**Practicing *t*-tests using R-commander**

Open the file Anthropometrics\_2010\_2013.RData.

**Problem 1.** Check whether the mean hip circumference (variable hip) is 90 cm in the population.

Run: *Graphs/Histogram and Statistics/Means/Single Sample t-test/* and type *mu=90*

Name of the test:.one-sample t-test............

H0: μ=90, the mean circumference in the population is 90 cm…………..

Ha:… μ≠90……..

Assumption:…the sample is drawn from a normally distributed population………….

Graphical check of the assumption:…the distribution is symmetric, perhaps more peaky than the normal curve………..



Result of the test:

Test statistic=…4.17……. df=……1840…….. p-value=…0,00003102(=3.102e-05)……….

Decision: …the population mean significantly differ from 90 at 5% level, we reject H0……………………..

Explain your decision:…p<0.05, also, the 95% confidence interval does not contain 90 …………………

> with(D1, (t.test(hip, alternative='two.sided', mu=90, conf.level=.95)))

One Sample t-test

data: hip

t = 4.1762, df = 1840, p-value = 3.102e-05

alternative hypothesis: true mean is not equal to 90

95 percent confidence interval:

90.50633 91.40298

sample estimates:

mean of x

90.95465

**Problem 2.** Compare the means of the two consecutive measurements of hip circumference (variables hip1 and hip2).

Run: *Statistics/Means/Paired t-test/*

To check assumptions *(*optional*): Data/Manage Variables in active dataset/Compute new variable/ name: variables, expression: hip1-hip2, then Graphs/Histogram*

Name of the test:...paired t-test..........

H0: μ1=μ2 or μdifference=0. The two population means are equal…………..

Ha:… μ1≠μ2 or μdifference≠0………..

Assumption:…the sample of differences is drawn from a normal population………….

Graphical check of the assumption: not too good but still acceptable…..

> D1$diff <- with(D1, hip1- hip2)

> with(D1, Hist(diff, scale="frequency", breaks="Sturges", col="darkgray"))



Result of the test:

Test statistic=…4.6852…. df=…1840….. p-value=…3.002e-06…….

Decision: …p<0.05, the difference is significant at 5% level. ……………………..

Explain your decision:… p<0.05. Also, the 95% CI does not contain 0. It is difficult to believe that while the sample means are so close to each other, the difference is still statistically significant. We might interpret it as statistically significant but biologically not meaningful, as the confidence interval for the mean difference is (0.11-0.27) mm…………………

> numSummary(D1[,c("hip1", "hip2")], statistics=c("mean", "sd", "IQR",

+ "quantiles"), quantiles=c(0,.25,.5,.75,1))

mean sd IQR 0% 25% 50% 75% 100% n

hip1 91.10016 9.919429 12.0 48 85 90.3 97.0 137 1841

hip2 90.90418 9.818108 11.1 47 85 90.0 96.1 138 1841

> with(D1, (t.test(hip1, hip2, alternative='two.sided', conf.level=.95,

+ paired=TRUE)))

Paired t-test

data: hip1 and hip2

t = 4.6852, df = 1840, p-value = 3.002e-06

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

0.1139416 0.2780193

sample estimates:

mean of the differences

0.1959804

**Problem 3.** Check whether the two consecutive measurements of body height have the same population-mean.

Name of the test:...paired t-test..........

H0: μ1=μ2 or μdifference=0. The two population means are equal…………..

Ha:… μ1≠μ2 or μdifference≠0………..

Assumption:…the sample of differences is drawn from a normal population………….

Graphical check of the assumption:…seems not normally distributed. We continue because t-tests are robust, in case of small departure from normality assumption they are still valid………..



Result of the test:

Test statistic=…-0.0159…. df=…1840……….. p-value=…0.9873……….

Decision: …p>0.05, the difference is not significant at 5% level……………………..

Explain your decision:…p>0.05, and the 95% CI contains 0. The test statistic is small (smaller than the t-value in the table). We do not reject H0…………………

Paired t-test

data: height1 and height2

t = -0.0159, df = 1840, p-value = 0.9873

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.03849070 0.03787147

sample estimates:

mean of the differences

-0.0003096143

**Problem 4.** Compare the mean hip circumference of boys and girls!

*Run: Graphs/Histogram/chose variable Plot by groups/* choose variable*. Test: Statistics/Means/Independent Samples t-test*

Name of the test:.two-sample t-test (Independent samples t-test)............

H0: μ1=μ2. The two population means are equal…………..

Ha:… μ1≠μ2 ………..

Assumption:…both samples are drawn from a normal population AND the calculation of the test statistic depends also on equality of variances.………….

Graphical check of the assumption:…the histograms are similar to a Gauss-curve, the boxplots are symmetric. The variances are similar, so the formula assuming equal variances could be used, however, the formula assuming unequal variances is also appropriate…..

 

Result of the test:

Test statistic=…3.7709……. df=…1706.864……….. p-value=…0.0001682……….

Decision: …the difference is statistically significant at 5% level, we reject H0……………………..

Explain your decision:…p<0.05, the difference is statistically significant. This could be caused by the high sample size. The difference between population means is maximum 2.64 cm (we are 95% confident). This is biologically not meaningful.…………………

> numSummary(D1[,"hip"], groups=D1$gender, statistics=c("mean", "sd", "IQR",

+ "quantiles"), quantiles=c(0,.25,.5,.75,1))

mean sd IQR 0% 25% 50% 75% 100% data:n

Male 91.90466 10.326135 13.00000 52 85 91.33333 98.00000 137.3333 841

Female 90.16369 9.283978 10.33333 47 85 89.66667 95.33333 135.3333 998

> t.test(hip~gender, alternative='two.sided', conf.level=.95, var.equal=FALSE,

+ data=D1)

Welch Two Sample t-test

data: hip by gender

t = 3.7709, df = 1706.864, p-value = 0.0001682

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

0.8354342 2.6464920

sample estimates:

mean in group Male mean in group Female

91.90466 90.16369

Open the file „BEER.Rdata. The following table contains the name and the meaning of the variables. State which variables are categorical (factors) and which of them are continuous.

|  |  |  |
| --- | --- | --- |
| Variable name | Meaning | values |
| rating | Rated Quality of Beer | Very good-Good-Fair |
| beer |  |  |
| origin |  | USA-Canada\_France\_Holland\_Mexico |
| avail | Availability in the U.S. | National-Regional |
| price | Price per 6-pack |  |
| cost | Cost per 12 Fluid Ounces |  |
| calories | Calories per 12 Fluid Ounces |  |
| sodium | Sodium per 12 Fluid Ounces in mg |  |
| alcohol | Alcohol by Volume (in %) |  |
| class | Price Class | Not given-Superpremium-Premium-Popular |
| light | Rated Quality of Beer | NONLIGHT-LIGHT |

**Problem 5.** Compare the mean calorie-content of light and non-light beers!

Name of the test:.two-sample t-test (Independent samples t-test)............

H0: μ1=μ2. The two population means are equal…………..

Ha:… μ1≠μ2 ………..

Assumption:…both samples are drawn from a normal population AND the calculation of the test statistic depends also on equality of variances.………….

Graphical check of the assumption:…the sample size is too small to see check the form of the histogram. The box plots are little skewed. The variances seem to be different.………..

 

Result of the test:

Test statistic=…5.8981…. df=……6.395…….. p-value=…0.0008381……….

Decision: ……the difference is statistically significant at 5% level, light beers have less calories in means…………..

Explain your decision:…p<0.05, the test statistic is big, the confidence interval does not contain 0………

> numSummary(BEER[,"calories"], groups=BEER$light, statistics=c("mean", "sd",

+ "IQR", "quantiles"), quantiles=c(0,.25,.5,.75,1))

mean sd IQR 0% 25% 50% 75% 100% data:n

NONLIGHT 150.2143 8.310445 8.5 136 144.75 149 153.25 175 28

LIGHT 98.0000 23.050669 23.0 68 84.50 99 107.50 135 7

> t.test(calories~light, alternative='two.sided', conf.level=.95,

+ var.equal=FALSE, data=BEER)

Welch Two Sample t-test

data: calories by light

t = 5.8981, df = 6.395, p-value = 0.0008381

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

30.87244 73.55613

sample estimates:

mean in group NONLIGHT mean in group LIGHT

150.2143 98.0000

**Problem 6.** Compare the mean price of light and non-light beers!

H0: μ1=μ2. The two population means are equal…………..

Ha:… μ1≠μ2 ………..

Assumption:…both samples are drawn from a normal population AND the calculation of the test statistic depends also on equality of variances.………….

Graphical check of the assumption:…normality cannot be checked. Variances are different. Formula for unequal variances (Welch test) should be used.………..



Result of the test:

Test statistic=…1.7688……. df=…31.912……….. p-value=……0.08649…….

Decision: …the difference is not statistically different at 5% level, we do not reject H0……………………..

Explain your decision:…p>0,05. Although the sample means seem to be far from each other, but with we cannot show that this difference is statistically significant. …………………

> numSummary(BEER[,"price"], groups=BEER$light, statistics=c("mean", "sd",

+ "IQR", "quantiles"), quantiles=c(0,.25,.5,.75,1))

mean sd IQR 0% 25% 50% 75% 100% data:n

NONLIGHT 3.115357 1.2403956 1.3275 1.59 2.465 2.62 3.7925 7.19 28

LIGHT 2.675714 0.2186974 0.1800 2.29 2.590 2.73 2.7700 2.99 7

> t.test(price~light, alternative='two.sided', conf.level=.95,

+ var.equal=FALSE, data=BEER)

Welch Two Sample t-test

data: price by light

t = 1.7688, df = 31.912, p-value = 0.08649

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.06671186 0.94599758

sample estimates:

mean in group NONLIGHT mean in group LIGHT

3.115357 2.675714

**Problem 7.** Compare the variances of the price of light and non-light beers!

Run: *Statistics/Variances/Two Variances F test*

Name of the test:.F-test............ H0:…σ12=σ22…, the two variances are equal…….. Ha:… :…σ12≠σ22.

Result of the test: test statistic=…32.1687……. df=…27 and 6……….. p-value=…0.0003062……….

Decision: …the two variances are significantly different at 5% level……………………..

Explain your decision:……………………

> var.test(price ~ light, alternative='two.sided', conf.level=.95, data=BEER)

F test to compare two variances

data: price by light

F = 32.1687, num df = 27, denom df = 6, p-value = 0.0003062

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

6.321938 94.023580

sample estimates:

ratio of variances

32.16867

**Problem 8**. Open the file „BEFAFTER.RData”. Compare the mean body weights before and after the diet.

Name of the test:...paired t-test..........

H0: μ1=μ2. The two population means are equal…………..

Ha:… μ1≠μ2 ………..

Assumption:……the difference sample is drawn from a normal population……….

Graphical check of the assumption:…the assumption is violated. The histogram is skewed………..



Result of the test:

Test statistic=…2.1548……. df=…8……….. p-value=…0.0633……….

Decision: …the difference is not statistically different at 5% level. ……………………..

Explain your decision:…based on this information, we cannot show at 5% level, that the diet is effective . Perhaps it is rely not effective or it is effective but we need more data to show it.…………………

Paired t-test

data: before and after

t = 2.1548, df = 8, p-value = 0.0633

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-1.018145 30.040367

sample estimates:

mean of the differences

14.51111