) **outliers**  (ii) **no outliers**.

(b) Statement. Choose one.
Chapter 5. Hypothesis Testing (LECTURE NOTES 10)

Figure 5.11: Normal probability plots, boxplots for company returns

i. $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_1 : \sigma_1^2 > \sigma_2^2$

ii. $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_1 : \sigma_1^2 < \sigma_2^2$

iii. $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_1 : \sigma_1^2 \neq \sigma_2^2$

(c) Test.
Since $\alpha = 0.05$, $n_1 = 10$, $n_2 = 9$,
With $n_1 - 1 = 10 - 1 = 9$ df and $n_2 - 1 = 9 - 1 = 8$ df,
2 × chance observed $s_1^2/s_2^2 \approx 9.01$ or more, if $\sigma_1^2 = \sigma_2^2$, is

$$p\text{-value} = 2 \cdot P \left( F \geq \frac{s_1^2}{s_2^2} \right) \approx P (F \geq 9.01) \approx$$

(i) 0.01  (ii) 0.04  (iii) 0.11.
Level of significance $\alpha = (i) 0.01$  (ii) 0.05  (iii) 0.11.

s1 <- sqrt(var(small)); s1; n1 <- length(small); n1
s2 <- sqrt(var(large)); s2; n2 <- length(large); n2
var2.F.test <- function(s1, n1, s2, n2, s.null, signif.level, type) {
  F.test.statistic <- s1^2/s2^2
  if(type=="right") {
    F.crit <- qf(signif.level,n1-1,n2-1,lower.tail=FALSE) # F, n1,n2 df, alpha critical value
    p.value <- 1-pf(F.test.statistic,n1-1,n2-1)
  }
  if(type=="two.sided") {
    F.crit <- qf(signif.level/2,n1-1,n2-1,lower.tail=FALSE) # F, n1,n2 df, alpha/2 critical value
    p.value <- 2*(1-pf(F.test.statistic,n1-1,n2-1)) # twice area to right
  }
  dat <- c(s.null, F.crit, F.test.statistic, p.value)
  names(dat) <- c("s.null", "F crit value", "F test stat", "p value")
  return(dat)
Section 6. Comparing Two Means (LECTURE NOTES 10)

\[ \text{var2.F.test(s1, n1, s2, n2, 1, 0.05, "two.sided") # approx 2-variance F-test, two-sided} \]

<table>
<thead>
<tr>
<th>s null</th>
<th>F crit value</th>
<th>F test stat</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000000000</td>
<td>4.357233065</td>
<td>9.010753229</td>
<td>0.005040243</td>
</tr>
</tbody>
</table>

(d) Conclusion.
Since p-value = 0.01 < \( \alpha = 0.05 \),
(i) do not reject  (ii) reject null guess: \( H_0 : \sigma_1^2 = \sigma_2^2 \).
Sample \( \frac{s_1^2}{s_2^2} \) indicates population \( \frac{\sigma_1^2}{\sigma_2^2} \)
(i) is greater than  (ii) equals  (iii) does not equal 1: \( H_0 : \sigma_1^2 \neq \sigma_2^2 \).
In other words, data indicates SD in small company returns
(i) is greater than  (ii) equals  (iii) does not equal SD in large company returns.

5.6 Comparing Two Means

Let \( \bar{x}_1 \) and \( \bar{x}_2 \) be the means of two independent samples of size \( n_1 \) and \( n_2 \) from two populations and means \( \mu_1 \) and \( \mu_2 \). The 2-sample Z-test for \( \mu_1 - \mu_2 \), with known variances \( \sigma_1^2 \) and \( \sigma_2^2 \), is

\[
    z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}},
\]

or, with unknown variances but where it is assumed \( \sigma_1^2 = \sigma_2^2 \), the 2-sample T-test for \( \mu_1 - \mu_2 \), is

\[
    t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},
\]

where the pooled standard deviation estimate is

\[
    s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}},
\]

with \( n_1 + n_2 - 2 \) degrees of freedom or, with unknown variances but where it is assumed \( \sigma_1^2 \neq \sigma_2^2 \),

\[
    t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}},
\]

with the following \( r \) degrees of freedom (round down),

\[
    r = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1} \left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2-1} \left(\frac{s_2^2}{n_2}\right)^2},
\]
where either the underlying distribution of both samples are normal with no outliers or if both random sample sizes large \((n_1 \geq 30, n_2 \geq 30)\). Also, if the two samples are dependent or paired, the test statistic for the difference in two means \(\mu_d\) is

\[
t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}},
\]

where either the underlying distribution of differences is normal with no outliers or the random sample size is large \((n \geq 30)\).

Exercise 5.6 (Comparing Two Means)

1. Test for \(\mu_1 - \mu_2\), independent samples, unknown \(\sigma_1^2 \neq \sigma_2^2\): progesterone.

A study is conducted to determine cellular response to progesterone in females. Blood cells from four females are injected with progesterone; blood cells from four different females are, for comparison purposes, left untreated. Test if average progesterone response is greater than average control response at 5%. Assume normality with no outliers.

<table>
<thead>
<tr>
<th>female</th>
<th>progesterone (1)</th>
<th>female</th>
<th>control (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.85</td>
<td>5</td>
<td>5.23</td>
</tr>
<tr>
<td>2</td>
<td>2.28</td>
<td>6</td>
<td>1.21</td>
</tr>
<tr>
<td>3</td>
<td>1.51</td>
<td>7</td>
<td>1.40</td>
</tr>
<tr>
<td>4</td>
<td>2.12</td>
<td>8</td>
<td>1.38</td>
</tr>
</tbody>
</table>

\(f_1 \leftarrow c(5.85, 2.28, 1.51, 2.12)\)

\(f_2 \leftarrow c(5.23, 1.21, 1.40, 1.38)\)

(a) Statement.

i. \(H_0 : \mu_1 - \mu_2 = 0\) versus \(H_1 : \mu_1 - \mu_2 < 0\)

ii. \(H_0 : \mu_1 - \mu_2 = 0\) versus \(H_1 : \mu_1 - \mu_2 > 0\)

iii. \(H_0 : \mu_1 - \mu_2 = 0\) versus \(H_1 : \mu_1 - \mu_2 \neq 0\)

(b) Test.

<table>
<thead>
<tr>
<th>female</th>
<th>progesterone (1)</th>
<th>female</th>
<th>control (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.85</td>
<td>5</td>
<td>5.23</td>
</tr>
<tr>
<td>2</td>
<td>2.28</td>
<td>6</td>
<td>1.21</td>
</tr>
<tr>
<td>3</td>
<td>1.51</td>
<td>7</td>
<td>1.40</td>
</tr>
<tr>
<td>4</td>
<td>2.12</td>
<td>8</td>
<td>1.38</td>
</tr>
</tbody>
</table>

average \(\bar{x}_1 = \frac{5.85 + 2.28 + 1.51 + 2.12}{4} = 2.94\)  \(\bar{x}_2 = 2.305\)

SD \(s_1 = 1.97\)  \(s_2 = 1.95\)
Section 6. Comparing Two Means (LECTURE NOTES 10)

\[ x_{\text{bar}1} \leftarrow \text{mean}(f1); \quad x_{\text{bar}1}; \quad s1 \leftarrow \text{sqrt}(\text{var}(f1)); \quad s1; \quad n1 \leftarrow \text{length}(f1); \quad n1 \]
\[ x_{\text{bar}2} \leftarrow \text{mean}(f2); \quad x_{\text{bar}2}; \quad s2 \leftarrow \text{sqrt}(\text{var}(f2)); \quad s2; \quad n2 \leftarrow \text{length}(f2); \quad n2 \]

> x_{\text{bar}1} \leftarrow \text{mean}(f1); \quad x_{\text{bar}1}; \quad s1 \leftarrow \text{sqrt}(\text{var}(f1)); \quad s1; \quad n1 \leftarrow \text{length}(f1)
[1] 2.94
[1] 1.968163

> x_{\text{bar}2} \leftarrow \text{mean}(f2); \quad x_{\text{bar}2}; \quad s2 \leftarrow \text{sqrt}(\text{var}(f2)); \quad s2; \quad n2 \leftarrow \text{length}(f2)
[1] 2.305
[1] 1.951862

Test statistic of \( \bar{x}_1 - \bar{x}_2 = 0 \) is
\[
t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(2.94 - 2.305) - 0}{\sqrt{\frac{1.97^2}{4} + \frac{1.95^2}{4}}} =
\]
(i) 0.458 (ii) 2.93 (iii) 4.56,
with degrees of freedom
\[
df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{1 - \frac{s_1^2}{n_1^2}} + \frac{1 - \frac{s_2^2}{n_2^2}}{1} = \frac{\left(\frac{1.97^2}{4} + \frac{1.95^2}{4}\right)^2}{1 - \frac{1.97^2}{4}} + \frac{1}{1} - \frac{1.95^2}{4}
\]
(i) 4 (ii) 5 (iii) 6,
and so chance observed \( \bar{x}_1 - \bar{x}_2 = 0.635 \) or more, if \( \mu_1 - \mu_2 = 0 \), is p-value
\[P(\bar{x}_1 - \bar{x}_2 \geq 0.635) = P\left(\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \geq \frac{(2.94 - 2.305) - 0}{\sqrt{\frac{1.97^2}{4} + \frac{1.95^2}{4}}}\right) \approx P(t \geq 0.458)
\]
equals (i) 0.13 (ii) 0.25 (iii) 0.33.

\[
f1 \leftarrow c(5.85, 2.28, 1.51, 2.12)\]
\[
f2 \leftarrow c(5.23, 1.21, 1.40, 1.38)\]
\[
x_{\text{bar}1} \leftarrow \text{mean}(f1); \quad s1 \leftarrow \text{sqrt}(\text{var}(f1)); \quad n1 \leftarrow \text{length}(f1)\]
\[
x_{\text{bar}2} \leftarrow \text{mean}(f2); \quad s2 \leftarrow \text{sqrt}(\text{var}(f2)); \quad n2 \leftarrow \text{length}(f2)\]
\[
\text{mean2.t.test} \leftarrow \text{function}(x_{\text{bar}1}, s1, n1, x_{\text{bar}2}, s2, n2, m.null, \text{signif.level}, \text{var.type}, \text{type}) {\}
\]
\[
\quad \text{if(\text{var.type}=="diff.var")} {\}
\quad \quad \text{df} \leftarrow \left(\frac{(s1^2/n1 + s2^2/n2)^2}{(1/(n1-1))*(s1^2/n1)^2 + (1/(n2-1))*(s2^2/n2)^2}\right)\]
\[
\quad \quad \text{t.test.statistic} \leftarrow (x_{\text{bar}1}-x_{\text{bar}2}-m.null)/(\text{sqrt}(s1^2/n1+s2^2/n2))\]
\quad \}
\]
\[
\quad \text{if(\text{type}=="right")} {\}
\quad \quad \text{t.crit} \leftarrow -1*qt(\text{signif.level}, \text{df})\]
\quad \quad \text{p.value} \leftarrow 1-\text{pt}(\text{t.test.statistic}, \text{df})\]
\quad \}
\]
\[
\quad \text{dat} \leftarrow c(m.null, x_{\text{bar}1}-x_{\text{bar}2}, \text{df}, \text{t.crit}, \text{t.test.statistic}, \text{p.value})\]
\[
\quad \text{if(\text{type}=="two.sided") \text{names(dat)} \leftarrow c("m.null", "mean.diff", "df", "lower t crit", "upper t crit", "t test stat")\}
\quad \quad \text{if(\text{type}=="two.sided") \text{names(dat)} \leftarrow c("m.null", "mean.diff", "df", "t crit value", "t test stat", "p value")\}
\quad \}
\]
\[
\quad \text{return(dat)}\]
\]
\# approximate 2-mean t-test for \( \mu_1 - \mu_2 \), different variance
\[
\text{mean2.t.test}(x_{\text{bar}1}, s1, n1, x_{\text{bar}2}, s2, n2, 0, 0.05, "\text{diff.var}", \text{"right")}\]
\]
\[
\text{m.null} \quad \text{mean.diff} \quad \text{df} \quad \text{t crit value} \quad \text{t test stat} \quad \text{p value}
\begin{array}{cccccc}
0.000 & 0.635 & 6.000 & 1.943 & 0.458 & 0.331
\end{array}
\]
Level of significance \( \alpha = (i) 0.01 \) (ii) 0.05 (ii) 0.10.
(c) **Conclusion.**

Since p-value = 0.33 > α = 0.05,

(i) **do not reject** (ii) **reject** null guess: $H_0: \mu_1 - \mu_2 = 0$.

Sample average difference $\bar{x}_1 - \bar{x}_2$ indicates population difference $\mu_1 - \mu_2$ (i) **is less than** (ii) **equals** (iii) **is greater than** 0: $H_0: \mu_1 - \mu_2 = 0$.

In other words, progesterone population mean cellular response (i) **is less than** (ii) **equals** (iii) **is greater than** (iv) **is different from** control population mean cellular response.

(d) **Population, Sample, Statistic, Parameter.** Match columns.

<table>
<thead>
<tr>
<th>terms</th>
<th>infection example</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>populations</td>
</tr>
<tr>
<td>(b)</td>
<td>samples</td>
</tr>
<tr>
<td>(c)</td>
<td>statistic</td>
</tr>
<tr>
<td>(d)</td>
<td>parameter</td>
</tr>
<tr>
<td></td>
<td>(a) cellular response, 8 females</td>
</tr>
<tr>
<td></td>
<td>(b) $\bar{x}_1 - \bar{x}_2$</td>
</tr>
<tr>
<td></td>
<td>(c) cellular response, all females</td>
</tr>
<tr>
<td></td>
<td>(d) $\mu_1 - \mu_2$</td>
</tr>
</tbody>
</table>

2. **Inference for difference in dependent means, $\mu_d$: cellular response.**

A study is conducted to determine cellular response to progesterone in females. Blood cells from female 1 are broken into two groups. One group of these blood cells are injected with progesterone; the other group, the control, is, for comparison purposes, left untreated. Blood cells of females 2, 3 and 4 are handled in same way. Assume normality with no outliers. Test if mean progesterone response **greater** than mean control response at 5%.

<table>
<thead>
<tr>
<th>female</th>
<th>progesterone (1)</th>
<th>control (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.85</td>
<td>5.23</td>
</tr>
<tr>
<td>2</td>
<td>2.28</td>
<td>1.21</td>
</tr>
<tr>
<td>3</td>
<td>1.51</td>
<td>1.40</td>
</tr>
<tr>
<td>4</td>
<td>2.12</td>
<td>1.38</td>
</tr>
</tbody>
</table>

$f_1 \leftarrow c(5.85, 2.28, 1.51, 2.12)$

$f_2 \leftarrow c(5.23, 1.21, 1.40, 1.38)$

(a) **Statement.**

If mean progesterone response, $\mu_1$, is **greater** than average control response, $\mu_2$, $\mu_1 > \mu_2$, difference in responses must be greater than zero, $\mu_d = \mu_1 - \mu_2 > 0$, so

i. $H_0 : \mu_d = 0$ versus $H_1 : \mu_d > 0$

ii. $H_0 : \mu_d = 0$ versus $H_1 : \mu_d < 0$
iii. $H_0: \mu_d = 0$ versus $H_1: \mu_d \neq 0$

(b) Test.

<table>
<thead>
<tr>
<th>female</th>
<th>progesterone (1)</th>
<th>control (2)</th>
<th>differences, $d_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.85</td>
<td>5.23</td>
<td>$d_1 = 5.85 - 5.23 = 0.62$</td>
</tr>
<tr>
<td>2</td>
<td>2.28</td>
<td>1.21</td>
<td>1.07</td>
</tr>
<tr>
<td>3</td>
<td>1.51</td>
<td>1.40</td>
<td>0.11</td>
</tr>
<tr>
<td>4</td>
<td>2.12</td>
<td>1.38</td>
<td>0.74</td>
</tr>
</tbody>
</table>

Average of differences is
\[ \bar{d} = \frac{0.62+1.07+0.11+0.74}{4} = (i) \; 0.355 \; (ii) \; 0.431 \; (iii) \; 0.635 \]

standard deviation of differences,
\[ s_d \approx (i) \; 0.398 \; (ii) \; 0.931 \; (iii) \; 1.522, \]

so test statistic of $\bar{d} = 0.635$ is
\[ t = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}} = \frac{0.635 - 0}{0.398/\sqrt{4}} = \]

(i) 1.42 (ii) 2.93 (iii) 3.19.

with $n - 1 = 4 - 1 = (i) \; 1 \; (ii) \; 2 \; (iii) \; 3$ degrees of freedom,
so chance observed $\bar{d} = 0.635$ or more, if $\mu_d = 0$, is
\[ p\text{-value} = P(\bar{d} \geq 0.635) = P \left( \frac{\bar{d} - \mu_0}{s_d/\sqrt{n}} \geq \frac{0.635 - 0}{0.398/\sqrt{4}} \right) \approx P(t \geq 3.19) \approx \]

(i) 0.013 (ii) 0.025 (iii) 0.075.

mean1.t.test <- function(x.bar,m.null,s,n,signif.level,type) {
  t.test.statistic <- (x.bar-m.null)/(s/sqrt(n))
  if(type=="right") {
    t.crit <- -1 qt(signif.level, n-1)
    p.value <- 1 pt(t.test.statistic, n-1)
  }
  dat <- c(m.null, x.bar, t.crit, t.test.statistic, p.value)
  names(dat) <- c("m.null", "x.bar", "t crit value", "t test stat", "p value")
  return(dat)
}
mean1.t.test(x.bar.pair, 0, s.pair, n.pair, 0.05, "right") # m: mean, s: SD, n: sample size, paired t-test

Level of significance $\alpha = (i) \; 0.01 \; (ii) \; 0.05 \; (iii) \; 0.10$.

(c) Conclusion.
Since $p\text{-value} = 0.025 < \alpha = 0.050$,
(i) do not reject (ii) reject null guess: $H_0 : \mu_d = 0$. 

Sample average difference \( \bar{d} \) indicates population average difference \( \mu_d \)
(i) is less than (ii) equals (iii) is greater than 0: \( H_1 : \mu_d > 0. \)
In other words, progesterone population mean cellular response
(i) is less than (ii) equals (iii) is greater than (iv) is different from
control population mean cellular response.

(d) Comment: dependent samples
Control blood samples (i) depend on (ii) are independent of
progesterone–treated blood samples. These dependent samples are paired
within females. In general, sampling is dependent if individuals in one
sample are used to determine individuals in other sample. A depen-
dent test does control for differences due to different females, whereas an
independent test does not do this.

5.7 Goodness-of-Fit Tests

We perform a goodness of fit test using test statistic

\[
\chi^2_{obs} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}, \quad i = 1, \ldots, k,
\]

which is approximately chi-square, \( k - 1 \) df, provided expected frequencies \( E_i \geq 5 \)
and where \( E_i = np_i \) and \( O_i \) are the number of data values in category \( i, i = 1, \ldots, k. \)

Exercise 5.7 (Goodness-of-Fit Tests)

   Age distribution of a random sample of 463 people living in Uppsala, a city in
   Sweden, is compared to age distribution to all of Sweden. Test if age distribution
   in Uppsala is different from age distribution for all of Sweden at \( \alpha = 0.05. \)

<table>
<thead>
<tr>
<th>age</th>
<th>Uppsala (out of 463)</th>
<th>Sweden (% total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>under 5</td>
<td>47</td>
<td>6.7%</td>
</tr>
<tr>
<td>5 to 16</td>
<td>75</td>
<td>14.1%</td>
</tr>
<tr>
<td>16 to 65</td>
<td>296</td>
<td>69.5%</td>
</tr>
<tr>
<td>over 65</td>
<td>45</td>
<td>9.7%</td>
</tr>
</tbody>
</table>

\( \text{obs} \leftarrow \text{c}(47, 75, 296, 45) \)
\( \text{p} \leftarrow \text{c}(0.067, 0.141, 0.695, 0.097) \)
\( \text{labels} \leftarrow \text{c}("\text{under 5"","5 to 16","16 to 65","over 65")} \)
\( \text{options(digits=3)} \) # print to 3 digits of accuracy

(a) Statement. Choose one or more.
Section 7. Goodness-of-Fit Tests (LECTURE NOTES 10) 207

i. \( H_0 \): Uppsala age distribution same as Sweden  
   versus \( H_1 \): Uppsala age distribution different from Sweden

ii. \( H_0 \): \( p_1 = 0.067 \), \( p_2 = 0.141 \), \( p_3 = 0.695 \), \( p_4 = 0.097 \)  
    versus \( H_1 \): Uppsala age distribution different from null

iii. \( H_0 \): \( p_1 = 0.067 \), \( p_2 = 0.141 \), \( p_3 = 0.695 \), \( p_4 = 0.097 \)  
     versus \( H_1 \): at least one \( p_i \) not equal to null

No matter how this question is worded, null hypothesis for test is always the same (as expected distribution) and alternative hypothesis is always different (from expected distribution).

(b) Test.

<table>
<thead>
<tr>
<th>age</th>
<th>( O_i )</th>
<th>( E_i = n \times p_i )</th>
<th>( \frac{(O_i - E_i)^2}{E_i} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>under 5</td>
<td>47</td>
<td>( 463 \times 0.067 \approx 31 )</td>
<td>8.2</td>
</tr>
<tr>
<td>5 to 16</td>
<td>75</td>
<td>( 463 \times 0.141 \approx 65 )</td>
<td>1.5</td>
</tr>
<tr>
<td>16 to 65</td>
<td>296</td>
<td>( 463 \times 0.695 \approx 322 )</td>
<td>2.1</td>
</tr>
<tr>
<td>over 65</td>
<td>45</td>
<td>( 463 \times 0.097 \approx 45 )</td>
<td>0</td>
</tr>
</tbody>
</table>

observed test statistic is \( \chi^2 = \sum \frac{(O-E)^2}{E} = 8.2 + 1.5 + 2.1 + 0 = 11.7 \)  
(ii) 18.3  (iii) 23.4,
with degrees of freedom

\[
\text{number of rows} - 1 = 4 - 1 =
\]

(i) 1  (ii) 3  (iii) 4 df,

and so p-value is

\[
p\text{-value} = P(\chi^2 \geq 11.7) =
\]

(i) 0.01  (ii) 0.08  (iii) 0.10.

goodness.fit.test <- function(obs, p, labels, signif.level) {
  total <- sum(obs); exp <- total*p; k <- length(obs)
  chi.contrib <- (obs - exp)^2/exp
  tbl <- as.table(rbind(obs,exp,chi.contrib))
  dimnames(tbl) <- list(type <- c("observed","expected","chi2.contrib"), categories <- labels)
  print(tbl)
  chi2.test.statistic <- sum(chi.contrib)
  chi2.crit <- qchisq(signif.level,k-1,lower.tail=FALSE)
  p.value <- 1-pchisq(chi2.test.statistic,k-1)
  dat <- c(chi2.crit, chi2.test.statistic, p.value)
  names(dat) <- c("chi2 crit value", "chi2 test stat", "p value")
  print(dat)
}
goodness.fit.test(obs, p, labels, 0.05) # goodness of fit chi2-test

<table>
<thead>
<tr>
<th>( \text{age} )</th>
<th>( \text{under 5} )</th>
<th>( \text{5 to 16} )</th>
<th>( \text{16 to 65} )</th>
<th>( \text{over 65} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>observed</td>
<td>4.70e+01</td>
<td>7.50e+01</td>
<td>2.96e+02</td>
<td>4.50e+01</td>
</tr>
<tr>
<td>expected</td>
<td>3.10e+01</td>
<td>6.53e+01</td>
<td>3.22e+02</td>
<td>4.49e+01</td>
</tr>
<tr>
<td>chi2.contrib</td>
<td>8.23e+00</td>
<td>1.45e+00</td>
<td>2.07e+00</td>
<td>1.76e-04</td>
</tr>
</tbody>
</table>

\( \text{chi2 crit value} \) \( \text{chi2 test stat} \) \( \text{p value} \)

7.81473 11.74350 0.00832

Level of significance \( \alpha = \) (i) 0.01  (ii) 0.05  (iii) 0.10.
(c) Conclusion.
Since \( p\)-value = 0.01 < \( \alpha = 0.05 \),
(i) do not reject  (ii) reject \( H_0 \): Uppsala different from Sweden.
Sample statistic \( \chi^2 \) indicates age distribution in Uppsala
(i) same as  (ii) different from
age distribution for all of Sweden.

(d) Understanding a goodness of fit test.
Goodness of fit test involves comparing observed data with expected data. Expected data is data we derive assuming age distribution in Uppsala is equal to age distribution in all of Sweden. In other words, if observed data set and generated expected data sets are “close” to one another, this indicates age distribution in Uppsala
(i) same as  (ii) different from
age distribution for all of Sweden.

(e) Table used to compare observed to expected.

<table>
<thead>
<tr>
<th>age</th>
<th>( O )</th>
<th>( E = n \times p_i )</th>
<th>( \frac{(O-E)^2}{E} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>under 5</td>
<td>47</td>
<td>( 463 \times 0.067 \approx 31 )</td>
<td>8.2</td>
</tr>
<tr>
<td>5 to 16</td>
<td>75</td>
<td>( 463 \times 0.141 \approx 65 )</td>
<td>1.5</td>
</tr>
<tr>
<td>16 to 65</td>
<td>296</td>
<td>( 463 \times 0.695 \approx 322 )</td>
<td>2.1</td>
</tr>
<tr>
<td>over 65</td>
<td>45</td>
<td>( 463 \times 0.097 \approx 45 )</td>
<td>0</td>
</tr>
</tbody>
</table>

Expected number under age 5 in Uppsala
if age distribution in Uppsala same as Sweden.
If 6.7% of Sweden under age 5, then in a town of size 463 people,
\( 463 \times 0.067 \approx (i) \ 24 \  \ (ii) \ 31 \  \ (ii) \ 49 \)
of these people should be age 5. In other words, in Uppsala, with 463 people, we would expect 31 people to be under age of 5 if age distribution in Uppsala same as age distribution in all of Sweden.

(f) Large \( \chi^2 \) test statistic indicates different distribution..
The \( \chi^2 \) test statistic is one possible way to measure how close observed and expected data sets are to one another. Since \( \sum \frac{(O-E)^2}{E} = 11.8 \) appears to be is large, indicates age distribution in Uppsala (i) same as  (ii) different from age distribution for all of Sweden.

2. Goodness of fit test: peas, equally likely.
Given dataset of 556 observed frequencies of various types of peas, test whether or not proportion of round–yellow, wrinkled–yellow, round–green and wrinkled–green peas occurs with equal frequency or not at 5%.

<table>
<thead>
<tr>
<th>type</th>
<th>round–yellow</th>
<th>wrinkled–yellow</th>
<th>round–green</th>
<th>wrinkled–green</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequency</td>
<td>315</td>
<td>101</td>
<td>108</td>
<td>32</td>
</tr>
</tbody>
</table>
Section 7. Goodness-of-Fit Tests (LECTURE NOTES 10)

\[ \text{obs} \leftarrow \text{c}(315,101,108,32) \]
\[ \text{p} \leftarrow \text{c}(0.25,0.25,0.25,0.25) \]
\[ \text{labels} \leftarrow \text{c}("\text{round-yellow}", "\text{wrinkled-yellow}", "\text{round-green}", "\text{wrinkled-green}") \]
\[ \text{options(digits=3)} \] # print to three digits of accuracy

(a) **Statement.** Choose one or more.

i. \( H_0 : \text{observed pea type distribution same as expected 9 : 3 : 3 : 1 distribution} \)
   versus \( H_1 : \text{observed pea type distribution different from expected} \)

ii. \( H_0 : p_1 = \frac{1}{4}, p_2 = \frac{1}{4}, p_3 = \frac{1}{4}, p_4 = \frac{1}{4} \)
   versus \( H_1 : \text{observed pea type distribution different from null} \)

iii. \( H_0 : p_1 = \frac{1}{4}, p_2 = \frac{1}{4}, p_3 = \frac{1}{4}, p_4 = \frac{1}{4} \)
   versus \( H_1 : \text{at least one } p_i \text{ not equal to null} \)

(b) **Test.**

<table>
<thead>
<tr>
<th>pea type</th>
<th>( O_i )</th>
<th>( E_i )</th>
<th>( \frac{(O_i - E_i)^2}{E_i} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>round–yellow</td>
<td>315</td>
<td>( \left(\frac{1}{4}\right) 556 = 139 )</td>
<td>222.85</td>
</tr>
<tr>
<td>wrinkled–yellow</td>
<td>101</td>
<td>139</td>
<td>10.388</td>
</tr>
<tr>
<td>round–green</td>
<td>108</td>
<td></td>
<td>6.9137</td>
</tr>
<tr>
<td>wrinkled–green</td>
<td>32</td>
<td>139</td>
<td></td>
</tr>
</tbody>
</table>

The observed test statistic is
\[ \chi^2 \approx \sum \frac{(O_i - E_i)^2}{E_i} = 222.85 + 10.388 + 6.9137 + 82.367 \approx 311 \]

(i) 311  (ii) 323  (iii) 345,

with degrees of freedom

\[ \text{number of rows} - 1 = 4 - 1 = \]

(i) 1  (ii) 3  (iii) 4 df,

and so p–value is
\[ \text{p–value} = P(\chi^2 \geq 323) = \]

(i) 0.00  (ii) 0.05  (iii) 1.72.

```
goodness.fit.test(obs, p, labels, 0.05) # goodness of fit chi2-test
```

```
round-yellow wrinkled-yellow round-green wrinkled-green
observed 315.00 101.00 108.00 32.00
expected 139.00 139.00 139.00 139.00
chi2.contrib 222.85 10.39 6.91 82.37
chi2crit 7.81
critvalue 7.81
chiteststat 322.52
pvalue 0.00
```

Level of significance \( \alpha = \)

(i) 0.01  (ii) 0.05  (iii) 0.10.

(c) **Conclusion.**

Since p–value = 0.00 < \( \alpha = 0.05 \),

(i) do not reject  (ii) reject null: \( H_0 : \text{observed same as expected ratio} \)

Observed data indicates pea type distribution

(i) same as  (ii) different from

the hypothesized equal pea type distribution.
3. **Goodness of fit test: pea defects, Poisson.**

Test if number of defects, \( I \), on any given pea follows a Poisson distribution where each pea averages \( \lambda = 0.8 \) defects, \( p_i = \frac{0.8^i}{i!} e^{-0.8} \). In the dataset of 556 peas below, 250 peas each have no (0) defects, for example. Assume \( \alpha = 0.05 \).

<table>
<thead>
<tr>
<th>defects</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>( \geq 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequency</td>
<td>250</td>
<td>180</td>
<td>100</td>
<td>26</td>
</tr>
<tr>
<td>Poisson ( p_i )</td>
<td>0.45</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\text{obs} <- c(250,180,100,26)
\]
\[
\text{lam} <- 0.8 \quad \# \text{assumed ave number of defects}
\]
\[
\text{p} <- c(dpois(0,lam), dpois(1,lam), dpois(2,lam), 1-ppois(2,lam)); p
\]
\[
\text{labels} <- c("0","1","2", "ge3")
\]

\[ [1] \quad 0.4493 \quad 0.3595 \quad 0.1438 \quad 0.0474 \]

(a) **Statement.** Choose one or more.

i. \( H_0 \) : observed pea type distribution same as expected Poisson distribution versus \( H_1 \) : observed pea type distribution different from expected

ii. \( H_0 \) : \( p_1 = 0.45, p_2 = 0.36, p_3 = 0.14, p_4 = 0.05 \)

   versus \( H_1 \) : observed pea type distribution different from null

iii. \( H_0 \) : \( p_1 = 0.45, p_2 = 0.36, p_3 = 0.14, p_4 = 0.05 \)

   versus \( H_1 \) : at least one \( p_i \) not equal to null

(b) **Test.**

<table>
<thead>
<tr>
<th>defects</th>
<th>( O_i )</th>
<th>( E_i = np_i )</th>
<th>( \frac{(O_i - E_i)^2}{E_i} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>250</td>
<td>0.45 \cdot 556 \approx 250</td>
<td>0.00</td>
</tr>
<tr>
<td>1</td>
<td>180</td>
<td>200</td>
<td>1.97</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td></td>
<td>5.03</td>
</tr>
<tr>
<td>( \geq 3 )</td>
<td>26</td>
<td>26.4</td>
<td></td>
</tr>
</tbody>
</table>

observed test statistic is

\[
\chi^2_{\text{obs}} = \sum \frac{(O - E)^2}{E} = 0.00 + 1.97 + 5.03 + 0.01 \approx
\]

(i) **0.31**  (ii) **5.67**  (iii) **7.01**, with degrees of freedom

number of rows \(- 1 = 4 - 1 =

(i) **1**  (ii) **3**  (iii) **4** df,

and so critical value at \( \alpha = 0.05 \) is

\[
\chi^2 =
\]

(i) **5.81**  (ii) **6.81**  (iii) **7.81**.
goodness.fit.test(obs, p, labels, 0.05) # goodness of fit chi2-test

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>ge3</th>
</tr>
</thead>
<tbody>
<tr>
<td>observed</td>
<td>2.50e+02</td>
<td>1.80e+02</td>
<td>1.00e+02</td>
<td>2.60e+01</td>
</tr>
<tr>
<td>expected</td>
<td>2.50e+02</td>
<td>2.00e+02</td>
<td>7.99e+01</td>
<td>2.64e+01</td>
</tr>
<tr>
<td>chi2.contrib</td>
<td>1.20e-04</td>
<td>1.97e+00</td>
<td>5.03e+00</td>
<td>5.11e-03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>chi2 crit value</th>
<th>chi2 test stat</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.8147</td>
<td>7.0102</td>
<td>0.0716</td>
</tr>
</tbody>
</table>

(c) **Conclusion.**

Since $\chi^2_{obs} = 7.01 < \chi^2_{crit} = 7.81$,

(i) **do not reject**  (ii) **reject** null: $H_0$ : observed same as expected.

Observed data indicates number of pea defects

(i) **same as**  (ii) **different from**

number of pea defects which following a Poisson distribution where $\lambda = 0.8$.

(d) **Check assumptions.**

All $E_i$ (i) **are**  (ii) **are not** greater than 5.