1. **(10 points) Do exercise 15.6.8. A study was conducted to evaluate the relative efficacy of supplementation with calcium versus calcitriol in the treatment of postmenopausal osteoporosis [259]. Calcitriol is an agent that has the ability to increase gastrointestinal absorption of calcium. A number of patients withdrew from this study prematurely due to the adverse effects of treatment which include thirst, skin problems, and neurologic symptoms. The relevant data appear below.**

Treatment	Withdrawal		Total
	Yes	No	
Calcitriol	27	287	314
Calcium	20	288	308
Total	47	575	622

(a) Compute the sample proportion of subjects who withdrew from the study in each treatment group.
(b) Test the null hypothesis that there is no association between treatment group and withdrawal from the study at the 0.05 level of significance. What is the p-value of the test?
(c) Do you reject or fail to reject the null hypothesis? What do you conclude?

Solution:

a. Sample proportion of subjects who withdrew from the study in Calcitriol group $= 27 / 314 = 0.086$
Sample proportion of subjects who withdrew from the study in Calcium group $= 20 / 308 = 0.065$

b. H_0: There is no significant association between treatment group and withdraw group
H_α: There is significant association between treatment group and withdraw group
The expected frequencies of the trial groups can be found using the following equation:

$$Expected\ Frequency = \frac{(Row\ Total)(Column\ Total)}{Overall\ Total}$$

$$EF\ of\ Calcitriol\ for\ who\ withdrew = \frac{(314)(47)}{622} = 23.73$$

$$EF\ of\ Calcitriol\ for\ who\ didn't\ withdrew = \frac{(314)(575)}{622} = 290.27$$

$$EF\ of\ Calcium\ for\ who\ withdrew = \frac{(308)(47)}{622} = 23.27$$

$$EF\ of\ Calcium\ for\ who\ didn't\ withdrew = \frac{(308)(575)}{622} = 284.73$$

The test Statistic Value: $\chi^2 = \sum \frac{(O-E)^2}{E}$

$$\chi^2 = \frac{(27-23.73)^2}{23.73} + \frac{(287-290.27)^2}{290.27} + \frac{(20-23.73)^2}{23.73} + \frac{(288-284.73)^2}{284.73}$$

$$\chi^2 = 0.451 + 0.037 + 0.586 + 0.038 = 1.112$$

Degrees of Freedom: $df = (r-1)(e-1) = (2-1)(2-1) = 1$
When the significance level $\alpha = 0.05$, the p-value is 0.292

c. Since $0.292 > 0.05$, we fail to reject the null hypothesis and can be concluded that there is no significant association between treatment group and withdraw group.

2. **(15 points) Do exercise 15.6.17. In a study of HIV infection among women entering the New York State prison system, 475 inmates were cross classified with respect to HIV seropositivity and their histories of intravenous drug use [268]. These data are saved in a dataset called *prison*. The indicators of seropositivity are saved under the variable name *hiv*, and those of intravenous drug use under *ivdu*.**
(a) Among women who have used drugs intravenously, what proportion are HIV positive? Among women who have not used drugs intravenously, what proportion are HIV positive?
(b) At the 0.05 level of significance, test the null hypothesis that there is no association between history of intravenous drug use and HIV seropositivity. What is the p-value?
(c) What do you conclude?
(d) Estimate the odds ratio for being HIV positive for women who have used intravenous drugs versus those who have not.
(e) Construct a 95% confidence interval for this odds ratio.

Solution:

a. Sample proportion of HIV positive among women who use intravenously
$$= 61 / 136 = 0.449$$
Sample proportion of HIV positive among women who don't use intravenously
$$= 27 / 339 = 0.080$$

b. H_0: There is significant association between history of intravenous use and HIV
H_α: There is no significant association between history of intravenous use and HIV
Using a Chi-Squared test for association: $\chi^2 = 87.502$

c. When the significance level $\alpha = 0.05$, the p-value is 0.00001
Since 0.00001 < 0.05, we don't have enough evidence to reject the null hypothesis. It can be concluded that there is a significant association between a history of intravenous drug use and positive HIV tests.

d. Odds Ratio $= \frac{\frac{P(postive+intravenous\ use)}{1-P(positive+intravenous\ use)}}{\frac{P(positive+no\ use)}{1-P(positive+no\ use)}}$
$$= \frac{(61)(312)}{(27)(75)} = 9.39$$

e. The 95% confidence interval:
Standard deviation = 0.262
z-score = 1.96
$$CI\ for\ 2-values = [2.240 - 1.96(0.262), 2.240 + 1.96(0.262)] = [1.726, 2.752]$$
$$CI\ for\ OR = [e^{1.726}, e^{2.752}] = [5.618, 15.674]$$

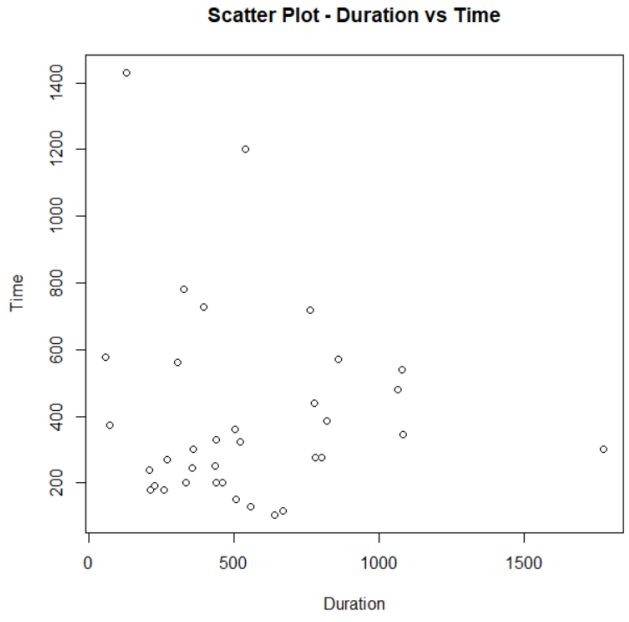
3. **(20 points) Do exercise 16.5.6. Thirty-five patients with ischemic heart disease, a suppression of blood flow to the heart, took part in a series of tests designed to evaluate the perception of pain. In one part of the study, the patients exercised until they experienced angina or chest pain; time until the onset of angina and the duration of the attack were recorded. The observations are saved in the dataset *ischemic_hd* [275]. Time to angina in seconds is saved under the variable name time; the duration of angina, also in seconds, is saved under the name duration.**
   (a) Create a two-way scatter plot of duration of angina versus time to angina.
   (b) In the population of patients with ischemic heart disease, does there appear to be any evidence of a linear relationship between time to angina and the duration of the attack?
   (c) Does the duration of angina tend to increase or decrease as time to angina increases?
   (d) Construct histograms for both time to angina and duration of angina. Do these measurements appear to be normally distributed?
   (e) Compute the Spearman rank correlation.
   (f) Test the null hypothesis

$$Ho: p = 0$$

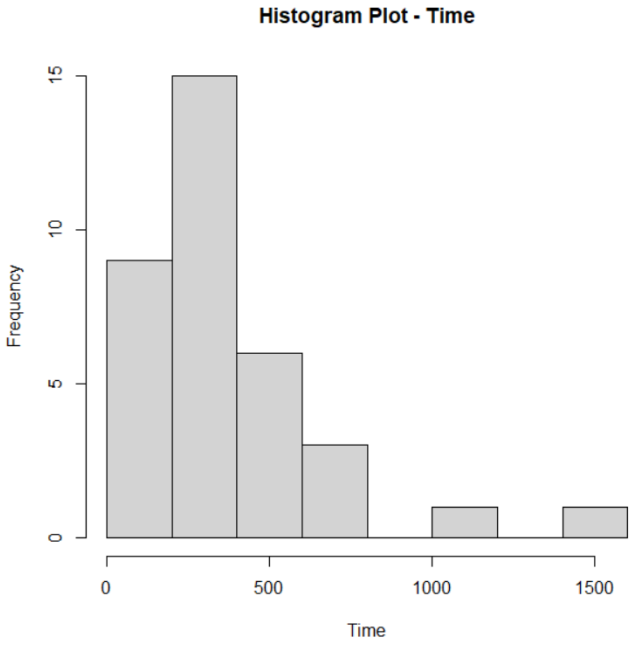
   at the 0.05 level of significance. Do you reject the null hypothesis? What do you conclude?
   (g) Using R, again test the null hypothesis that the population correlation is equal to 0. What do you conclude?
   (h) Use permutation tests on the Pearson and Spearman correlation, respectively. Use 100,000 sampled permutations to get the p-values for these two tests.

Solution:

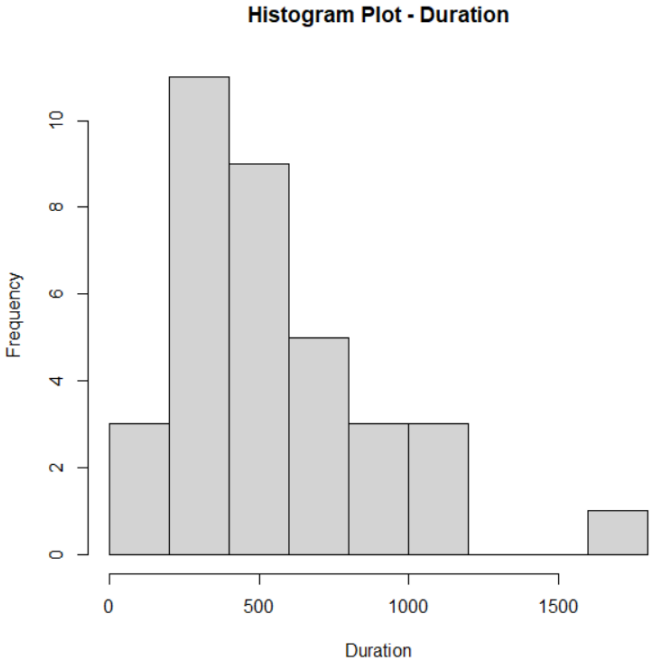
a.



b. Yes. There does seem to be proof in the ischemic heart disease patient group that there is a linear connection between the time to angina and the length of the incident. This is because all of the scatter plot's plotted elements make a diagonal line that runs from the lower left corner to the upper right corner.
c. The duration of angina tends to increase as time to angina increases because the association coefficient is negative.
d. Histogram of Time



Histogram of Duration



Above histograms, doesn't seem to be normally distributed.

e. Spearman rank correlation can be computed by using following formula:

$$r_s = 1 - \frac{6\Sigma d_i^2}{n(n^2-1)}$$

$$r_s = 1 - \frac{(6)(6598)}{35(35^2-1)} = 1 - \frac{39588}{42840} = 1 - 0.9241 = 0.0759$$

f. H_0: $\rho = 0$

H_α: $\rho \neq 0$

$\alpha = 0.05$

Test statistic: $t = r\sqrt{\frac{n-2}{1-r^2}}$

$$t = \left(-0.074\sqrt{\frac{35-2}{1-(-0.074)^2}}\right) = (-0.074)(5.76) = -0.426$$

Degrees of Freedom: $df = n - 2 = 35 - 2 = 33$

The critical value of t for a two-tailed test at significance level 0.05 and 33 degrees of freedom is ± 2.035 , from the standard t-table.

Therefore, we fail to reject the null hypothesis and it can be concluded that the correlation coefficient is not significant.

g. Using the same hypothesis and significance level.

Test statistic: $t_s = r_s\sqrt{\frac{n-2}{1-r_s^2}}$

$$t_s = \left(0.0759\sqrt{\frac{35-2}{1-(0.0759)^2}}\right) = (0.0759)(5.76) = 0.437$$

Degrees of Freedom: $df = n - 2 = 35 - 2 = 33$

The critical value of t for a two-tailed test at significance level 0.05 and 33 degrees of freedom is ± 2.035 , from the standard t-table.

Therefore, we fail to reject the null hypothesis and it can be concluded that the correlation coefficient is not significant.

h. R code:

```
> perm.cor.test <- function(x, y, n.perm=100000, cor.method="spearman")
+ {
+ n <- length(x)
+ T.obs <- cor(x, y, method=cor.method)
+ T.perm = rep(NA, n.perm)
+ for(i in 1:n.perm)
+ {y.perm = sample(y, n, replace=F)
+ T.perm[i] = cor(x, y.perm, method=cor.method)}
+ mean(abs(T.perm) >= abs(T.obs))
+ }
>
> perm.cor.test(ischemic_hd$time, ischemic_hd$duration, cor.method = "pearson")
[1] 0.66554
> perm.cor.test(ischemic_hd$time, ischemic_hd$duration, cor.method = "spearman")
[1] 0.66365
```