Statistical methods: Homework 3

Agata Ciesielski

November 5, 2018

1 9.2.6

Ring Lardner was one of this country's most popular writers during the 1920s and 1930s. He was also a chronic alcoholic who died prematurely at the age of forty-eight. The following table lists the life spans of some of Lardner's contemporaries (39). Those in the sample on the left were all problem drinkers; they died, on the average, at age sixty-five. The twelve (sober) writers on the right tended to live a full ten years longer. Can it be argued that an increase of that magnitude is statistically significant? Test an appropriate null hypothesis against a one-sided H1. Use the 0.05 level of significance. (Note: The pooled sample standard deviation for these two samples is 13.9.)

Authors Noted for Alcohol Abuse		**Authors Not Noted for Alcohol Abuse**	
Name	**Age at Death**	**Name**	**Age at Death**
Ring Lardner	48	Carl Van Doren	65
Sinclair Lewis	66	Ezra Pound	87
Raymond Chandler	71	Randolph Bourne	32
Eugene O'Neill	65	Van Wyck Brooks	77
Robert Blenchley	56	Samuel Elliot Morrison	89
J.P. Marquand	67	John Crowe Ransom	86
Dashiell Hammett	67	T.S. Eliot	77
e.e. cummings	70	Conrad Aiken	84
Edmund Wilson	77	Ben Ames Williams	
Average:	65.2	Henry Miller	88
		Archibald MacLeish	90
		James Thurbur	67
		Average	75.5

Answer:

H$_0$: $\mu_{alc} = \mu_{N-alc}$
H$_1$: $\mu_{alc} < \mu_{n-alc}$

Our test statistic:

$$\frac{\mu_{alc} - \mu_{n-alc}}{s_p\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{65.22 - 75.5}{13.9\sqrt{\frac{1}{9} + \frac{1}{12}}} = -1.68$$

Our critical value is $t_{.05,19} = -1.729$
Since our test statistic is not less that -1.72, we do not reject H_0.

2 9.2.12

Suppose that $H_0 : \mu X = \mu Y$ is being tested against $H_1 : \mu X \neq \mu Y$,where σ_X^2 and σ_Y^2 are known to be 17.6 and 22.9, respectively. If $n = 10$, $m = 20$, $x = 81.6$, and $y = 79.9$, what P-value would be associated with the observed Z ratio?

Answer:

$\mathbf{H}_0 : \mu X = \mu Y$
$\mathbf{H}_1 : \mu X \neq \mu Y$
Our test statistic is calculated:

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_x^2}{N} + \frac{s_y^2}{M}}} = \frac{81.6 - 79.9}{\frac{17.6}{10} + \frac{22.9}{20}} = 0.997$$

The P value we use R:

```
> pnorm(.997, lower.tail=FALSE)
 [1] 0.1593823
> #two tailed
> 2*pnorm(.997, lower.tail=FALSE)
 [1] 0.3187645
```

Therefore our p value is 0.3187645.

3 9.2.20

Two popular forms of mortgage are the thirty-year fixed-rate mortgage, where the borrower has thirty years to repay the loan at a constant rate, and the adjustable rate mortgage (ARM), one version of which is for five years with the possibility of yearly changes in the interest rate. Since the ARM offers less certainty, its rates are usually lower than those of fixed-rate mortgages. Test this hypothesis at the $\alpha = 0.01$ level using the following sample of mortgage offerings for a loan of $250,000. Do not assume the variances are equal.

Answer:

30 Year Fixed: $\bar{x}_{fixed} = 3.673$ $sd_{fixed} = .153$ n=10 ARM: $\bar{x}_{arm} = 3.338$ $sd_{arm} = .216$ n=10
$\mathbf{H}_0 : \bar{x}_{fixed} = \bar{x}_{arm}$
$\mathbf{H}_1 : \bar{x}_{fixed} \neq \bar{x}_{arm}$

We use Welch's t-test to calculate our test statistic.

$$\frac{\bar{x}_{fixed} - \bar{x}_{arm}}{\sqrt{\frac{sd_{fixed}}{n} + \frac{sd_{arm}}{m}}} = \frac{3.673 - 3.338}{\sqrt{\frac{.152^2}{10} + \frac{.216^2}{10}}} = 4$$

Now we calculate the degrees of freedom.

$$df = \frac{\frac{s_1^1}{n} + \frac{s_2^2}{m}}{\frac{\frac{s_{fixed}^2}{n}^2}{n-1} + \frac{\frac{s_{arm}^2}{n}^2}{n-1}} = \frac{\frac{.153^2}{10} + \frac{.216^2}{10}}{\frac{\frac{.152^2}{10}}{9} + \frac{\frac{.216^2}{10}^2}{9}} = 16$$

Pvalue= .000516, We reject the null hypothesis. Therefore the average rates are not the same between fixed and floating mortgages.

4 9.3.4

In a study designed to investigate the effects of a strong magnetic field on the early development of mice (7), ten cages, each containing three 30-day-old albino fe-male mice, were subjected for a period of 12 days to a magnetic field having an average strength of 80 Oe/cm. Thirty other mice, housed in ten similar cages, were not put in the magnetic field and served as controls. Listed in the table are the weight gains, in grams, for each of the twenty sets of mice/

In Magnetic Field		**Not In Magnetic Field**	
Cage	**Weight Gain (g)**	**Cage**	**Weight Gain (g)**
1	22.8	11	23.5
2	10.2	12	31.0
3	20.8	13	19.5
4	27.0	14	26.2
5	19.2	15	26.5
6	9.0	16	25.2
7	14.2	17	24.5
8	19.8	18	23.8
9	14.5	19	27.8
10	14.8	20	22.0

Test whether the variances of the two sets of weight gains are significantly different. $Let\alpha = 0.05$. For the mice in the magnetic field, $s_X = 5.67$; for the other mice, $s_Y = 3.18$.

Answer:

$\mathbf{H}_0 : \sigma_x = \sigma_y$
$\mathbf{H}_1 : \sigma_x \neq \sigma_y$

Remember, $\frac{s_x}{s_y} \sim F_{m-1,n-1}$.
Therefore, our test statistic is:

$$\frac{s_y^2}{s_x^2} = \frac{3.18^2}{5.67^2} = .314$$

We need to get our critical values: $m - 1 = 9$
$n - 1 = 9$
$F_{.025,9,9} = .248$
$F_{.975,9,9} = 4.03$

Since .314 is in the interval (.248, 4.03) we do not reject the null hypothesis.

5 9.4.4

If flying saucers are a genuine phenomenon, it would follow that the nature of sightings (that is, their physical characteristics) would be similar in different parts of the world. A prominent UFO investigator compiled a listing of ninety-one sightings reported in Spain and eleven hundred seventeen reported elsewhere. Among the information recorded was whether the saucer was on the ground or hovering. His data are summarized in the following table (95).Let pS and pNS denote the true probabilities of "Saucer on ground" in Spain and not in Spain, respectively. Test $H_0 : p_S = p_{NS}$ against a two-sided H_1.Let $\alpha = 0.01$.

	In Spain	**Not in Spain**
Saucer on ground	53	705
Saucer hovering	38	412

Test whether the variances of the two sets of weight gains are significantly different. $Let\alpha = 0.05$. For the mice in the magnetic field, $s_X = 5.67$; for the other mice, $s_Y = 3.18$.

Answer:

$\mathbf{H}_0 : p_S = p_{NS}$
$\mathbf{H}_1 : p_S \neq p_{NS}$

Our estimates
$p_S = \frac{53}{91} = .582$
$p_S \frac{705}{1117} = .6311$

Our pooled estimates: $\hat{p} = \frac{758}{1208} = .627$
Our critical value $z_{\frac{\alpha}{2}} = z_{.005} = -2.58$

Now we calculate our test statistic,

$$\frac{.582 - .6311}{\sqrt{\frac{.627(1-.627)}{91} + \frac{.627(1-.627)}{1208}}} = \frac{-.049}{.0527} = -.93$$

.93 is in the interval (-2.58, 2.58). Therefore we do not reject H_0.

6 9.5.2

Male fiddler crabs solicit attention from the oppo-site sex by standing in front of their burrows and waving their claws at the females who walk by. If a female likes what she sees, she pays the male a brief visit in his bur-row. If everything goes well and the crustacean chemistry clicks, she will stay a little longer and mate. In what may be a ploy to lessen the risk of spending the night alone, some of the males build elaborate mud domes over their burrows. Do the following data (229) suggest that a male's time spent waving to females is influenced by whether his burrow has a dome? Answer the question by constructing and interpreting a 95% confidence interval for $\mu_X - \mu_Y$. Use the value $s_p = 11.2$.

% of Time Spend Waving to Females	
Males with Domes, x_i	**Males without Domes,** y_i
100.0	76.4
58.6	84.2
93.5	96.5
83.6	88.8
84.1	85.3
	79.1
	83.6

Test whether the variances of the two sets of weight gains are significantly different. $Let \alpha = 0.05$. For the mice in the magnetic field, $s_X = 5.67$; for the other mice, $s_Y = 3.18$.

6.1 Answer:

$$\left(\bar{x} - \bar{y} - t_{.025.m+n-2} * s_p * \sqrt{\frac{1}{n} + \frac{1}{m}}, \bar{x} - \bar{y} - t_{.975.m+n-2} * s_p * \sqrt{\frac{1}{n} + \frac{1}{m}}\right)$$
$$\leftrightarrow$$
$$\left(83.86 - 88.84 - 2.281\sqrt{\frac{1}{5} + \frac{1}{7}}, 83.86 - 88.84 + 2.281\sqrt{\frac{1}{5} + \frac{1}{7}}\right)$$
$$\leftrightarrow$$
$$(-15.49, 13.73)$$

implies,

$$(0, 13.73)$$

7 9.5.6

Construct a 95% confidence interval for the $\frac{\sigma_x^2}{\sigma_y^2}$ based on the data in Case Study 9.2.1. The hypothesis test referred to tacitly assumed that the variances were equal. Does that agree with your confidence interval? Explain.

Answer:

$$(F_{1-.025,m-1,n-1}\frac{s_x^2}{s_y^2}, F_{025,m-1,n-1}\frac{s_x^2}{s_y^2})$$

$$(.2383 * (\frac{.0002103}{.0000955}), 4,8232 * (\frac{.0002103}{.0000955})$$

$$C.I. = (0.525, 10.621)$$

8 9.5.10

Construct an 80% confidence interval for the difference $p_m - p_w$ in the nightmare frequency data summarized in Case Study 9.4.2.

Answer:

$$CI = (\bar{x} - \bar{y} - z_{\frac{\alpha}{2}} s_x, \bar{x} - \bar{y} + z_{\frac{\alpha}{2}} s_x)$$

$$CI = (0.031 - .0645, 0.031 + .0645)$$

$$CI = (-0.0335, .0955)$$

9 10.2.6

Based on his performance so far this season, a baseball player has the following probabilities associated with each official at-bat:

Outcome	Probability
Out	.713
Single	.270
Double	.010
Triple	.002
Home Run	.005

Answer:

$$P(X_1 = 2, X_2 = 2, X_3{+}1, X_4 = 0, X_5 = 0) = \frac{5!}{2!2!1!0!0!}.270^2 * .713^2 .010^1 * 1 * 1 = .011$$

10 10.3.6

A number of reports in the medical literature suggest that the season of birth and the incidence of schizophrenia may be related, with a higher proportion of schizophrenics being born during the early months of the year. A study (78) following up on this hypothesis looked at 5139 persons born in England or Wales during the years 1921?1955 who were admitted to a psychiatric ward with a diagnosis of schizophrenia. Of these 5139, 1383 were born in the first quarter of the year. Based on census figures in the two countries, the expected number of persons, out of a random 5139, who would be born in the first quarter is 1292.1. Do an appropriate χ^2 test with ? = 0.05.

Answer:

$\mathbf{H}_0$: Incidence of schizophrenia is uniformly distributed across the year.
$\mathbf{H}_1$: Incidence of schizophrenia is not uniformly distributed.
We have 1 degree of freedom. Therefore the $\chi^2_{.025} = 3.841$

Calculating the test statistic.

$$\hat{p} = np = 1292.1$$

$$r = \sum \frac{(d_i - n\hat{p}^2)}{n\hat{p}} = \frac{(1383 - 1292.1)^2}{1292.1} = 6.39$$

Since 6.39¿3.841 we reject h_0. The incidence of admission for the disease is not uniformly distributed.

11 10.3.9

Records kept at an eastern racetrack showed the following distribution of winners as a function of their starting-post position. All one hundred forty-four races were run with a field of eight horses.

Starting Post	1	2	3	4	5	6	7	8
Number of Winners	32	21	19	20	16	11	14	11

Test the goodness-of-fit hypothesis that horses are equally likely to win from any starting position. Let $\alpha = 0.05$.

Answer:

$\mathbf{H}_0$: The data fits the uniform distribution across starting places.
$\mathbf{H}_1$: The data does not fit the distribution across places.
Our critical value is $\chi_{.95,7} = 14.06$ $E[x] = \frac{144}{8}$

$$\sum \frac{(x_i - E[x])^2}{E[x]} = \frac{(32-13)^2}{18} + \frac{(21-13)^2}{18} + \frac{(19-13)^2}{18} + \frac{(20-13)^2}{18} + \frac{(16-13)^2}{18}$$

$$+\frac{(11-13)^2}{18}+\frac{(14-13)^2}{18}+\frac{(11-13)^2}{18}=18.22$$

$18.22 > 14.06$. Therefore the data does not fit the distribution well and reject H_0.

12 10.4.10

In American football a turnover is defined as a fumble or an intercepted pass. The table below gives the number of turnovers committed by the home team in four hundred forty games. Test that these data fit a Poisson distribution at the 0.05 level of significance.

Turnovers	**Numbers Observed**
0	75
1	125
2	126
3	60
4	34
5	13
6+	7

Answer:

$\mathbf{H}_0$: The data fits the poisson distribution.
$\mathbf{H}_1$: The data does not fit the poisson distribution.

We need to estimate λ

$$\hat{\lambda}=\sum_1^6 f_i * n_i=\frac{800}{440}=1.81$$

Our degrees of freedom $df = n-k-l = 6-1-1 = 4$ Our critical value $\chi_4^2 = 9.4877$

Now our test statistic:

$$\sum\frac{(x-\lambda)^2}{\lambda}=3.52$$

Since 3.52¡9.48, we fail to reject H_0. The data fits the poisson distribution well.

13 10.5.6

High blood pressure is known to be one of the major contributors to coronary heart disease. A study was done to see whether or not there is a significant relationship between the blood pressures of children and those of their fathers (96). If such a relationship did exist, it might be possible to use one group to screen for high-risk individuals in the other group. The subjects were ninety-two eleventh graders, forty-seven males and forty-five females, and their fathers. Blood pressures for both the children and the fathers were categorized as belonging to either the lower, middle, or upper third of their respective distributions. Test whether or not the blood pressures of children can be considered to be independent of the blood pressures of their fathers. Let $\alpha = 0.05$.

Answer:

$\mathbf{H}_0$: Blood Pressure of child and father are independent.
$\mathbf{H}_1$: Blood Pressure of child and father are not independent.

Degrees of freedom, $df = (3-1)*(3-1) = 4$ Critical Value $\chi^2_{.95,4} = 9.488$

$$k = \sum \frac{(o-e)^2}{e} = \frac{(x_{ij} - n\hat{p}\hat{q})^2}{n\hat{p}\hat{q}} = .7461 + .0199 +$$
$$.5547 + .0294 + .0044 + .0609 + 1.2504 + .0070 + 1.1414 = 3.81$$

Since $3.81 < 9.48$ we fail to reject H_o and the blood pressure of the child and parent for the assumption that they are independent well.

14 R PROGRAMMING ASSIGNENT

We have 1 occurrence of the sample means being less than -.33.

Therefore our p value is $p = \frac{1}{1000} = .0001$

R Programming Assignment:

```
relcomp <- function(a, b) {

  comp <- vector()

  for (i in a) {
    if (i %in% a && !(i %in% b)) {
      comp <- append(comp, i)
    }
  }

  return(comp)
}

arm<-c(2.923,3.385,3.154,3.363,3.226,3.283,3.427,3.437,3.746,3.438)
fxd<-c(3.525,3.625,3.383,3.625,3.661,3.791,3.941,3.781,3.660,3.773)
diff<-mean(arm)-mean(fxd)
combined <- c(arm, fxd)
diffavg <- c()
lt<-0
for (i in 1:1000){
  smp <- sample(1:20,10)
  nsmp <-relcomp(1:20, smp)
```

```
  set1 <- combined [smp]
  set2 <- combined [nsmp]
  diffavg = c(diffavg, mean(set1)-mean(set2) )
  if (mean(set1)-mean(set2) < diff){
    lt <-lt+1
  }
}
```