Homework 0A (Review?)
due? Jan. 10

The following problem set is based from information in Appendix A. This is review or new material depending on which statistics course you took as a pre-requisite.

1. (From A.3) A certain automobile manufacturer equips a particular model with either a six-cylinder engine or a four-cylinder engine. Let X_1 and X_2 be fuel efficient for independently and randomly selected six-cylinder and four-cylinder cars, respectively. The means and standard deviations are as follows:

$$E\{X_1\} = 22 \qquad \sigma\{X_1\} = 1.2$$
$$E\{X_2\} = 26 \qquad \sigma\{X_2\} = 1.5$$

a) Calculate the mean and standard deviation of $D = X_1 - X_2$.

$E\{D\} = E\{X_1\} - E\{X_2\} = 22 - 26 = -4$

$\sigma\{D\} = \sqrt{\sigma^2\{X_1\} + \sigma^2\{X_2\}} = \sqrt{1.2^2 + 1.5^2} = 1.921$

b) Calculate the mean and standard deviation of $T = X_1 + X_2$.

$E\{T\} = E\{X_1\} + E\{X_2\} = 22 + 26 = 48$

$\sigma\{T\} = \sqrt{\sigma^2\{X_1\} + \sigma^2\{X_2\}} = \sqrt{1.2^2 + 1.5^2} = 1.921 = \sigma\{D\}$

c) Calculate the mean and standard deviation of $A = T/2$.

$$E\{A\} = \frac{E\{T\}}{2} = \frac{48}{2} = 24$$
$$\sigma\{A\} = \frac{\sigma\{T\}}{2} = \frac{1.921}{2} = 0.9605$$

2. (A.4) Suppose the tensile strength of type-A steel has a distribution of $\mathcal{N}(\mu = 105 \text{ ksi}, \sigma^2 = 64 \text{ ksi}^2)$ and the tensile strength of type-B steel has a distribution of $\mathcal{N}(\mu = 100 \text{ ksi}, \sigma^2 = 36 \text{ ksi}^2)$. Assume that the strengths of the two types of steel are independent. Let T = A + B.

a) Describe the probability distribution of T.

Referring to 1,

$E\{T\} = E\{A\} + E\{B\} = 105 + 100 = 205$

$\sigma\{T\} = \sqrt{\sigma^2\{A\} + \sigma^2\{B\}} = \sqrt{64 + 36} = 10$

Since both A and B are normally distributed, T = A + B is also normally distributed so

$T \sim \mathcal{N}(\mu = 205 \text{ ksi}, \sigma^2 = 100 \text{ ksi}^2)$

b) Find the probability that the total strength of these two types of steel exceeds 200 ksi?

$$P(T > 200) = 1 - P(T \le 200) = 1 - P\left(Z \le \frac{200 - 205}{10}\right) = 1 - P(Z \le -0.5) = 1 - 0.3085 = 0.6915$$

c) Find the probability that the total strength is between 190 and 200 ksi?

$$P(190 < T < 200) = P(T < 200) - P(T < 190) = P\left(Z < \frac{200 - 205}{10}\right) - P\left(Z < \frac{190 - 205}{10}\right)$$
$$= P(Z < -0.5) - P(Z < -1.5) = 0.3085 - 0.0668 = 0.2417$$

d) Obtain the 10th percentile of the probability distribution of T. Interpret this quantity.

P(Z < z) = 0.10 ==> z = -1.28

$$z = \frac{t - E\{T\}}{\sigma\{T\}} \Rightarrow t = E\{T\} + z\sigma\{T\} = 205 + (-1.28)(10) = 192.2$$

3. (A.3) Suppose that the pH of a certain chemical compound is 5.00 and the pH measured by a randomly selected beginning chemistry student is 5.00 with variance 0.04.

a) For a random sample of 55 students, what is the probability that the sample mean will be within ±0.01 of the population mean?

$E\{\bar{P}\} = E\{P\} = 5.00$

$$\sigma\{\bar{P}\} = \frac{\sigma\{P\}}{\sqrt{n}} = \frac{0.2}{\sqrt{55}} = 0.0270$$

$$P(4.99 < P < 5.01) = P(P < 5.01) - P(P < 4.99) = P\left(Z < \frac{5.01 - 5.00}{0.0270}\right) - P\left(Z < \frac{4.99 - 5.00}{0.0270}\right)$$
$$= P(Z < 0.37) - P(Z < -0.37) = 0.6443 - 0.3557 = 0.2886$$

b) Would the probability be increased by 20 percent if the sample size were increased by 20 percent, to n = 66? Be specific, but do not perform any calculations.

$E\{\bar{P}\} = E\{P\} = 5.00$

$$\sigma\{\bar{P}\} = \frac{\sigma\{P\}}{\sqrt{n}} = \frac{0.2}{\sqrt{66}} = 0.0246$$

Therefore, the probability will be between two larger numbers so it will be increased. But it will not be increased by 20%.

$$P(4.99 < P < 5.01) = P(P < 5.01) - P(P < 4.99) = P\left(Z < \frac{5.01 - 5.00}{0.0246}\right) - P\left(Z < \frac{4.99 - 5.00}{0.0246}\right)$$
$$= P(Z < 0.41) - P(Z < -0.41) = 0.6591 - 0.3409 = 0.3182$$

4. (A.6) It is important in a safe workplace that workers are not asked to perform tasks, such as lifting, that exceed their capabilities. The following data is for the maximum weight of lift (MAWL, kg) for five randomly selected healthy males with ages from 18 – 30.

25.8	**36.6**	**26.3**	**21.8**	**27.2**

Let μ be the mean MAWL weight for healthy males in this age range. Assume that MAWL are normally distributed.

a) Test the hypothesis that H_0: μ ≤ 22 versus H_a: μ > 22 at a significance level of 0.05.

$\bar{y}$ = 27.54, s = 5.471

$$t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}} = \frac{27.54 - 22}{5.471/\sqrt{5}} = 2.264$$

df = 4, $t^c = t_4(1 - 0.05)$ = 2.132 or p-value = P(T > 2.264) = 0.0431 (I got this from SAS, code follows)

2.264 ≥ 2.132 OR 0.041 ≤ 0.05 ==> reject H_0

The data strongly supports the claim (P = 0.0431) that the maximum weight of lift for the healthy males with ages from 18 – 30 is above 22 kg.

SAS input code:

```
*problem 4a;
data lift;
input MAWL @@;
datalines;
25.8 36.6 26.3 21.8 27.2
;
run;

proc print data=lift; run;

*H0: mu0, side=u ==> upper test, Ha: mu > mu0. If you wanted a lower test
     (Ha: mu < mu0), you would use side=l;
proc ttest data=lift alpha=0.05 H0=22 side=u;
   var MAWL;
run;
quit;
```

SAS Output

The TTEST Procedure
Variable: MAWL

N	Mean	Std Dev	Std Err	Minimum	Maximum
5	27.5400	5.4706	2.4465	21.8000	36.6000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
27.5400	22.3243	Infty	5.4706	3.2776	15.7202

DF	t Value	Pr > t
4	2.26	0.0431

In addition, there are diagnostic graphs.

b) Construct a 99% confidence interval for μ.

$$\bar{y} \pm t_c s\{\bar{y}\} = 27.54 \pm (4.604)\frac{5.471}{\sqrt{5}} = 27.54 \pm 11.2646 \Longrightarrow (16.28,38.80)$$

$t_c = t_4(1 - \frac{0.01}{2}) = 4.604$

SAS input

```
*alpha = 0.01 ==> 1 - alpha = 99;
proc ttest data=lift alpha=0.01;
   var MAWL;
run;
```

SAS Output

The TTEST Procedure
Variable: MAWL

N	Mean	Std Dev	Std Err	Minimum	Maximum
5	27.5400	5.4706	2.4465	21.8000	36.6000

Mean	99% CL Mean		Std Dev	99% CL Std Dev	
27.5400	16.2759	38.8041	5.4706	2.8383	24.0489

DF	t Value	Pr > |t|
4	11.26	0.0004

Plus the diagnostic graphs.

5. (A.7, A.9) The maximum lean angle (the furthest a subject is able to lean and still recover in one step) for a random sample of 10 younger females (21 – 29 years) has an average of μ_Y = 30.7 and a standard deviation of σ_Y = 2.751 and the data for a random sample of 5 older females (67 – 81 years) has an average of μ_O =16.2 and a standard deviation of σ_O = 4.438. Assume that the two sets of observations constitute independent random samples from normal populations.

a) Test the hypothesis that H_0: $\sigma_Y^2 = \sigma_O^2$ versus H_a: $\sigma_Y^2 \neq \sigma_O^2$ at a significance level of 0.10.

$$F = \frac{s_Y^2}{s_O^2} = \frac{2.751^2}{4.438^2} = 0.3842$$

df{numerator} = 10 – 1 = 9, df{denominator} = 5 – 1 = 4
F(1-0.10;4,9) = 3.94, since 0.3842 < 0.3.94 ==> we fail to reject H_0

The data does not strongly support the claim (P = 0.14846) that the maximum lean angle for younger females is different from the angle for older females.

SAS input

```
data Ftest;
Fs = 2.751**2/4.438**2;
dfn = 9; dfd = 4; alpha=0.05;
Pvalue=1-PROBF(abs(Fs),dfn,dfd)
run;

proc print data=Ftest; run;
```

SAS Output

Obs	Fs	dfn	dfd	alpha	Pvalue
1	0.38424	9	4	0.05	0.14846

b) Assuming a common variance, calculate a 90% confidence interval for $\mu_Y - \mu_O$.

$\bar{y}_Y - \bar{y}_O \pm t_c s\{\bar{y}_Y - \bar{y}_O\}$
We can assume a common variance since in part a) we showed that the variances were the same.

df = $n_y + n_O - 2 = 13$, $t_{13}(1 - \frac{0.1}{2}) = 1.771$

$$s^2 = \frac{(n_Y - 1)s_Y^2 + (n_O - 1)s_O^2}{n_Y + n_O - 2} = \frac{(9)2.751^2 + (4)4.438^2}{13} = 11.300$$

$$s\{\bar{y}_Y - \bar{y}_O\} = \sqrt{s^2\left(\frac{1}{n_Y} + \frac{1}{n_O}\right)} = \sqrt{(11.300)\left(\frac{1}{10} + \frac{1}{5}\right)} = 1.841$$

30.7 - 16.2 ± (1.771)(1.841) = 14.5 ± 3.261 ==> (11.2, 17.8)

Using method where we assume that the variances are not the same.

$$se_Y^2 = \frac{s_Y^2}{n_Y} = \frac{2.751^2}{10} = 0.7568, se_O^2 = \frac{s_O^2}{n_O} = \frac{4.438^2}{5} = 3.939$$

$$df = \frac{(se_Y^2 + se_O^2)^2}{\frac{se_Y^4}{n_Y - 1} + \frac{se_O^4}{n_O - 1}} = \frac{(0.7568 + 3.939)^2}{\frac{0.7568^2}{10-1} + \frac{3.939^2}{5-1}} = \frac{22.051}{3.943} = 5.59 \Longrightarrow 5$$

$t_5(1 - \frac{0.1}{2}) = 2.015$

$$s\{\bar{y}_Y - \bar{y}_O\} = \sqrt{\frac{s_Y^2}{n_Y} + \frac{s_O^2}{n_O}} = \sqrt{\frac{2.751^2}{10} + \frac{4.438^2}{5}} = 2.167$$

30.7 - 16.2 ± (2.015)(2.167) = 14.5 ± 4.367 ==> (10.1, 18.7)

To calculate the exact value for the critical value of t use the following:

SAS input

```
data tcrit;
*p = 1 - alpha/2, df = degrees of freedom;
alpha = 0.05;
p = 1 - alpha/2;
df = 5.59;
CritVal = TINV(p,df);

proc print data=tcrit; run;
```

SAS Output

Obs	alpha	p	df	CritVal
1	0.05	0.975	5.59	2.49108