Math 339: Exam #2

Eliot Stanton

1

a) $\bar{y} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$

$\bar{y} = 85.1 \quad s = 3.96 \quad n = 10$

$\alpha = .05 \quad |t_{.025}| = |t_{.975}| = 2.26$

$85.1 \pm 2.26\left(\frac{3.96}{\sqrt{10}}\right) = \boxed{[82.27, 87.93]}$

```
> #1a
>
> ff = c(88,82,82,85,78,91,83,85,88,89)
> ybar= mean(ff)
> ybar
[1] 85.1
> s= sd(ff)
> s
[1] 3.95671
> t= qt(.975,9)
> t
[1] 2.262157
> n=10
> ybar-t*(s/sqrt(n))
[1] 82.26954
> ybar+t*(s/sqrt(n))
[1] 87.93046
```

b) $\left[\frac{(n-1)s^2}{\chi^2_{1-(\alpha/2)}}, \frac{(n-1)s^2}{\chi^2_{\alpha/2}}\right]$

$s^2 = 15.66 \quad n = 10$

$\chi^2_{.975} = 19.02 \quad \chi^2_{.025} = 2.70$

$\left[\frac{9 \cdot 15.66}{19.02}, \frac{9 \cdot 15.66}{2.70}\right] = \boxed{[7.41, 52.18]}$

```
> #1b
>
> s2 = var(ff)
> s2
[1] 15.65556
> chi975 = qchisq(.975,9)
> chi975
[1] 19.02277
> chi025 = qchisq(.025,9)
> chi025
[1] 2.700389
> 9*s2/chi975
[1] 7.406914
> 9*s2/chi025
[1] 52.17766
```

c) In part a, I used "small-sample CI for a mean" methods, which involves the t-distribution. In part b, I used "CI for variance" methods which involves the chi-square distribution. For both methods, I must assume that the sample is randomly-selected from a Normal population - otherwise the distributions and methods would not be appropriate.

2

a) $(\mu_2 - \mu_1) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

$\alpha = .01 \quad |z_{.005}| = |z_{.995}| = 2.576$

$2 \pm 2.576 \cdot .640 = \boxed{[.35, 3.65]}$

```
> #2a
>
> mu2 = 12.75
> mu1 = 10.75
> mud = mu2-mu1
> z = qnorm(.995)
> z
[1] 2.575829
```

```
> s1 = 4
> s2 = 5
> n = 100
> se = sqrt((s1^2/n)+(s2^2/n))
> se
[1] 0.6403124
> mud-z*se
[1] 0.3506645
> mud+z*se
[1] 3.649336
```

b) $V = \dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(df_1 = n-1, df_2 = n_2 - 1)$ $\quad V = \dfrac{S_1^2}{S_2^2} \cdot \dfrac{\sigma_2^2}{\sigma_1^2}$

V is a pivotal quantity for $\dfrac{\sigma_2^2}{\sigma_1^2}$ because:

- V is a function of the data (used to calculate S_1^2 and S_2^2) and the unknown parameter $\dfrac{\sigma_2^2}{\sigma_1^2}$, where $\dfrac{\sigma_2^2}{\sigma_1^2}$ is the only unknown
- the probability distribution of V doesn't depend on $\dfrac{\sigma_2^2}{\sigma_1^2}$
  ↳ $F(n_1-1, n_2-1)$ depends only on the sample sizes

c) if $P(a \le V \le b) = 1-\alpha$, $a = F_{\alpha/2}$ and $b = F_{1-\alpha/2}$ because $V \sim F(n_1-1, n_2-1)$

$$1-\alpha = P\left(F_{\alpha/2} \le \frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2} \le F_{1-\alpha/2}\right) = P\left(\frac{S_2^2}{S_1^2} F_{\alpha/2} \le \frac{\sigma_2^2}{\sigma_1^2} \le \frac{S_2^2}{S_1^2} F_{1-\alpha/2}\right)$$

so, a $(1-\alpha) \times 100\%$ CI for $\dfrac{\sigma_2^2}{\sigma_1^2}$ is $\boxed{\left[\dfrac{S_2^2}{S_1^2} F_{\alpha/2}, \dfrac{S_2^2}{S_1^2} F_{1-\alpha/2}\right]}$

d) $\left[\dfrac{S_2^2}{S_1^2} F_{.005}, \dfrac{S_2^2}{S_1^2} F_{.995}\right]$

```
> #2d
>
> f005 = qf(.005,n-1,n-1)
> f995 = qf(.995,n-1,n-1)
> (s2^2/s1^2)*f005
[1] 0.9270997
> (s2^2/s1^2)*f995
[1] 2.63338
```

$F_{.005} = .593 \quad F_{.995} = 1.69$

$\left[\dfrac{25}{16} \times .593, \dfrac{25}{16} \times 1.69\right] = \boxed{[.93, 2.63]}$

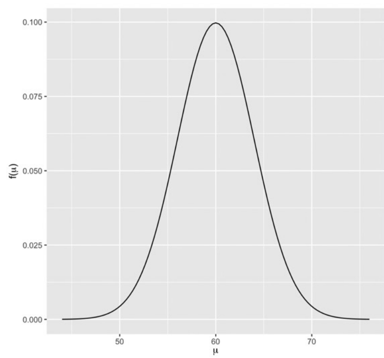
This 99% CI does not suggest that σ_1^2 and σ_2^2 are different. Because it's an interval for $\dfrac{\sigma_2^2}{\sigma_1^2}$, we check to see if the CI contains 1, since $\dfrac{\sigma_2^2}{\sigma_1^2}$ equals 1 if the two variances are equal. This interval contains 1 so it doesn't suggest different variances between the two groups

$\boxed{3}$

a) $\mu \sim \text{Normal}(\theta, \tau^2)$ $\quad \theta = 60$

I choose $\tau = 2$, so $\mu \sim \text{Normal}(60, 2^2)$.

This distribution is centered at 60 with nearly all area concentrated between 50 and 70. I chose this because we're fairly certain the temp will be around 60, so I wanted values 55-65 to be pretty likely with a small variance otherwise.



b) $n = 10 \quad Y_i | \mu \sim N(\mu, \sigma^2 = 5^2) \quad \bar{y} = 65$

$$\mu | \underset{\sim}{y} \sim N\left(\theta \frac{\sigma^2}{n\tau^2 + \sigma^2} + \bar{y}\frac{n\tau^2}{n\tau^2+\sigma^2}, \frac{\tau^2\sigma^2}{n\tau^2+\sigma^2}\right)$$

$\theta = 60 \quad \tau = 2 \quad \sigma = 5$

$$\mu | \underset{\sim}{y} \sim N\left(\frac{60 \times 5^2}{10 \times 2^2 + 5^2} + \frac{65 \times 10 \times 2^2}{10 \times 2^2 + 5^2}, \frac{2^2 \times 5^2}{10 \times 2^2 + 5^2}\right)$$

$$= N(63.08, 1.54)$$

$$(2.5^{th} \text{pctl}, 97.5^{th} \text{pctl}) = [60.65, 65.51]$$

```
> #3b
>
> n=10
> theta=60
> tau=2
> ybar=65
> sigma=5
> denominator=n*tau^2+sigma^2
> postmean = theta*sigma^2/denominator + ybar*n*tau^2/denominator
> postvar = (tau^2*sigma^2)/denominator
> postmean
[1] 63.07692
> postvar
[1] 1.538462
```

```
> qnorm(.025,postmean,sqrt(postvar))
[1] 60.64589
> qnorm(.975,postmean,sqrt(postvar))
[1] 65.50796
```

c) $\mu \sim N(60, 50^2) \quad Y_i | \mu \sim N(\mu, 5^2)$

$\tau = 50$

$$\mu | \underset{\sim}{y} \sim N\left(\frac{60 \times 5^2}{10 \times 50^2 + 5^2} + \frac{65 \times 10 \times 50^2}{10 \times 50^2 + 5^2}, \frac{50^2 \times 5^2}{10 \times 50^2 + 5^2}\right)$$

$$= N(65.00, 2.50)$$

$$(2.5^{th} \text{pctl}, 97.5^{th} \text{pctl}) = [61.90, 68.09]$$

```
> #3c
>
> tau=50
> denominator=n*tau^2+sigma^2
> postmean = theta*sigma^2/denominator + ybar*n*tau^2/denominator
> postvar = (tau^2*sigma^2)/denominator
> postmean
[1] 64.995
> postvar
[1] 2.497502
> qnorm(.025,postmean,sqrt(postvar))
[1] 61.89758
> qnorm(.975,postmean,sqrt(postvar))
[1] 68.09243
```

The interval is wider because the prior is less informative - it has a much higher variance, 2500 instead of 4. We know less about what values μ might take, so we need a wider interval in order to ensure we still have a 95% chance of μ being in it.

4

$\hat{\theta}_1 = \min(Y)$ $\quad n=16$

$\hat{\theta}_2 = \bar{Y} - 2.5 \times s$

a) $\hat{\theta}_1$ estimates the minimum number of chips in a bag, for the whole population.

$\hat{\theta}_2$ estimates the number of chips in a bag from the 5^{th} percentile

(well, technically it's closer to estimating the 6^{th} percentile)

b) w/ $\hat{\theta}_1$: [1087, 1135]

w/ $\hat{\theta}_2$: [944.12, 1088.44]

```
> #4b
>
> chips_ahoy = c(1219, 1214, 1087, 1200, 1419, 1121, 1325, 1345,
+                1244, 1258, 1356, 1132, 1191, 1270, 1295, 1135)
>
> set.seed(339) # Run this for reproducibility!
> boot_dist1 = replicate(10000, {
+   boot_samp1 = sample(chips_ahoy,16,replace=TRUE)
+   min(boot_samp1)
+ })
>
> quantile(boot_dist1,
+          probs=c(.025,.975))
 2.5% 97.5%
 1087  1135
>
> boot_dist2 = replicate(10000, {
+   boot_samp1 = sample(chips_ahoy,16,replace=TRUE)
+   mean(boot_samp1)-2.5*sd(boot_samp1)
+ })
>
> quantile(boot_dist2,
+          probs=c(.025,.975))
     2.5%     97.5%
 944.1202 1088.4406
```

c) The findings do not support Nabisco's claim. While the interval built with $\hat{\theta}_1$ has a lower bound of 1087, above 1000 like they claim, it's a very bad estimator.

$\hat{\theta}_1$, the sample minimum, can only take on values in our original sample, since the bootstrap procedure only samples values from the original sample. The sample minimum is, unsurprisingly, 1087, so this first interval was bound to imply that the population minimum is greater than 1000.

$\hat{\theta}_2$, on the other hand, is more resistant to outliers and thus better in a small sample. Plus, it's a continuous variable (whereas the minimum only discretely takes on certain separate values), and it uses mean and standard deviation to consider what the 5^{th}-6^{th} percentile would be in a population that can include values lower than sampled.

Because the interval built with $\hat{\theta}_2$ has a lower bound of 944.12 chips, we do not have evidence that Nabisco's claim is true.