Lab 7 suggested solution (with comments)

Zhangzhang (Bacon)

1. compare means of two groups, using p-value

H_0: mean weight of baby born by smoker mother = mean weight of baby born by nonsmoker mother. Or simply put: $\mu_{smoker} - \mu_{nonsmoker} = 0$.

H_a: $\mu_{nonsmoker} > \mu_{smoker}$. (so here we use one-sided test)

Performing the t-test for comparing means of two groups, I get the following result in fathom:

Test of 1000Births. Compare Means

First attribute (numeric): weight
Second attribute (numeric or categorical): smoker

Sample count of **smoker** = **nonsmoker**: **873**
Sample count of **smoker** = **smoker**: **126**
Sample mean of **weight** where **smoker** = **nonsmoker**: **7.14427**
Sample mean of **weight** where **smoker** = **smoker**: **6.82873**
Standard deviation of **weight** where **smoker** = **nonsmoker**: **1.51868**
Standard deviation of **weight** where **smoker** = **smoker**: **1.38618**
Standard error of **weight** where **smoker** = **nonsmoker**: **0.0513996**
Standard error of **weight** where **smoker** = **smoker**: **0.123491**
Alternative hypothesis: The population mean of **weight** where **smoker** = **nonsmoker** **is greater than** that where **smoker** = **smoker**

One-sided test

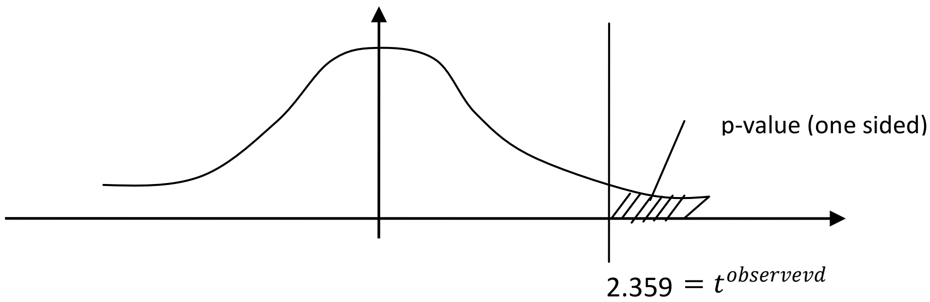
The test statistic, Student's t, using **unpooled variances** , is **2.359**. There are **171.325** degrees of freedom.

If it were true that the population mean of **weight** where **smoker** = **nonsmoker** were equal to that of **weight** where **smoker** = **smoker** (the null hypothesis), and the sampling process were performed repeatedly, the probability of getting a value for Student's t **this great or greater** would be **0.0097**.

Reading from the fathom output, P-value = 0.0097 < 0.05. So the evidence from observed data is significant enough to reject the null hypothesis that the mean weights are equal for smoker group vs. non-smoker group.

Recall: how is p-value computed? (compare two groups) It is computed by evaluating how extreme/rare the observed difference is compared to the null hypothesis.
- Firstly, we compute the observed difference $\bar{y}_{nonsmoker} - \bar{y}_{smoker}$ = 7.144 - 6.828 = 0.316 lbs.
- Secondly, compute the Standard Error. SE for two groups : $SE = \sqrt{\frac{s^2_{nonsmoker}}{n_{nonsmoker}} + \frac{s^2_{smoker}}{n_{smoker}}}$.
- Thirdly, compute the **observed** t-score. To get t-score, the observed difference is “rescaled” or “standardized” by subtracting the null difference (which is 0) and being divided the Standard Error (SE). $t^{observed} = 2.359$.
- Finally, compute p-value. With d.f. = 171.325 also provided (don’t worry about the formula of d.f. for two-group comparison), we **reversely** look up in t-table the proportion larger than $t^{observed}$ (or smaller than $t^{observed}$ when it is negative). Then p-value is 2 times this proportion if we perform 2-sided test; p-value equals this proportion if it is 1-

sided test.



Side note: if we perform 2-sided test, then p-value simply *doubles*. This is because we measure the proportion either larger than 2.359 or smaller than -2.359. (see the fathom result below)

Alternative hypothesis: The population mean of **weight** where **smoker** = **nonsmoker** **is not equal to** that where **smoker** = **smoker**

The test statistic, Student's t, using **unpooled variances**, is **2.359**. There are **171.325** degrees of freedom.

If it were true that the population mean of **weight** where **smoker** = **nonsmoker** were equal to that of **weight** where **smoker** = **smoker** (the null hypothesis), and the sampling process were performed repeatedly, the probability of getting a value for Student's t **with an absolute value this great or greater** would be **0.019**.

2. compare means of two groups: using p-value

Test of 1000Births. Compare Means

First attribute (numeric): weight

Second attribute (numeric or categorical): mature

Sample count of **mature** = **mature mom**: **133**
Sample count of **mature** = **younger mom**: **867**
Sample mean of **weight** where **mature** = **mature mom**: **7.12556**
Sample mean of **weight** where **mature** = **younger mom**: **7.09723**
Standard deviation of **weight** where **mature** = **mature mom**: **1.65908**
Standard deviation of **weight** where **mature** = **younger mom**: **1.48548**
Standard error of **weight** where **mature** = **mature mom**: **0.143861**
Standard error of **weight** where **mature** = **younger mom**: **0.0504495**
Alternative hypothesis: The population mean of **weight** where **mature** = **mature mom** **is not equal to** that where **mature** = **younger mom**

The test statistic, Student's t, using **unpooled variances**, is **0.1858**. There are **166.08** degrees of freedom.

If it were true that the population mean of **weight** where **mature** = **mature mom** were equal to that of **weight** where **mature** = **younger mom** (the null hypothesis), and the sampling process were performed repeatedly, the probability of getting a value for Student's t **with an absolute value this great or greater** would be **0.85**.

$H_0: \mu_M - \mu_Y = 0$. $H_a: \mu_M - \mu_Y \neq 0$.

p-value = 0.85 > 0.05. So the evident is not significant enough. We fail to reject the null hypothesis, that baby's weight for mature moms and young moms is the same.

3. compare proportions of two groups: using p-value

Test of 1000Births. Compare Proportions

Attribute (categorical): premie

Attribute (categorical or grouping): mature

In **premie where mature is mature mom** **23** out of **133**, or **0.172932**, are **premie**
In **premie where mature is younger mom** **129** out of **867**, or **0.148789**, are **premie**
Alternative hypothesis: The population proportion for **premie** in **premie where mature is mature mom** **is greater than** that for **premie** in **premie where mature is younger mom**

The test statistic, **z**, is **0.7221**.

If it were true that the two population proportions were equal (the null hypothesis), and the sampling process were performed repeatedly, the probability of getting a value of **z** **this great or greater** would be **0.24**.

Note: This probability was computed using the normal approximation.

$H_0: p_M - p_Y = 0$. $H_a: p_M - p_Y > 0$. p_M means the proportion of premature baby for mature mom. p_Y means the proportion of premature baby for young mom.

p-value = 0.24 > 0.05. So at 5% level of significance, we fail to reject the null hypothesis that mature moms and young moms are equally likely to have premature babies. In other words, we do NOT support the claim that the proportion of premature births for mature moms is significantly higher than the premature birth rate for younger moms.

(If you use one-sided test and get a p-value larger than 0.5, then probably you chose the wrong side.)

4. compare means of two groups: using confidence interval

Estimate of 1000Births. Difference of Means

First attribute (numeric): weight

Second attribute (numeric or categorical): smoker

Interval estimate for the difference of means of **weight** for **smoker** = **nonsmoker** and **smoker**.

Sample count of **smoker** = **nonsmoker**: **873**
Sample count of **smoker** = **smoker**: **126**
Sample mean of **weight** where **smoker** = **nonsmoker**: **7.14427**
Sample mean of **weight** where **smoker** = **smoker**: **6.82873**
Standard deviation of **weight** where **smoker** = **nonsmoker**: **1.51868**
Standard deviation of **weight** where **smoker** = **smoker**: **1.38618**
Standard error of **weight** where **smoker** = **nonsmoker**: **0.0513996**
Standard error of **weight** where **smoker** = **smoker**: **0.123491**

Based on the samples and using **unpooled variances** , the **95.0** % confidence interval for the mean(**weight** where **smoker** = **nonsmoker**) - mean(**weight** where **smoker** = **smoker**) is **0.315542** plus or minus **0.264031** ranging from **0.0515117** to **0.579573**.

If the sampling process were performed repeatedly, the confidence intervals generated would capture the population difference of means **95.0** % of the time.

$H_0: \mu_{smoker} - \mu_{nonsmoker} = 0$. $H_a: \mu_{nonsmoker} \neq \mu_{smoker}$.

The null difference 0 is not included in the confidence interval [0.05, 0.58]. So we reject the null. The conclusion is the same as question 1.

Recall: how is confidence interval computed? (compare two groups)
- Firstly, we compute the observed difference $\bar{y}_{nonsmoker} - \bar{y}_{smoker}$ = 7.144 - 6.828 = 0.316 lbs.
- Secondly, compute the Standard Error. SE for two groups : $SE = \sqrt{\frac{s^2_{nonsmoker}}{n_{nonsmoker}} + \frac{s^2_{smoker}}{n_{smoker}}}$.

- Thirdly, compute **reference** t-score for 95% confidence interval. With d.f. = 171.325 also provided, we look up in the t-table, to find the reference t-score $t_{0.025,\ 171}$. Note that the observed difference of 0.316 was NOT used in computing this t-score.
- Finally, compute margin of error and confidence interval.

Margin of error = SE * $t_{0.025,\ 171}$

Confidence interval = $[\text{observed difference} - \text{Margin of error},\ \text{observed difference} + \text{Margin of error}]$

= [0.316 – margin of error , 0.316 + margin of error]

What is the relationship between p-value and confidence interval?

Graphically, hypothesis testing is about evaluating *how far is observed data from the null*. See the two figures below to see two methods of hypothesis testing (or inference): using p-value vs. using confidence interval.

Observed data is far from the null

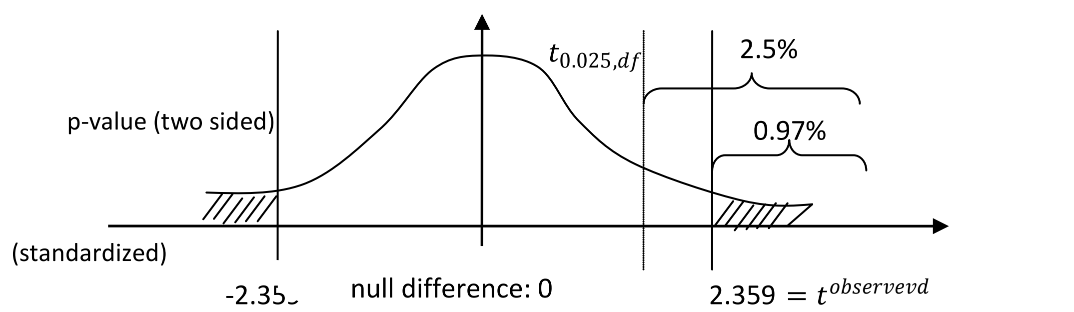
⇔ observed difference (or other test statistic) rarely happens under the null (chance model)

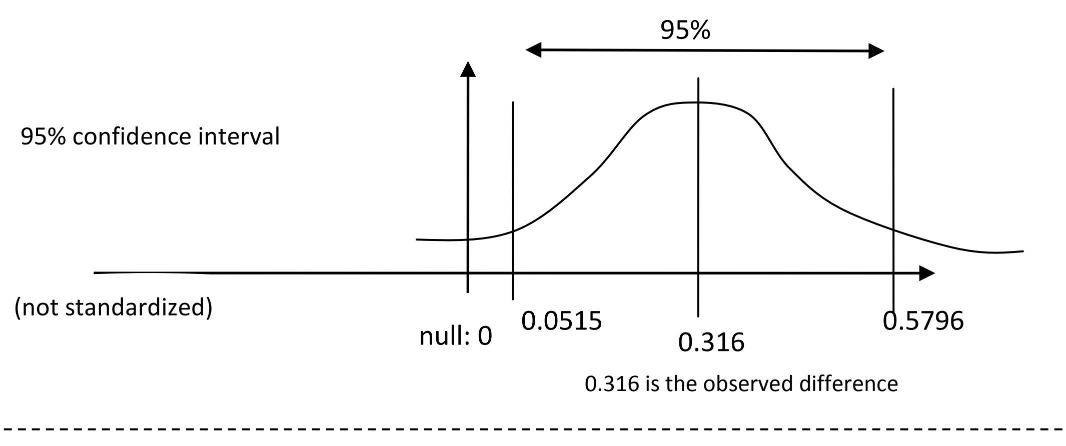
⇔ p-value is small

⇔ t-score / z-score is large (far from **0**) . Since in t-score we already subtracted the null difference, so we always compare t-score with 0.

⇔ 95% Confidence interval *around the observed difference* does NOT include the null value (it may or may not be 0 !!) .

⇔ Reject the null.





Numerically, in two-sided test for two groups, using p-value is equivalent to using confidence interval.

Recall: 95% Confidence interval = $\text{observed difference} \pm \text{SE} * t_{0.025,\ df}$

Let's see what it means when the null difference is not included in the 95% CI. To simplify, only consider the case when null difference is to the left of the confidence interval.

Null difference < observed difference $-$ SE $* \; t_{0.025,\; df}$

$\Leftrightarrow \dfrac{\text{observed difference} - \text{Null difference}}{SE} > t_{0.025,\; df}$

$\Leftrightarrow \quad t^{observed} \;=>\; t_{0.025,\; df}$

$\Leftrightarrow$ p-value < (0.025 *2) = 0.05

The last step is because we know when d.f. is fixed, larger t-score means smaller right-tail. So tail < 0.025, and for two sided test p-value = 2* tail.

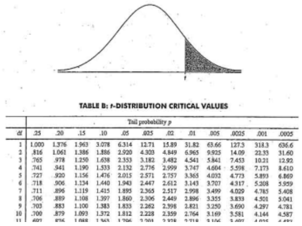


TABLE B: t-DISTRIBUTION CRITICAL VALUES

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

In test-of-mean for one group, the conclusion is the same. Feel free to check it as an exercise.

5. compare means of two groups: using confidence interval

Estimate of 1000Births. Difference of Means

First attribute (numeric): gained

Second attribute (numeric or categorical): premie

Interval estimate for the difference of means of **gained** for **premie** = **full term** and **premie**.

Sample count of **premie** = **full term**: **827**
Sample count of **premie** = **premie**: **145**
Sample mean of **gained** where **premie** = **full term**: **31.1258**
Sample mean of **gained** where **premie** = **premie**: **25.8**
Standard deviation of **gained** where **premie** = **full term**: **14.3008**
Standard deviation of **gained** where **premie** = **premie**: **13.0919**
Standard error of **gained** where **premie** = **full term**: **0.497286**
Standard error of **gained** where **premie** = **premie**: **1.08722**

Based on the samples and using **unpooled variances** , the **95.0** % confidence interval for the mean(**gained** where **premie** = **full term**) - mean(**gained** where **premie** = **premie**) is **5.32576** plus or minus **2.35689** ranging from **2.96887** to **7.68264**.

If the sampling process were performed repeatedly, the confidence intervals generated would capture the population difference of means **95.0** % of the time.

$H_0: \mu_F - \mu_P = 0 \,.\, H_a: \mu_F \neq \mu_P$.

The null difference 0 is not included in the 95% CI. So we reject the null. The conclusion is that mothers with full term babies are likely to gain significantly more weight.

6. test the mean of one group: using confidence interval

Estimate of 1000Births. Estimate Proportion

Attribute (categorical): sexbaby

Interval estimate for population proportion of **male** in **sexbaby**

In the sample **497** out of **1000**, or **0.497**, are **male** .

Based on the sample, the **95.0** % confidence interval for the population proportion of **male** in **sexbaby** is from **0.466** to **0.528**.

If the sampling process were performed repeatedly, the confidence intervals generated would capture the population proportion **95.0** % of the time.

$H_0: p_{Male} = 0.52 \,.\, H_a: p_{Male} \neq 0.52$

The null value 0.52 is included in the 95% CI. So we fail to reject the null at 5% significance level. This supports the claim by the North Carolina report.