STAT 509 HW 6 Solution

1. (a) 95% confidence interval for μ: $\bar{Y} \pm t_{49, 0.975} \frac{s}{\sqrt{50}}$

$$738.66 \pm 2.009575 \times 155.80/\sqrt{50}$$

$$[694.3821, \quad 782.9379]$$

We are 95% confident that the population mean lifetime is between 694.3821 and 782.9379 hours.

Yes, the customer should go ahead with the purchase arrangment because the advertised mean lifetime: 750 hours is within the 95% confidence interval.

(b) ① Normality assumption for population distribution can be checked by qq-plot. qq plot shows linear pattern in general, except a mild violation on the right tail. However, the t-distribution is robust to normality, which means even though the population distributio is mildly non-normal, the confidence interval can still be used to estimate the population mean μ.

② Data is collected from a random Sample.

(c) 95% confidence interval for σ: $\left(\sqrt{\frac{(n-1)s^2}{\chi^2_{49, 0.975}}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{49, 0.025}}}\right)$

$$\left(\sqrt{\frac{49 \times 155.80^2}{70.22241}}, \sqrt{\frac{49 \times 155.80^2}{31.55492}}\right)$$

$$(130.145, \quad 194.1476)$$

We are 95% confident that the population standard deviation is between 130.145 and 194.1476 hours.

Assumptions: ① Data is collected from a random sample.
② Normality assumption for population distribution

NO, I am not concerned. We used a random sample and qq-plot shows linear pattern in general.

2. see attached.

3. (a) population: All pregnant women.

(b) 95% confidence interval for p:

$$\hat{p} \pm Z_{0.975}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\frac{125}{999} \pm 1.959964\sqrt{\frac{\frac{125}{999}(1-\frac{125}{999})}{999}}$$

$$(0.1046083,\ 0.145642)$$

We are 95% confident that the population proportion of pregnant women who smoked during the last 3 months of pregnancy is between 10.46083% and 14.5642%.

Rule of thumb: $n\hat{p} = 999 \times \frac{125}{999} = 125 \geq 5$

$n(1-\hat{p}) = 999 \times (1-\frac{125}{999}) = 874 \geq 5$

The confidence interval is valid.

(c) Margin of error: $B = Z_{1-\alpha/2}\sqrt{\frac{P_0(1-P_0)}{n}} = 0.01$

Let's use $P_0 = \hat{p} = \frac{125}{999}$

$$Z_{0.995}\sqrt{\frac{\frac{125}{999}(1-\frac{125}{999})}{n}} = 0.01$$

$$2.575829 \times \sqrt{\frac{\frac{125}{999}(1-\frac{125}{999})}{n}} = 0.01$$

$$n = \left(\frac{2.575829}{0.01}\right)^2 \times \frac{125}{999}(1-\frac{125}{999})$$

$$= 7263.142$$

We need to sample 7263 pregnant women to meet these requirements.

4. (a) $1-\alpha = P\left(-t_{n-1,\alpha/2} < \dfrac{Y^* - \bar{Y}}{S\sqrt{1+1/n}} < t_{n-1,\alpha/2}\right)$

$1-\alpha = P\left(-t_{n-1,\alpha/2}\cdot S\sqrt{1+1/n} < Y^* - \bar{Y} < t_{n-1,\alpha/2}\cdot S\sqrt{1+1/n}\right)$

$1-\alpha = P\left(\bar{Y} - t_{n-1,\alpha/2}\cdot S\sqrt{1+1/n} < Y^* < \bar{Y} + t_{n-1,\alpha/2}\cdot S\sqrt{1+1/n}\right)$

Thus $\bar{Y} \pm t_{n-1,\alpha/2}\, S\sqrt{1+\frac{1}{n}}$ is a $100(1-\alpha)$ percent prediction interval for the new observation Y^*.

(b) $\bar{Y} \pm t_{n-1,\alpha/2}\, S\sqrt{1+\frac{1}{n}}$

$738.66 \pm 2.009575 \times 155.80 \cdot \sqrt{1+\frac{1}{50}}$

$(422.4528, \quad 1054.867)$

We are 95% confident that, the life time of the next light bulb is between 422.4528 and 1054.867 hours.

(c) The above interval is wider because not only do we need to account for the variation σ^2/n in estimating the population mean μ, but we additionally need to account for the variation σ^2 attached to the new value Y^*.

Homework 6—R code attachment

Problem 1

```
# Problem 1(a)
# Get upper 0.025 quantile from t(49) distribution
> qt(0.975,49)
[1] 2.009575
```

```
# Problem 1(c)
# Get upper/lower 0.025 quantile from chi-square(49) distribution
> qchisq(0.975,49) # upper
[1] 70.22241
> qchisq(0.025,49) # lower
[1] 31.55492
```

Problem 2

```
# Problem 2(a)
# Enter data
impression.data.mm = c(22.4,23.6,24.0,24.9,25.5,25.6,25.8,26.1,26.4,
       26.7,27.4,27.6,28.3,29.0,29.1,29.6,29.7,29.8,29.9,30.0,30.4,30.5,
       30.7,30.7,31.0,31.0,31.4,31.6,31.7,31.9,31.9,32.0,32.1,32.4,32.5,
       32.5,32.6,32.9,33.1,33.3,33.5,33.5,33.5,33.5,33.6,33.6,33.8,33.9,
       34.1,34.2,34.6,34.6,35.0,35.2,35.2,35.4,35.4,35.4,35.5,35.7,35.8,
       36.0,36.0,36.0,36.1,36.1,36.2,36.4,36.6,37.0,37.4,37.5,37.5,38.0,
       38.7,38.8,39.8,41.0,42.0,42.1,44.6,48.3,55.0)
```

We use the t.test function to get the confidence interval directly (for the population mean):

```
> t.test(impression.data.mm,conf.level=0.95)$conf.int
[1] 32.21951 34.52025
attr(,"conf.level")
[1] 0.95
```

Interpretation: We are 95 percent confident that the population mean deepest impression μ is between 32.2 and 34.5 mm.

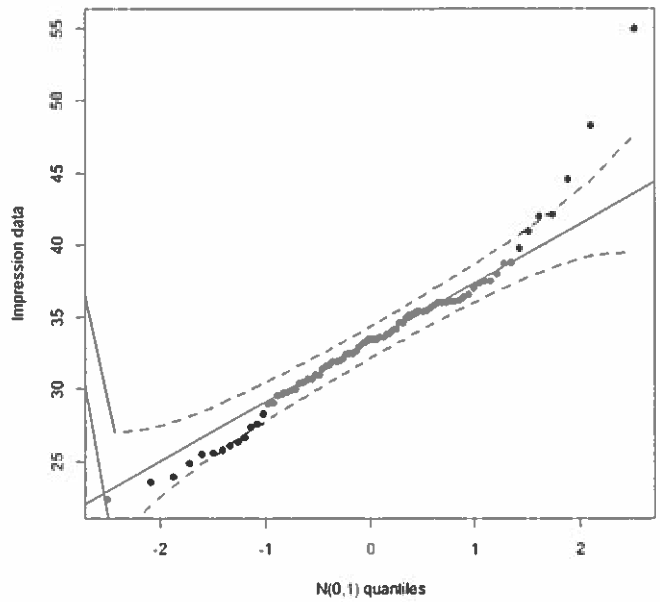
Note: Consider the collection of all times this projectile would be fired at this clay armor model. The population mean μ is the mean of all deepest impressions that would be recorded.

Assumptions: This interval requires that

- The sample of impressions is a random sample
- The impression measurements are normally distributed.

We would have to assume that this sample was a random sample (hopefully true!). To check the normality assumption, we can look at a normal qq-plot for the data (see next page).

Normality looks pretty reasonable here. Again, there may be some very slight departures, but nothing to be too concerned about. Remember, confidence intervals for population means (based on the t distribution) are robust to slight normality departures.



(b) Your co-worker is misinterpreting what the interval in part (a) means. A confidence interval for a population mean is just that—**it says where we think the population mean falls.** It does not say anything else, and it certainly doesn't say anything about where individual measurements fall.

In short, your co-worker either doesn't understand what a confidence interval is, or s/he is extremely sloppy in his/her interpretation (and ultimately grossly incorrect).

(c) Recall that I wrote an R function to write a confidence interval for the population variance (see R code web site).

```
# I wrote this function to do it
var.interval = function(data,conf.level=0.95){
      df = length(data)-1
      chi.lower = qchisq((1-conf.level)/2,df)
      chi.upper = qchisq((1+conf.level)/2,df)
      s2 = var(data)
      c(df*s2/chi.upper,df*s2/chi.lower)
      }

# CI for population variance
> var.interval(impression.data.mm)
[1] 20.89197 38.67669
```

Interpretation: We are 95 percent confident that the population variance deepest impression σ^2 is between 20.89 and 38.68 $(mm)^2$.

To get a confidence interval for the population standard deviation σ, simply take square roots:

```
> sqrt(var.interval(impression.data.mm))
[1] 4.570773 6.219058
```

Interpretation: We are 95 percent confident that the population standard deviation deepest impression σ is between 4.6 and 6.2 mm.

Assumptions: This interval requires that

- The sample of impressions is a random sample
- The impression measurements are normally distributed.

The normality assumption (based on the qq-plot in part (a)), looks pretty reasonable here. However, if we are concerned about the normal assumption, we should be very cautious when interpreting confidence intervals for σ^2 or σ. Confidence intervals for variances (or standard deviations) are not robust to normality departures.

Problem 3

```
# Problem 3(b)
# Get upper 0.025 quantile from standard normal distribution
> qnorm(0.975,0,1)
[1] 1.959964

# Problem 3(c)
# Get upper 0.005 quantile from standard normal distribution
> qnorm(0.995,0,1)
[1] 2.575829
```