Problem 1

Answer: A bootstrap confidence interval constructed based on a biased sample will still yield an unbiased estimate for the population parameter of interest.

Problem 2

Answer:
▷ Version 1: Each observation in one data set is subtracted from the average of the other data set's observations.
▷ Version 2: The data can't be considered paired data because the days for which we have Intel data may be different from the days for which we have Southwest Airlines data.

Problem 3

Answer: Sampling distribution

Problem 4

Answer:
▷ Version 1: The study's authors are 95% confident that babies born to non-smoking mothers are on average 0.2 to 0.9 pounds heavier than babies born to smoking mothers.
▷ Version 2: Since 0 is apparently an unusual value for the statistic, then at the 5% significance level we would fail to reject a null hypothesis that claims that the fathers' and mothers' average IQs are equal.

Problem 5

Answer:
▷ Version 1: Because the standard error estimate may not be accurate.
▷ Version 2: A confidence interval based on this sample is not accurate since the sample size is small.

Problem 6

Answer: It becomes more normal looking.

Problem 7

Answer: Tom has a one-sided alternative hypothesis and should do a paired t-test.

Problem 8

Answer:
▷ Version 1: Between 0.02 and 0.05.
▷ Version 2: $T < -1.71$

```
# Version 1
n <- 26
t <- 2.485
p_value <- 2 * pt(t, n - 1, lower.tail = F)
print(p_value)
```

```
0.0200048010978179
```

```
# Version 2
print_p_value <- function(t, n) {
  p_value <- pt(t, n - 1)
  cat("T =", t, "--- p-value =", p_value, "\n")
}

n <- 25
ts <- c(-1.71, -1.32, 1.32, 1.71, 1.96)
for (t in ts) {
  print_p_value(t, n)
}
```

```
T = -1.71 --- p-value = 0.05008269
T = -1.32 --- p-value = 0.09964372
T = 1.32 --- p-value = 0.9003563
T = 1.71 --- p-value = 0.9499173
T = 1.96 --- p-value = 0.969147
```

Problem 9

Answer: F-test (ANOVA)

Problem 10

Answer:
▷ Version 1: There should be at least 10 successes and 10 failures.
▷ Version 2: Side-by-side box plots showing roughly equally sized boxes for each group.

Problem 11

Answer: 1.87

```
p <- 1 - 0.0767
df1 <- 7
df2 <- 189
f_stat <- round(qf(p, df1, df2), digits = 2)
print(f_stat)
```

```
1.87
```

Problem 12

Answer:
▷ Version 1: At least two group means are significantly different from each other.
▷ Version 2:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$
$$H_A : \text{at least one } \mu_i \text{ is different}$$

Problem 13

Answer:
▷ Version 1: Decrease
▷ Version 2: 1.023

```
# Version 2
n1 <- 18
s1 <- 3.4

n2 <- 18
s2 <- 2.7

sp <- sqrt(((n1 - 1) * s1^2 + (n2 - 1) * s2^2) / (n1 + n2 - 2))
std_err <- sp * sqrt(1 / n1 + 1 / n2)
signif(std_err, digits = 4)
```

```
[1] 1.023
```

Problem 14

Answer:
▷ Version 1: $\alpha' = 0.005$
▷ Version 2: $K = 10$

The adjusted significance level is given by

$$\alpha' = \frac{\alpha}{K}, \tag{1}$$

where K can be obtained from the number of groups $k = 5$ as follows:

$$K = \frac{k(k-1)}{2} = \frac{5(5-1)}{2} = 10. \tag{2}$$

Using this result and $\alpha = 0.05$, we get

$$\alpha' = \frac{0.05}{10} = 0.005. \tag{3}$$