STAT 420

1. A store sells "16-ounce" boxes of *Captain Crisp* cereal. A random sample of 9 boxes was taken and weighed. The results were the following (in ounces):

15.5	16.2	16.1	15.8	15.6	16.0	15.8	15.9	16.2

Assume the weight of cereal in a box is normally distributed.

a) Compute the sample mean $\bar{x}$ and the sample standard deviation s.

$$\sum x = 143.1, \qquad \sum x^2 = 2275.79, \qquad \sum (x - \bar{x})^2 = 0.50.$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{143.1}{9} = \mathbf{15.9}.$$

$$s^2 = \frac{\sum x_i^2 - \left(\sum x_i\right)^2 / n}{n-1} = \frac{2275.79 - (143.1)^2 / 9}{8} = \frac{0.5}{8} = 0.0625.$$

OR

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{0.5}{8} = 0.0625.$$

$$s = \sqrt{s^2} = \sqrt{0.0625} = \mathbf{0.25}.$$

```
> x = c(15.5,16.2,16.1,15.8,15.6,16.0,15.8,15.9,16.2)
```

```
> mean(x)
[1] 15.9
```

OR

```
> sum(x)/length(x)
[1] 15.9
```

```
> sd(x)
[1] 0.25
```

OR

```
> sqrt(var(x))
[1] 0.25
```

OR

```
> sqrt(sum((x-mean(x))^2)/(length(x)-1))
[1] 0.25
```

b) Construct a 95% confidence interval for the overall average weight of boxes of *Captain Crisp* cereal.

σ is unknown. $n = 9$ - small. The confidence interval : $\overline{\mathrm{X}} \pm \mathrm{t}_{\alpha/2} \dfrac{s}{\sqrt{n}}$.

$\alpha = 0.05$ $\alpha/2 = \mathbf{0.025}$.

number of degrees of freedom $= n - 1 = 9 - 1 = \mathbf{8}$. $\mathrm{t}_{\alpha/2} = \mathbf{2.306}$.

$15.9 \pm 2.306 \cdot \dfrac{0.25}{\sqrt{9}}$ $\mathbf{15.9 \pm 0.192}$ **(15.708 ; 16.092)**

```
> t.test(x,alternative=c("two.sided"),conf.level=0.95)

        One Sample t-test

data:  x
t = 190.8, df = 8, p-value = 6.372e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 15.70783 16.09217
sample estimates:
mean of x
     15.9
```

```
OR         > qt(0.975,8)
           [1] 2.306004
           > mean(x)-qt(0.975,8)*sd(x)/sqrt(9)
           [1] 15.70783
           > mean(x)+qt(0.975,8)*sd(x)/sqrt(9)
           [1] 16.09217
```

c) The company that makes *Captain Crisp* cereal claims that the average weight of its box is at least 16 ounces. Use a 0.05 level of significance to test the company's claim. What is the p-value of this test?

Claim: $\mu \ge 16$

$H_0 : \boldsymbol{\mu \ge 16}$ vs. $H_1 : \boldsymbol{\mu < 16}$. Left - tailed.

σ is unknown. $n = 9$ – small.

$$T = \frac{\overline{X} - \mu_0}{s / \sqrt{n}} = \frac{15.9 - 16}{0.25 / \sqrt{9}} = \mathbf{-1.2}.$$

Rejection Region: $T < -t_{\alpha}$

number of degrees of freedom

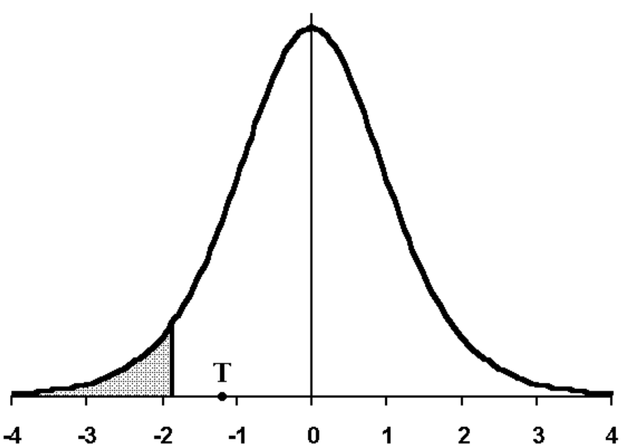
$n - 1 = 9 - 1 = \mathbf{8}$.

$-t_{0.05} = \mathbf{-1.860}$.

The value of the test statistic is **not** in the Rejection Region.

Do NOT Reject H_0

at $\alpha = 0.05$.



OR

	Upper-tail Probability			
df	…	0.15	0.10	…
8	…	1.108	1.397	…

$T = -1.2$.

0.10 < p-value < 0.15.

P-value > α.

Do NOT Reject H_0 at $\alpha = 0.05$.

The t Distribution

r	$t_{0.40}$	$t_{0.25}$	$t_{0.20}$	$t_{0.15}$	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.02}$	$t_{0.01}$	$t_{0.005}$
8	0.262	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355

```
> t.test(x,mu=16,alternative=c("less"),conf.level=0.95)

        One Sample t-test

data:  x
t = -1.2, df = 8, p-value = 0.1322
alternative hypothesis: true mean is less than 16
95 percent confidence interval:
     -Inf 16.05496
sample estimates:
mean of x
     15.9
```

OR

```
> pt(sqrt(9)*(mean(x)-16)/sd(x),8)
[1] 0.1322336
```

P-value > α. **Do NOT Reject H_0 at $\alpha = 0.05$.**

2. Assume that the distributions of X and Y are $\mathrm{N}(\mu_1, \sigma^2)$ and $\mathrm{N}(\mu_2, \sigma^2)$, respectively. Given the $n = 6$ observations of X,

70, 82, 78, 74, 94, 82

and the $m = 8$ observations of Y,

64, 72, 60, 76, 72, 80, 84, 68

find the p-value for the test $\mathrm{H}_0: \mu_1 = \mu_2$ vs. $\mathrm{H}_1: \mu_1 > \mu_2$.

$\bar{x} = 80$, $s_x^2 = 68.8$. $\bar{y} = 72$, $s_y^2 = 64$.

$$s_{\text{pooled}}^2 = \frac{(6-1)\cdot 68.8 + (8-1)\cdot 64}{6+8-2} = 66 \qquad s_{\text{pooled}} \approx 8.124$$

Test Statistic:

$$\mathrm{T} = \frac{(\overline{\mathrm{X}} - \overline{\mathrm{Y}}) - \delta_0}{s_{\text{pooled}} \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{(80-72)-0}{8.124 \cdot \sqrt{\frac{1}{6} + \frac{1}{8}}} \approx \mathbf{1.82337}.$$

$n + m - 2 = 6 + 8 - 2 = \mathbf{12}$ d.f.

	Upper-tail Probability			
df	…	0.05	0.025	…
12	…	1.782	2.179	…

T = 1.82337.

0.025 < p-value < 0.05.

```
> x = c(70,82,78,74,94,82)
> y = c(64,72,60,76,72,80,84,68)
```

```
> t.test(x,y,alternative=c("greater"),var.equal=TRUE)

        Two Sample t-test

data:  x and y
t = 1.8234, df = 12, p-value = 0.04662
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.1802451       Inf
sample estimates:
mean of x mean of y
       80        72
```

OR

```
> Spooled2 = ((6-1)*var(x)+(8-1)*var(y))/(6+8-2)
> Spooled2
[1] 66
> test_stat = (mean(x)-mean(y))/sqrt(Spooled2*(1/6+1/8))
> test_stat
[1] 1.823369
> 1-pt(test_stat,6+8-2)
[1] 0.04661961
```

3. Consider the model:

$$X_{11}, X_{12}, \ldots, X_{1n} \text{ are i.i.d. } N(\mu_1, \sigma^2)$$

$$X_{21}, X_{22}, \ldots, X_{2n} \text{ are i.i.d. } N(\mu_2, \sigma^2)$$

Assume that $\mu_1 = 6,\ \mu_2 = 5,\ \sigma^2 = 4,\ n = 25$.

Let $\overline{X}_1 = \frac{1}{n}\sum_{i=1}^{n} X_{1i}$, $\overline{X}_2 = \frac{1}{n}\sum_{i=1}^{n} X_{2i}$, $D = \overline{X}_1 - \overline{X}_2$.

a) Find $P(0 < D < 2)$.

$$D = \overline{X}_1 - \overline{X}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma^2}{n} + \frac{\sigma^2}{n}) = N(6-5, \frac{4}{25} + \frac{4}{25})$$

$$D \sim N(1, 0.32) \qquad \frac{D-1}{\sqrt{0.32}} = Z \sim N(0,1)$$

$$P(0 < D < 2) \approx P(-1.77 < Z < 1.77) = 0.9616 - 0.0384 = \mathbf{0.9232}.$$

OR

```
> z = 1/sqrt(0.32)
> z
[1] 1.767767
> pnorm(z)-pnorm(-z)
[1] 0.9229001
```

b) Empirical distribution of D:

Generate $S = 1000$ datasets for each of group 1 and group 2.

For each of the $s = 1:1000$ datasets, compute $d_s = \overline{x}_{1s} - \overline{x}_{2s}$.

Make a histogram for the 1000 values of d.

What is the proportion of values of d (among the 1000 values of d generated) that are between 0 and 2?

```
> N = 25
> mu1 = 6
> mu2 = 5
> std = 2
```

```
> S=1000
> count = 0
> diffall = c(1:S)
>
> for (i in 1:S){
+ x1 = rnorm(N, mu1, std)
+ x2 = rnorm(N, mu2, std)
+ diffall[i] = mean(x1) - mean(x2)
+ if ((diffall[i] > 0) && (diffall[i] < 2)){
+ count = count + 1}}
>
> count/S
[1] 0.925
>
> hist(diffall)
```