AMCS 210 - Applied Statistics and Data Analysis
Hernando Catequista Ombao
Applied Mathematics and Computational Sciences Program
Computer, Electrical and Mathematical Sciences & Engineering (CEMSE) Division
King Abdullah University of Science and Technology (KAUST)

HOMEWORK 3

Henrique Aparecido Laureano

Fall Semester 2017

Contents

Exercise 1

Exercises on reading tables of distributions.

(a)

Suppose that a population of scores has a normal distribution with mean $\mu = 80$ and variance $\sigma^2 = 9$. Use the standard normal table to determine the 2.5-th, 5-th, 25-th, 50-th, 75-th, 95-th and 97.5-th percentile of this distribution of scores. Compare the answers you obtained manually by that provided by R (use the help function to determine if you need to use pnorm, rnorm, dnorm, etc.)

Solution:

Table 1: In the second column (left to right) we have the values from the standard normal table; in the third column we have the percentiles obtained manually, after the percentiles provided by R and in the end the difference between them (we used two decimal places).

Percentile	Z	$Q_i = Z \cdot \sigma + \mu$	R = `qnorm(`Z`, 80, sqrt(9))`	Diff = Q_i - R
2.5-th	-1.96	$-1.96 \cdot \sqrt{9} + 80 = 74.12$	74.12	0
5-th	-1.64	$-1.64 \cdot \sqrt{9} + 80 = 75.08$	75.07	0.01
25-th	-0.67	$-0.67 \cdot \sqrt{9} + 80 = 77.99$	77.98	0.01
50-th	0.00	$0 \cdot \sqrt{9} + 80 = 80$	80	0
75-th	0.67	$0.67 \cdot \sqrt{9} + 80 = 82.01$	82.02	-0.01
95-th	1.64	$1.64 \cdot \sqrt{9} + 80 = 84.92$	84.93	-0.01
97.5-th	1.96	$1.96 \cdot \sqrt{9} + 80 = 85.88$	85.88	0

(b)

Determine the 90-th, 95-th and 97.5-th percentile of a χ^2-distribution with

(i.) degrees of freedom equal to 5.

Solution:

Table 2: In the second column (left to right) we have the percentiles provided by the χ^2 table; in the third column we have the percentiles provided by R and in the end the difference between them (we used three decimal places).

Percentile	χ^2 (using the table)	`qchisq(`χ^2`, 5, lower.tail = FALSE)`	Difference
90-th	$\chi^2_{0.10} = 9.236$	9.236	0
95-th	$\chi^2_{0.05} = 11.070$	11.07	0
97.5-th	$\chi^2_{0.025} = 12.833$	12.833	0

(ii.) with degrees of freedom equal to 10.

Solution:

Table 3: In the second column we have the percentiles provided by the χ^2 table; in the third column we have the percentiles provided by R and in the end the difference between them (used three decimal places).

Percentile	χ^2 (using the table)	`qchisq(`χ^2`, 10, lower.tail = FALSE)`	Difference
90-th	$\chi^2_{0.10} = 15.987$	15.987	0
95-th	$\chi^2_{0.05} = 18.307$	18.307	0
97.5-th	$\chi^2_{0.025} = 20.483$	20.483	0

(c)

Determine the 2.5-th, 95-th and 97.5-th percentile of a t-distribution with

(i.) degrees of freedom equal to 5.

Solution:

Table 4: In the second column we have the percentiles provided by the table, using the cumulative probability; in the third column we have the percentiles provided by R and in the end the difference between them (we used three decimal places).

Percentile	cum. prob (using the table)	R = `qt(cum. prob, 5)`	Difference
2.5-th	$t_{0.025} =$ -2.571	-2.571	0
95-th	$t_{0.95} = 2.015$	2.015	0
97.5-th	$t_{0.975} = 2.571$	2.571	0

(ii.) with degrees of freedom equal to 1000.

Solution:

Table 5: In the second column we have the percentiles provided by the table, using the cumulative probability; in the third column we have the percentiles provided by R and in the end the difference between them (we used three decimal places).

Percentile	cum. prob (using the table)	R = `qt(cum. prob, 1000)`	Difference
2.5-th	$t_{0.025} =$ -1.962	-1.962	0
95-th	$t_{0.95} = 1.646$	1.646	0
97.5-th	$t_{0.975} = 1.962$	1.962	0

(d)

As the degrees of freedom of a t-distribution increases, show that the percentiles becomes closer to that of the standard normal distribution. You can choose a few percentiles to demonstrate this (e.g., 2.5-th, 25-th, 75-th, 97.5-th).

Solution:

In the Table 6 we see that for every percentile, conform the degrees of freedom increase, the difference between the t-distribution and the Standard Normal distribution is smaller, being zero for very high degrees, like 1000.

Table 6: Percentiles of a t-distribution and standard normal distribution for seven different degrees of freedom in four different percentiles. In the last column we have the absolute difference between them.

Percentile	Degrees of freedom	t-distribution	Standard Normal distribution	|Difference|
2.5-th	5	-2.57	-1.96	0.61
2.5-th	10	-2.23	-1.96	0.27
2.5-th	15	-2.13	-1.96	0.17
2.5-th	20	-2.09	-1.96	0.13
2.5-th	25	-2.06	-1.96	0.10
2.5-th	30	-2.04	-1.96	0.08
2.5-th	1000	-1.96	-1.96	0.00
25-th	5	-0.73	-0.67	0.05
25-th	10	-0.70	-0.67	0.03
25-th	15	-0.69	-0.67	0.02
25-th	20	-0.69	-0.67	0.01
25-th	25	-0.68	-0.67	0.01
25-th	30	-0.68	-0.67	0.01
25-th	1000	-0.68	-0.67	0.00
75-th	5	0.73	0.67	0.05
75-th	10	0.70	0.67	0.03
75-th	15	0.69	0.67	0.02
75-th	20	0.69	0.67	0.01
75-th	25	0.68	0.67	0.01
75-th	30	0.68	0.67	0.01
75-th	1000	0.68	0.67	0.00
97.5-th	5	2.57	1.96	0.61
97.5-th	10	2.23	1.96	0.27
97.5-th	15	2.13	1.96	0.17
97.5-th	20	2.09	1.96	0.13
97.5-th	25	2.06	1.96	0.10
97.5-th	30	2.04	1.96	0.08
97.5-th	1000	1.96	1.96	0.00

We can also see that the difference between this distributions for small degrees of freedom is smaller for percentiles more closer to the median, the 50-th percentile, and bigger when the percentile is more closer to the borders (zero and 100-th percentile).

(e)

Determine the 90-th, 95-th and 97.5-th percentile of a F-distribution with

(i.) numerator degrees of freedom $df_n = 1$ and denominator degrees of freedom $df_d = 10$.

Solution:

Table 7: In the third column we have the percentiles provided by the table; in the fourth column we have the percentiles provided by R and in the end the difference between them (used two decimal places).

Percentile	p (table)	F^*	`qf(1 - p, 1, 10)`	Difference
90-th	0.1	3.29	3.29	0
95-th	0.05	4.96	4.96	0
97.5-th	0.025	6.94	6.94	0

(ii.) numerator degrees of freedom $df_n = 5$ and denominator degrees of freedom $df_d = 31$.

Solution:

Table 8: In the third column we have the percentiles provided by the table; in the fourth column we have the percentiles provided by R and in the end the difference between them (used two decimal places).

Percentile	p (table)	F^*	`qf(1 - p, 5, 31)`	Difference
90-th	0.1	2.05	2.04	0.01
95-th	0.05	2.53	2.52	0.01
97.5-th	0.025	3.03	3.01	0.02

In the table we don't have $df_d = 31$, so we approximate to 30.

Exercise 2

Recall that many of the test statistics are compared to a null distribution which is often standard normal (normal with mean equal 0 and variance equal to 1). We did

not provide any theoretical justification for this in class. In this exercise you will need to conduct some simulation studies that illustrate this.

(a)

In testing for the equality of population proportions, the test statistic based on random samples of sizes and proportions n_A and $\hat{\pi}_A$ from population A and n_A and $\hat{\pi}_B$ from population B is

$$T_1 = \frac{\hat{\pi}_A - \hat{\pi}_B}{\sqrt{\frac{\hat{\pi}_A(1-\hat{\pi}_A)}{n_A} + \frac{\hat{\pi}_B(1-\hat{\pi}_B)}{n_B}}}.$$

When $H_0 : \pi_A = \pi_B$ is true then the reference distribution for T_1 is approximately equal to that of a standard normal when both sample sizes n_A and n_B are sufficiently large. Conduct a simulation study to determine an empirical distribution of T_1 when the null hypothesis is correct for each pair of sample sizes. Try a few pairs, say, (i.) $n_A = n_B = 30$, (ii.) $n_A = 25$, $n_B = 40$.

Solution:

```r
# <code r> ========================================================================== #
# empty vector of size 1e4 (# of replications) to store the test statistics
ts <- numeric(1e4)
# function to generate the reference distribution
ref.dist <- function(pia = .67, pib = .67, na, nb, title){
  # choosed population proportions: 0.67
  t1 <- function(pia, pib, na, nb){ # test statistic
    (pia - pib) / sqrt( (pia*(1-pia)/na) + (pib*(1-pib)/nb) )
  }
  for (i in 1:length(ts)) { # computing the reference distribution
    hpia <- mean(rbinom(na, 1, pia))
    hpib <- mean(rbinom(nb, 1, pib))
    ts[i] <- t1(pia = hpia, pib = hpib, na = na, nb = nb)
  }
  # histogram
  hist(ts, freq = FALSE, col = "gray60", las = 1, xlab = "", ylab = "", main = NA
       , axes = FALSE) ; axis(side = 1, at = -4:4)
  lines(density(ts), col = "#0080ff", lwd = 3)
  abline(v = c(quantile(ts, .025), mean(ts), quantile(ts, .975))
         , col = 2, lwd = 2)
  mtext(side = 3, text = title, adj = 0, cex = 1.3)
}
par(mfrow = c(1, 2), mar = c(2, 2, 4, 2))
# generating the reference distribution
ref.dist(na = 30, nb = 30, title = bquote(paste(n[A], " = " , n[B], " = 30")))
ref.dist(na = 25, nb = 40, title = bquote(paste(n[A], " = 25, " , n[B], " = 40")))
# </code r> ========================================================================= #
```

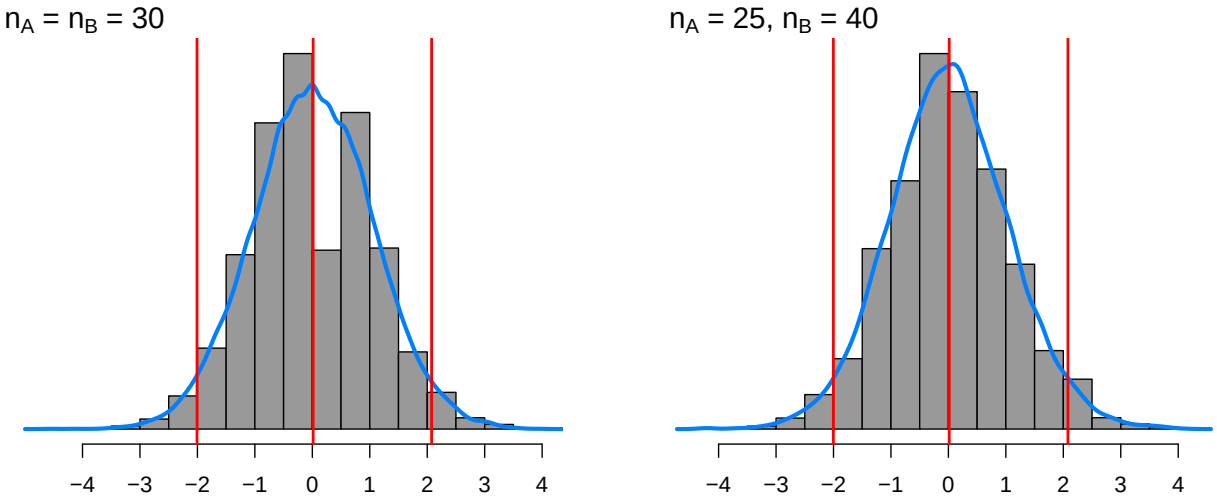

Figure 1: Empirical distribution of T_1 when the null hypothesis in correct for two different pair of sample sizes. In blue we have the estimated density for this distributions and in red the mean, the 2.5th and 97.5th percentiles.

We can see in Figure 1 that in both simulation scenarios we found a bell shape behavior, with mean around zero. In red we have the mean, the 2.5th and 97.5th percentiles, and in both cases this percentiles are very close to -2, 2, which is a sign that the reference distributions are very similar to a Standard Normal distribution, because we know that in a Standard Normal we have 95% of the data between $\pm$ 1.96 (very close to 2). The mean close to zero is another sign that the reference distributions are very similar to a Standard Normal distribution, since in a Standard Normal the mean is zero.

(b)

In testing for the equality of the means of two Gaussian populations with common and known variance σ^2, the test statistic based on the sample means $\bar{X}_A$ and $\bar{X}_B$ is

$$Z = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\sigma^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}.$$

Solution:

```
# <code r> ========================================================================== #
# empty vector of size 1e4 (# of replications) to store the test statistics
zs <- numeric(1e4)
# function to generate the reference distribution
ref.dist <- function(mua = 24, mub = 24, s2 = 32, na, nb, title){
  # choosed population means: 24 ; choosed population variances: 32
  z <- function(mua, mub, s2, na, nb){ # test statistic
    (mua - mub) / sqrt( s2 * ((1/na) + (1/nb)) )
  }
```

```r
  for (i in 1:length(zs)) { # computing the reference distribution
    hmua <- mean(rnorm(na, mua, sqrt(s2)))
    hmub <- mean(rnorm(nb, mub, sqrt(s2)))
    zs[i] <- z(mua = hmua, mub = hmub, s2, na = na, nb = nb)
  }
  # histogram
  hist(zs, freq = FALSE, col = "gray60", las = 1, xlab = "", ylab = "", main = NA
       , axes = FALSE) ; axis(side = 1, at = -4:4)
  lines(density(zs), col = "#0080ff", lwd = 3)
  abline(v = c(quantile(zs, .025), mean(ts), quantile(zs, .975))
         , col = 2, lwd = 2)
  mtext(side = 3, text = title, adj = 0, cex = 1.3)
}
par(mfrow = c(1, 2), mar = c(2, 2, 4, 2))
# generating the reference distribution
ref.dist(na = 30, nb = 30, title = bquote(paste(n[A], " = " , n[B], " = 30")))
ref.dist(na = 25, nb = 40, title = bquote(paste(n[A], " = 25, " , n[B], " = 40")))
# </code r> ========================================================================== #
```

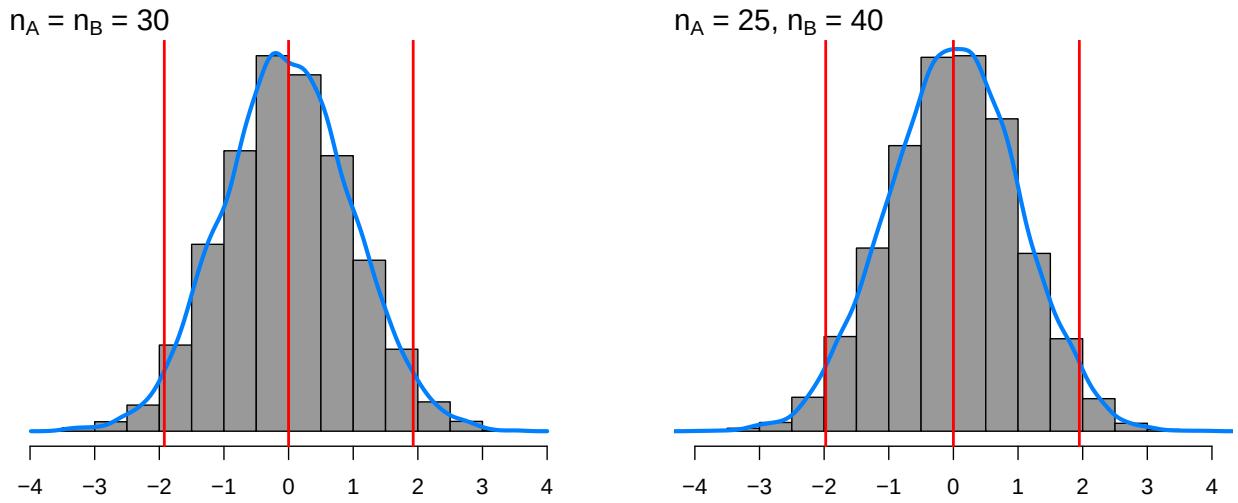

Figure 2: Empirical distribution of Z when the null hypothesis in correct for two different pair of sample sizes. In blue we have the estimated density for this distributions and in red the mean, the 2.5th and 97.5th percentiles.

We can see in Figure 2 that in both simulation scenarios we found a bell shape behavior, with mean around zero and with 2.5th and 97.5th percentiles very close to -2, 2. Which is a sign that the reference distributions are very similar to a Standard Normal distribution.

(c)

When the common variance in the two normal distribution case is not known it has to be estimated. Denote the sample variances for group A to be

$$S_A^2 = \frac{1}{n_A - 1} \sum_{i=1}^{n_A} (X_i^A - \bar{X}_A)^2.$$

The sample variance for group B is defined in a similar way. Then the pooled variance estimate is a weighted average of these two sample variances

$$S_P^2 = W_A S_A^2 + W_B S_B^2$$

where the weights are $W_A = \frac{n_A - 1}{(n_A - 1) + (n_B - 1)}$. The test statistic for comparing the population means is similar to that of Z above but with σ^2 replaced by its estimator S_P^2 because it is not known. Thus,

$$T = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{S_P^2 \left(\frac{1}{n_A} + \frac{1}{n_B} \right)}}.$$

The distribution of T under the null hypothesis is t with degrees of freedom equal to $d = (n_A - 1) + (n_B - 1)$. Now conduct simulations that demonstrate this theoretical result. You may choose whatever settings you like. When you simulate your data, you need to specify the variance σ^2 but when you compute your test statistic you need to pretend that you don't actually know this.

Solution:

```
# <code r> ========================================================================== #
# empty vector of size 1e4 (# of replications) to store the test statistics
ts <- numeric(1e4)
# function to generate the reference distribution
ref.dist <- function(mua = 24, mub = 24, s2 = 32, na, nb, title){
  # choosed population means: 24 ; choosed population variances: 32
  t <- function(xa, xb, mua, mub, na, nb){ # test statistic
    sa2 <- sum((xa - mua)**2) / (na - 1)
    sb2 <- sum((xb - mub)**2) / (nb - 1)
    wa <- (na - 1) / ((na - 1) + (nb - 1))
    wb <- (nb - 1) / ((na - 1) + (nb - 1))
    sp2 <- wa * sa2 + wb * sb2 # pooled variance estimate
    (mua - mub) / sqrt( sp2 * ((1/na) + (1/nb)) )
  }
  for (i in 1:length(ts)) { # computing the reference distribution
    xa <- rnorm(na, mua, sqrt(s2))
    xb <- rnorm(nb, mub, sqrt(s2))
    hmua <- mean(xa)
    hmub <- mean(xb)
    ts[i] <- t(xa = xa, xb = xb, mua = hmua, mub = hmub, na = na, nb = nb)
  }
  # histogram
  hist(ts, freq = FALSE, col = "gray60", las = 1, xlab = "", ylab = "", main = NA
       , axes = FALSE) ; axis(side = 1, at = -4:4)
```

```
  lines(density(ts), col = "#0080ff", lwd = 3)
  abline(v = c(quantile(ts, .025), mean(ts), quantile(ts, .975))
         , col = 2, lwd = 2)
  mtext(side = 3, text = title, adj = 0, cex = 1.3)
}
par(mfrow = c(1, 2), mar = c(2, 2, 4, 2))
# generating the reference distribution
ref.dist(na = 30, nb = 30, title = bquote(paste(n[A], " = " , n[B], " = 30")))
ref.dist(na = 25, nb = 40, title = bquote(paste(n[A], " = 25, " , n[B], " = 40")))
# </code r> ============================================================================ #
```

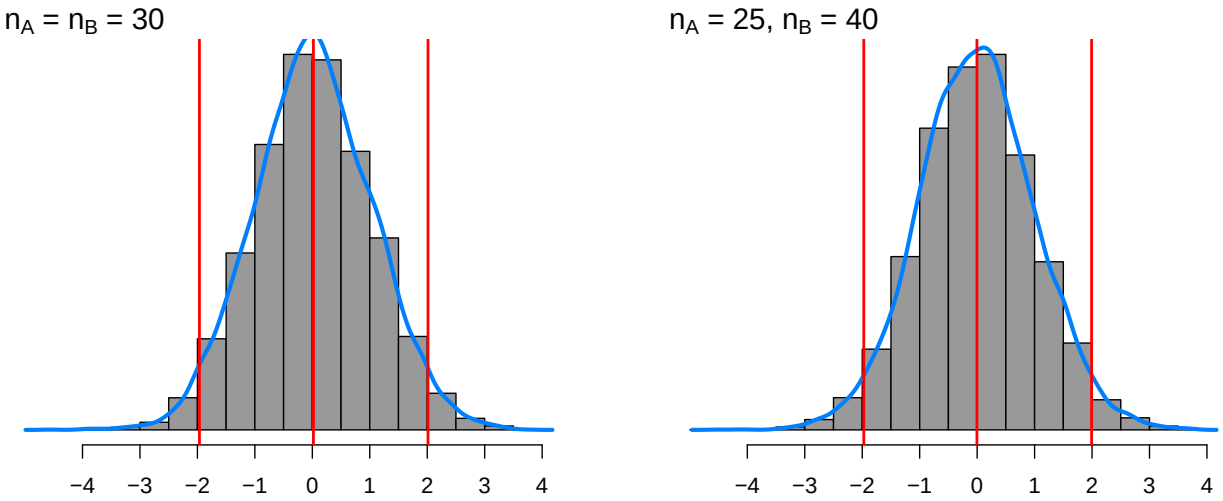

Figure 3: Empirical distribution of T when the null hypothesis in correct for two different pair of sample sizes. In blue we have the estimated density for this distributions and in red the mean, the 2.5th and 97.5th percentiles.

We can see in Figure 3 that in both simulation scenarios we found a bell shape behavior, with mean around zero and with 2.5th and 97.5th percentiles very close to -2, 2. Which is a sign that the reference distributions are very similar to a Standard Normal distribution.

Exercise 3

Prior to the 2016 US Presidential elections, a political scientist conducted a study to determine if there was any association between support for then-candidate DJ Trump and the view on imposing severe restriction on immigration. The results of the survey based on a sample of $n = 500$ respondents are displayed below.

	Support restriction	Do not support restriction	Total
Support DJT	150	50	200
Do not support DJT	100	200	300

(a)

What is the proportion of participants who support the immigration restriction and at the same time also support DJ Trump?

Solution:

$$\text{Proportion} = \frac{\text{Support restriction} \cap \text{Support DJ Trump}}{n} = \frac{150}{500} = 0.3.$$

(b)

What is the marginal sample proportion of DJT supporters? What is the marginal sample proportion of the restriction on immigration?

Solution:

$$\text{Proportion of DJT supporters} = \frac{\text{Support DJT}}{n} = \frac{200}{500} = 0.4.$$
$$\text{Proportion of the restriction on immigration} = \frac{\text{Support restriction}}{n} = \frac{150 + 100}{500} = 0.5.$$

(c)

Among those who do not support the restriction, what is the proportion of DJT supporters? What is the odds of selecting a participant who support DJT?

Solution:

$$\text{Proportion} = \frac{\text{DJT supporters} \mid \text{Do not support the restriction}}{\text{Do not support restriction}} = \frac{50}{250} = 0.2.$$
$$\text{Odds} = \frac{50/250}{1 - 50/250} = 0.25.$$

(d)

Conduct a formal test of association between support for DJT and support for restriction on immigration. Note that a formal test should include

(i.) the null and alternative hypotheses.

Solution:

H_0: There isn't not a association between support for DJT and whether or not someone support restriction on immigration.

H_a: There is a association between support for DJT and whether or not someone support restriction on immigration.

(ii.) the test stastistic.

Solution:

χ^2 Test.

$$E_{\text{Support DJT, Support restriction}} = \frac{200 \cdot 250}{500} = 100$$
$$E_{\text{Support DJT, Don't support restriction}} = \frac{200 \cdot 250}{500} = 100$$
$$E_{\text{Don't support DJT, Support restriction}} = \frac{300 \cdot 250}{500} = 150$$
$$E_{\text{Don't support DJT, Don't support restriction}} = \frac{300 \cdot 250}{500} = 150$$

Observed and expected counts are often presented together in a contingency table. In the table below, expected values are presented in parentheses.

	Support restriction	Do not support restriction	Total
Support DJT	150 (100)	50 (100)	200
Do not support DJT	100 (150)	200 (150)	300
Total	250	250	500

χ^2 Test Statistic:

$$\begin{aligned}\chi^2 &= \sum \frac{(O-E)^2}{E} \\ &= \frac{(150-100)^2}{100} + \frac{(50-100)^2}{100} + \frac{(100-150)^2}{150} + \frac{(200-150)^2}{150} \\ &= 25 + 25 + 16.6666667 + 16.6666667 \\ &= 83.3333333\end{aligned}$$

(iii.) its null distribution and the rejection region based on the probability of Type I error of $\alpha = 0.05$.

Solution:

The χ^2 test statistic is 83.33 with df = (number of rows - 1) $\cdot$ (number of columns - 1) = (2 - 1) $\cdot$ (2 - 1) = 1.

Using the table, we find that for a probability of Type I error of $\alpha = 0.05$, with 1 degree of freedom, the rejection region is determined by the value 3.841, which is much smaller than the value of the test statistic. Therefore, we have significant statistical evidence to don't accept H_0. We have statistical evidence that there is a association between support for DJT and whether or not someone support restriction on immigration.

Exercise 4

Assume that children's score in the KidScore data set has a $N(\mu_1, \sigma^2)$ distribution if the mother has graduated from high school, and $N(\mu_2, \sigma^2)$ if the mother has not graduated from high school. Using the random sample collected in the KidScore dataset,

(a)

Draw the boxplots for each of the two samples. Compare and contrast these two boxplots.

Solution:

```
# <code r> ============================================================================ #
path <- "~/Dropbox/CLASS-DROPBOX/BOOK-DATA/"
da <- read.table(paste0(path, "KidScore.txt"), header = TRUE, sep = ",")
da$momHs <- with(da, factor(momHs, labels = c("No", "Yes")))

boxplot(kidScore ~ momHs, da, las = 1, xlab = "Mother Completed High School?"
        , main = "Child's Test Score at Age 3")
# </code r> ============================================================================ #
```

Figure 4: Boxplot for the child's test score at age 3 by mothers that completed, or not, the high school.

We can see in the Figure 4 that childrens of mothers that completed the high school present biggest medians and smallest variance, than the childrens of mothers that didn't completed the high school.

(b)

Compute the sample means and variances for each of the two groups.

Solution:

```
# <code r> ============================================================================ #
library(plyr)
ddply(da, .(momHs), summarise, Mean = mean(kidScore), Variance = var(kidScore))
# </code r> ============================================================================ #
```

```
  momHs     Mean Variance
1    No 77.54839 509.5764
2   Yes 89.31965 362.8828
```

(c)

Compute the pooled variance estimate S_P^2.

Solution:

$$\text{Pooled variance:} \qquad S_P^2 = W_{\text{momHs: No}} S^2_{\text{momHs: No}} + W_{\text{momHs: Yes}} S^2_{\text{momHs: Yes}}.$$

$$S^2_{\text{momHs: No}} = \frac{1}{n_{\text{momHs: No}} - 1} \sum_{i=1}^{n_{\text{momHs: No}}} \left(X_i^{\text{momHs: No}} - \bar{X}_{\text{momHs: No}} \right)^2 = 509.5764376.$$

$$W_{\text{momHs: No}} = \frac{n_{\text{momHs: No}} - 1}{(n_{\text{momHs: No}} - 1) + (n_{\text{momHs: Yes}} - 1)} = \frac{92}{92 + 340} = 0.212963.$$

$$S^2_{\text{momHs: Yes}} = \frac{1}{n_{\text{momHs: Yes}} - 1} \sum_{i=1}^{n_{\text{momHs: Yes}}} \left(X_i^{\text{momHs: Yes}} - \bar{X}_{\text{momHs: Yes}} \right)^2 = 362.8828187.$$

$$W_{\text{momHs: Yes}} = \frac{n_{\text{momHs: Yes}} - 1}{(n_{\text{momHs: Yes}} - 1) + (n_{\text{momHs: No}} - 1)} = \frac{340}{340 + 92} = 0.787037.$$

$$\boxed{S_P^2 = 108.520908 + 285.6022184 = 394.1231264.}$$

(d)

Conduct a formal test for comparing the means μ_1 and μ_2. Again - you already should know the complete information that is needed to conduct this test.

Solution:

We will test the **hypothese** that the **difference** between the mean of the child's test score at age 3 of mothers that completed the high school **are not statiscally significant** to the mean of the child's test score at age 3 of mothers that didn't completed the high school.

$$\text{Hypotheses :} \qquad H_0 : \mu_{\text{momHs: No}} = \mu_{\text{momHs: Yes}}, \qquad H_a = \mu_{\text{momHs: No}} \neq \mu_{\text{momHs: Yes}}.$$

$$\text{Test Statistic:} \quad T = \frac{\bar{X}_{\text{momHs: No}} - \bar{X}_{\text{momHs: Yes}}}{\sqrt{S_P^2 \left(\frac{1}{n_{\text{momHs: No}}} + \frac{1}{n_{\text{momHs: Yes}}} \right)}}$$
$$= \frac{77.5483871 - 89.3196481}{\sqrt{394.1231264 \cdot \left(\frac{1}{93} + \frac{1}{341} \right)}}$$
$$= -5.0685161$$

Reference distribution: t-Student with $(n_{\text{momHs: No}} - 1) + (n_{\text{momHs: Yes}} - 1) = 432$ degrees of freedom. We will also consider an $\alpha = 0.05$.

For this probability of Type I error and this amount of degrees of freedom, the critical value is

```
# <code r> ============================================================================ #
qt(.975, 432)
# </code r> ========================================================================== #
```

```
[1] 1.965471
```

$t_{0.025,432} = -1.965471$, $t_{0.975,432} = 1.965471$. Extremely closer to a Standard Normal.

As our Test Statistic is less than the critical value, we don't accept the H_0.

We have statistical evidence to don't accept H_0, in other words, we have evidence that exist difference statiscally significant between the mean of the child's test score at age 3 of mothers that completed the high school and the mean of the child's test score at age 3 of mothers that didn't completed the high school.

■