NAME →

ISyE 6739 — Test 3 Solutions — Summer 2016

This is a take-home test. But please limit the total work time to less than about 4hours.

1. Let X be the outcome of a 4-sided die toss. Find $\mathsf{Var}(\ell\text{n}(X)+1)$.

   Solution: By the law of the Unconscious Statistician,

   $$\mathsf{E}[\ell\text{n}(X)] = \sum_{i=1}^{4} \ell\text{n}(i)\mathsf{P}(X=i) \;=\; \frac{1}{4}\sum_{i=1}^{4}\ell\text{n}(i) \;=\; 0.7945.$$

   Similarly,

   $$\mathsf{E}[\ell\text{n}^2(X)] \;=\; \sum_{i=1}^{4} \ell\text{n}^2(i)\mathsf{P}(X=i) \;=\; 0.9023.$$

   Then

   $$\mathsf{Var}(\ell\text{n}(X)+1) \;=\; \mathsf{Var}(\ell\text{n}(X)) \;=\; \mathsf{E}[\ell\text{n}^2(X)] - (\mathsf{E}[\ell\text{n}(X)])^2 \;=\; 0.2711. \quad \diamondsuit$$

2. Suppose that the lifetime of a transistor is exponential with a mean of 10 years. Further suppose that the transistor has already survived 20 years. Find the probability that it will not fail in the next 10 years.

   Solution: By the memoryless property,

   $$\mathsf{P}(X \ge 30 | X \ge 20) \;=\; \mathsf{P}(X \ge 10) \;=\; e^{-\lambda x} \;=\; e^{-1} \;=\; 0.3679. \quad \diamondsuit$$

3. Suppose that $X \sim c + \text{Exp}(\lambda)$, where the time units are in years. X could represent the lifetime of a lightbulb that is *guaranteed* to last at least c years. Find the median of X, that is, the point m such that $\mathsf{P}(X \le m) = \mathsf{P}(X > m) = 0.5$.

   Solution: Let $Y \sim \text{Exp}(\lambda)$. Then

   $$0.5 \;=\; \mathsf{P}(X \le m) \;=\; \mathsf{P}(c+Y \le m) \;=\; \mathsf{P}(Y \le m-c) \;=\; 1 - e^{-\lambda(m-c)}.$$

This is true iff

$$-\lambda(m-c) \;=\; \ell\text{n}(0.5) \quad \text{iff} \quad m \;=\; c+\frac{\ell\text{n}(2)}{\lambda}. \quad \diamondsuit$$

4. Suppose X_1, X_2, and X_3 are i.i.d. Bernoulli(p) random variables, which represent the functionality of three network components. Think of a signal passing through a network, where $X_i = 1$ if the signal can successfully get through component i, for $i = 1, 2, 3$ (and $X_i = 0$ if the signal is unsuccessful). Let's consider two set-ups:

   (*) X_1, X_2, X_3 have $p = 0.95$ and are hooked up in *series* so that a signal getting through the network has to pass through all components 1 AND 2 AND 3.

   (**) X_1, X_2, X_3 have $p = 0.8$ and are hooked up in *parallel* so that a signal getting through the network only has to pass through components 1 OR 2 OR 3.

   Which series is more reliable, i.e., more likely to permit a signal to pass through — (*) or (**)?

   Solution:

$$\mathsf{P}(*) \;=\; \mathsf{P}\left(\bigcap_{i=1}^{3}(X_i=1)\right) \;=\; \prod_{i=1}^{3}\mathsf{P}(X_i=1) \;=\; 0.8574.$$

   Meanwhile, by inclusion-exclusion and the fact that the X_i's are i.i.d., we have

$$\begin{aligned}
\mathsf{P}(**) &= \mathsf{P}\left(\bigcup_{i=1}^{3}(X_i=1)\right) \\
&= \mathsf{P}(X_1=1)+\mathsf{P}(X_2=1)+\mathsf{P}(X_3=1) \\
&\quad -\mathsf{P}(X_1=1\cap X_2=1)-\mathsf{P}(X_1=1\cap X_3=1)-\mathsf{P}(X_2=1\cap X_3=1) \\
&\quad +\mathsf{P}(X_1=1\cap X_2=1\cap X_3=1) \\
&= 3\mathsf{P}(X_1=1)-3\mathsf{P}^2(X_1=1)+\mathsf{P}^3(X_1=1) \\
&= 3p-3p^2+p^3 \;=\; 0.992.
\end{aligned}$$

   Or, now that I think of it, we could've gotten this result way more quickly by

$$\mathsf{P}(**) \;=\; \mathsf{P}\left(\bigcup_{i=1}^{3}(X_i=1)\right) \;=\; 1-\mathsf{P}\left(\bigcap_{i=1}^{3}(X_i=0)\right) \;=\; 1-(1-p)^3 \;=\; 0.992.$$

   In any case, (**) is more reliable. $\diamondsuit$

5. TRUE or FALSE? If X is *any* normal distribution, then about 95% of all observations from X will fall within two standard deviations of the mean.

   Solution: TRUE. $\diamondsuit$

6. TRUE or FALSE? The normal quantile value $\Phi^{-1}(0.975) = 1.96$.

   Solution: TRUE. $\diamondsuit$

7. Suppose $X \sim \mathsf{Nor}(1, 1)$. Find c such that $\mathsf{P}(-c \le X \le c) = 0.95$. Careful — this may require a little trial-and-error.

   Solution: Standardizing with $Z \sim \mathrm{Nor}(0,1)$, we have

$$0.95 \;=\; \mathsf{P}(-c \le X \le c) \;=\; \mathsf{P}(-c-1 \le Z \le c-1) \;=\; \Phi(c-1) - \Phi(-c-1).$$

   Now we have to find the value of c that satisfies this equation. I'll go to Excel and use the `NORM.S.DIST` function to evaluate $\Phi(c-1) - \Phi(-c-1)$ for various values of c. After a little trial and error, I find that $c = 2.65$ pretty much does the trick. $\diamondsuit$

8. Suppose that X and Y are the scores that a Georgia Tech student and his twin who goes to the Univ. of Georgia will receive, respectively, on the same math test. Further suppose that X is $\mathsf{Nor}(90, 100)$, Y is $\mathsf{Nor}(60, 100)$ and $\mathsf{Cov}(X,Y) = 50$. Find the probability that the GT kid will beat the UGA kid by at least 40 points. (You can assume that $X - Y$ is normal.)

   Solution: Note that $\mathsf{E}[X - Y] = 30$ and

$$\mathsf{Var}(X-Y) = \mathsf{Var}(X) + \mathsf{Var}(Y) - 2\mathsf{Cov}(X,Y) = 100.$$

   Therefore, $X - Y \sim N(30, 100)$. This implies that

$$\mathsf{P}(X - Y > 40) = \mathsf{P}\left(Z > \frac{40-30}{\sqrt{100}}\right) = \mathsf{P}(Z > 1) = 0.1587. \quad \diamondsuit$$

9. If $X_1, \ldots, X_{400}$ are i.i.d. from some distribution with mean 0 and variance 1600, find the approximate probability that the sample mean $\bar{X}$ is between -2 and 2.

   Solution: By the Central Limit Theorem, $\bar{X}$ will be approximately normal. Let's find its mean and variance. To this end,

$$\mathsf{E}[\bar{X}] = \mathsf{E}[X_i] = 0$$

   and

$$\mathsf{Var}(\bar{X}) = \frac{\mathsf{Var}(X_i)}{n} = \frac{1600}{400} = 4.$$

   Thus, $\bar{X} \approx \mathsf{Nor}(0, 4)$, and so

$$\mathsf{P}(-2 \leq \bar{X} \leq 2) \approx \mathsf{P}(-1 \leq Z \leq 1) = 2\Phi(1) - 1 = 2(0.8413) - 1 = 0.6826. \quad \diamondsuit$$

10. TRUE or FALSE? Consider *any* i.i.d. sequence of random variables having finite variance. Then the Central Limit Theorem says that a properly standardized sample mean can be approximated by a standard normal random variable as the sample size becomes large.

    Solution: TRUE. $\diamondsuit$

11. Short Research Question: Go to The Internets and write up a couple of paragraphs on who "invented" the Central Limit Theorem (there are several reasonably correct answers).

12. If $X \sim \chi^2(5)$, find $\mathsf{P}(X < 11.07)$.

    Solution: 0.95. $\diamondsuit$

13. TRUE or FALSE? $t_{0.025,8} > z_{0.025}$.

    Solution: TRUE, since $t_{0.025,8} = 2.306 > 1.96 = z_{0.025}$. In fact, since the t-distribution has fatter tails than the standard normal, you don't even need tables for this one. $\diamondsuit$

14. Suppose $T \sim t(343)$. What's $\mathsf{P}(T < 1)$?

    Solution: Because of the high degrees of freedom, $\mathsf{P}(T < 1) \approx \mathsf{P}(Z < 1) = 0.8413$. $\diamondsuit$

15. TRUE or FALSE? $\mathsf{P}(F(5,3) < F_{0.95,5,3}) = \mathsf{P}(F(5,3) < 1/F_{0.05,5,3})$.

    Solution: FALSE, since $F_{0.95,5,3} = 1/F_{0.05,3,5}$ (flip the d.f.). $\diamondsuit$

16. Suppose $X_1, \ldots, X_6 \sim \mathsf{Nor}(3, 9)$, $Y_1, \ldots, Y_7 \sim \mathsf{Nor}(-3, 2)$, and everything is independent. Let S_X^2 and S_Y^2 denote the sample variances of the X_i's and Y_j's, respectively. Name the distribution (with parameter(s)) of S_Y^2/S_X^2.

    Solution: $S_X^2 \sim \sigma_X^2 \chi^2(5)/5$ and $S_Y^2 \sim \sigma_Y^2 \chi^2(6)/6$, where $\sigma_X^2 = \mathsf{Var}(X_i) = 9$ and $\sigma_Y^2 = \mathsf{Var}(Y_j) = 2$. Therefore,

    $$\frac{S_Y^2}{S_X^2} \sim \frac{2\chi^2(6)/6}{9\chi^2(5)/5} \sim \frac{2}{9} F(6, 5). \quad \diamondsuit$$

17. Suppose $X_1, \ldots, X_n$ are i.i.d. $\text{Exp}(\lambda)$.

    (a) TRUE or FALSE? The sample mean $\bar{X}$ is unbiased for the mean $1/\lambda$.

    Solution: TRUE. (The sample mean is always unbiased for the true mean.) $\diamondsuit$

    (b) TRUE or FALSE? $1/\bar{X}$ is unbiased for λ.

    Solution: FALSE. (It's almost unbiased, but not quite; in fact, see the next question.) $\diamondsuit$

    (c) Find $\mathsf{E}[1/\bar{X}]$. (This might take a little work if you do it from scratch.)

    Solution: Let's first look at $Y = \sum_{i=1}^{n} X_i$ (so that $\bar{X} = Y/n$). Since Y is the sum of i.i.d. $\text{Exp}(\lambda)$ random variables, we know that $Y \sim \text{Erlang}_n(\lambda)$, and so

Y has p.d.f.

$$f_Y(y) = \frac{\lambda^n y^{n-1} e^{-\lambda y}}{(n-1)!}, \quad y > 0,$$

Then by LOTUS,

$$\begin{aligned}
\mathsf{E}[1/\bar{X}] &= n\mathsf{E}[1/Y] \\
&= n \int_0^\infty \frac{1}{y} \frac{\lambda^n y^{n-1} e^{-\lambda x}}{(n-1)!}\, dy \\
&= \frac{n\lambda}{n-1} \int_0^\infty \frac{\lambda^{n-1} y^{n-2} e^{-\lambda x}}{(n-2)!}\, dy \\
&= \frac{\lambda n}{n-1},
\end{aligned}$$

since the thing inside the integral is the p.d.f. of the $\text{Erlang}_{n-1}(\lambda)$ distribution. ♢

(d) TRUE or FALSE? $1/\bar{X}$ is the MLE for λ.

Solution: TRUE. ($\bar{X}$ is the MLE for the mean $1/\lambda$, so the result follows by invariance of MLE's.) ♢

18. Suppose X_1, X_2, X_3 are i.i.d. $\mathsf{Nor}(\mu, \sigma^2)$, and we observe $X_1 = 7$, $X_2 = 1$, and $X_3 = 4$.

(a) What is the sample variance of the X_i's?

Solution:

$$S^2 = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1} = 9. \quad ♢$$

(b) What is the maximum likelihood estimate of σ^2?

Solution: $\bar{X} = 4$ and $n = 3$. Thus,

$$\widehat{\sigma^2} = \frac{n-1}{n} S^2 = 6. \quad ♢$$

(c) What is the maximum likelihood estimate of $\mathsf{P}(X_i > 5)$?

Solution: Since

$$\mathsf{P}(X_i > 5) \;=\; \mathsf{P}\left(Z > \frac{5-\mu}{\sigma}\right) \;=\; 1 - \Phi\left(\frac{5-\mu}{\sigma}\right) \;=\; \Phi\left(\frac{\mu-5}{\sigma}\right),$$

invariance gives us

$$\widehat{\mathsf{P}}(X_i > 5) \;=\; \Phi\left(\frac{\hat{\mu}-5}{\hat{\sigma}}\right) \;=\; \Phi\left(\frac{\bar{X}-5}{\sqrt{\frac{(n-1)S^2}{n}}}\right) \;=\; \Phi\left(\frac{4-5}{\sqrt{6}}\right) \;=\; 0.3415. \quad \diamondsuit$$

19. Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. Geom(p). Thus, for all i, we have $\mathsf{P}(X_i = k) = (1-p)^{k-1}p$, for $k = 1, 2, \ldots, n$. What is the maximum likelihood estimate of p?

Solution: $\hat{p} = 1/\bar{X}$. $\diamondsuit$

In case you missed it, here's the proof. The likelihood function is

$$L(p) \;=\; \prod_{i=1}^{n} \mathsf{P}(X_i = x_i) \;=\; (1-p)^{\sum_{i=1}^{n} x_i - n} p^n,$$

so that

$$\ln(L(p)) \;=\; \left(\sum_{i=1}^{n} x_i - n\right) \ln(1-p) + n \ln(p).$$

Thus, setting

$$\frac{d}{dp}\ln(L(p)) \;=\; -\frac{(\sum_{i=1}^{n} x_i - n)}{1-p} + \frac{n}{p} \;=\; 0,$$

we have (after a little algebra) $\hat{p} = 1/\bar{X}$, as promised. $\diamondsuit$

20. Suppose X_1, X_2, X_3 are i.i.d. Unif$(\theta, 0)$, and we observe $X_1 = -7$, $X_2 = -1$, and $X_3 = -4$. What is the maximum likelihood estimate of θ?

Solution: By symmetry from an example in class, we have $\hat{\theta} = \min_i X_i = -7$. $\diamondsuit$

21. Suppose X_1, X_2, X_3 are i.i.d. $\mathsf{Nor}(\mu, 16)$. Define two estimators for μ: $T_1 \equiv X_1 - X_2 + X_3$ and $T_2 \equiv (4X_1 + 3X_2 + X_3)/8$. Which of T_1 or T_2 has the smaller MSE?

    Solution: It is easy to show that $\mathsf{E}[T_1] = \mathsf{E}[T_2] = \mu$, so that both estimators are unbiased for μ. Thus, in this case, $\mathsf{MSE}(T_1) = \mathsf{Var}(T_1)$ and $\mathsf{MSE}(T_2) = \mathsf{Var}(T_2)$.

    First, $\mathsf{Var}(T_1) = \mathsf{Var}(X_1 - X_2 + X_3) = \mathsf{Var}(X_1) + \mathsf{Var}(X_2) + \mathsf{Var}(X_3) = 48$.

    Similarly,

    $$\mathsf{Var}(T_2) = \mathsf{Var}\left(\frac{4X_1 + 3X_2 + X_3}{8}\right) = \frac{16\mathsf{Var}(X_1) + 9\mathsf{Var}(X_2) + \mathsf{Var}(X_3)}{64} = 6.5.$$

    Therefore, T_2 has lower MSE. $\diamondsuit$.

22. Which family member is actually an estimation method?

    (a) CAT
    (b) DAD
    (c) MOM
    (d) BRO
    (e) SIS

    Solution: (c) MOM (method of moments). $\diamondsuit$

23. Suppose $X_1, \ldots, X_{10}$ are i.i.d. normal with unknown mean and *known* variance $\sigma^2 = 49$. Further suppose that $\bar{X} = -50$ and $S^2 = 60$.

    (a) Find a 99% two-sided confidence interval for μ.

    Solution: You can ignore the red herring information about S^2 (since we know that $\sigma^2 = 49$).

    $$\begin{aligned}\mu \in \bar{X} \pm z_{\alpha/2}\sqrt{\sigma^2/n} &= -50 \pm 2.576\sqrt{49/10} \\ &= -50 \pm 5.70 \\ &= [-55.70, -44.30]. \quad \diamondsuit\end{aligned}$$

(b) Find a 99% two-sided confidence interval for $2\mu - 4$.

Solution: For any confidence interval for the mean μ with lower and upper bounds L and U, we have

$$1 - \alpha = \mathsf{P}(L \le \mu \le U) = \mathsf{P}(2L - 4 \le 2\mu - 4 \le 2U - 4).$$

Using the L and U bounds from Question 23a, we have

$$2\mu - 4 \in [-115.40,\ -92.60]. \quad \diamondsuit$$

24. Suppose $X_1, \ldots, X_{10}$ are i.i.d. normal with unknown mean and *unknown* variance σ^2. Further suppose that $\bar{X} = -50$ and $S^2 = 60$.

(a) Find a 99% two-sided confidence interval for μ.

Solution: Since $t_{\alpha/2,n-1} = t_{0.005,9} = 3.250$, we have

$$\begin{aligned} \mu \in \bar{X} \pm t_{\alpha/2,n-1}\sqrt{S^2/n} &= -50 \pm 3.250\sqrt{60/10} \\ &= -50 \pm 7.96 \\ &= [-57.96,\ -42.04]. \quad \diamondsuit \end{aligned}$$

(b) Is your answer to Question 24a wider or narrower than your answer to Question 23a? Why?

Solution: They're wider. First of all, the sample variance S^2 happens to be larger than the actual variance σ^2 (by bad luck). Second, we pay a penalty for not knowing the variance by having to use the t quantile instead of the z quantile. $\diamondsuit$

(c) Find a 99% two-sided confidence interval for σ^2.

Solution:

$$\sigma^2 \in \left[\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}},\ \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right] = \left[\frac{540}{23.59},\ \frac{540}{1.735}\right] = [22.89,\ 311.24]. \quad \diamondsuit$$

25. Suppose that $X_1, \ldots, X_n$ are i.i.d. Bernoulli with unknown mean p, and that we have carried out a preliminary investigation suggesting $p \approx 0.9$. How big would n have to be in order for a two-sided 99% confidence interval to have a half-length of 0.01? (Give the smallest such number.)

Solution: $z^2_{\alpha/2}\dfrac{\hat{p}(1-\hat{p})}{n} \le \varepsilon^2$. So,

$$n \;\ge\; z^2_{\alpha/2}\frac{\hat{p}(1-\hat{p})}{\varepsilon^2} \;=\; \frac{2.576^2 \cdot 0.9 \cdot 0.1}{(0.01)^2} \;=\; 5973 \quad \text{(rounded up)}. \quad \diamondsuit$$

26. After collecting your set of observations, what happens to the length of a confidence interval for the mean as the confidence level moves from 90% to 95%?

(a) It increases.
(b) It decreases.
(c) It stays the same.
(d) You can't tell.

Solution: (a) It increases. $\diamondsuit$

27. Consider i.i.d. normal observations $X_1, \ldots, X_{10}$ with unknown mean μ and unknown variance σ^2. What is the *expected width* of the usual 95% two-sided confidence interval for σ^2? You can keep your answer in terms of σ.

Solution: The confidence interval is of the form

$$\sigma^2 \in \left[\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right].$$

Thus, the length is

$$L \;=\; \left[\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}} - \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}}\right],$$

and so the expected length is

$$\mathsf{E}[L] \;=\; (n-1)\left[\frac{1}{\chi^2_{0.975,9}} - \frac{1}{\chi^2_{0.025,9}}\right]\cdot \mathsf{E}[S^2] \;=\; 9\left[\frac{1}{2.700} - \frac{1}{19.02}\right]\sigma^2 \;=\; 2.86\sigma^2. \quad \diamondsuit$$

28. Suppose we conduct an experiment to test to see if people can throw farther right- or left-handed. We get 20 people to do the experiment. Each throws a ball right-handed once and throws a ball left-handed once, and we measure the distances. If we are interested in determining a confidence interval for the mean difference in left- and right-handed throws, which type of c.i. would we likely use?

    (a) z (normal) confidence interval for differences
    (b) pooled t confidence interval for differences
    (c) paired t confidence interval for differences
    (d) χ^2 confidence interval for differences
    (e) F confidence interval for differences

    Solution: (c) $\diamondsuit$

29. TRUE or FALSE? We reject the null hypothesis if we are given statistically significant evidence that it is false.

    Solution: TRUE. $\diamondsuit$

30. The quantity α is known as

    (a) P(Type I error)
    (b) P(Type II error)
    (c) level of significance
    (d) $\mathsf{P}(\text{Reject } H_0 \mid H_0 \text{ is true})$

    Solution: (a), (c), and (d). $\diamondsuit$

31. Suppose that we examine the IQs of 50 Justin Bieber concert attendees. We assume that the IQs are normally distributed with a standard deviation of 10. Suppose that the sample mean turns out to be 82. Test the null hypothesis that the mean IQ of the attendees is at least 90. Use $\alpha = 0.05$.

    Solution: The null hypothesis is $H_0 : \mu \geq 90$. The test statistic is

$$Z_0 = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}} = \frac{82 - 90}{\sqrt{100/50}} = -5.66. \quad \diamondsuit$$

Since $Z_0 < -z_\alpha = -1.645$, we reject H_0. (We didn't even really need tables for this, since it's so obvious.) $\diamondsuit$

32. Referring to Question 31, how many observations should we take if we want the probability of a Type II error to be 0.10 when μ happens to equal 87?

Solution: Since this is a one-sided test,

$$n \approx \frac{\sigma^2(z_\alpha + z_\beta)^2}{\delta^2} = \frac{\sigma^2(z_{0.05} + z_{0.10})^2}{(87-90)^2} = \frac{100(1.645 + 1.28)^2}{9} = 95.1 = 96. \quad \diamondsuit$$

33. Suppose we want to compare the means of two normal populations, both of which have *unknown but approximately equal variances.* We take $n = 6$ observations from the first population and find that the sample mean and sample variance are $\bar{x} = 50$ and $s_x^2 = 120$. We take $m = 5$ observations from the second population and find that the sample mean and sample variance are $\bar{y} = 75$ and $s_y^2 = 100$. Test the hypothesis that $\mu_x = \mu_y$ with $\alpha = 0.05$, i.e., either accept or reject.

Solution: First of all, the pooled variance is

$$S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2} = 11.11.$$

Then the CI is

$$\begin{aligned}
\mu_x - \mu_y &\in \bar{X} - \bar{Y} \pm t_{\alpha/2,n+m-2}\sqrt{S_p^2\Big(\frac{1}{n} + \frac{1}{m}\Big)} \\
&= -25 \pm t_{0.025,9}\sqrt{11.11\Big(\frac{1}{6} + \frac{1}{5}\Big)} \\
&= -25 \pm 2.262(6.383) \\
&= -25 \pm 14.44.
\end{aligned}$$

Since the CI doesn't contain 0, we *reject* $H_0 : \mu_x = \mu_y$. $\diamondsuit$

34. Short Research Question: Go to The Internets and write up a couple of paragraphs on what a "goodness-of-fit" hypothesis test is.

35. Let's see if weight depends on height. We consider 6 i.i.d. people.

Height (in.)	70	74	62	66	71	60
Weight (lbs.)	180	210	146	175	165	150

(a) Fit a regression line to this data and report your estimates for β_0 and β_1.

Solution: After the usual algebra (which I did in Minitab because I'm lazy), we obtain the model $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = -74.5 + 3.66x$. $\diamondsuit$

(b) What is the expected weight of a person who is 72 inches tall?

Solution: $\hat{y}|x = -74.5 + 3.66(72) = 188.7$. $\diamondsuit$

(c) Give a 95% confidence interval for β_1.

Solution: First of all, note that $\bar{x} = 67.17$ and $S_{xx} = \sum_{i=1}^n x_i^2 - n\bar{x}^2 = 148.83$. Moreover,

$$\begin{aligned} \hat{\sigma}^2 &= \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2} \\ &= \frac{\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n-2} \\ &= 182.8. \end{aligned}$$

The desired confidence interval is therefore

$$\begin{aligned} \beta_1 &\in \hat{\beta}_1 \pm t_{\alpha/2,n-2} \frac{\hat{\sigma}}{\sqrt{S_{xx}}} \\ &= \hat{\beta}_1 \pm t_{0.025,4} \frac{\hat{\sigma}}{\sqrt{S_{xx}}} \\ &= 3.66 \pm (2.776)(1.11) \\ &= 3.66 \pm 3.08. \quad \diamondsuit \end{aligned}$$