

# CS 361 Spring 2018

## Homework 7

Nathaniel Murphy (njmurph3)

### 7.3

Let us first compute  $\text{mean}(\{x\})$  and  $\text{stderr}(\{x\})$ .

$$\text{mean}(\{x\}) = \frac{1}{N} \sum_{x=1}^N x_i = \frac{199}{10} = 19.9$$

$$\text{stdunbiased}(\{x\}) = \sqrt{\frac{\sum_{i=1}^N (x_i - \text{mean}(\{x\}))^2}{N-1}} = \sqrt{\frac{110.9}{9}} = 3.51$$

$$\text{stderr}(\{x\}) = \frac{\text{stdunbiased}(\{x\})}{\sqrt{N}} = \frac{3.51}{\sqrt{10}} = 1.11$$

$$t = \frac{\mu - \text{mean}(\{x\})}{\text{stderr}(\{x\})} = \frac{25 - 19.9}{1.11} = 4.59$$

Using the test statistic  $t$ , we use a two sided test. Obtaining a p-value between 0.001 and 0.002, which is less than 0.05, so we may conclude that  $P(\text{popmean}(\{x\}) = 25)$  is extremely unlikely, so we reject the claim.

### 7.6

$$N_1 = 35, \bar{x}_1 = 300, s_1 = 30, \mu_1 = ?, \sigma_1 = ?$$

$$N_2 = 30, \bar{x}_2 = 400, s_2 = 100, \mu_2 = ?, \sigma_2 = ?$$

**Claim:**  $\mu_1 = \mu_2$

$$s_{\text{ed}}^2 = \frac{s_1^2}{N_1} + \frac{s_2^2}{N_2} = \frac{30^2}{35} + \frac{100^2}{30} = 359.05$$

$$s = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_{\text{ed}}^2}} = \frac{300 - 400}{\sqrt{359.05}} = -5.277$$

Degrees of freedom:

$$\frac{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}{\left(\frac{s_1^2}{N_1}\right)^2 + \left(\frac{s_2^2}{N_2}\right)^2} = \frac{359.05^2}{19.45 + 3831.42} = \frac{128916.90}{3850.87} \approx 33$$

Since we have  $33 > 30$  degrees of freedom, we may use a z-table for a normal distribution. We find that a two tailed test with  $z = 5.277$  has probability approximately zero, so we may conclude that the two populations have different means: specifically that the mean weight of female mice is less than the mean weight of a male mouse.

## 7.7

(a)

$$\begin{aligned}\mathbb{E}[L^{(k)}] &= \text{popmean}(L) \\ \text{std}(L^{(k)}) &= \frac{\text{popstd}(L)}{\sqrt{k}}\end{aligned}$$

(b)

$$\begin{aligned}\mathbb{E}[F^{(s)}] &= \text{popmean}(F) \\ \text{std}(F^{(s)}) &= \frac{\text{popstd}(F)}{\sqrt{s}}\end{aligned}$$

(c)

$$\begin{aligned}\mathbb{E}[F^{(s)} - 2L^{(k)}] &= \mathbb{E}[F^{(s)}] - 2\mathbb{E}[L^{(k)}] = \text{popmean}(F) - 2 \cdot \text{popmean}(L) = 0 \\ \text{std}(F^{(s)} - 2L^{(k)}) &= \sqrt{(\text{std}(F^{(s)}))^2 + 4 \cdot (\text{std}(L^{(k)}))^2} = \sqrt{\frac{\text{popstd}^2(F)}{s} + \frac{4 \cdot \text{popstd}^2(L)}{k}} \\ \text{stderr}(F^{(s)} - 2L^{(k)}) &= \frac{\text{std}(F^{(s)} - 2L^{(k)})}{s + k - 1} = \frac{1}{s + k - 1} \sqrt{\frac{\text{popstd}^2(F)}{s} + \frac{4 \cdot \text{popstd}^2(L)}{k}}\end{aligned}$$

(d)

We see that in our samples of lean mice,  $L$ , we observe an average weight of exactly half the average of the fatty mice,  $F$ . It follows that every confidence interval of the form

$$\text{mean}(\{F\}) - 2 \cdot \text{mean}(\{L\}) - \epsilon < \text{mean}(\{F\}) - 2 \cdot \text{mean}(\{L\}) < \text{mean}(\{F\}) - 2 \cdot \text{mean}(\{L\}) + \epsilon$$

will contain 0 for all  $\epsilon > 0$ , so we cannot reject the claim that  $\text{popmean}(\{F\}) = 2 \cdot \text{popmean}(\{L\})$ .

## 7.8

$$N = 2009, \hat{p} = \frac{983}{2009} = 0.4893$$

Recall that the test for comparing whether two proportions are equal is given by the following formula:

$$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{N}}}$$

We will reject the claim that a boy is born with probability  $p = 0.5$  if our test statistic  $z > 1.96$  or  $z < -1.96$  as our test is a 2-tail test (because our claim contains  $=$ ).

$$\frac{0.4893 - 0.5}{\sqrt{\frac{0.5(0.5)}{2009}}} = -0.9592$$

We see that  $-1.96 < -0.9592 < 1.96$ , so we may not reject our claim that  $p = 0.5$ .

## 7.9

(a)

Assess the evidence that survival is independent of passenger ticket class. A chi squared p-value less than 0.05 would imply that the data is not correlated.

After preprocessing the data so that ticket class is a number, we group the data and find observed frequencies and expected frequencies for each category. The observed and expected tables are as below.

**Observed:**

	1st class	2nd class	3rd class
Perish	129	161	573
Survive	193	119	138

**Expected:**

	1st class	2nd class	3rd class
Perish	211	184	467
Survive	110	95	243

Using these observed and expected frequencies, we obtain a chi-square test statistic of 172.863 and a corresponding p-value of  $1.78 \times 10^{-35}$ . We therefore reject the claim that survival is dependent of passenger ticket class.

(b)

Assess the evidence that survival is independent of passenger gender. A chi squared p-value less than 0.05 would imply that the data is not correlated.

After preprocessing the data so that passenger sex is a number (0 or 1), we group the data and find observed frequencies and expected frequencies for each category. The observed and expected tables are as below.

**Observed:**

	<b>Male</b>	<b>Female</b>
<b>Perish</b>	709	154
<b>Survive</b>	142	308

**Expected:**

	<b>Male</b>	<b>Female</b>
<b>Perish</b>	559.34	303.66
<b>Survive</b>	291.66	158.34

Using these observed and expected frequencies, we obtain a chi-square test statistic of 332.06 and a corresponding p-value of  $1.14 \times 10^{-71}$ . We therefore reject the claim that survival is dependent of passenger ticket class.

## 7.10

**(a)**

Assess the evidence that income category is independent of sex. A chi squared p-value less than 0.05 would imply that the two variables are not correlated.

After preprocessing the data so that income category and sex are numbers, we group the data and find observed and expected frequencies for each category listed below.

**Observed:**

	<b>Male</b>	<b>Female</b>
<b>Income &lt; 50k</b>	15128	9592
<b>Income <math>\geq</math> 50k</b>	6662	1179

**Expected:**

	<b>Male</b>	<b>Female</b>
<b>Income &lt; 50k</b>	16542.76	8177.24
<b>Income <math>\geq</math> 50k</b>	5247.24	2593.76

Using these observed and expected frequencies, we obtain a chi-square test statistic of 1518.89 with a corresponding p-value of approximately 0, so we can conclude that income is independent of gender.

**(b)**

Assess the evidence that income category is independent of education level. A chi squared p-value less than 0.05 would imply that the two variables are not correlated.

After preprocessing the data so that income category is, we group the data by college graduate (education-level  $\geq 11$ ) and non-college graduate (education-level  $< 11$ ) and find observed and expected frequencies for each category listed below.

**Observed:**

	Non college graduate	College graduate
Income < 50k	18739	5981
Income $\geq$ 50k	3306	4535

**Expected:**

	Non college graduate	College graduate
Income < 50k	16736.35	7983.65
Income $\geq$ 50k	5308.65	2532.35

Using these observed and expected frequencies, we obtain a chi-square test statistic of 3081.21 with a corresponding p-value of approximately 0, so we can conclude that income is independent of gender.